

# Nordiska studier i lexikografi

## 11

Rapport från Konferensen om lexikografi i Norden  
Lund 24–27 maj 2011

Redigerad av

*Birgit Eaker*

*Lennart Larsson*

*Anki Mattisson*

Skrifter utgivna av Nordiska föreningen för lexikografi  
Skrift nr 12

I samarbete med Språkrådet i Norge och  
Svenska Akademiens ordboksredaktion

**DANSK SPROGNÆVN**

2012/11

Nordiska studier i lexikografi, nr 11  
© Nordiska föreningen för lexikografi och författarna, 2012  
Redaktion: Birgit Eaker, Lennart Larsson och Anki Mattisson  
Layout: eddy.se ab, Visby 2012  
Typsnitt: Adobe Garamond  
Tryckeri: Elanders Sverige AB, Mölnlycke 2012  
ISBN: 978-91-85333-42-4  
ISSN: 0803-9313



**norden**

Nordisk Kulturfond



**NORDPLUS**



**SVENSKA  
AKADEMIEN**

**LETTERSTEDTSKA  
FÖRENINGEN**



Rapporten kan beställas från

Språkrådet  
Postboks 8107 Dep  
NO-0032 Oslo  
e-post: [post@sprakradet.no](mailto:post@sprakradet.no)

# Innehåll

Mot en hållbar lexikografi – En balansgång mellan återanvändning och förnyelse <i>Godelieve Laureys</i>	13
Biskopar, ordböcker och verkligheten – Vad är lexikografihistoria bra för? <i>Bo Ralph</i>	30
What has the OED become? <i>John Simpson</i>	55
Strategier for funktionsbaseret betydningsordning <i>Birger Andersen</i>	77
Flødeboller, mælkebøtter og smørkringler Om leksikalsk-semantiske problemstillinger i arbejdet med Ømålsordbogen <i>Lena Wienecke Andersen</i>	92
Koordinering i behandlingen af faste ordforbindelser <i>Torben Arboe</i>	103
Norsk Ordbok på nett: Utfordringar og løysingar <i>Sturla Berg-Olsen &amp; Åse Wetås</i>	115
Breaking away from tradition: Linking a database of inflection to an electronic dictionary <i>Kristín Bjarnadóttir</i>	128

Nye kilder til dansk ordbogshistorie <i>Simon Skovgaard Boeck</i>	138
Nordisk barneordbok på nett <i>Torbjörg Breivik</i>	150
Hur blir en traditionell, tryckt ordbok en webbordbok? <i>Ilse Cantell &amp; Caroline Sandström</i>	157
Maintaining consistency of monolingual verb entries with interannotator agreement <i>Silvie Cinková, Lenka Smejkalová, Anna Vernerová, Jonáš Thál &amp; Martin Holub</i>	169
Ellips – ett beskrivningsproblem i en historisk ordbok <i>Birgit Eaker</i>	181
Fra alfabet til begrep: Bokmålsordboka og NorNet <i>Ruth Vatvedt Fjeld, Rune Lain Knudsen &amp; Julie M. Torjusen</i>	190
Ømålsordbogens beskrivelse af verbalbøjning <i>Asgerd Gudiksen</i>	202
Ivar Aasen og svensk leksikografi <i>Kjetil Gundersen &amp; Dagfinn Worren</i>	212
Ordbog over Dansk Talesprog (ODT) – værksteds- præsentation af en ordbogsportal <i>Carsten Hansen &amp; Martin H. Hansen</i>	224
Islandsk-tsjekkisk / Tsjekkisk-islandsk ordbok (IT/II) En biskopal, bidireksjonal ordbok i stort lommeformat <i>Helgi Haraldsson</i>	236
Ensam fras søker kontekst: Översättarens hantering av språkprov i en tvåspråkig ordbok <i>Ylva Hellerud</i>	243



Massive ordbokspalter i transparent presentasjon – NAOB på nett: samspillet mellom datastruktur, innhold og brukergrensesnitt <i>Petter Henriksen</i>	251
Russin i kakan i Jesper Swedbergs Svensk Ordbok <i>Lars Holm</i>	263
Norsk dialektatlas i digital utgåve og språkdata på kart <i>Jon Holmen &amp; Oddrun Grønvik</i>	273
SAOLHist – alla upplagor av SAOL i en och samma databas <i>Louise Holmer</i>	287
Data og repræsentativitet i ordbogsarbejdet <i>Henrik Hovmark</i>	296
De første trykte islandske ordbøger <i>Gunnlaugur Ingólfsson</i>	309
Etymologien til ordet ”jordmor” <i>Ernst Håkon Jahr</i>	318
Parallellkorpusen som resurs i lexikografiskt arbete <i>Håkan Jansson</i>	328
Bruklighetsoppgifter i ordböcker: ett retoriskt perspektiv <i>Lars-Erik Johansson</i>	340
ISLEX – en flersproget nordisk ordbog <i>Halldóra Jónsdóttir &amp; Þórdís Úlfarsdóttir</i>	353
Adverb og adverbialer: En forsømt ordklasse i ordbøkene <i>Jón Hilmar Jónsson</i>	367
Ordbogsmanuskripter og et historisk ordbogsarbejde <i>Guðrún Kvaran</i>	377

Varför inte dub(b)lettformer? Om SAOB som normativ ordbok <i>Lennart Larsson</i>	385
Atypiske lemmaer i Det Norske Akademis Store Ordbok Noen betraktninger om deres plass og funksjon i en moderne diakron og deskriptiv ordbok <i>Hanne Lauvstad</i>	396
Ejendomsordbogen fransk/dansk: ny integreret e-ordbog <i>Patrick Leroyer &amp; Liselotte Kruse</i>	405
Netordbøger – hvordan finder brugerne frem til dem, og hvad gør de, når de har fundet dem? <i>Henrik Lorentzen &amp; Liisa Theilgaard</i>	418
Hapax i Olof Verelius ordbok från 1681 <i>Anton Lundqvist</i>	431
Lexikografisk beskrivning av ortnamn <i>Jonas Löfström</i>	443
Från Svenska ord (Lexin) 3 till Svenska ord 4 <i>Sven-Göran Malmgren</i>	454
Farväl till en stor grupp adverb En breddad prepositionskategori i ordboksartiklar? <i>Nina Martola</i>	466
Språkteoriers avtryck i SAOB <i>Pär Nilsson</i>	479
Nettordbøker og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok <i>Christian-Emil Ore</i>	488
Dansk leksikografi med polsk: past, present and future <i>Marcin Overgaard Ptaszynski</i>	500

Ein-, to- eller fleirspråkleg ordbok? <i>Margunn Rauset, Anna Helga Hannesdóttir &amp; Aldís Sigurðardóttir</i>	512
A corpus based method for a diachronic study of the central vocabulary of New Norwegian <i>Daniel Ridings &amp; Oddrun Grønvik</i>	524
Ord i kombination En undersökning av några adjektiv som uttrycker 'hög grad' <i>Bodil Rosqvist</i>	534
Wienerbröd och andra finesser. Den nya svensk-litauiska ordboken <i>Erika Sausverde</i>	546
Orientering om arbeidet med Norsk-russisk juridisk ordbok <i>Aud Anna Senje</i>	555
Et informationsvidenskabeligt serviceeftersyn af Medicin.dk <i>Henrik Køhler Simonsen</i>	563
A och O om akademiska ord Om framtagning av en svensk akademisk ordlista <i>Emma Sköldberg &amp; Sofie Johansson Kokkinakis</i>	575
Svordomar i ett lexikografiskt perspektiv <i>Ulla Stroh-Wollin</i>	586
Folklig dialektlexikografi 1991–2011 <i>Eva Thelin</i>	596
Udtale i netordbøger – hvad, hvordan, hvor meget? <i>Lars Trap-Jensen</i>	610
Ordfrekvens och ordboksutrymme <i>Bo-A. Wendt</i>	621

# Mot en hållbar lexikografi

## – En balansgång mellan återanvändning och förnyelse

*Godelieve Laureys*

### 1. Inledning

Begreppet *hållbar* är på frammarsch. Det används inom ett brett spektrum av aktiviteter. Det bär även en viss prägel av att vara ett modeord och uttrycker i vissa fall närmast en allmän positiv betydelse. Idag skall man göra saker och ting *på ett hållbart sätt*.<sup>1</sup> Inte bara skall man *odla, fiska, producera, förvalta, använda, utveckla, planera, hantera, nyttja, sköta, ta hand om, bevara, bygga, tillverka* ”*på ett hållbart sätt*”, utan man skall också *resa, jobba, flyga, investera pengar, laga mat, steka, träna, äta, gå ned i vikt, dricka, sola sig, duscha, klä sig, förbereda sig för ett barn*, eller kort sagt *leva* ”*på ett hållbart sätt*”<sup>2</sup>. Den mest frekventa och mest lexikaliserade kollokationen är *hållbar utveckling*, ett begrepp som har kommit in i språket som en lånöversättning av engelskans *sustainable development*. Det kan vara intressant att notera att begreppet, som myntades av miljöforskaren och essäisten Lester Brown på det tidiga 1980-talet, vann världsvid anklång i samband med Brundtland-rapporten, *Vår felles framtid* (1987). *Hållbar utveckling* återges på danska med *bæredygtig udvikling*, på norska med *bærekraftig utvikling* och på finska med *kestävä kehitys* eller med det abstrakta substantivet *kestävyyys*. Substantivet *hållbarhet* används även på svenska ofta för att beteckna *hållbar utveckling*.

Begreppet avser i första hand en jämvikt mellan ekologiska, ekonomiska och sociala krav och vädjar om en långsiktig och medveten förvaltning och exploatering av världens naturliga resurser med tanke på kommande generationers behov och välfärd och pläderar för en vertikal solidaritet över tiden. Från den ekologiska domänen har begreppet förts vidare till de sociala och ekonomiska domänerna, där det även handlar om en långsiktig och genomtänkt hushållning

---

<sup>1</sup> Mer än en halv miljon hits på Google, maj 2011.

<sup>2</sup> Alla nämnda kollokationer är högre frekventa på Google.

med mänskliga och materiella resurser. På en ännu högre abstraktionsnivå innebär hållbarhetsidealet en strävan att åstadkomma bättre och varaktiga resultat med minskad resursåtgång och att skapa resurser som förblir tillgängliga för det framtida samhället. Då tangerar vi begrepp som *långsiktig effektivitet* och *optimering*.

Det är självfallet i den sistnämnda abstrakta bemärkelsen av *ökad effektivitet, för att åstadkomma bättre resultat med minskad resursåtgång* som begreppet *hållbarhet* används inom forskning i allmänhet och inom lexikografi i synnerhet. Nyckelbegrepp är: långsiktigt perspektiv, varaktiga investeringar i motsättning till engångsinsatser, framtidsinriktade och användarvänliga produkter, återanvändning av materiella och mänskliga resurser, överförbara resultat, robusta verktyg. Någon direkt miljöeffekt är det i vårt sammanhang nog knappast tal om, annat än att datoriseringen inom lexikografien torde bidra till ett papperslöst samhälle.

I följande danska citat används begreppet i en betydelse som i alla fall inte avses här: ”Og ordbøgerne er faktisk handy og praktiske at håndtere. Indbindingen er solid med et omslag i blødt plastic, hvilket øger *holdbarheden* i udsatte situationer. Alle tre ordbøger er udformet så de kan tåle at blive brugt i marken.” (Hovmark 2005: 15)

I mitt bidrag vill jag försöka belysa på vilket sätt hållbarhetsprincipen kan tillämpas i lexikografien och hur begreppet har utvecklats genom tiden. Jag urskiljer därvid tre steg i utvecklingen.

## 2. En ordbok åt gången

Ordböcker är redskap som hjälper oss att få inblick i både uttryckssidan och innehållssidan av ett språks lexikaliska element. Det finns som bekant många typer av ordböcker, som antingen fokuserar på vissa delar av ett språks lexikon eller som är avsedda för vissa funktioner och/eller som riktar sig till olika användargrupper. Trots alla möjliga specifikationer vad gäller typologi, funktion och användargrupp är ordböcker alltid i viss mån generiska verktyg, som skall kunna användas i flera olika sammanhang. Bortsett från vissa ordböcker som just vill fånga det nya och kanske det flyktiga i språket, som t.ex. nyordsordböcker eller slangordböcker, och bortsett från slutna vokabulärer som t.ex. terminologiska listor och glossarier, tenderar ordböcker överlag att beskriva språket i ett registermässigt, stilistiskt och tidsmässigt brett perspektiv. På så sätt kan ordböcker sägas ha hållbarhet som en inneboende egenskap.

Förr i tiden kom ordböcker oftast till som alster av en enskild författare. Ofta var det tal om ett mångårigt arbete eller till och med ett livsverk. De tidi-

gaste ordböckerna i Norden var latinska glossarier och översättningsordböcker från latinet. På 1600-talet, då intresset för folkspråken vann terräng, uppstod de första monolinguala ordböckerna för våra nuvarande nationalspråk. Även dessa ordböcker stod i tacksamhetsskuld till den latinska traditionen, således att många uttryck och exempel har förts vidare från ordbok till ordbok och levt sitt bästa liv i den lexikografiska traditionen. Detta belyses mycket tydligt i Anna Helga Hannesdóttirs avhandling om den enspråkiga svenska lexikografins utveckling (Hannesdóttir 1998).

Existerande ordböcker har länge utgjort den första tillgången för en lexikograf och därför innehåller ordböcker ofta ackumulerad kunskap. Det framgår till och med att vissa svagheter eller direkta fel på så sätt förs vidare från den ena ordboken till den andra. Detta är i och för sig inte så märkvärdigt eftersom all forskning bygger vidare på och förhåller sig till andras resultat. Men eftersom ordböcker kartlägger ett språks eller ett delspråks lexikaliska element på systematisk nivå, finns det bara hänvisningar till tidigare ordböcker i de fall där det är tal om en omarbetning eller en reviderad version av en redan existerande ordbok och inte när det gäller långods i enskilda belägg. Strävan efter att kartlägga det kanoniska språkbruket, vilket metaforiskt uttryckt faller under "allemansrätten", gör ordbokstexter till ett slags inventarier, som inte avspeglar en individuell författares eller talares idiosynkratiska språkbruk, utan som representerar det kollektiva språkbruket. Ordböcker innehåller m.a.o. en förteckning över ett språksamhälles gemensamma arvegods, som självfallet är *open source*. Att lexikografer använder existerande ordböcker som källa kan även leda till vissa frågor angående den intellektuella egendomsrätten. Det som är upphovsrättsligt skyddat är själva urvalet både på mikro- och på makronivå. Den unika selektionen av språkligt material, baserad på visa redaktionella principer, är *intellectual property*<sup>3</sup>. Det är det som utgör resultatet av en noggrann samlings- och systematiseringsprocess, de enstaka uppslagsorden är däremot *share ware*. När det gäller metaspråkliga uppgifter, såsom betydelsesuméer, definitioner och grammatisk annotation möter vi samma problem. Ingen redaktör kan sägas vara ägaren av pluralisformen till substantivet *flicka*, men när det t.ex. gäller orddefinitioner finns det tydliga upphovsrättigheter.

Frågan gäller även exempelaterialet och autentiska korpusexempel. Vissa korpusar är *open source*, andra är licensierade och får inte användas utan tillstånd. De korpusar som har kommit till stånd med offentliga medel är överlag tillgängliga för forskarsamhället och ofta även för en bred publik. Frågan är om

<sup>3</sup> Vissa ordboksförfattare har därför tagit upp s.k. fantomord eller virusord, dvs fiktiva ord som kan spåras för att påvisa plagiat.

dessas material också får användas av privata företag i kommersiellt syfte. Många ordboksredaktörer kommer på så sätt att arbeta inom en gråzon.

### 3. Från ordbok till ordboksserie

På 1900-talet kom det fart i ordboksproduktionen. Ett antal förlag började specialisera sig på ordböcker. Det ledde till olika ordboksserier, som skulle förstärka förlagets varumärke. Seriebildningen var ett kraftigt medel i marknadsföringsstrategin. Serierna blev också alltmer specialiserade. Särskilt gäller detta för tvåspråksordböcker, som delades upp i skolordböcker, rese- eller mini-ordböcker, medelstora ordböcker och för vissa språk även stora professionella ordböcker.

Dessa ordboksserier har inte alltid representerat en enhetlig redaktionspolitik. Visserligen har ett antal ordböcker kommit till i förlagets regi, dvs. de blev utarbetade på själva förlaget av interna redaktörer, men ett ännu större antal ordböcker har utarbetats på beställning av externa redaktörer. I vissa fall köptes helt enkelt färdiga manuskript eller kom det först till en överenskommelse, när manuskriptet var färdigt i stora delar. Det som faktiskt mest höll ihop en serie var pärmens utformning, vilket är helt i överensstämmelse med dåvarande marknadsstrategi.

På 1990-talet, då den första stora fusionsvågen inom förlagsvärlden ägde rum, började de stora förlagen att lansera tanken på en egen förlagspolitik vad gäller ordböcker. Kvaliteten i de många nya uppköpta produkterna skulle granskas och en viss profilering på innehållssidan trängde sig på. Detta sammanföll med att datoriseringen gjorde sitt inträde i branschen. Tanken på att samköra olika databaser och göra dem inbördes kompatibla växte fram. Drömmen om en masterdatabas eller ett s.k. moder-bestånd, som ett antal olika ordböcker kan härledas ur, skymtade vid horisonten. Men ännu var det en lång väg att gå. Det fanns inga standardiserade softwarenormer. Datorer var mestadels stand alone-maskiner eller var kopplade till lokala nätverk, det manuella arbetet i post-redigeringsfasen sparades sällan elektroniskt och mycket gick förlorat på vägen. Det är den period då förfarandena kan betecknas som datorstödda, inte som datorbaserade eller datorstyrda. Datorerna var trogna, noggranna och flitiga medarbetare, som i bästa fall utförde det som redaktörerna hade bestämt, men som varken påverkade redaktionsprinciperna eller ordbokens utformning.

Som exempel kan nämnas Gyldendals *røde ordbøger* och Norstedts *randiga serie*. Det danska förlaget Gyldendal har utgett ordböcker alltsedan 1798. Den första *røde ordbog* utkom 1902 som ett svar på Nordisk Förlags blå ordboksserie.

Ordbøgerne var indbundet i rødt lærred i modsætning til den uindbundne udgave fra Nordisk Forlag – og så var de endda en krone billigere. De kostede den nette sum af

1,50 kr. Den første røde ordbog var Dansk-Norsk-Engelsk Ordbog. I 1902 udkom desuden Dansk-Tysk ordbog – derefter blev ordbogsserien udbygget med alle de vigtige sprog i løbet af de følgende år. I dag tæller serien mere end 30 titler fordelt på 24 sprog, fra svensk til persisk, polsk og tyrkisk. De røde ordbøger blev en øjeblikkelig succes og blev hurtigt en af Gyldendals vigtigste udgivelseslinjer. Det samlede salg ligger på ca. 14 millioner eksemplarer, og serien er i dag et af de bedst kendte danske varemærker.<sup>4</sup>



© Gyldendal

I Sverige har Norstedts profilerat sig som ledande ordboks-förlag. Varumärket används också som en kvalitetsgaranti, som det framgår av följande citat:

”Ordboken är utvecklad av Norstedts, som har en lång tradition av ordboksutgivning. Norstedts randiga ordböcker är kända av de flesta, och ordböckerna borgar för kvaliteten.”<sup>5</sup>



#### 4. Återanvändning av ordboksmaterial och generiska redaktionsprinciper

Sedan millennieskiftet har den traditionella grafiska ordboken så småningom varit på väg att bli en nischprodukt. De första elektroniska ordböckerna, som distribuerades på CD-rom på 1990-talet, avvek dock inte stort från sina grafiska modeller, förutom att layouten var mer översiktlig för att man inte längre skulle ta hänsyn till utrymme. Själva mediet hade blivit annorlunda, det gick snabbare att slå upp ett ord eller ett uttryck, man kunde bläddra med en browser, men sökningsprocedurerna var och är desamma som i grafiska ordböcker. Man kan exempelvis inte selektera vissa informationskategorier, utan man får presenterat

<sup>4</sup> Citat från Gyldendals Forlags hemsida <http://ordbog.gyldendal.dk/> maj 2011.

<sup>5</sup> Citat från pärmen till den Svensk-Nederländska ordboken. Se även (Laureys & De Groot 1996).



allt på skärmen precis som det står i den grafiska ordboken. Men uppkomsten av elektroniska ordböcker utgör ändå ett första exempel på återanvändning: ur samma underliggande intellektuella material härleds två olika produkter.

Vi fick emellertid vänta på de nätbaserade lexikaliska databaserna och ordböckerna för att verkligen kunna ta det tredje steget fullt ut. Här rymms helt andra möjligheter för att lagra, presentera, distribuera och uppdatera lexikalisk information. Webbordböcker är inte lineära och rymmer därför mycket större möjligheter att på ett flexibelt sätt komma åt just den information man är intresserad av. Från användarens synvinkel bjuder dessa ordböcker på mycket större flexibilitet. Man kan selektera vissa typer eller delmängder av information som man är intresserad av och skapa sig en ordbok med en skraddarsydd profil. Informationskategorier som man inte har användning för gömmer man undan, men vid behov kan användaren ändå klicka sig fram till den bortsorterade informationen. Allt oftare lagras ordboken numera under arbetets gång den önskade profilen<sup>6</sup> och förvandlas på så sätt till ett dynamiskt verktyg.

Samtidigt kan användaren också på ett associativt sätt hoppa runt i ordboken genom att klicka sig fram från ett uppslagsord till ett annat. De nyaste webbordböckerna är multimediala och till texten kopplas bild, ljud och video. I samma applikation finns också tillgång till elektroniska korpusar och inbyggda verktyg (konkordanser, förteckningar över frekventa kollokationer o.d.). Den kvalificerade användaren kan uppehålla sig länge på webben och navigera sig fram genom interna länkar mellan olika ordboksartiklar och genom externa länkar föras direkt till hypertexter och/eller hypermedier. Vi närmar oss begreppet ”hyperlexikon”, såsom det beskrivs i ”Handbok i lexikografi” (Svensén 2004: 525) och såsom det på ett visionärt sätt lanserades av Sue Atkins redan på 1990-talet (Atkins 1994 och 1996).

Sedan digitaliseringen och särskilt sedan uppkomsten av nätbaserade ordböcker har även hela reviderings- och uppdateringsprocessen fått en helt annan karaktär. Det lexikografiska materialet kan uppdateras löpande och revisionen är en permanent process. En ordbok är inte längre någon färdig och statisk produkt, som ligger oförändrad till förfogande tills en ny reviderad upplaga utkommer. Tvärtom, så kan ändringsförslag samlas in och exempelvis läggas ut på nätet med några få dagars eller kanske veckors mellanrum. Dessutom kan det tas hänsyn till användarnas kommentarer och förslag, så att produkten ständigt utvecklas och görs mera anpassad till användarnas krav och behov. Detta dyna-

---

<sup>6</sup> Fuzzy SF (Simultaneous Feedback) är t.ex. ett system, utarbetat av Gilles-Maurice De Schrijver, som automatiskt strukturerar informationen i en ordbok på grundval av användarens sökstrategier. På så sätt bildas dynamiska ”ultimate customised dictionaries” för varje individuell användare (De Schrijver & Prinsloo 2001).

miska tillvägagångssätt innebär en viss interaktion mellan redaktörerna och användarna. Det finns utrymme för dialog och *user based content* på chattforum eller genom andra kanaler. Dessutom kan redaktörer också erhålla nya insikter genom att studera användarnas faktiska beteende och sökstrategier, som kan rekonstrueras ur logfilerna. Sistnämnda heuristiska förfarande skapar nya möjligheter att fokusera på det som användarna verkligen söker efter (vilka ord? vilka informationstyper?). Ur logfilerna kan man också härleda hur användarna går tillväga, när de inte genast hittar det de letar efter. Analysen av denna sökprocess kan ligga till grund för en optimering av lemmasektionen och för etableringen av nya länkar och datorstyrda instruktioner t.ex. i form av pop ups.

Inom språkteknologin har det på senare tid gjorts ihärdiga försök att utveckla standardiseringsnormer som underlättar återanvändning av vissa verktyg och material<sup>7</sup>. Förr hindrades återanvändning av att resurserna hade utvecklats inom ramen för ett bestämt projekt i ett idiosynkratiskt format och med bestämda skraddarsydd datastrukturer. I det följande beskriver jag två konkreta exempel på återanvändning: det nordiska Lexin-projektet och mina egna erfarenheter från verkstadsgolvet i samband med framtagningen av olika bilinguala ordböcker: nederländska-svenska/danska/finska.

#### 4.1 LEXIN

Ett bra exempel på återanvändning är det svenska LEXIN-projektet, som redan i ett tidigt stadium på 1980-talet koncipierade en serie monodirektionella bilinguala ordböcker för invandrare. Alla utarbetades med samma svenska källspråksdatabas som utgångspunkt. Modellen övertogs senare i det utvidgade nordiska LEXIN-projektet. Tanken bakom denna form för återanvändning var i första hand inspirerad av eller rent av framtvungad av den begränsade marknaden för en stor del av invandrarordböckerna. Därtill kom den ideologiska principen att ordböckerna skulle vara lättillgängliga och billiga. I det nordiska Lexin-projektet har återanvändningsprincipen tillämpats på två plan. Dels tillämpar man även för den danska och den norska versionen samma princip med en gemensam källspråksdatabas och dels har den svenska ordbasen i stort legat till grund för den danska resp. norska makro- och mikrostrukturen.

Naturligtvis är det i och för sig ingen nyhet att ett ordboksprojekt i ett nordiskt land utnyttjar arbete som tidigare utförts inom ordboksprojekt i andra nordiska länder. Men möjligen innebär arbetet på nordiska invandrarordböcker det hittills tydligaste exemplet

---

<sup>7</sup> En del standarder har utvecklats inom ramen för europeiska samarbetsprogram som EAGLES och ISLE. Andra viktiga instanser i detta samband är ISO/T37, W3C samt LISA (Localising Industry Standards Association).

på möjligheterna till sådant samarbete, i syfte att så långt som möjligt undvika dubbelarbete. (Bergenholtz & Malmgren 1999: 1)

#### 4.2 Hub and Spoke-modellen tillämpad inom bilingual lexikografi

Under de senaste femton åren har jag som huvudredaktör successivt arbetat med tre bilinguala ordböcker, som har nederländska som källspråk och resp. svenska, danska och finska som målspråk. (Laureys & De Groot, 1996; Laureys et al 2004; Laureys & Moisisio 2012.) På så sätt har jag kunnat fokusera på de metodiska och organisatoriska aspekterna av ett ordboksprojekt och jag har fattat ett särskilt intresse för de metalexikografiska perspektiven. Ovannämnda tre ordböcker uppvisar ett antal likheter och skillnader. Likheter kommer till synes i de färdiga produkterna, eftersom ordböckerna återspeglar de gemensamma målsättningar och lexikografiska principer som ligger till grund för dem. Likheter har sin logiska förklaring i att alla tre ordboksparerna skall uppfylla samma funktionskrav, att de är avsedda för parallella målgrupper, har identisk ambitionsnivå vad angår mikro- och makrostruktur, är bifokala samt är utarbetade enligt i stort samma redaktionsprinciper. Alla tre ordböcker omfattar ca 45 000 uppslagsord och ca 75 000 exempel, de är korpusbaserade och behandlar det moderna ordförrådet både i tal och skrift.

Utgångspunkten för den nederländska makro- och mikrostrukturen i dessa ordböcker utgör det s.k. RBN (*Referentiebestand Nederlands*), en generisk korpusbaserad relationell ordbas, utarbetad på uppdrag av *Nederlandse Taalunie*<sup>8</sup> med det specifika syftet att bilda underlag för ett antal bilinguala ordböcker med nederländska som källspråk. I detta syfte har även redigeringsprogrammet OMBI utvecklats. OMBI är skraddarsytt till RBN-strukturen och är utstyrt med en avancerad vändfunktion på LU-nivå. LU definieras som "lexikalisk enhet", dvs. en viss betydelsevariant av ett visst uppslagsord.<sup>9</sup> Varje lexikalisk enhet (LU) i källmaterialet länkas i databasen till en eller fler översättningsekvivalenter på målspråket och på detta sätt etableras en kontrastiv databas på LU-nivå. Programmet tillåter redaktören att när som helst byta källspråk och målspråk. RBN omfattar inte bara lexikaliska data (uppslagsord och exempel) utan också olika informationskategorier och annotationer. Informationen för varje uppslagsord omfattar förutom den semantiska profilen, uppgifter om stavning, morfologi, syntaktiska egenskaper och mönster, pragmatik, stilistiska särdrag

<sup>8</sup> "The Dutch Language Union", det gemensamma Nederländska Språkrådet för Nederländerna och Flandern.

<sup>9</sup> Vid monosemiska uppslagsord sammanfaller alltså formenheten (FU) med den lexikaliska enheten (LU).

samt geografisk tillhörighet. Dessa annotationer överstiger den information som tas upp i grafiska eller elektroniska ordböcker, men är till stor hjälp under redaktionsprocessen och berikar även databasen för framtida lexikologisk forskning. För att kunna använda RBN till en konkret ordbok måste redaktionen därför först fastställa selektionsprinciper med avseende på makro- och mikrostrukturen. Med tanke på ordbokens funktion(er) och på målgruppens behov måste ett antal uppslagsord strykas och andra läggas till och likaså måste ett relevant urval göras ur det stora antalet exempel i RBN.

I samband med det nederländsk-danska ordboksprojektet lades ned mycket tid och pengar på en adekvat selektion av uppslagsord och exempel för att åstadkomma en relevant makro- och mikrostruktur för den tilltänkta målgruppen. Denna investering kan ses som produktutveckling av RBN-materialet. Det var just denna tanke på produktutveckling som gjorde att jag bestämde mig för att återanvända materialet och dra nytta av denna investering i ett nytt ordboksprojekt. I projektbeskrivningen för den nederländsk-finska ordboken framsattes förslaget att återanvända den variant av RBN, som vi hade utarbetat i samband med det danska projektet<sup>10</sup>. Till grund för detta beslut låg dock inte bara ekonomiska övervägningar utan också en djärvare syn på återanvändning. Från början har tanken varit att komma underfund med vilka möjligheter som ligger i att framställa råmaterial till en bilingual ordbok för ett nytt språkpar på grundval av en multilingual databas. Jag hänvisar i detta samband till *Hub and Spoke*-modellen, lanserad av Willy Martin (2001) (2003) och (2004). Jag ville pröva denna modell i praktiken genom att senare försöka länka ihop den danska och den finska målspråksdatabasen och på semi-automatisk väg framställa en experimentell version av en finsk-dansk ordbok.

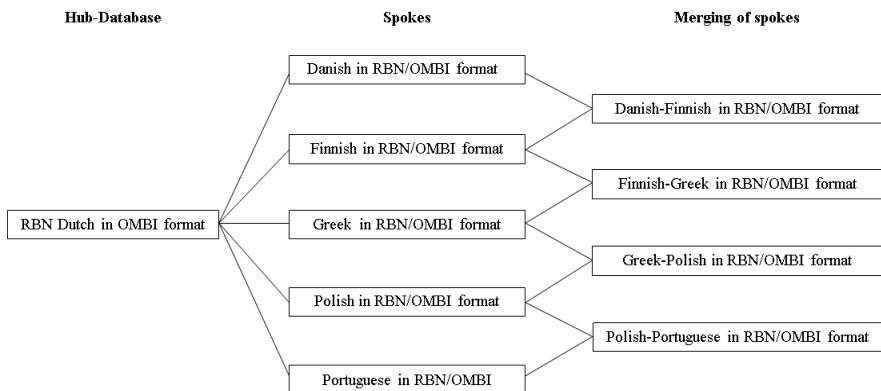
I överensstämmelse med ambitionen om att optimera redaktionsprocessen har vi dessutom försökt att återanvända existerande finskt material. Vi erhöll en licens på den elektroniska underliggande databasen till *Suomi-saksa opiskelusanakirja*, som ägs av WSOY-förlag. Materialet har konverterats så att det kunde laddas ned direkt i OMBI. Ovannämnda ordbok, som omfattar 51 000 uppslagsord, är en typisk inlärningsordbok för finstalande som skall lära sig tyska. Fördelen med detta förfarande var att vår databas fylldes från starten med 51 000 finska uppslagsord och ett stort antal exempel. Förutom detta material

---

<sup>10</sup> Det skall betonas att vi inte blint reproducerade den danska varianten av RBN-databasen, då vi satte igång med den nederländsk-finska ordboken. Redaktionen fick självfallet ta hänsyn till den nya målgruppen och till den kontrastiva relevansen för språkparet nederländska-finska. Men även efter denna bearbetning kan man räkna med att ca 80% av det nederländska materialet är gemensamt för den nederländsk-danska och den nederländsk-finska ordboken.

blev även en annan viktig resurs i förväg laddad ned i databasen. Från KOTUS<sup>11</sup> erhöll vi en licens på den annoterade makrostrukturen av den stora monolinguala finska ordboken *Suomen kielen perussanakirja*, som omfattar 93 000 uppslagsord försedda med grammatisk information. I denna ordbok är alla substantiv och verb försedda med en kod, som hänvisar till flexionstabellerna. Dessa grammatiska upplysningar fogades automatiskt till de finska uppslagsorden i vår databas.

I och med att den danska resp. den finska ordbokstexten i stort kan betraktas som en översättning av identiska nederländska uppslagsord och exempel, torde det i princip vara möjligt att länka ord och strängar från båda målspråken till varandra. Utgångspunkten för experimentet med en finsk-dansk ordbok utgör alltså de elektroniska kontrastiva lexikaliska databaser för språkparen nederländska/danska och nederländska/finska, som utarbetades i samband med redaktionen av ordböckerna för motsvarande språkpar. Under hänvisning till *Hub and Spoke*-modellen har jag experimenterat med att foga ihop de två kontrastiva databaserna i en integrerad trespråkig databas. I den integrerade databasen kan det på ett entydigt sätt etableras en länk mellan en dansk och en finsk ordsträng på grund av att de danska resp. de finska uppslagsorden och exemplen är markerade som översättningar av bestämda nederländska uppslagsord och exempel. Jag har med andra ord betraktat materialet i den danska och den finska databasen som två *spokes* länkade till samma *hub*.



Visualisering av Hub and Spoke-Modell

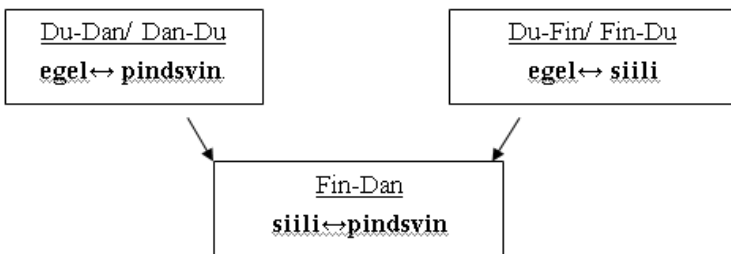
Jag påvisar här den stora fördelen med att både den danska och den finska databasen blev utarbetade på grundlag av samma RBN-version, vilket medför att de

<sup>11</sup> Kotimaisten kielen keskus (Institutet för de inhemska språken), Helsingfors.

båda lexikaliska databaserna er maximalt parallella. Därmed är det viktigaste kravet för en meningsfull och framgångsrik *merging* uppfyllt: strukturen i databasen er identisk och de semantiska och grammatiska specifikationerna följer samma mall. Därutöver förekommer ofta identiska exempel. Det innebär att den danska och den finska delmängden i databasen utgör en ideell utgångspunkt för en experimentell länkning, som siktar mot en finsk-dansk kontrastiv databas, som i sin tur kan utgöra en rå version av en finsk-dansk ordbok. Den bakomliggande tanken är hypotesen att, om en finsk LU och en dansk LU är länkade till en och samma LU i den nederländska databasen, så finns det en rimlig sannolikhet att den danska och finska LU även inbördes är adekvata översättningsekvivalenter. Spåret är den gemensamma länk som den finska och den danska LU har till en och samma nederländsk LU. Om A är lika med B och C är lika med A, så råder det även likhet mellan A och C.

Jag har försökt att på semi-automatisk väg härleda ett antal artiklar för en tilltänkt finsk-dansk ordbok utgående ifrån de två föreliggande bilinguala databaserna. Med hjälp av några exempel vill jag illustrera tillvägagångssättet och de uppnådda resultaten. Jag vill visa möjligheterna och begränsningarna med denna procedur och i synnerhet utpeka vilka zoner i ordbokstexten som ägnar sig bäst till semi-automatisk länkning.

Det enklaste fallet är ett monosemiskt nederländskt ord som matchar ett monosemiskt danskt och ett monosemiskt finskt ord. I normalfallet torde det danska och det finska ordet också vara översättningsekvivalenter. I exemplet fungerar det nederländska LU *egel* som en *hub* och den finska och danska översättningen som *spokes*.



Mera invecklat blir det när vi har att göra med polysemiska uppslagsord på danska och finska. För det danska ordet *bølge* finns fem olika betydelser (LU) i databasen, som emellertid svarar mot två olika – i sin tur – polysemiska uppslagsord i finskan, *aalto* och *laine*.

**aalto** [101.] [I] <NOUN>

- [veden pinnan kohouma] *bolge a. ~~(500) aallot b. ~~(500) korkea aalto c. korkeat aallot høje bolger d. ~~(500) raivoavat aallot e. keinua aalloilla ride på bølgerne f. vaipua aaltoihin forsvinde i bølgerne~~~~~~*
- [värähdysliike] (fys) *bolge a. pitkät aallot langbolge b. lyhyet aallot kortbolge c. lähettää lyhyillä aalloilla NIHIL*
- [kuv: aallon kaltainen ilmiö] *bolge a. uusi aalto NIHIL b. vihreä aalto grøn bolge c. kuumat aallot hedetur*

**laine** [48.] <NOUN>

- [aalto] *bolge a. valaa öljyä laineille gyde olie på vandene*
- [aaltoa muistuttava kuvio] *bolge a. ~~(500) jonkun hiukset ovat laineilla b. (500) kam-mata laineita hiuksiin c. muotoilla hiukset laineille lave bolger i håret~~*

Det mest komplicerade fallet är då de olika LU tillhörande ett polysemiskt finskt uppslagsord är länkade till LU som tillhör olika (polysemiska) danska uppslagsord. I dessa fall är distansen mellan de två berörda språkens lexikaliska struktur som störst<sup>12</sup>.

**kello** [101.] <NOUN>

- [aikaa mittaava laite] *ur a. katsoa kelloa se på uret b. kello lyö yksitoista klokken slår elleve c. kello edistää/on edellä uret er foran d. kello jätättää/on jäljessä uret er bagud e. kelloni jätättää/edistää kaksi minuuttia mit ur er to minutter bagefter/foran f. kello seisoo uret er gået i stå g. ~~(500) kello on pysähtynyt h. vetää kello trække uret op i. siirtää kelloa eteenpäin NIHIL j. asettaa kello oikeaan aikaan stille uret k. ~~(500) asettaa kellot samaan aikaan l. ~~(500) kellon koneisto~~~~~~*
- [ajankohta] *time a. kysyä kelloa NIHIL b. paljonko/mitä kello on? Hvad er klokken? c. ~~(500) voitteko sanoa, paljonko kello on? d. ~~(500) kello on yksi e. kello on viisi klokken er fem f. kello viideltä klokken fem g. tasan kello kaksitoista NIHIL h. nukkua kellon ympäri sove i tolv timer~~~~*
- [soittolaite] *klokke a. soittaa (kirkon)kelloja ringe med (kirke)klokkerne b. kellot soivat NIHIL c. soittaa (ovi)kelloa ringe på d. ~~(500) soittaa kelloa stemme dorklokker som leg e. kello soi det ringer på døren f. ~~(500) hänelle tuli toinen ääni kelloon g. ~~(500) sitten hänelle tuleekin toinen ääni kelloon h. nyt on toinen ääni kellossa nu får piben en anden lyd i. ei vahinko tule kello kaulassa en ulykke kan ske hurtigt, det kan pludseligt gå galt~~~~~~*

Ovanstående exempel, som är genererade på semi-automatiskt vis från den integrerade trespråkiga databasen visar att en hel del översättningar faller på plats

<sup>12</sup> Ännu mera extremt är naturligtvis de fall där det inte föreligger någon översättningsekivalent alls på grund av att ett visst betydelseinnehåll inte är lexikaliserat i ett av de båda språken. I så fall finns det helt enkelt ingen länk, utan uppslagsordets betydelse på språk A återges med en parafraaserande beskrivning på språk B.

med en gång och kan återanvändas i en finsk-dansk ordbok. Detta gäller både *lexikaliska enheter* (LU) och *exempel* (EU). De fall där systemet inte genererar någon översättningsekivalent alls markeras med *NIHIL*. Det implicerar att den nederländska motsvarigheten i systemet inte fungerar som en *hub*, i och med att den endast är länkad till det finska materialet och inte till det danska. Uppslagsordet eller exempelmeningen härstammar i sådana fall alltså inte från RBN. Detta kan bero på att det är en nederländsk översättning av ett finskt uppslagsord eller en finsk exempelmening som antingen har tillfogats av de finska redaktörerna eller som har genomgått drastiska förändringar. De finska exempel som har försetts med koden <500> däremot har sällats ut som uppslagsord eller exempelmeningar i den finsk-nederländska delen och måste följaktligen avlägnas från den tilltänkta finsk-danska ordboken. Det är finskt material som befinner sig i databasen enbart som översättning av nederländska uppslagsord och exempel, men som inte kvalificerar sig som källspråkmaterial beroende på t.ex. låg frekvens, udda lexikalisering eller brist på kulturell specificitet. Länken är i så fall unidirektionell<sup>13</sup>. De finska ord(strängarna) visas här bara för att illustrera proceduren.

På grundval av ett antal experimentella körningar kan vi dra slutsatsen att det är möjligt att på ett semi-automatiskt sätt härleda en rå version av en finsk-dansk ordbok ur en trespråkig databas, där nederländska som *hub* är länkad till danska och till finska som *spokes*. På den utarbetade finsk-danska databasen kan självfallet senare vändfunktionen i OMBI tillämpas, så att vi även får en rå version av en dansk-finsk ordbok. Den stora vinsten är hur som helst att all relevant grammatisk information och alla stilistiska och pragmatiska specifikationer redan är befintliga för danska och finska uppslagsord. Den semi-automatiska länkningen genererar alltså även de befintliga annotationerna till uppslagsorden, de lexikaliska enheterna samt exemplen. Självfallet finns det en hel del efterredigering att utföra, men arbetsinsatsen för en ny ordbok reduceras ansevärt och återanvändningen av de befintliga resurserna ger en god utdelning.

---

<sup>13</sup> The editorial team opted for an integrated contrastive bi-directional database, from which both the Dutch-Finnish and Finnish-Dutch volumes are derived. To this end all translation links in the database have been marked according to their directionality status: *bi-directional*, if the translation is univocal, adequate and relevant in both directions. In that case both the Dutch and the Finnish LUs and EUs are represented in both volumes either as a source or as a target item; *unidirectional Du→Fi*, if the Finnish item only figures as a translation unit, but does not qualify as a LU or a EU in the Finnish-Dutch volume; *unidirectional Fi→Du*, if the Dutch item is represented only as a translation unit in the Finnish-Dutch volume and must be prohibited from appearing in the Dutch-Finnish direction. (Laureys 2007: 301.)



Ingen eller endast marginell efterredigering krävs när det gäller monosemiska ord, lexikaliserade flerordsfraser och idiom samt enkla icke-kontextuella exempelfraser. Tillsammans utgör dessa fall i själva verket drygt 70% av ordbokstexten. Betydligt mera efterredigering behövs när det gäller polysemiska substantiv och verb, de allra flesta adjektiv och överlag när det gäller korpusexempel. Det ska härvid beaktas att översatta exempel inte är genuint korpusbaserade och att det föreligger en viss risk för bias mot hub-språket. Dessutom ska den kontrastiva relevansen prövas för varje nytt språkpar. När det gäller funktionsord, högfrekventa substantiv och verb, som oftast förevisar en hög grad av polysemi, duger däremot resultaten av den semi-automatiska länkningen knappast till en direkt bearbetning.

## 5. Integrerad lexikalisk databas eller ”hyperlexikon”

Eftersom det både är dyrt och tidskrävande att sammanställa lexikaliska databaser, är det önskvärt att försöka reducera kostnaderna och att försöka förlänga ”livstiden”, dvs. *the usable life time* (Banski & Przepiorkowski 2011:34 och Simons & Bird 2008:16) för dessa resurser. “To be sustainable, the results of our work must transcend computer environments, communities of practice, domains of application, and especially the passage of time.” (Bird & Simons, 2008:4). Bird & Simons (2004:4) anger sex nödvändiga och tillräckliga villkor för en hållbar användning av språkliga resurser: ”the resource must be extant, discoverable, available, interpretable, portable and relevant”. Det läggs stora krav dels på tillgängligheten och den formella aktualiseringen av språkliga resurser och dels på den innehållsmässiga relevansen, vilket kräver en ständig ajourföring av resurserna.

Avslutningsvis fokuserar jag på de nya utmaningar som ligger i strävan efter hållbarhet inom lexikografin. En integrerad lexikalisk databas är den moderna lexikografins svar på hållbarhetskravet. En integrerad databas förutsätter att man optimerar arbetet och utför det så vitt möjligt ”en gång för alla”. För att åstadkomma en integrerad databas får man tänka långsiktigt. Databasen skall kunna ligga till grund för många olika ordböcker och lexikografiska verktyg. Den skall vara robust och återanvändbar. Därtill skall designen vara öppen och dynamisk och samtidigt skall den konformera sig till fasta och kontrollerbara standarder, som beskrivs explicit i metadokument och som ständigt kan valideras. Man använder sig av kalibrerade verktyg enligt internationella normer. Etiketterna är entydiga och standardiserade, så att de alltid svarar mot en viss infor-

mationstyp och tvärtom<sup>14</sup>. Databasstrukturen är enkel och entydig. Layoutmässigt har den integrerade basen en flat struktur, så att layout och formgivning kan bestämmas per produkt eller produktserie. Input i strukturen är medieneutral. Alla produkter framställs med hjälp av en transparent redaktionsmanual, så att produktionsprocessen snabbas upp och görs mera effektiv.

Databasen företer en modulär mjukvarudesign, vilket innebär att den rymmer möjligheten att en framtida redaktör skall kunna selektera relevanta kategorier på alla nivåer inom ramen för ett konkret projekt, utan att hierarkin eller den överordnade strukturen förstörs. Databasen består m.a.o. av olika deldatabaser, som kan definieras och härledas på beställning. Samtidigt finns det länkar till andra databaser och deldatabaserna emellan. Denna s.k. interoperabilitet torde vara en markant fördel med en integrerad databas.

En ordbok utvecklas inte längre som ett isolerat projekt, utan som en del av ett samlat övergripande projekt. En enskild ordbok blir i så fall ett derivat ur en annoterad lexikalisk databas. Det innebär att varje redaktionsgrupp slipper göra en hel del arbete, som redan är gjort av andra. Feedback från användarna, kan, om det är relevant, generaliseras och förbättringar eller uppdateringar sker på en gång för alla derivat.

Särskild uppmärksamhet kräver procedurerna för förvaltningen av materialet. För att databaserna skall kunna användas inom ramen för nya projekt måste de inte bara vara lättillgängliga, utan även beskrivas på ett transparent sätt i relevanta metadokument.<sup>15</sup> Kontinuerlig uppdatering av databaserna och säkring av tillgängligheten av data med nya versioner av relevant software är en förutsättning för en permanent och stabil tillgång i snabbt förändrande omgivningar.

Stora insatser med avseende på infrastrukturella resurser och databasarkitektur har gjorts inom ramen för det europeiska projektet CLARIN<sup>16</sup> och inom

<sup>14</sup> Ett viktigt begrepp i detta samband är *persistent identifier*. Med *persistent identifier* avses en permanent, unik och entydig etikett som identifierar en digital resurs (text, audio, video, databas etc) helt oberoende av dess fysiska befintlighet eller placering. En *persistent identifier* garanterar att resursen alltid kan hittas och konsulteras via nätet.

<sup>15</sup> Det läggs mycket vikt på kodning, långsiktig digital arkivering och på tillgänglighetskrav. Viktiga initiativ har tagits inom ramen av ELSNET (European Network of Excellence in Speech and Language Technology).

<sup>16</sup> "CLARIN (Common Language Resources and Technologies Infrastructure) is committed to establishing an integrated and interoperable research infrastructure and technology for language resources and its technology. It aims at removing the current fragmentation, to offer a stable, persistent, accessible and extendable infrastructure and therefore enabling e-Humanities". (Broeder et al. 2008: 1457.)

DAM-LRprojektet<sup>17</sup> och mer nyligen även inom OLAC (Open Language Archives Community)<sup>18</sup>.

Vi är på väg mot en mer integrerad, en mera användarvänlig, en mera funktionell, effektiv, flexibel och hållbar lexikografi. Perspektivet är långsiktigt, vägen är utstakad, nödvändigheten påvislig, men det finns mycket arbete kvar. Lexikografen i Norden står vid en skiljeväg. Med den uppbyggda erfarenheten och expertisen, med viljan till ett öppet samarbete har vi alla möjligheter att bidra på ett positivt sätt till denna utveckling. Av flera sektionsföredrag under konferensen har det framgått att utvecklingen av webbordböcker och hyperlexikon är i full gång i alla de nordiska länderna.

## LITTERATUR

- Atkins, Sue & Fillmore, Charles J. et al., 1994 : The Dictionary of the Future: a Hypertext Database. Presentation and on-line demonstration at the Xerox-Acquilex Symposium on the Dictionary of the Future. Uriage.
- Atkins, Sue, 1996: Bilingual Dictionaries: Past, Present and Future. I: Euralex'96 Proceedings, (eds.) Gellerstam, M. , J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C. R. Pappmehl. Gothenburg, S. 515-590.
- Bergenholtz, Henning & Malmgren, Sven-Göran (red), 1999: Förord. I: Lexico Nordica 6, S.1-3.
- Broeder, Daan, Nathan, David, Strömquist, Sven & van Veenendaal, Remco, 2008: Building a federation of Language Resource Repositories; the DAM-LR project and its continuation within CLARIN. [www.lrec-conf.org/proceedings/lrec2008/pdf/370\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/370_paper.pdf), S.1457-1461.
- Brundtland, Gro Harlem, 1987: Vår felles framtid. Oslo (även som: Our Common Future: The World Commission on Environment and Development. Oxford.
- De Schryver, Gilles-Maurice & Prinsloo, D. J., 2001 : Fuzzy SF: Towards the ultimate customised dictionary. I: Studies in Lexicography 11.1, S. 97–111.
- Hannesdóttir, Anna Helga, 1998: Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga. Meijerbergs arkiv för svensk ordforskning 23. Göteborg.
- Hovmark, Henrik, 2005: Tre franske miniordbøger til rejsebrug og daglig tale. I: Lexico Nordica 12, S.13-33.

<sup>17</sup> The DAM-LR project aimed at investigating the needs and laying the basis for establishing a federation of language archives that can offer their users a shared virtual space of resource location facilities and resource access. (Broeder et al. 2008: 1457.)

<sup>18</sup> OLAC står för Open Language Archives Community. "OLAC is an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by: (i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resources." (Simons & Bird 2008:8.)

- Kielitoimiston sanakirja, 2006. (Uusittu ja laajennettu versio Suomen kielen perussanakirjasta). Kotimaisten kielten tutkimuskeskuksen julkaisuja 140. Helsinki.
- Laureys, Godelieve, 2007: Optimizing Procedures for the Making of Bilingual Dictionaries and the Concept of Linking Contrastive Lexical Databases. I: *International Journal of Lexicography* 20. 3, S. 295-311.
- Laureys, Godelieve & De Groot, Hans, 1996: *Van Dale Handwoordenboek Zweeds-Nederland /Nederlands-Zweeds*. Utrecht (även som: *Norstedts nederländsk-svensk/svensk-nederländska ordbok*. Stockholm).
- Laureys, Godelieve et al., 2004: *Gyldendals Hollandsk-Dansk ordbog*. Köpenhamn.
- Laureys, Godelieve & Moisis, Maritta, 2012: *Prisma Groot Woordenboek Nederlands-Fins en Fins-Nederlands/ Hollanti-Suomi ja Suomi-Hollanti Suursankirja*. Houten – Antwerpen.
- Martin, Willy, 2001: Linkability, the Hub-and-Spoke Model and the (semi-)automatic derivation of a German-French Dictionary. I: H. E. Wiegand (ed.). *Studien zur zweisprachigen Lexikographie VI, Germanistische Linguistik* 16. Hildesheim: Olms, S. 67-91.
- Martin, Willy, 2003: Lexicography, Lexicology, Linking and the Hub-and-Spoke Model. I: W. Botha (ed.). *Festschrift D. Van Schalkwijk*. Stellenbosch, S. 237-249.
- Martin, Willy, 2004: SIMuLLDA, the Hub-and-Spoke Model and Frames or How to Make the Best of Three Worlds? I: *International Journal of Lexicography*, 17 (2): 175-187.
- Nathan, David, 2006: Proficient, permanent, or pertinent: aiming for sustainability. I: Barwick, L. and Thieberger, N., (eds.). *Sustainable Data from Digital Fieldwork*. The University of Sydney, S. 57-68.
- Simons, Garry F., Bird, Steven, 2008: Toward a global infrastructure for the sustainability of language resources. *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation. PACLIC 22*. S.87-100.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm.

*Godelieve Laureys*

Institutionen för nordiska språk vid universitetet Gent  
 godelieve.laureys@ugent.be

# Biskopar, ordböcker och verkligheten

## – Vad är lexikografihistoria bra för?

*Bo Ralph*

På EURALEX-konferensen i Leeuwarden sommaren 2010 höll Patrick Hanks ett sektionsföredrag med lexikografihistoriskt ämne. Han berättade med ohöljd förtjusning om vilka fynd man kunde göra i gamla ordböcker. Man fick ett intryck av att han för egen del hade gjort en stor upptäckt och att han nu ville dela med sig av den.

Möjligen är det så att lexikografihistorien internationellt verkligen spelat en undanskymd roll. För ett decennium sedan skrev Reinhard R.K. Hartmann en bok med titeln *Teaching and Researching Lexicography* (2001). I den finns ett litet kapitel med rubriken ”Historical and critical perspectives”, där Hartmann inledningsvis (2001: 39) citerar sig själv från ett tidigare verk:

*Most dictionaries have forerunners, and all have imitators: an understanding of the historical foundations of dictionary-making is therefore one of the preconditions of further progress in academic lexicography. [Emfas/kursiv tillagd här.]*

Som synes tilldelar Hartmann här lexikografihistorien en mycket central roll. Det kan väl därför finnas goda argument för att ta upp lexikografihistoria på en nordisk lexikografikonferens. Nordiska lexikografer är dock inte lika yrvakna som en del andra tycks vara inför vad lexikografihistorien kan erbjuda. Det är ganska många i Norden som länge har intresserat sig för den historiska aspekten på lexikografi eller åtminstone har varit inne på ämnet någon gång. Exempelvis är det naturligt att man intar en historisk hållning även på metaplanet när man arbetar med en historisk ordbok, såsom redaktionen på Svenska Akademiens ordbok gör, och som man även gör på Gammeldansk Ordbog osv. Gamla ordböcker är välkända för dessa redaktörer redan därför att de konsulteras regelbundet i det dagliga arbetet.

Hartmann, som nämndes alldeles nyss, definierar det övergripande syftet med lexikografihistoria på följande sätt (2001: 39; kursiv tillagd):

The historical study of dictionaries [...] is an occupation pursued by those who want to find out *how these products and their creators have contributed to cultural development, social progress and scientific knowledge.*

Troligen stämmer detta på ett abstrakt plan, men man kan säkert ha olika skäl till sitt intresse för lexikografihistoria; tyngdpunkten kan under alla omständigheter variera.

## Perspektiv på lexikografihistoria

Låt oss dröja något ytterligare vid Hartmann, som ger ett allmänt schema för hur den lexikografihistoriska forskningen enligt hans mening skall gå till (2001: 40; fetstil i originalet):

- Dictionary history [...] can be pursued along (at least) seven different perspectives [...]:
- (1) the **universal–anthropological approach** is aimed at finding out how reference technologies have developed around the world throughout human history [...];
  - (2) the **diachronic–linguistic approach** is aimed at finding out how languages have changed over time and how these periods have been covered in reference works [...];
  - (3) the **cultural–historical approach** is aimed at finding out how lexicographic principles have been developed within various intellectual or national traditions [...];
  - (4) the **genre–specific historical approach** is aimed at finding out how different types of dictionaries have evolved for different purposes [...];
  - (5) the **genealogical approach** is aimed at finding out how different reference works are ‘family-related’ [...];
  - (6) the **personal–biographical approach** is aimed at finding out how individual dictionary compilers have coped with their tasks and what innovations they have brought to lexicography [...];
  - (7) the **historiographic approach** is aimed at finding out how (meta)lexicographic theories have advanced over time [...].

Av dessa sju aspekter ger Hartmann egentligen bara exempel på vad som åstadkommit med hänsyn till (2)–(5). Under punkten (1) studeras hur mänskligheten har förmedlat information genom tiderna. Med (2) menar Hartmann dokumentationen av de lexikaliska förändringarna i ett språk från ett stadium till ett annat i den diakroniska utvecklingen. Med (3) avses hur utövande lexikografers arbete och produkter relaterar till tidens kulturella kontext, så att författarna/kompilatorerna/redaktörerna och deras ordböcker sätts in i en kulturell tradition. Med (4) fokuseras den historiska utveckling som en särskild typ av ordböcker genomgått. Med (5) avses beroendeförhållandena mellan olika ordböcker (”familjerelationer”). Punkt (6) tar sikte på individuella lexikografer och

deras egen historia. (7) syftar till en historisk beskrivning av metalexikografi. Det är uppenbart att denna sista aspekt i sin tur kan indelas i flera olika underavdelningar.

Hartmann tycks mena att föga har utträttats på de nämnda områdena över huvud taget, mindre på vissa än på andra. Det kan här noteras dels att han i första hand koncentrerar sig på den engelskspråkiga lexikografin, dels att situationen även där har ändrats under senare år (jfr Cowie 2009). Om vi begränsar oss till Norden kan vi konstatera att det åstadkommit en hel del på det lexikografihistoriska området generellt, men det är uppenbart att vi inte nått ut internationellt tillräckligt effektivt.

Därför kan Hartmanns indelning i det följande tas till utgångspunkt för en genomgång av några olika aspekter på lexikografihistoria. Till stor del kommer synpunkterna att baseras på erfarenheter som inhämtats inom ett ganska omfattande lexikografihistoriskt projekt, som jag tillsammans med Sven-Göran Malmgren var ansvarig för och flera andra deltog i under en följd av år under 1900-talets slut, med början i mitten av 1980-talet, och som vi fortfarande håller på att dra slutsatser av när tillfälle ges. Detta är ett sådant tillfälle. Av nödvändighet kommer konkreta exempel att hämtas från svensk lexikografihistoria, eftersom det är där vi har vår erfarenhet, men det mesta som följer är troligen överförbart även till övriga Norden. Hartmanns indelning bibehålls.

(1) The **universal–anthropological approach** (intellectual history).

Hartmanns uttolkning lyder (2001: 41): ”The first perspective studies the way humankind has transmitted information throughout the ages”. Han sammanfattar: ”This is definitely a desideratum for further research”, men detta lämnas här åt sidan eftersom inte Hartmann utvecklar saken vidare.

(2) The **diachronic-linguistic approach** (language history).

Detta är alltså den första punkt som Hartmann kommenterar någorlunda utförligt. ”The second perspective is associated with efforts to document the (mainly lexical) changes from one stage in the diachronic development of a language to another”, säger han (2001: 41). Vad som här avses är ordböcker som dokumenterar det lexikaliska systemet under ett givet språkstadium. Sådana verk är i den svenska traditionen exempelvis K.F. Söderwalls Ordbok öfver svenska medeltids-språket och Ordbok över svenska språket utgiven av Svenska Akademien (SAOB; från 1521). De båda verken är avsedda att komplettera varandra och tillsammans täcka det svenska ordförrådets hela historia.

För lagspråket hänvisar Söderwall i sin tur till Carl Johan Schlyters Ordbok till Samlingen af Sweriges Gamla Lagar, den stora svit av lageditioner som Schlyter gav ut mellan 1827 och 1869. Ordboken trycktes 1877, alltså ett halvsekel

efter det att utgivningen inletts. Av samma slag som Söderwalls ordbok är i andra nordiska länder exempelvis Gammeldansk Ordbog, Johan Fritzners Ordbog over Det gamle norske Sprog, på sitt sätt Sveinbjörn Egilssons *Lexicon poeticum antiquæ linguæ septentrionalibus m.fl.* Söderwalls ordbok har ganska nyligen beskrivits av Ulrika Djärv, i en doktorsavhandling som försvarades i Stockholm (Djärv 2009).

(3) The **cultural-historical approach** (cultural history).

Detta perspektiv ”relates the work and product of practising lexicographers to the cultural context of the time, thus placing the compilers and their dictionaries into a particular tradition” (2001: 42). Hartmann ger flera exempel på sådana ordböcker som gäller det brittiska språkområdet. Även denna aspekt tar Ulrika Djärv upp i sin avhandling om Söderwalls ordbok, och förutsättningarna kring SAOB har kartlagts i olika sammanhang. Att placera in olika ordböcker i den nationella traditionen var också ett bärande moment i det lexikografihistoriska projekt som nämndes nyss och som kallades just ”Lexikografisk tradition i Sverige”. Till denna fråga återkommer vi.

(4) The **genre-specific historical approach** (genre-specific history).

Hartmann fortsätter (2001: 43–44): ”The fourth perspective focuses on the historical development of a particular genre or type of dictionary” [...] ”One of the most interesting and complex reference works is the bilingual dictionary, as it involves two languages and their associated cultural environments and the codification of translation equivalents established between them, usually on the basis of language contact over relatively long stretches of time.” Här framhäver alltså Hartmann de tvåspråkiga ordböckernas betydelse i mycket uppskattande ordalag. Tvåspråkig lexikografi har inte alltid haft någon hög prestige, åtminstone inte i Sverige.

Man kan kanske härleda denna låga värdering ur 1800-talets syn på språkvetenskap, där ett historiskt synsätt var det enda som ansågs garantera vetenskaplighet. Det var därför arbetet med Svenska Akademiens ordbok under 1800-talet stagnerade efter att ha startat som ett typiskt 1700-talsprojekt, ett utslag av detta århundrades koncentration på stora uppslagsverk inriktade på samtidens förhållanden. Från början var Akademiens sikte mycket riktigt inställt på en stor synkronisk ordbok (även om denna formulering givetvis inte användes). När SAOB kom igång på allvar under 1890-talet var i stället det historiska perspektivet en självklarhet.

Man kan också lägga märke till hur Adolf Noreen karakteriserar tvåspråkiga ordböcker i sitt stora översiktsverk *Vårt språk* (1903: 49): ”praktiska hjälpredor, bland hvilka visserligen några mycket värdefulla, för öfversättning till främ-



mande språk”. Men lexikografi är över huvud taget inte någon särskilt uppskattad disciplin vid denna tid. På ett annat ställe säger Noreen om ordboken generellt (1903: 49; spärr i originalet):

En osystematisk och såtillvida ovetenskaplig uppställning af grammatikens innehåll, företagen i det ”praktiska” syftet att åstadkomma en lättskött uppslagsbok och därför till utgångspunkt tagande ordförrådet, ordnadt efter någon yttre synpunkt, vanligen den alfabetiska ordningsföljden hos de skrifna ordens begynnelsebokstäfver, kallas o r d b o k eller l e x i k o n.

Det är klart att lexikograferna inte har mycket att hämta när en sådan dominerande nationell auktoritet som Noreen uttrycker sig så förklenande om lexikografin.

Här kommer man för övrigt mycket nära den extrema syn på den lexikaliska delen av språket som en del allmänlingvister senare företrätt (t.ex. Bloomfields och i hans efterföljd Chomskys välkända karakteristik: ”a list of irregularities”). Orsaken till att lexikografi tidvis haft lågt anseende bland språkvetare anas bakom ett mera sentida uttalande: ”Lexicographers did not make a sincere effort to acquaint themselves with linguistic theories and apply them in their dictionaries” (Al-Kasimi 1977: 3).

Längre fram i sin översikt talar Hartmann om önskvärdheten av ”more attention to the interdependence of the monolingual and the bilingual dictionary traditions” (2001: 47). Om just detta skrev Anna Hannesdóttir redan 1998 en doktorsavhandling på mer än 500 sidor. Syftet framgår av titeln: Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga. Ett slående exempel på det förhållande som denna titel åsyftar är Anders Fredrik Dalins Ordbok öfver svenska språket (1850–1855), som delvis påverkats av samme Dalins tidigare utgivna tvåspråkiga ordböcker (framför allt fransk–svenska).

Lena Rogström specialstuderade samma år de första tvåspråkiga ordböckerna mellan svenska och engelska (såväl svensk–engelska som engelsk–svenska). Resultaten publicerades i doktorsavhandlingen Jacob Serenius lexikografiska insats (mer än 400 sidor). Monica Johansson tog sig an en del av den latin–svenska (och delvis den svensk–tyska och tysk–svenska) ordbokstraditionen i ännu en doktorsavhandling på över 400 sidor: *Lexicon Lincopense*. En studie i lexikografisk tradition och svenskt språk vid 1600-talets mitt.

För svenskan gäller således Hartmanns omdömen dåligt. Däremot har han rätt när han säger: ”Genres other than the general monolingual and bilingual dictionary (such as pedagogical dictionaries, thesauri and technical reference books) have not yet received much attention from lexicographic historians” (Hartmann 2001: 44).

(5) The **genealogical approach** (genealogical history).

”The filiational (or ‘family-tree’) relations between dictionaries, their mutual influences and derivations have intrigued many scholars” (Hartmann 2001: 44). En grundläggande fråga är: ”Are subsequent dictionaries in any one tradition just variations on a basic text?” En sådan fråga framstår som avgörande i de fall en ordbok används som exempelvis källa för språkhistoriska resonemang. Har vi flera oberoende vittnesmål om ett ords existens och egenskaper vid en viss tidpunkt, eller går alla uppgifter tillbaka på en enda – kanske mindre pålitlig – källa?

Denna typ av undersökningar upptog ganska mycket utrymme i de tre nämnda avhandlingarna och projektet som helhet (se t.ex. Hannesdóttir & Ralph 1989 och Ralph 1993). Beroendet av andra ordböcker gör sig gällande inom den nationella traditionen men återspeglar också påverkan utifrån. Det är en vanlig situation att en berömd ordbok av främmande ursprung har bildat mönster för de första svenska ordböckerna i sitt slag (inom såväl den latin-svenska som den tvåspråkiga och enspråkiga traditionen). Exempelvis har Adelungs tyska ordbok (1774–1786) efterliknats av många. Abel Boyers engelsk-franska (1702) har använts av Jacob Serenius vid utarbetandet av hans engelsk-svenska och svensk-engelska ordböcker. Längre fram kan uppburna svenska ordböcker av något slag ha varit förlagor för senare svenska ordböcker, oavsett språkkombinationer. Så uppges Abraham Sahlstedts svenska ordbok (1773; med latin som metaspråk) ha legat till grund för en stor senare fransk-svensk ordbok (Holmberg 1795; se titelsidan). Franska Akademiens enspråkiga ordbok har konsulterats flitigt av Anders Fredrik Dalin i hans fransk-svenska ordbok (Dalin 1843); detta arbete har i sin tur utnyttjats i Dalins senare enspråkiga svenska ordbok (1850–55). Så kan drag från olika nationella traditioner ha spritts över språkgränserna.

(6) The **personal-biographical approach** (personal history).

Hartmann ger flera exempel på biografier över enskilda lexikografer och deras arbete. På samma sätt lämnar de nyss nämnda avhandlingarna delvis fylliga uppgifter om svenska lexikografer och deras gärning. Så kan Lena Rogströms avhandling uppfattas som en monografi över Jacob Serenius och hans lexikografiska insats; Monica Johansson beskriver Nicolaus Grubbs och Jonas Petri Gotthus arbete med *Lexicon Lincopense* mot en biografisk bakgrund; Anna Hannesdóttir tar upp ett dussintal lexikografiska författarskap till ganska utförlig behandling. Liknande uppgifter finner man om Jesper Swedberg i Lars Holms svit av arbeten över dennes lexikografiska insatser (senast Holm 2009), och om Samuel Schultze hos Sture Hast (1990). Samma sak gäller Ulrika Djärvs avhandling om Söderwall och hans ordbok.

(7) The **historiographic approach** (metalexigraphic history).

”The historiography of lexicography is still in its infancy, and a historical account of metalexigraphy has hardly been attempted”, säger Hartmann (2001: 46). Detta gäller alltså inte längre för det engelska språkområdet och inte heller för Norden, i det senare fallet inte ens när Hartmann skrev så. Bland annat tar de nämnda avhandlingarna särskilt sikte på den lexikografiska strukturen i de ordböcker som undersöks, och varje ordboks plats inom traditionen definieras med hjälp av sådana analyser. Femte delen av serien Handbücher zur Sprach- und Kommunikationswissenschaft (HSK 5, 1989–1991) handlar delvis också om lexikografihistoria. Lars Holm och Hans Jonsson skrev det svenska partiet (HSK 5: 1933–1943). Härtill skulle kunna läggas en lång rad specialstudier som tillkom inom ramen för det nämnda projektet, de flesta publicerade som tidskriftsartiklar eller konferensbidrag i olika sammanhang.

Andra viktiga lexikografihistoriska bidrag är exempelvis de redan antydda Noreen (1903), Holm (1986 etc.), Hast (1990), vidare Larsson (2003, jämte flera senare arbeten), Djärv (2009) m.fl. I Danmark har exempelvis Simon Skovgaard Boeck belyst lexikografihistorien i flera studier (framför allt Boeck 2009). Äldre ordböcker har återutgivits i både Danmark och Island. Isländsk lexikografihistoria har behandlats i flera sammanhang av bland andra Guðrún Kvaran och Ingólfur Gunnlaugsson (t.ex. i förordet till respektive utgåva). Som redan noterats är det således inte lika befogat att klaga över bristen på lexikografihistoriskt intresse i Norden som på många andra håll.

## Lexikografihistoriens funktioner

Som framgår av Hartmanns indelning kan lexikografihistoria aktualiseras i skiftande sammanhang. De olika infallsvinklar man kan välja får växlande betydelse beroende på det syfte man har med att anlägga ett visst lexikografihistoriskt perspektiv. Från rent språkvetenskaplig utgångspunkt är det naturligt att närma sig lexikografihistoria från två håll: självklart från den *lexikografiska* sidan, men också från den *språkhistoriska*. Den språkvetare som likt Adolf Noreen tycker att lexikografi är en språkvetenskapligt trivial verksamhet är kanske inte särskilt imponerad av lexikografihistoria heller, men helt annorlunda blir det för den som har ett *inomlexikografiskt* perspektiv.

Ur rent lexikografisk synvinkel kan man nämligen se beskrivningen av lexikografihistorien som en del av disciplinens akademisering. För bara ett tjugotal år sedan talade en nordisk auktoritet som Dag Gundersen om att lexikografen (då) hade börjat akademiseras. Detta var i samband med att Nordiska föreningen för lexikografi bildades. Med det menade Gundersen att lexikografi bör-

jade bli en självständig, erkänd universitetsverksamhet. För detta krävs naturligtvis en aktiv teori- och metoddiskussion, så att verksamheten står på fast vetenskaplig grund, men dit hör också en förståelse av disciplinens framväxt och dess plats inom ett större akademiskt och allmänt socialt sammanhang.

Vad lexikografihistorien kan bidra med i dessa avseenden är i första hand en allmän beskrivning av verksamhetens historiska utveckling: vilka ordböcker och lexikografer som funnits och betytt något särskilt, en karakteristik (typologisering, klassificering) av existerande lexikografiska arbeten, en värdering av såväl ordböckerna som lexikografernas insatser samt deras respektive betydelse för den fortsatta utvecklingen, ömsesidiga beroendeförhållanden osv. Med andra ord: det som Hartmann tar upp som punkt (3), (4), (5) och (6), men kanske delvis på ett annat sätt än han tänker sig.

En möjlighet som inte skall underskattas är den, att nutida lexikografer i den äldre lexikografin kan finna uppslag som kan leda till förbättringar. Det är lätt att vi förblindas av vår egen överlägsna tekniska standard, så att skillnaderna mellan vår tids och en äldre tids lexikografi överbetonas. Förhoppningsvis skall lexikografer ha lärt sig något under några hundra år, så att vi nu verkligen är skickligare än pionjärerna, även bortsett från tekniken, men det hindrar inte att en och annan briljant idé från det förflutna kan ha kommit bort på vägen fram genom historien. Anna Hannesdóttir (2000a) ger exempel på sådana bortglömda grepp. På så vis kan lexikografihistorien ge *inspiration* till nutida lexikografi.

Inte minst visar lexikografihistorien hur traditioner växer fram och etableras. Man ifrågasätter inte oavbrutet hur en ordbok skall utformas, utan när man verkar inom en tradition tenderar man i stället att göra som man blivit lärd. På så sätt uppnås konstans och stabilitet, men det finns ingen garanti för att nytänkande odlas.

Det har exempelvis aldrig utarbetats någon stort upplagd ordbok som direkt korresponderar med en grammatik. När Gustaf III inrättade Svenska Akademien år 1786 skrev han själv dess stadgar. ”Thy åligger äfven Academien at utarbete en Svensk ordabok och Grammatica”, står det på framskjuten plats (Svenska Akademiens stadgar § 23). Men Akademien dröjde med att lyda order. Om man bortser från en mindre framgångsrik språklära publicerad 1836 utkom grammatiken först 1999, ordboken är ännu inte avslutad. Och de är framför allt inte anpassade till varandra. Däremot skriver Abraham Sahlstedt i sin ordbok (1773, förordet: 9):

Grammatikans och Ordbokens nära förening lär jag icke behöfwa widlöftigt omtala. Hwar och en wet, at det är icke nog, at lära et ords bemärkelse, utan at af Grammatikan inhämta des slag, böjelse och öfriga bruk. Läsaren finner altså i Ordboken alt hwad af den Grammatiska delen i Språket dit hörer.

Sahlstedt levde upp till sina egna föresatser och förverkligade det som Svenska Akademien ännu idag inte har åstadkommit: en ordbok och en grammatik som korresponderar med varandra. Ordboken (Sahlstedt 1773) och grammatiken (Sahlstedt 1769) förelåg tryckfärdiga ungefär samtidigt, men tryckningen av ordboken drog ut på tiden.

Man kan fråga sig var Sahlstedt fick sin idé ifrån. Tidens stora internationella lexikografer, engelsmannen Samuel Johnson och tysken Johann Christoph Adelung, har visserligen båda i sina ordböcker (Johnson 1755 och Adelung 1774–1786) inkluderat en grammatisk inledning, men Johnsons är bara rudimentär och Adelungs är senare än Sahlstedts. Vi får kanske räkna med att Sahlstedt hade en egen genial tanke.

Lexikografihistoria har också en roll att spela *utanför själva lexikografien*. Hartmann betraktar i sin första punkt lexikografien som en aspekt av människans tekniker för att överföra information, i detta fall om ordförrådet och dess egenskaper. Om man anser att lexikografi är en form av språkvetenskap är lexikografihistoria av definition en del av språkhistoriografien. Om man antar att språkhistoria är något som är värt att sysselsätta sig med, kan lexikografihistorien rimligen bidra till den språkhistoriska beskrivningen. Men förhållandet mellan lexikografihistoria och språkhistoria är något mer komplicerat än man kanske först tänker på.

Äldre ordböcker kan vara *symptom på språkets förändring*, dvs. de kan i sin beskrivning avspegla samtida språkförhållanden. Genom en jämförelse mellan ordböcker från olika skeden kan en förändringsprofil i idealfallet tecknas. Det som ligger närmast till hands att tänka på är då att ordböckerna ger besked om språkssystemets egenskaper i äldre tid, genom att de var och en beskriver språkliga förhållanden i sin samtid, som nu är förgången tid. Det krävs i normalfallet att språkhistorikern analyserar olika situationer och drar språkhistoriskt relevanta slutsatser. Men då och då kan man också mera direkt hämta information som i sig är av diakronisk natur. Så är det när ordböcker lämnar kommentarer om att vissa bruk är ålderdomliga, eller att man jämför med äldre, ”nu” övergivna språkliga strukturer osv.

Ordböcker kan emellertid också ha spelat en rent aktiv roll i språkhistorien. Under äldre perioder har de varit viktiga inslag i en pågående *standardiseringsprocess*. En sådan pågår ju alltid, mer eller mindre märkbart, men för svenskans del var den som intensivast före 1900-talet. I den viktiga språkliga standardisering som 1700-talet på många sätt medförde var ordböckerna mycket centrala, vilket också framgår av den drastiskt ökande tryckverksamheten under detta århundrade.

Einar Haugen har lanserat en modell för språkplanering, som också kan fungera väl som underlag för en beskrivning av ett språks standardiseringspro-

cess. Detta har Anna Hannesdóttir påpekat redan i sin avhandling (1998: 11 ff.), som nämndes nyss (vidareutvecklat i bl.a. Hannesdóttir 2000b). Haugens modell ser ut såsom bild 1 visar (jfr Haugen 1987: 64).

	Form (policy planning)	Function (cultivation)
Society (Status planning)	1) Selection (decision procedures)	3) Implementation (educational spread)
	a) Identifications of problem	a) Corrections procedures
	b) Allocation of norms	b) Evaluation
Language (Corpus planning)	2) Codification (standardization procedures)	4) Elaboration (functional development)
	a) Graphization	a) Terminological modernization
	b) Grammatication	b) Stylistic development
	c) Lexication	

Bild 1. *Faser i språkplaneringsprocessen (enligt Haugen 1987).*

Som synes anses etableringen av ett gemensamt ordförråd utgöra ett särskilt moment i en språkgemenskaps konsolidering (*lexication*). Ordböcker ger dock vanligen besked om mer än själva orduppsättningen. Genom medvetna val av ordens uppslagsformer, böjning, konstruktionssätt och annat påverkar lexikograferna i äldre tid standardens utseende på ett mycket påtagligt sätt. Olika lexikografer träffar länge olika val, beroende på ideal, egen ålder och personlig bakgrund, olika erfarenheter osv., men efterhand börjar uppgifterna konvergera.

Det är svårt att tänka sig en effektiv standardisering av ett skriftlöst språk, men i det ögonblick skrift införs startar också standardiseringen. Denna går fram på olika nivåer, från fastställande av en *gemensam teckenuppsättning* och *fast alfabetisk ordning*, över ett homogeniserat böjnings- och ordfogningssystem, alltså en *grammatik*, till mer eller mindre bjudande föreställningar om olika *text-typers* status i olika sammanhang och hur text skall utformas grafiskt i tryck. Idag tar ordböcker sällan upp alla dessa frågor explicit, men så sker ofta i äldre tid. Ordböckerna har därför även påverkat den grammatiska utvecklingen, särskilt i standardspråket. Under alla omständigheter fungerar vissa ordböckers specifika lösningar som mönster för andra språkbrukare och för andra lexikografer. De blir en del av lexikografitraditionen.

Äldre ordböcker kan således tjäna som *källa* till olika slags kunskap, antingen för att de ger direkt besked genom *explicita* omnämmanden eller för att man *implicit* kan utläsa en hel del om tidens språk ur de uppgifter som lämnas.

Det språkhistoriska källäget är dramatiskt olika om det finns samtida ordböcker från den tid man undersöker eller ej, även om dessa skulle vara lexikografiskt bristfälliga och kräva försiktigt handhavande.

Till det som lexikografihistorien kan belysa utanför det rent lexikografiska området hör också *språkpedagogiken* och dess historia. Man kan i grova drag dela in den svenska lexikografihistorien i följande perioder (jfr Ralph 2000: 19):

- (1) latintraditionen (med syfte att träna svenskar i latin) fram till ca 1700;
- (2) tvåspråkig lexikografi (med andra syften än att lära svenskar latin) från ca 1700;
- (3) enspråkig svensk lexikografi från ca 1850;
- (4) lexikografi i nya medier från ca 2000.

Eftersom hela den latinska perioden syftar till inläring av latin är huvudsyftet här genomgående pedagogiskt. Även de tvåspråkiga ordböckerna har först och främst ett inläringssyfte, vid sidan av syftet att vara en praktisk hjälp vid översättning. Att de kommit att fungera som viktiga standardiseringsinstrument är en sekundär effekt. I den mån enspråkiga ordböcker är till för att befästa normen kan även dessa sägas ha ett pedagogiskt syfte.

I praktiken framstår emellertid de olika ordböckernas pedagogiska syfte olika starkt. Det kommenteras endast sällan uttryckligen, då vanligen i förorden (men observera i synnerhet Lindfors 1815, förordet). Om man inte är medveten om pedagogikens tidigare framställningsformer kan man till och med missa poängen fullständigt. Så är hela kategorin språkexempel ett utslag av den klassiska retorikens betoning av språkliga och moraliska exempel (Ralph 2000: 13). Språkproven i de äldsta, av latintraditionen beroende ordböckerna hämtas från de klassiska auktorerna med hänsyn till deras ändamålsenlighet i de angivna avseendena – de formuleras inte på grundval av författarens/kompilatorns/redaktörens egen språkkänsla.

Ordböcker har genom alla tider varit viktiga pedagogiska hjälpmedel. I vilken utsträckning och på vad sätt de använts är dåligt kartlagt. I vilken mån det pedagogiska syftet varit avsiktligt och kanske för upphovsmannen det primära har egentligen aldrig utforskats systematiskt.

Vi har sett hur lexikografihistoria kan användas i olika sammanhang. Att den kan sprida ljus över rådande förhållanden borde vara självklart. Om dagens lexikografer undviker att inta en överlägsen attityd gentemot äldre tiders kolleger kan man kanske inte bara få ett välgörande perspektiv på nutida lösningar utan också rentav inspiration till förbättringar. Vidare kan lexikografihistorien belysa språkhistorien i allmänhet, framför allt standardiseringsprocessen. Denna har tilldragit sig förvånansvärt liten uppmärksamhet i svensk språkvetenskap,

såsom Ulf Teleman påpekat (2002, förordets inledning). (I Norge har det varit helt annorlunda.) Att språkpedagogiken har åtskilligt att hämta i lexicografihistorien är uppenbart. Det ger också en fingervisning om att nya sidoeffekter kan väntas om det lexicografihistoriska studiet förs vidare och intensifieras.

## Den icke-språkliga kontexten

Hartmann tangerar på flera ställen ett problem utan att riktigt tydligt formulera det, långt mindre att ange sätt att angripa det. Närmast kommer han då han flyktigt beskriver sin ”cultural-historical approach”: ”how lexicographic principles have been developed within various intellectual or national traditions” (2001: 40). Jag skall nu gå in ganska utförligt på det problemet, som för mig framstår som helt avgörande, om lexicografihistoria skall kunna bidra med relevant belysning av lexicografiska eller andra frågor. Det kan verka vara en lång utviking, men jag hoppas att relevansen skall framgå efterhand.

Anders Fredrik Dalins Ordbok öfver svenska språket är den första kompletta enspråkiga svenska ordboken. Den är bibliografiskt besvärlig, men 1855 hade sista häftet kommit ut, och publiceringen påbörjades alldeles avgjort år 1850. Året dessförinnan, 1849, föddes August Strindberg, av många betraktad som Sveriges störste författare någonsin. År 2012 uppmärksammas 100-årsdagen av hans död på olika sätt, med konferenser världen över, högtidstal och tillställningar, som han själv hade kunnat skriva underbara parodier på. Den gamle rebellen, som alla tyckte var stökig så länge han levde, har blivit en väl integrerad del av det nationella kulturarvet.

Det hann Strindberg faktiskt bli före sin död. När han fyllde 63 år den 22 januari 1912 fick han motta en nationalgåva, som hade 20 000 bidragsgivare. När den överlämnades den 2 mars skänkte han bort alltihop till fattiga, sjuka och arbetslösa. Arbetarrörelsen bestämde sig för att han var deras man. Arbetarungdomen hade hyllat den sjuke författaren på hans födelsedag med ett fackeltåg. När han dog den 14 maj 1912 hade han själv bestämt att begravningen skulle äga rum tidigt på morgonen för att undvika nyfikna. När han söndagen den 19 maj bars till den sista vilan av åtta arbetare hade de sällskap av en mångtusenhövdad skara, som ringlade genom staden till Stockholms Nya kyrkogård. Kungahus, regering och riksdag, universitet, författare, skådespelare, ett hundratal arbetarföreningar hörde till dem som var representerade. Svenska kyrkans kanske största ande någonsin (näst Den Helige, får man väl tillägga), den blivande och så småningom legendariske ärkebiskopen Nathan Söderblom, avslutade med Jesu Bergspredikan: ”Saliga äro de som hungra och törsta efter rättfärdighet, ty de skola bliva mättade.”



När Dalin dog knappt fyrtio år tidigare, mitt i sommaren 1873, året efter det att Strindberg slutgiltigt givit upp studentlivet och bestämt sig för att bli litteratör på heltid, var han fyra år äldre än Strindberg blev. Han fick den begravning Strindberg hade tänkt sig. Ingen nationalinsamling där inte, inget väsen. Den döve Dalin lämnade det jordiska lika stillsamt som han varit en del av det, den idoge lexikografens lott i jämförelse med den uppburne författarens, kan man tycka. Han fick trösta sig med Bergspredikans ord, detsamma som lästes över Strindberg: ”Saliga äro de som hungra och törsta efter rättfärdighet, ty de skola bliva mättade.”

Dalin har förvisso en grundmurad ställning i svensk lexikografihistoria precis som Strindberg i svensk litteraturhistoria, men Dalin och Strindberg hade veterligen aldrig något med varandra att göra. Det kan verka irrelevant att dra fram Strindberg i ett sådant här sammanhang, men det finns en poäng. Under de sista fem–sex åren av sitt liv utvecklade Strindberg nämligen en intensiv, närmast manisk, aktivitet med språklig inriktning, och det är detta som är relevant i sammanhanget. Bitvis tar sig hans språkliga sammanställningar formen av ett slags lexikografi. Men först kan möjligen några korta biografiska notiser och hållpunkter i hans litterära och intellektuella utveckling vara till hjälp. Därigenom hörsammas också på sätt och vis Hartmanns uppmaning att skriva biografier över lexikografer.

Strindberg föddes alltså mitt i nittonde seklet, i Stockholm, där han också dog 63 år senare. Han är en av de främsta Stockholmskildrarna i svensk litteratur, men han studerade i Uppsala. Han tänkte ta examen i medicin. Studierna gick trögt och blev resultatlösa. Han ägnade sig helst åt skönlitteraturen. Hans dröm var att bli skådespelare, och han försökte skriva pjäser, men länge med begränsad framgång. Den första utlandsresan gick sommaren 1869 till Köpenhamn, där den dansk-isländske bildhuggaren Bertel Thorvaldsens verk inspirerade honom till hans första uppförda skådespel, *I Rom*. Under de närmaste åren svängde han från skådespeleriet helt över till skrivandet. Våren 1872 lämnade han för alltid Uppsala.

Strindberg började arbeta som journalist och skrev en artikel om Kungliga Biblioteket, som väckte uppmärksamhet. Han fick en tjänst som amanuens vid biblioteket, där han verkade mellan 1874 och 1882 och grundlade en livslång beundran för dess chef, den ganska egensinnige Gustaf Klemming. Denne gjorde också en storstilad insats som textutgivare; han var en av de drivande krafterna bakom Svenska Fornskrift-sällskapet. Vid biblioteket fick Strindberg i uppgift att ordna de kinesiska samlingarna, varför han lärde sig en del kinesiska.

Strax före nyår 1877 gifte han sig med den blivande skådespelerskan Siri von Essen. 1879 slog han igenom med romanen *Röda rummet*, som också markerar det moderna genombrottet i svensk litteraturhistoria. På hösten 1883

reste hela familjen till Paris för ett kortare besök, men det blev en sexårig rundresa genom Europa, som slutade i Danmark. Strindberg knöt kontakter med Bjørnstjerne Bjørnson, Jonas Lie, bröderna Brandes och Friedrich Nietzsche. Den juridiska processen kring novellsamlingen *Giftas* 1884 gjorde honom anti-religiös och antifeministisk, han drogs mot socialism och Nietzsches nihilism. 1887 påbörjades den långdragna och uppsplitande skilsmässan, 1892 gick han för andra gången i landsflykt.

Under hela denna tid skrev han febrilt sina mest naturalistiska arbeten. Ett mycket kort äktenskap med österrikiskan Frida Uhl byttes mot ensamhet, miss-tänksamhet och förföljelsemani. Hans tidigare kausala materialism övergick i en dragning till buddhism och katolicism. Hans relativa vetenskaplighet ersätts av tro på en magiskt-religiös världs- och verklighetstolkning. Åren 1892–97 fullbordar han en intellektuell, moralisk och religiös helomvändning. Under denna tid är han upptagen av förment naturvetenskapliga studier, men han dras lika mycket till alkemister, ockultister, teosofier och hypnotisörer. Han börjar tro på andar, demoner och makter, som han anser styra hans liv.

Efter denna kris, som han själv skildrat i skriften *Inferno* (1897), återfår han sin litterära produktivitet. Han skriver nu några av sina största dramer. Men det är inte bara skönlitteraturen som lockar honom. Under de sista åren av sitt liv, ungefär 1907–1912, var Strindberg besatt av ett storslaget språkligt projekt. Hans allmänna tanke är den bibliska: ett ursprungligt gemensamt mänskligt språk har blivit uppdelat i ett stort antal ömsesidigt oförståeliga språk; det första var hebreiska – eller, kanske, en föregångare som hebreiskan senare utvecklats ur.

Vad den än berodde på utlöste Strindbergs fascination för språkliga ting ett oavbrutet läsande och skrivande i ärendet under en femårsperiod. Ytligt sett rör det sig om ett slags språkvetenskap. Det fanns hundratals böcker om språkliga ämnen i hans bibliotek när han dog. 1907 bad han sin förläggare om hjälp med att skaffa allt som kunde uppbringas om gamla skriftsystem, grammatiska beskrivningar av olika språk osv. Resultatet av hans språkstudier har huvudsakligen avrapporterats i artiklar som samlats i hans s.k. »Blå böcker» (En blå bok I–IV, 1907–1908, mer än 1 150 sidor totalt) samt en serie monografier. Dessa återfinns i litteraturlistan som Strindberg 1910a, 1910b, 1911a, 1911b samt 1912.

Vid sidan av det publicerade materialet, som också omfattade en mängd tidsningsartiklar, finns det flera tryckfärdiga manuskript, ca 600 sidor, samt mer än 1 000 sidor och 2 000 mindre pappersbitar med handskrivna anteckningar av språklig natur, som bevarats i författarens personliga arkiv. Allt detta vittnar om en enorm aktivitet, som också kan följas i Strindbergs korrespondens från den

här tiden. Just *korrespondens* är ett nyckelord, som vi strax skall se, men i en annan betydelse än 'brevskrivning'.

Strindberg är naturligtvis ingen språkforskare i ordets rätta bemärkelse. Han har själv en ambivalent inställning: han är övertygad om att hans språkvetenskapliga iakttagelser är av stort värde, han är noga med att hans förläggare skall sända arbetena vitt och brett till specialisterna – han söker bekräftelse hos auktoriteterna på området, samtidigt som han hånar deras trångsynthet. Teoretiskt, metodologiskt och inte minst terminologiskt sätter han sig över alla regler. Han har betydande språkkunskaper rent praktiskt, sedan ungdomens studier och sina ständiga resor i främmande land, men han är självsvaldig och vill inte ta rättelse.

Orsaken är att Strindberg egentligen är ute i helt andra ärenden än rent språkliga. Han söker samband av alla de slag, och i språkens värld finner han dem på det mest häpnadsväckande sätt. Låt oss bara se på ett enda exempel, hämtat ur *Världs-Språkens Rötter* (1911a). Med det som återges på bild 2 menar Strindberg följande:

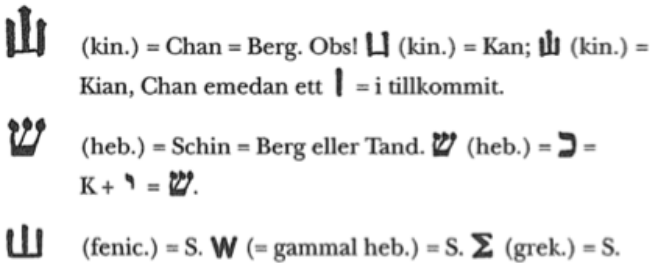


Bild 2. Uppställning hämtad ur *Världs-Språkens Rötter* (1911a: 9).

- (1) Det kinesiska tecknet "Kan" är ekvivalent med hebreiskt kaf (= /k/) – därför att de är lika till formen, med 90° vridning.
- (2) Det kinesiska "Kian"- eller "Chan"-tecknet är ett "Kan"-tecken med ett vertikalt streck tillagt; det senare är likvärdigt med det latinska alfabetets vokaltecken <i>/<I>.
- (3) Det hebreiska "Schin"-tecknet (dvs. shin) är ett kaf, svängt ett kvarts varv och med jod tillagt, enligt Strindberg.
- (4) Kinesiskt "Chan" och hebreiskt "Schin" är fonetiskt ganska lika och semantiskt besläktade, eftersom båda betyder 'berg' ('tand'). (Hebreiska tecken har mnemotekniska namn, precis som exempelvis runorna.)
- (5) När en tonlös velar klusil /k/ följs av en hög vokal, uttalas den i den palatala delen av munnens uttalsrymd, som en konsekvens av uttalsapparatusens anpassning till den följande palatala vokalen. Detta kan senare ge upphov

till en palatal affrikata eller frikativa (t.ex. [č] eller [ʃ]). Ett sådant ljud är den ordinledande konsonanten i ”Schin” (ʃ).

- (6) Det feniciska tecknet för <s> liksom det gammalhebreiska samt grekiskans versala sigma är alla formellt likartade, samtidigt som /s/ är fonetiskt besläktat med de nämnda palatala konsonanterna.

Strindberg relaterar tecken och språkljud i olika språk till varandra, delvis med hjälp av etablerade språkvetenskapliga principer, delvis genom rent fabulerande. Hans analys är en fri blandning av semantiska, morfologiska och fonetisk-fonologiska argument. Han rör sig ohämmat över gränsen mellan tal och skrift, och han skiljer inte ens på något systematiskt sätt mellan genetiskt besläktade och helt obesläktade språk. I sin avhandling säger Karl-Åke Kärnell (1962: 258):

De språkvetenskapliga studier, som Strindberg upptog ett stycke in på 1900-talet, ligger helt i linje med hans naturvetenskapliga. Han är monist i lingvistikens som i naturvetenskapen; liksom han en gång sökte grundämnenas ursprung i ett enda ämne, vätet, så vill han där upphöja hebreiskan till språkens moder. Man måste, som han säger med en karakteristisk vändning, ”gå till roten, till rötterna, och börja försoningen, där förbistringen började”.

Han anspelar naturligtvis på den bibliska historien om Babels torn, en symbol för människans arrogans eller hybris. Till följd av den drabbades människorna av språkförbistringen, som ett Guds straff. (Detta är åtminstone en möjlig tolkning.) På denna grund kom språk att jämföras med varandra i århundraden, i syfte att rekonstruera det gemensamma ursprunget.

Under medeltiden var intresset för språkjämförelse svagt – latinet upptog all språklig uppmärksamhet – men en helt ny syn på såväl latinet som de olika folkspråken utvecklades under renässansen. Latinet intog en naturlig del av det allmänna intresset för antiken, och folkspråken beskrevs grammatiskt med latinet som modell. Spekulationer om det första språket tog fart igen. Hebreiska antogs ofta vara detta eller åtminstone ett närbesläktat språk.

Även andra språk började tävla med hebreiskan om att vara urspråket. Ett favorittrick var att hävda att ett visst språks talare av något skäl inte hade deltagit i Babeltornets byggande och därför inte hade varit på plats när förbistringen inträffade. Därför hade de kunnat behålla sitt gamla språk. I Sverige hade tanken att sammanställa goternas namn med de gamla landskapsnamnen *Väster- och Östergötland* samt naturligtvis *Gotland* uppstått redan under medeltiden. När den siste svenske katolske ärkebiskopen, Johannes Magnus, hade flytt till Rom skrev han en fantastisk historia om svenskarnas kungar, *Historia de omnibus gothorum sveonumque regibus*, som brodern Olaus Magnus tryckte i Rom 1555, året efter Johannes död.

Med utgångspunkt i de uppgifter som på 500-talet lämnas av den gotiske historieskrivaren Jordanes, om att goterna skulle ha utvandrat från ön Scandza, hävdar Johannes Magnus att goterna ursprungligen bodde i Sverige, att svenskarna härstammade från Noas son Jafet, att svenskarna fortfarande talade det äldsta språket i världen, och att goterna (dvs. svenskarna) såsom romarnas bese-grare var den romerska kulturens rättmätiga arvtagare och sålunda det äldste folket i världen.

Sådana idéer togs tacksamt upp när den unga militära stormakten Sverige skulle konstruera en ärorik förhistoria. Johannes Magnus historia översattes till svenska och lades till grund för alla slags spekulationer. Kulmen nåddes med Olaus Rudbecks *Atland*. Denna har vetenskapligt nydanande inslag, men lika mycket rör det sig om vilda fantasier av etymologisk natur.

Strindbergs metoder påminner om Rudbecks. Han nämner till och med Rudbeck då och då, inte utan entusiasm. Men Rudbeck hade i sin tur påverkats av Johannes Bureus, som i ännu högre grad var Strindbergs själsfrände. Bureus var i praktiken den förste riksantikvarien i Sverige, den förste grammatikern, dialektologen, runologen och möjligen också den förste lexikografen. Han efterlämnar ett fortfarande opublicerat manuskript på 600 sidor. Ingen har egentligen analyserat det noga.

Den mångsidige Bureus fick sällan något fullbordat. Han blev efterhand en religiös drömmare, han ägnade sig åt judisk kabbala, visionär ockultism, mysticism och teosofi. I det strängt lutherskt ortodoxa Sverige balanserade han hela tiden på gränsen till hädelse, ett av de svåraste brotten. Själv ansåg han sig vara Guds profet och gjorde beräkningar om apokalypsen. Han ville upptäcka de övergripande sambanden i universum och utvecklade en teori om ”korrespondenser”, som förutsade att varje detalj i evigheten var förutsagd i den fysiska världen. Det var bara fråga om att känna igen mönstren.

Georg Stiernhielm var Bureus lärjunge. Han var mer språkfilosof och poet. Han utgick från samma världsuppfattning som Bureus, inklusive svenskarnas gotiska relationer, men han menade att språket låg till grund för alla förklaringar. Det språkliga tecknet var i hans ögon ingalunda arbiträrt, som vi tror, och det var inte dess komponenter heller. De antogs i stället stå i ett mera direkt förhållande till det betecknade.

Stiernhielm var troligen inspirerad av Bureus paralleller mellan denna och den hinsides världen. Han ville undersöka sambanden mellan den materiella världen och dess språkliga symboler. Överensstämmelsen mellan de båda antogs vara tydligast i de språk som stod närmast urspråket, och hit hörde för Stiernhielm svenskan. Därför försöker han isolera dess mest primitiva ordrotter (lat. *radices universales*), som alla skulle bestå av meningsfulla element. Varje bokstav (eller ljud) relaterade till något särskilt begrepp, t.ex. A till ’ljus’, U till ’natt’ och

'mörker'. Ordrötterna kunde representera mer komplexa betydelser. Stora mängder av ord från olika språk insamlades och klassificerades enligt olika semantiska kategorier. På så vis kunde svenska ordrötter återfinnas på de mest fantastiska sätt i olika andra språk. Arbetsättet påminner om Strindbergs.

På 1600-talet var det många som var upptagna av ordrötter. Den tyske grammatikern Schottelius resonemang är symptomatiskt. Hans tanke var att varje språk har sin speciella natur och lagbundenhet, särskilt återspeglad i språkets "stamord". Dessa är ungefär ordrötterna, såsom dessa representeras av enstaviga ord och enstaviga böjningsformer, t.ex. imperativerna i de germanska språken. Ett givet språks relation till urspråket indikeras av hur många stamord i den angivna meningen som språket har. Schottelius fann mer ordrötter i tyskan än i andra språk, möjligen med undantag för hebreiskan. En ordbok som alltid förbryllat kommentatorerna är Jonas Rugmans *Mono-syllaba Is-landica* (1676), som upptar enbart enstaviga isländska ord, men detta är ett arbete med just det syfte som Schottelius också hade. Det är en lång lista över enstaviga ord i isländskan, utarbetad av den i Uppsala verksamme islänningen Rugman.

Man kan inte påvisa att Strindberg känt till just Schottelius, men Strindberg säger att han vill gå "till roten, till rötterna". Å andra sidan har språkstudiet inom den hebreiska traditionen – och även den indiska – alltid varit orienterad mot ordrötter, på grund av dessa språks struktur. För Strindberg är det angeläget att reducera ett språks ordförråd till ett begränsat antal ordrötter, därför att han antar att det är lättare att urskilja urspråket i dessa. I *Bibliska Egennamn* (1910a: 27) säger han: "Med ett ord, oavsett företräde i ålder, äga alla språk på jorden ett gemensamt, det må nu vara härlett eller ursprungligt, och de största lexikon hålla endast så många urljud och urbegrepp som dagarna i året".

Bland de auktoriteter Strindberg åberopar finns italienaren Alfredo Trombetti, som år 1905 publicerade sitt uppmärksammade arbete *L'unità d'origine del linguaggio*. Trombetti var en erkänd specialist på semitiska språk, men hans berömmelse vilar lika mycket på idén om alla språks gemensamma ursprung. Detta är fortfarande en omstridd fråga, men det är intressant att se hur aktivt Strindberg följde den samtida internationella språkforskningen. Man kan knappast kritisera honom för att han var i samklang med samtida specialister, även om en del av dessa nu står lägre i kurs.

Men det överordnade syftet med hans frenetiska språkstudier var inte språkligt över huvud taget. År 1907, när det tar verklig fart, skulle man just till Sverige återföra Emanuel Swedenborg från London, där han hade dött på 1770-talet. Swedenborg var vetenskapsman, teolog, spiritist och mystiker i förening. Strindbergs första blåa bok, som publicerades just 1907, dedicerades till Swedenborg, till "Läraren och Ledaren" från "Lärjungen", som Strindberg formulerar det.

Swedenborg konstruerade en teologi som var en ganska personlig version av renässansläran om analogier mellan olika aspekter av livet och verkligheten, t.ex. mellan människan och universum, mellan kroppsliga och andliga egenskaper, mellan Gud som skapande konstnär och människan som imitator osv. För Swedenborg var analogin mellan ljus, förnuft och gudomlig vishet avgörande. Hans lära om "korrespondenser" relaterar till renässansföreställningen om naturen som "Guds bok", som direkt korresponderar med mänskligt skrivande. Enligt Swedenborg kan språkliga uttryck förstås i en naturlig, andlig eller gudomlig mening.

Strindberg var också dragen till judisk kabbala. Syftet med kabbala var ursprungligen att upprätthålla den judiska skrifttraditionen men utvecklades senare till ett hemligt system för spekulationer. Dessa koncentrerades på Gud och Skapelsen. Eftersom Gud har skapat allt genom sitt ord, är det hebreiska språket och alfabetet laddat med särskild kraft, och som en följd uppstod olika typer av bokstavsmystik. Strindberg börjar den första av sina språkliga utredningar, *Bibliska Egennamn* (1910a: 7), med en påminnelse om dessa omständigheter: "I begynnelsen var Ordet ... Genom detta är allt gjort!" Ordet och språket var därmed direkt länkat till Skapelsen och Skaparen, vilket gör det till ett "Lingua Sancta" i Strindbergs ordalag. Det var för att skapa klarhet i förhållanden som dessa Strindberg blev "språkforskare".

## Lexikografi och verklighet

Strindberg har ambitioner att få sina förment språkvetenskapliga sammanställningar bedömda som just vetenskapliga. Han närmar sig medvetet sakkunskapen. När han sänder ett manus till Adolf Noreen och ber om ett utlåtande blir domen inte nådig. I ett vänligt formulerat men i sak förödande svar får han läsa: "Hjärtligt tack för vänliga sändningen, som härmed återgår! Att jag i de flesta fall ej kan dela Er uppfattning, kränker nog ej den, som sedan länge är van att stå 'Ensam' och kanske just därför – stark. / Med gammal och oförminskad beundran för den store språkkonstnären, om också icke språkforskaren, August Strindberg".

Kritikerkåren är mera kluven inför Strindbergs språkiakttagelser. Språkforskarna bland dem avfärdar kategoriskt hans "rön", många i ironiska ordalag. Men vissa kritiker (som borde vetat bättre) är positiva, väl på grund av en allmän beundran för Strindberg. Detta visar hur svårt det kan vara att bedöma något som presenteras i en oväntad form eller i ett oväntat sammanhang, även sådant som man tror sig veta något om. Håller vi oss till lexikografihistorien är det klart att det finns konkreta exempel på tänkande av det slag som här beskri-

vits. I svensk 1600-talslexikografi är Bureus och Stiernhielm pionjärer. Båda visar ansatser till det vi uppfattar som lexikografi. Bureus ordboksmanuskript har redan nämnts, Stiernhielm publicerade en volym som skulle bli inledningen till en fyllig ordbok (Stiernhielm 1643). Men hur skall detta bedömas, när det egentligen inte är lexikografi i vår mening som är deras mål?

Fram till de nämnda lexikografiska initiativen i Sverige är det latintraditionen som råder på det lexikografiska området. Detta är tydligt i 1500-talsarbetet *Vocabularum rerum vocabula* (1538). Elaus Petri Helsingius (1587) är först med svenska som källspråk, vartill latinska och grekiska ekvivalenter fogas. Dessa bidrag kan inte tillmätas någon autonom status inom den svenska lexikografihistorien. De hör till den starka (internationella) latinska traditionen. Denna har då redan länge dominerat på kontinenten. En svensk motsvarighet i det större formatet, men fortfarande väl förankrad i latintraditionen, är *Lexicon Lincopense* (1640).

Bureus och Stiernhielm frigör sig från latintraditionen. Det skulle kunna hävdas att de är de första nationellt svenska lexikograferna; deras mål är trots allt att skapa kunskap om ord i svenskan. Deras syfte är dessutom pedagogiskt liksom latintraditionen, men på ett helt annat sätt. De hade götiskt patriotiska avsikter. I detta hägn förblir lexikografen ett tag. Ett säreget exempel är Olaus Verelius märkliga *Index lingvæ veteris scytho-scandicæ sive gothicæ*, som trycktes år 1691, nio år efter Verelius död. Verelius *Index* innehåller enbart fornnordiska ord, inga aktuella svenska. Den fosterländska tanken var att ord hämtade från den nordiska medeltidslitteraturen skulle kunna återinkallas till tjänstgöring. Ordboken säger följaktligen inget om samtidens språk, men mycket om dess tänkande.

Från *Lexicon Lincopense* (1640) fortsätter latintraditionen in i 1700- och 1800-talet, alltid med en pedagogisk komponent, ännu tydligare hos Comenius, som översätts till svenska av Ericus Schroderus. Denne inkluderar en alfabetisk svensk ordlista hämtad från *Lincopensen*, som på så vis får förlängd effekt. Schenberg (1739) använder sig i sin latin-svenska ordbok däremot av Swedbergs opublicerade ordbok. Denna får på så vis lexikografihistorisk betydelse, trots att den inte tryckts förrän helt nyligen (Holm 2009).

Swedberg grundar sitt ordboksmanuskript på Bibelns ordförråd. Swedberg är en sammansatt natur med klart språkvetenskapliga böjelser. Men han är i första hand en kyrkans man, liksom flera andra lexikografer. I likhet med Grubb, Gothus, Spegel och Serenius är han biskop. Han är samtidigt något av en mystiker, precis som Bureus, och andeskådare. Påverkar detta hans lexikografi? På ett triviale plan är det uppenbarligen så: han har fått den nödvändiga lärdomen genom sin kyrkliga bana. Haqvin Spegel redovisar öppet att han sammanställt sin ordbok för att underlätta bibelläsningen för de nya sydsvenskar som före



1600-talets freder varit danskar. Lexikografernas patriotiska strävanden är ett återkommande drag.

Även 1700-talets tvåspråkiga lexikografi bedrivs av präster (varav några biskopar), inte av språkforskare. De ordböcker mellan samiska och svenska som tillkommer under perioden har delvis varit avsedda att stärka kristendomens ställning bland samerna. På samma sätt har det varit i Norge. Danmark visar paralleller i förhållande till Grönlands inuiter. Även när inte det direkta incitamentet till ordboksarbete varit religiöst nit kan det förväntas ha spelat in. Många av de tidiga ordboksredaktörerna har en världsbild och en språksyn som faktiskt påminner om Strindbergs. Swedberg är för övrigt den av Strindberg djupt beundrade Swedenborgs far.

Om vi till den religiösa tron lägger ideologi och allmän världs- och verklighetsuppfattning är det uppenbart att sådana faktorer haft avgörande betydelse för lexikografihistorien. I vissa fall har denna iakttagelse rent förklarande kraft. Exempelvis får man med det anlagda synsättet bättre förståelse för 1600-talets patriotiska lexikografi, som annars är helt förbryllande.

## Slutord

Vår syn på lexikografi är befriad från många av de vanföreställningar som äldre lexikografer bar på, men vi kan missa mycket av de äldre lexikografernas grundläggande syften, eftersom vi saknar deras världsbild. För oss är det naturligt att bedöma äldre lexikografer som om de vore kolleger, ute i samma ärenden som vi. Men i första hand var de präster, endast i andra hand lexikografer. Deras verklighet var under alla omständigheter en helt annan än vår, både den verklighet de levde i och den bild de gjorde sig av verkligheten. Den var inte i allt underlägsen vår. De levde exempelvis i ett mångspråkigt samhälle, som de erkände och bejakade. För dem var Haugens språkval verklighet (under momentet Selection i bild 1 ovan), en ständigt närvarande nödvändighet. Vi har ett mångspråkigt samhälle än idag, men många inbillar sig att det är enspråkigt. Så till vida har den götiska propagandan varit framgångsrik.

Vi kan ha lätt för att dra på munnen åt äldre lexikografers underligheter, när vi blir medvetna om dem. Men många urskiljer vi inte, och om vi gör det saknar vi ofta redskap att förstå dem rätt. Strindbergs extrema uppfattning märks tydligt eftersom den ligger sent i tiden, nära vår egen, och kontrasterar bjärt mot vår. Modern lexikografi slog igenom i Sverige med Dalins ordbok (1850–1855). Strindberg ligger i tiden på fel sida Dalin för att undgå upptäckt och nesligt avslöjande. Ett par generationer tidigare hade man kanske allmänt prisat hans

språkliga lärdom och skarpsinne. Det gör ju faktiskt några kritiker ännu ett drygt decennium in på 1900-talet.

I Bibeln är det Ordet som är alltings upphov. Ord är således inte – som för de flesta moderna lexikografer – enbart språkets praktiska byggstenar. Ordet är tillvarons kvintessens. Strindberg besjålas av samma djupa tro på Ordets betydelse. En lång rad prästmän och lexikografer i samma skepnad har genom århundraden rent praktiskt handskats med ord i ordböcker växelvis med att de grubblat över Ordets roll i tillvaron. Har de verkligen alltid kunnat hålla isär orden och Ordet? Strindberg kunde det uppenbarligen inte. Frågan är om inte Dalin är en av de första riktigt betydelsefulla svenska lexikografer som förmått detta i full utsträckning. Det finns en del i det han skriver som tyder på att han hade ett vetenskapligt sanningsideal som påminner om vår tids.

Det som här sagts är inte avsett att nedvärdera äldre lexikografers insatser eller att generellt ifrågasätta deras omdöme. Det är enbart menat som en påminnelse om hur farligt det är att glömma bort kontextens vikt, en bland språkvetare principiellt välkänd sanning. Men tänk än en gång på Bergspredikans ord: ”Saliga äro de som hungra och törsta efter rättfärdighet, ty de skola bliva mätade.” Nu för tiden är det naturligt att uppfatta *rättfärdighet* som en allmänt moralisk egenskap, men i Bergspredikan är det den specifikt religiösa (kristna) betydelsen som åsyftas, den som i Dalins ordbok definieras som ”Noggrant iakttagande af religionslärans bud”. Detta visste Strindberg, men har man inte denna betydelse högt i medvetandet förstår man inte bibelordet. Fatalt för en lexikograf? – Det är sådant detta har handlat om.

## LITTERATUR

- Adelung, Johann Christoph, 1774–1786: Versuch eines vollständigen grammatisch-kritischen Wörterbuchs der Hochdeutschen Mundart. I–IV ( ny uppl. 1793–1801).
- Al-Kasimi, Ali, 1977. Linguistics and Bilingual Dictionaries. Leiden.
- Boeck, Simon Skovgaard, 2009: Fire tematiske ordbøger fra dansk renæssance. København.
- Boyer, Abel, 1702: Dictionnaire royal, françois et anglois. A la Haye.
- Cowie, A.P. (ed.), 2008: The Oxford History of English Lexicography. I–II. Oxford.
- Dalin, A.F., 1843: Nytt fransyskt och svenskt lexikon, med utförlig fraseologi, ordens uttal då det afviker från allmänna reglor, samt tekniska och vetenskapliga termer; jemte ett bihang [...]. I–II. Stockholm.
- Dalin, A.F., 1850–1855. Ordbok öfver svenska språket. Stockholm. (Samlingar utgivna av Svenska fornskriftsällskapet: Svenska skrifter 91.)
- Djärv, Ulrika, 2009: Fornsvenskans lexikala kodifiering i Söderwalls medeltidsordbok. Stockholm.

- Egilsson, Sveinbjörn, 1860: *Lexicon poëticum antiquæ linguæ septentrionalibus*. Hafniae.
- Fritzner, Johan, 1883–1896: *Ordbog over Det gamle norske Sprog*. I–III. 2. utg. Christiania. (Band IV. Rettelser og tillegg ved Finn Hødnebo, 1972. Oslo, Bergen, Tromsø.)
- Hannesdóttir, Anna, 1998: *Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga*. (Meijerbergs arkiv för svensk ordforskning 23.) Göteborg.
- Hannesdóttir, Anna, 2000a: *Funktioner som försvann: det kontrastiva perspektivet i äldre svensk lexikografi*. I: *Nordiska studier i lexikografi 5. Rapport från Konferens om lexikografi i Norden, Göteborg 27–29 maj 1999*. Red. av Martin Gellerstam m.fl. Göteborg. (Meijerbergs arkiv för svensk ordforskning 27.) S. 3–17.
- Hannesdóttir, Anna, 2000b: *Ordböckernas roll i svenskans grammatiska standardisering*. I: *LexicoNordica 7–2000*. S. 23–40.
- Hannesdóttir, Anna & Ralph, Bo, 1989: *Early Dictionaries in Sweden. Traditions and Influences*. I: *Symposium on lexicography IV. Proceedings of the Fourth international symposium on lexicography April 20–22, 1988 at the University of Copenhagen*, ed. by Karl Hyldgaard-Jensen and Arne Zettersten. (*Lexicographica. Series Maior*. 26.) Tübingen. S. 265–279.
- Hartmann, Reinhard R.K., 2001: *Teaching and Researching Lexicography*. Harlow.
- Hast, Sture, 1990: *Samuel Schultze och hans svenskt Lexicon*. (Skrifter utgivna av Vetenskaps societeten i Lund 83.) Lund.
- Haugen, Einar, 1987: *Blessings of Babel. Bilingualism and Language Planning. Problems and Pleasures*. (Contributions to the Sociology of Language 46.) Berlin, New York & Amsterdam.
- Helsingius, Elaus Petri, 1587: *Synonymorum libellus*. Stockholmiae.
- Holm, Lars, 1986: *Jesper Swedbergs Svensk Ordabok – bakgrund och tillkomsthistoria*. Uppsala.
- Holm, Lars (utg.), 2009: [Utgåva av] *Jesper Swedberg, Svensk Ordabok*. (Skara stiftshistoriska sälls kamps skriftserie, nr 46.) Skara.
- Holm, Lars & Jonsson, Hans, 1990: *Swedish Lexicography*. I: *HSK 5:2*. S. 1933–1943. [J.C. Holmberg:] *Fransyskt och Svenskt Lexicon. Parallèle des langues Française & Suédoise; ou Dictionnaire François & Suédois [...]*. I–II. Stockholm 1795.
- HSK = *Handbücher zur Sprach- und Kommunikationswissenschaft (5. Wörterbücher/Dictionaries/Dictionnaires*. 1989–1991.) Berlin.
- Ingólfsson, Gunnlaugur, 1999: *Inngangur*. I: *Lexicon islandicum. Orðabók Guðmundar Andréssonar. Ný útgáfa*. Gunnlaugur Ingólfsson og Jakob Benediktsson önnuðust útgáfuna. (Orðfræðirit fyrri alda IV.) Reykjavík. S. IX–XXXV.
- Johansson, Monica 1997. *Lexicon Lincopense. En studie i lexikografisk tradition och svenskt språk vid 1600-talets mitt*. (Meijerbergs arkiv för svensk ordforskning 21.) Göteborg.
- Johnson, Samuel, 1755: *A Dictionary of the English Language [...]*. (To which are Prefixed, A History of the Language, and An English Grammar.)

- Kvaran, Guðrún, 1994: Inngangur. I: *Nucleus latinitatis*. Ný útgáfa. Guðrún Kvaran og Friðrik Magnússon sáu um útgáfuna. (Orðfræðirit fyrir alda III.) Reykjavík. S. IX–XXIX.
- Kärnell, Karl-Åke, 1962: *Strindbergs bildspråk*. En studie i prosastil. Stockholm etc.
- Larsson, Lennart, 2003: *Varifrån kom svenskan? Om den svenska vokabulären i en fyrspråkig ordbok utgiven i Riga 1705*. (Acta Academiae Regiae Gustavi Adolphi 81.). Uppsala.
- Lexicon lincopense* = Gothus, Jonas Petri, 1640: *Dictionarium Latino–Sveco–Germanicum*. Lincopiæ.
- Lindfors, Anders Otto, 1815–1824: *Fullständigt Swenskt och Latinskt Lexicon*. I–II. Lund.
- Noreen, Adolf, 1903: *Vårt språk*. Nysvensk grammatik i utförlig framställning. I. Lund.
- Ralph, Bo, 1992: *The older dictionaries as sources for Nordic language history*. I: *The Nordic Languages and General Linguistics 7*. Proceedings of the Seventh International Conference of Nordic and General Linguistics in Tórshavn, 7–11 August 1989. Vol. II. (Annales Societatis Scientiarum Færoensis. Suppl. XVII.) Tórshavn. S. 493–509.
- Ralph, Bo, 2000: *Svensk lexikografitradition*. I: *LexicoNordica 7–2000*. S. 5–22.
- Rogström, Lena, 1998: *Jacob Serenius lexikografiska insats*. (Meijerbergs arkiv för svensk ordforskning 22.) Göteborg.
- Rudbeck, Olaus, 1679–1701: *Atlant Eller Manheim*. I–IV. Upsalæ.
- Rugman, Jonas, 1676: *Mono-syllaba Is-landica*. Upsalæ.
- Sahlstedt, Abraham, 1769: *Swensk Grammatika efter det nu för tiden i språket brukliga sättet författad [...]*. Upsala.
- Sahlstedt, Abraham, 1773: *Swensk Ordbok Med Latinsk Uttolkning / Dictionarium Svecicum Cum interpretatione Latina*. Stockholm.
- SAOB = *Ordbok över svenska språket* utgiven av Svenska Akademien, 1893 ff. Lund.
- Schenberg, Petrus, 1747: *Lexicon latino–svecanum. Recognitum et Auctum*. Norcopiæ et Lincopiæ. (Första uppl. 1739.)
- Schlyter, Carl Johan, 1877: *Glossarium ad corpus iuris sueo-gotorum antiqui*. / *Ordbok till Samlingen af Sweriges Gamla Lagar*. Lund.
- Schottelius, Justus Georg, 1663: *Ausführliche Arbeit Von der Teutschen HauptSprache*. Braunschweig.
- Stiernhielm, Georg, 1643: *Gambla Swea- och Götha-måles fatebur*. Stockholm.
- Strindberg, August, 1907–1908: *En blå bok* I–IV. Stockholm.
- Strindberg, August, 1910a: *Bibliska Egennamn*. Med Ordfränder i klassiska och levande språk. Stockholm.
- Strindberg, August, 1910b: *Modersmålets Anor*. Svenska ordfränder i klassiska och levande språk. Stockholm.
- Strindberg, August, 1911a: *Världs-Språkens Rötter*. Stockholm.
- Strindberg, August, 1911b: *Kina och Japan*. Stockholm.
- Strindberg, August, 1912: *Kinesiska Språkets Härkomst*. I–II. Stockholm.
- Söderwall, K.F., 1884–1918: *Ordbok öfver svenska medeltids-språket*. I–II. Lund. (Supplement A–N av K.F. Söderwall, W. Åkerlund, K.G. Ljunggren 1953, O–Ö av K.G. Ljunggren och E. Wessén 1953–1973.)

Teleman, Ulf, 2002: Ära, rikedom och reda. Svensk språkvård och språkpolitik under äldre nyare tid. (Skrifter utgivna av Svenska språknämnden 85.) U.o.

Verelius, Olaus, 1691: Index lingvæ veteris scytho-scandicæ sive gothicæ. Upsalæ.

Vocabularum rerum vocabula cum sueca interpretatione, 1538. Stockholm.

*Bo Ralph*

Institutionen för svenska, Göteborg

bo.ralph@svenska.gu.se

# What has the *OED* become?

*John Simpson*

## Some background

Thirty years ago the *Oxford English Dictionary* was happily drifting towards obsolescence. The editors were engaged on supplementing the text of the original dictionary (1884-1928) by adding new words and meanings, recent ante-datings, and little else. The first edition of the *OED* was becoming increasingly enmeshed in a cocoon of cotton wool and moth balls.

The *OED* wouldn't have been the first dictionary to sparkle before running out of steam. But there was a feeling gathering pace in the late 1970s and early 1980s that this might not be the end for the *OED*. The cause for optimism was technological. From the 1960s we had been aware that computational power was being harnessed in the aid of linguistic and literary scholarship, especially in the production of concordances. As the 1970s wore on and the microcomputer made more people aware of the possibilities of computational manipulation, various publishers and editors at the University Press in Oxford began to think that the future of the *OED* might be linked to this new technology.

## Decision point

The decision point arrived in 1984, when the Delegates of the University Press approved plans for a Second Edition of the *OED*. The base text of the Second Edition was to be a machine-readable database, and so the existing text of the dictionary was pushed and pulled and squeezed so that it fitted into the database, and this was published in book form (twenty large volumes) in 1989. But this was only the beginning of the grand vision. Once the text was available on computer, we could envisage a number of ways in which it could be updated on the way to producing the Third (and comprehensively revised) Edition of the *OED*.

The old vision would have seen the production of supplement upon supplement of the dictionary, each one gradually making life more difficult for the user as yet more instalments had to be examined to ensure that all relevant material had been checked. The University Press was not prepared to countenance that route. Editorially the supplement model is problematic too, as the overall profile of an entry is not improvement by piecemeal additions – accessory coat hangers added to what was a self-consistent whole. Furthermore, the dictionary is too important for only part of it (say, the modern component) to be maintained. Modern vocabulary and word use depends on that of the past: the picture would misrepresent lexical development if we were to update the present at the expense of the past.

## What was the medium of the future?

We thought about what the *OED* could become, knowing that our ideas would certainly change as time went by. Our first plan was to use the machine-readable text of the dictionary as a platform from which to produce a revised CD-ROM (which represented the advanced technology of the day).

Whilst plans were being made for bringing this about, we started to tussle with the problems of a new editorial policy: one which would drag the *OED* from being a dinosaur of the nineteenth century to becoming an icon of the twenty-first.

As the twenty-first century loomed, we recognized that the role of the CD-ROM was dimming, and that its place was being taken by the world of the Internet. Fortunately the dictionary had originally been coded in a common-form mark-up language, SGML, and this could be updated in due course to new standards, such as XML. And then as the *OED* settled down to life online, we became aware of the potential of phones and other devices smaller than the pc or the laptop on which the *OED* might perform. The *OED* is in a position to metamorphose with the technology.

## The grand vision

The mechanics of the database are less important, though, than the vision. What could the *OED* become?

At first we thought of the *OED* as the central, pivotal database within a system of knowledge represented by text. Any user might encounter an issue which the *OED* might resolve, and would then leap from the text (of whatever sort)

currently being consulted into the *OED* for the resolution of the problem. As time went on, we started to modify this approach, and to see the *OED* as one of a number of major textual and other resources which might form a network of support for the user in various contexts.

## Changing the text

But before we look too far down the line, we should investigate what is happening to the dictionary on the editorial side.

The Second Edition of the *OED* consisted of 21,000 pages of closely printed, triple-column text: a substantial text covering 240,000 headwords and a total of around 600,000 lemmas (many of which were embedded in larger entries). The majority of the text dated from the period of the First Edition of the dictionary (1884-1928), with additional material supplementary to the main text added in 1933 and between 1972 and 1986. The text therefore represented editorial thought from various different eras.

Very early on we decided, along with the *OED*'s Advisory Committee of leading language scholars, that the only option available to us was to mount a comprehensive revision and update of the standing text. Anything short of this would not do justice to the fact that the words of the language had a beginning which needed reinvestigating, a middle period of development, and, in the cases of terms which have lapsed from use, a period of decline. Furthermore, words and meanings do not exist in a vacuum. They form part of a semantic network and the relationship between nodes of this network (the individual words and meanings) is in a constant state of flux. If we were to address only part of this vast network, we might do justice to individual words, but we would not do justice to the language.

As a result, all aspects of each entry are reviewed during the process of revision. It is surprising how often the form of the headword changes, in a language where we think spelling is reasonably consistent. But reviewing a word after one hundred years – which is often the case here – reveals significant changes in hyphenation practice, in the most familiar spelling of words which have entered English from other languages, and in innumerable other minor ways.

Much the same can be said about pronunciation. The First Edition of the dictionary employed a system of phonetic representation devised by the original editor, James Murray. This was maintained through the final supplement of 1972-86, but was replaced in 1989 by a variety of the International Phonetic Alphabet. In addition to the 'Received Pronunciation' provided by the *OED*,



the revised entries introduced a General American pronunciation as well as other regional varieties when particular words demanded this. And we have changed our speech patterns in many ways since the era of the First Edition of the dictionary. Both the typical quality of some vowels, and the stress-patterning of many words, has altered surreptitiously over the last hundred years, and the editors need to resolve these changes, without forgetting the older realizations.

Definitions are reviewed in the light of new evidence and changing resources; etymologies benefit from the discoveries of scholars over the past century, and from the reinvestigation carried out by the *OED*'s own etymological team. Changes, by any of the *OED*'s seventy or so editors, are typically submitted to external scholars for review. Perhaps the most significant changes occur in the realm of documentation, and approximately 60% of the words and meanings in the dictionary are now being provided with earlier attestations as a result of the work being done by editors and also by the wider research and general community interested in improving the documentary quality of the dictionary.

A series of 'before' and 'after' snapshots of *OED* entries ('Rewriting the *OED*') is currently available for consultation on the *OED*'s web site: [http://www.oed.com/staticfiles/Rewriting theOED.pdf](http://www.oed.com/staticfiles/Rewriting%20theOED.pdf).

This proceeds through each aspect of an entry, and explains what sort of changes have been made. Revised and updated entries are published every three months on the dictionary's home web site: [oed.com](http://www.oed.com). We can look at a short entry (in this case *market garden*) to obtain a detailed view of how it has changed during the process of revision.

*Market garden*: how has the entry altered after review and update?

This is how the entry appeared in the First and Second Editions of the *OED*:

**†market dame**, 'a strumpet' (Farmer); **†market-dasher** = *market-beater*; **market economy**, a system of economy which is subject to free competition; **market fair** *dial.*, a fairing or present from the market; **market-fish** *U.S.*, (a) fish, esp. cod-fish, of a marketable size (see quot.); (b) 'a corruption of *margate-fish*' (*Cent. Dict.* 1890); **market-fresh** *a. dial.* (see quot. 1841); **†market-friendship**, a friendship in business affairs; **market-fuddled** *a. dial.* = *market-fresh*; **market-garden**, a piece of land on which vegetables are grown for the market; hence *market-gardener*; **market gardening**, keeping a market garden; **†market geld**, **†geld** (see quotes.); **market-horse** (*slang*), 'a horse simply kept in the betting-lists for the purpose of being betted against' (*Hotten's Slang Dict.* 1874); **market-house**, a building in which a

We can see that the definition is tucked away alongside many other compounds of MARKET *n.*, and is separated physically from its accompanying illustrative quotations, which appear some distance further on in the entry for MARKET *n.*:

[cod] measuring less [than 22 in.], but weighing 3 lbs. or more. **1841** HARTSHORNE *Salop. Ant.* 498 \**Market-Fresh*, that dubious degree of sobriety with which farmers too commonly return home from market. **1651** HOBBS *Govt. & Soc.* i. §2. 4 If they meet for Traffique, 'a certain \*Market-friendship is begotten. **1895** 'M. E. FRANCIS' *Frieze & Fustian* 142 Jem 'was not by any means \*market-fuddled'. **1840** *Penny Cycl.* XVII. 96/1 [The land] would make excellent \*market-gardens. **1865** DICKENS *Mut. Fr.* ii. i, Where the railways still bestride the market-gardens that will soon die under them. **1826** *Westm. Rev.* V. 20 The fruit and vegetables of a \*market gardener, on which his subsistence depends. **1832** *Chambers's Edin. Jnl.* 7 Apr. 76/2 An industrious but rather unsuccessful market-gardener: 'from Kirkcaldy. **1839** *Penny Cycl.* XIII. 188/1 Market-gardeners, who raise vegetables: 'for the supply of the London markets. **1963** *Times* 22 Feb. 10/3 Sir Thomas Playford, orchardist and market gardener, has all the assurance even in the uncertain world of politics of a man who knows that he will reap what he sows. **1875** *Encycl. Brit.* I. 384/1 \**Market Gardening*... The growth of fruits and of culinary vegetables is in various parts of Great Britain an important department of farming. **1958** HAYWARD & HARARI tr. *Pasternak's Dr. Zhivago* i. vii. 218 It is not for the sake of market gardening that we are going all this enormous distance. ?**1500** in Blount *Law Dict.* (1670) s.v., Et valent per an. le Streteward & le \**Marketzeld* xviii. **1684** MANLEY *Cowell's Interpr.* (ed. 2), *Marketzeld*, more truly *Market-geld*, It signifies Toll of the Market.

Deep within this extensive block of quotations it is possible to establish that the First Edition of the *OED* provided two contextual examples of the expression *market garden*, the earliest from the *Penny Cyclopaedia* of 1840 (the same source provides the first attestation of *market gardener*), and another (from Dickens) dated 1865.

Nowadays, a moment's reflection will prompt the realization that a term such as *market garden* (and indeed *market gardener*) is unlikely to have first seen the light of day in a general-purpose encyclopaedia – no matter how up to date the encyclopaedia liked to think that it was. This is of course an issue with any historical dictionary with its origins set firmly in the nineteenth century. The language is so large, and the availability of researchers (even in the leisured days of the nineteenth century) so restricted, that it was often the case that evidence was taken from a reasonably modest selection of sources. This tended to give the impression that these sources (often canonical writers or standard reference texts) had more influence on lexical innovation than was actually the case. The problem can never be entirely overcome, but modern access to raw data on the language helps to ameliorate matters. Ultimately any dictionary relies on the good sense of its readers to interpret the information it provides.

So after one hundred years or so, how did *market garden* progress? Here is the revised entry as it first appeared in the year 2000:

**market garden** Brit. /,mɑːkɪt 'gɑːdn/, U.S. /'mɑrkət 'gɑrdən/ [< MARKET + GARDEN, after *market gardener*.] A piece of land on which fruit and vegetables are grown to be sold commercially. See also *market gardener*.

Now more commonly known in the United States as a truck-farm (see TRUCK 5).

**1811** J. MEASE *Picture of Philadelphia* 349 [There are] extensive market gardens, and several handsome summer retreats. **1865** DICKENS *Our Mutual Friend* I. ii. i. 164 Where the railways still bestride the market-gardens that will soon die under them. **1886** J. F. CONIGRAVE *S. Austral.* 80 The market gardens range from three or four acres in the rich peaty bottoms of gullies up to thirty, and even fifty acres. **1923** R.

HERRICK *Homely Lilla* 197, I can get a job in a market garden. 1948 G. D. H. BELL *Cultivated Plants of Farm* xiv. 129 Many of them are still regarded as market-garden crops. 1987 *Cornwall Rev.* July 12/1 The surrounding countryside is renowned for the earliest market garden produce in mainland Britain.

By this point we can see that the term has been deemed important enough to have its own entry (*market gardener* and *market gardening* have similarly been 'upgraded' to main-entry status). The headword form (spelling) has changed: there is now no hyphen in *market garden*, according to the majority of the evidence available to the editors. A formal etymology and pronunciation have been provided: in this case these are quite formulaic, as there appear to be no additional complications; but the systematic presence of etymologies and pronunciations facilitates subsequent computational analysis and interpretation. The definition has been altered slightly, at bottom in response to societal changes since the date of the First Edition of the dictionary. Finally a more extensive block of documentary evidence has been provided which shows, as we suspected, that the term *market garden* did not first see the light of day in the pages of the *Penny Cyclopaedia*, but has a prehistory in more common-or-garden texts. In this case it is evidenced from James Mease's *Picture of Philadelphia*, first published in 1811. The documentation is now taken well beyond Dickens and 1865 up to the present day, from a variety of types of sources (i.e. not simply canonical authors and reference texts).

One of the advantages of online publication is that we are able to respond to improvements suggested by readers of the dictionary. There must be many people interested in market gardening in the big world, because since the year 2000 the entry has been updated a number of times, and we can see that several improvements have been made.

market garden, *n.*Text size: **A** AView as: [Outline](#) | [Full entry](#)Quotations: [Show all](#) | [Hide all](#)

**Pronunciation:** Brit. /,mɑ:kɪt 'gɑ:dən/, U.S. /'mɑrkət ,gɑrdən/

**Etymology:** < MARKET *n.* + GARDEN *n.*, after MARKET GARDENER *n.*

A piece of land on which fruit and vegetables are grown to be sold commercially. See also MARKET GARDENER *n.*

[Thesaurus »](#)  
[Categories »](#)

Now more commonly known in the United States as a *truck-farm* (see TRUCK *n.* <sup>4</sup> Compounds 1b).

- 1727 B. LANGLEY *New Princ. Gardening* 26 Fourthly, the Market Garden, which produces all Sorts of Pulse.
- 1793 *Proc. Old Bailey* 20 Feb. 427/1 Several of them exactly matched with the tops where they were cut, found in the prosecutor's gardens, and the other gardens, which had no Mazarions growing in it, it being a market garden.
- 1811 J. MEASE *Picture of Philadelphia* 349 [There are] extensive market gardens, and several handsome summer retreats.
- 1864 DICKENS *Our Mutual Friend* (1865) I. ii. i. 164 Where the railways still bestride the market-gardens that will soon die under them.
- 1886 J. F. CONIGRAVE *S. Austral.* 80 The market gardens range from three or four acres in the rich peaty bottoms of gullies up to thirty, and even fifty acres.
- 1923 R. HERRICK *Homely Lilla* 197, I can get a job in a market garden.
- 1948 G. D. H. BELL *Cultivated Plants of Farm* xiv. 129 Many of them are still regarded as market-garden crops.
- 1987 *Cornwall Rev.* July 12/1 The surrounding countryside is renowned for the earliest market garden produce in mainland Britain.

[\(Hide quotations\)](#)

The principal improvement relates to the early documentation for the expression. As time passed after the first release of the revised entry we received notification that *market garden* could be found towards the end of the eighteenth century in the *Proceedings of the Old Bailey*. This remarkable source is available at <http://www.oldbaileyonline.org/>, and has already provided over forty new first references for the *OED* for words relating to everyday life and crime in London mainly in the eighteenth and nineteenth centuries. Examples include *marine store*, *rag-and-bone picker*, and *timekeeper* (in a sports context).

But then, once that example from 1793 had been installed as the earliest reference, we were notified about Batty Langley's use of the expression in his *New Principles of Gardening* (1727). Langley was something of an eccentric garden designer, and it is not surprising that a new term of this nature finds a place in his work. But even now a single new discovery might take our knowledge of the word back even further. At much the same time (2007) an equivalent quotation from Langley's text was added to the entry for *market gardening*. The other change to *market garden* since original revision is that the 1865 Dickens quotation has (along with many other quotations from the same source) been redated to 1864, in line with its first publication in periodical form.

These are the sorts of changes which are happening to *OED* entries both as they pass through the process of revision, and subsequently. The intention is to revise and maintain the *OED* as a dynamic resource which can respond rapidly both to actual changes in the language and to the emergence of new information about older features of the language.

## Benefits of online publication

The previous section has mentioned some of the benefits of online publication. There are others as well. Overall, the advantages of online publication include:

- the ability (on the part of editors) to be in control of the form of the entry at any stage in the editorial process
- the option (on the part of editorial managers) to select those entries or entry ranges for editing according to any of a number of criteria, from across the whole sweep of the dictionary (rather than in a narrow alphabetical range)
- the choice to publish only those entries which the publisher wishes to publish at any one time (for whatever reason)
- the use of tools which manage the workflow of the editorial process and report on any aspect of this
- the chance to re-edit material after it has been revised and published, in response to new discoveries
- the facility to alter the display, processing procedures, structure, tagging, or any other feature of the background software in response to external circumstances
- the maintenance of online communication with academic consultants and other contributors to the dictionary

These are some of the more traditional advantages of online working. We are finding out about others as we move onwards.

## How does this all fit in with the philosophy of the OED?

The *OED* is a historical dictionary; its original title page stated that it was based ‘on historical principles’. It looks at words over time, plotting changes as they occur. It looks back from the point or points at which a borrowed word enters English to the language from which it came, seeking for the semantic genes that it brought with it. It asks (silently, but emphatically all the same) why a word or a sense arose when it did: what were the conditions within the language and within the society which used that language that were favourable to the introduction of the term? And most importantly, it looks for logical consistency in the changes which a word experiences over time. We don’t expect changes to happen for no apparent reason. Changes cannot be predicted but they can normally be explained.

The word *cell* is helpful in explaining something of the philosophy of the *OED*. There are three nouns and one verb in English spelt *cell*, and we will look at the earliest of the nouns (containing senses such as a monk’s room, a small group of revolutionaries, and the fundamental unit of living organisms consisting of protoplasm encased within a membrane). Even a short review of the range of meanings of *cell* brings home to us the semantic breadth that the word encompasses. Words normally only achieve such breadth after many years of development in the language. And this is the case with *cell*. It was introduced into English in the Old English, pre-Conquest, period from Classical Latin *cella*, later reinforced by various French forms of the word. The ‘standard’ French form *celle* is no longer common in French, having been superseded earlier by *cellule* (from which we obtained yet another word). That, of course, is largely irrelevant once the word is embedded in English.

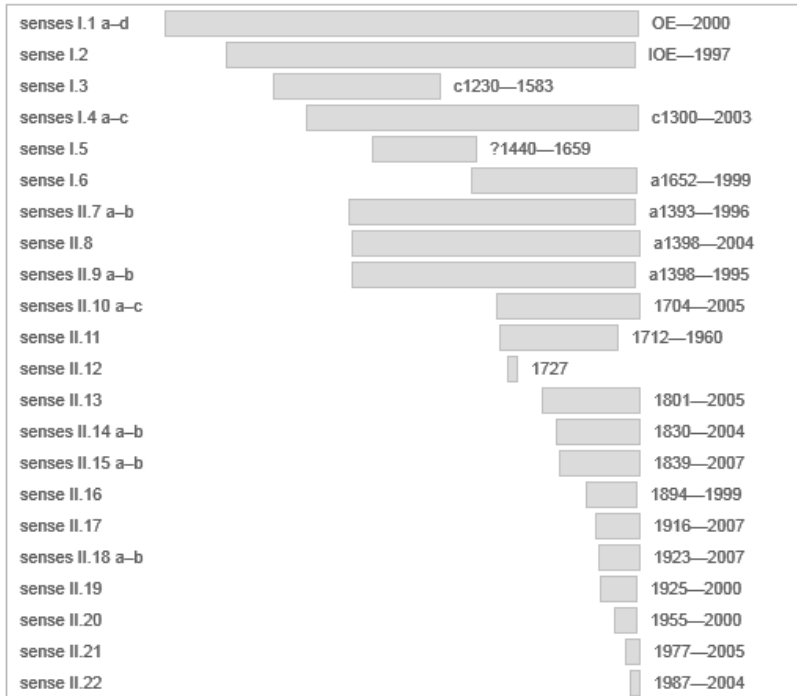
From our point of view it is very relevant that most of the meanings of the word to do with rooms or chambers already existed in Latin and French before the term took hold in English. We were simply taking over an existing word in existing meanings, and were not witnessing illogical, inexplicable change. Even the later senses ‘a honeycomb’ and ‘a compartment of the brain’ predate their English use in these donor languages. These meanings in turn seeped from Latin into later Romance languages across Europe. What we see as a major development in English was in fact happening throughout Europe along the established pathways of diffusion.

So even this pre-English section of the entry for *cell* touches on a fundamental principle of the *OED*: words very rarely arrive out of nowhere. They metamorphose, from other languages or in parallel with similar uses in other languages, or as a result of interaction within English, or in response to changes in the society or culture that uses them. What we are looking at is an interdepend-

ent network of lexical features. The editor's task is to piece together the individual aspects of a word, and let it take its place in the interlocking analysis of language represented by an online dictionary.

The *OED's* entry for *cell* includes a graphical summary of its development within English, available online. Here is that representation as it appeared in December 2011:

#### Timeline of sense development



The graph shows the twenty-two major sub-senses of *cell* (in the vertical axis) plotted against time (in the horizontal axis). The bar representing the earliest sense (“a dwelling consisting of a single chamber inhabited by a hermit or anchorite”) travels from the far left to the extreme right, depicting use all the way from the Old English period into the twenty-first century. The bar for the second sense (“a monastery or nunnery, generally of small size, dependent on some larger house”) is slightly indented – indicating that it is first recorded slightly later, in the late Old English period – and again ranges across to the present day.

The entry has been structured in two major branches: senses 1-6 cover cells which are rooms or chambers; senses 7-22 cover extended uses. The sub-senses have been ordered, within this structure, according to their date of earliest recorded use. As a result one would expect the graph to taper towards the bottom right-hand corner.

One surprising feature is that once a sub-sense starts, it normally continues through to the present day (even though in fact the sense may now be archaic or historical in use). There is not much evidence, for this particular common word, of sub-senses dying out on their march to the present day: this happens to two of the six 'room' senses and only one of the sixteen extended uses. Indeed, that particular extended use is a poorly attested use of uncertain meaning.

The first branch of meanings (senses 1-6) portrays a semantic development from a single room (inhabited by a hermit) to a monastery or nunnery which is dependent on a larger house. As the Middle Ages evolve, we find a new sense, taking aspects of these earlier uses – any of many single rooms in a monastery or nunnery occupied by a member of the religious community. The Middle English religious usages remain important features of the word, and we wait till the eighteenth century for the next major development of the 'room' meaning.

By the eighteenth century prison reformers were looking for a term to describe the new solitary rooms designed in forward-thinking prisons. *Cell* was the obvious term: prisoners were to be held isolated in rooms regarded as comparable to a monk's cell. Each individual cell was one of a large number of such chambers within the prison. Changes in society and its thinking about prisons fostered a change in the language which has endured to the present day.

The other branch of meanings for *cell* relates to things which are not room or chambers, but were considered to share some of the characteristics of these. The graph above shows the word *cell* had existed in English for several hundred years before this new branch developed. But after a while we start to find *cell* being applied in new circumstances: to any of the tiny compartments imagined to exist (according to medieval doctors) in the brain. The concept of a cell being one of a number of small compartments was simply transferred (in various European languages) from the context of buildings to the architecture of the brain. Similarly the word *cell* was applied to the compartments of a honeycomb – rather like the cells in a monastery seen through a reducing glass.

Throughout this development over time the *OED* expects to find a logical (or at least an understandable) connection between sequentially related sub-senses. The motivation for a new sense may be external (new technology, historical events, etc.), but the new sense is normally just an extension of something that already exists – with a new twist.



As early as the late Middle Ages *cell* took on a wider meaning in medical science, as it came to mean any large chamber in the structure of a tissue or organism. With the invention of the telescope there was further room for *cell* to expand, this time to describe the cavities and chambers scientists such as Hooke could see through their new magnifying devices. Once *cell* became associated with microscopic science, it experienced several metamorphoses in the Renaissance within the same semantic area until finally becoming the key word in the burgeoning biological sciences of the early nineteenth century: ‘the fundamental, usually microscopic, structural and functional unit of all living organisms, which consists of a small quantity of protoplasm enclosed within a membrane, typically contains a nucleus or nucleoid and other organelles and internal compartments, and is capable of utilizing energy, synthesizing proteins and other biomolecules, and (usually) replicating itself’. Subsequent changes took *cell* into the world of computing and mobile (‘cell’) phones.

*Cell*, therefore, has existed in the English language for well over one thousand years. During that time it has developed from a set of related meanings centred around ‘room’ or ‘chamber’, recorded first in the Old and Middle English periods, to become one of the key words of modern biological science and mobile-phone technology, by way of a linked set of sub-meanings which move us semantically through different compartments and cavities that were significant at different periods of the word’s history. Generally speaking, comparable developments can be identified throughout the vocabulary of any language observed over time. It’s one of the ways language works.

## Has the role of the editor changed over the last one hundred years?

There are many ways in which the editor’s job has remained virtually unchanged over the lifetime of the *OED*. However the documentary evidence for the language is obtained and analysed (on index cards or on-screen; by organizing slips of paper in different piles or through computational analysis of syntax and meaning), the editor still needs to apply much the same thought processes to the data today as one hundred years ago. Many little facts have to be reviewed in order to build the big picture of a word; apparent logical or chronological inconsistencies in semantic development have to be re-examined and explained; definitions have to be written elegantly but accurately; there is

not time to extend research on a word or a meaning beyond practical limits, or nothing will ever be completed. We work on the assumption that we complete as much work on a word as we can, as rapidly as we can, and then review and publish the result. There will be times when further information might improve the entry, but it would be impractical to extend the research to encompass this possible gain. Readers will identify gaps and will provide us with the intervening data. We can't do everything ourselves, and it would be presumptuous to think that we could.

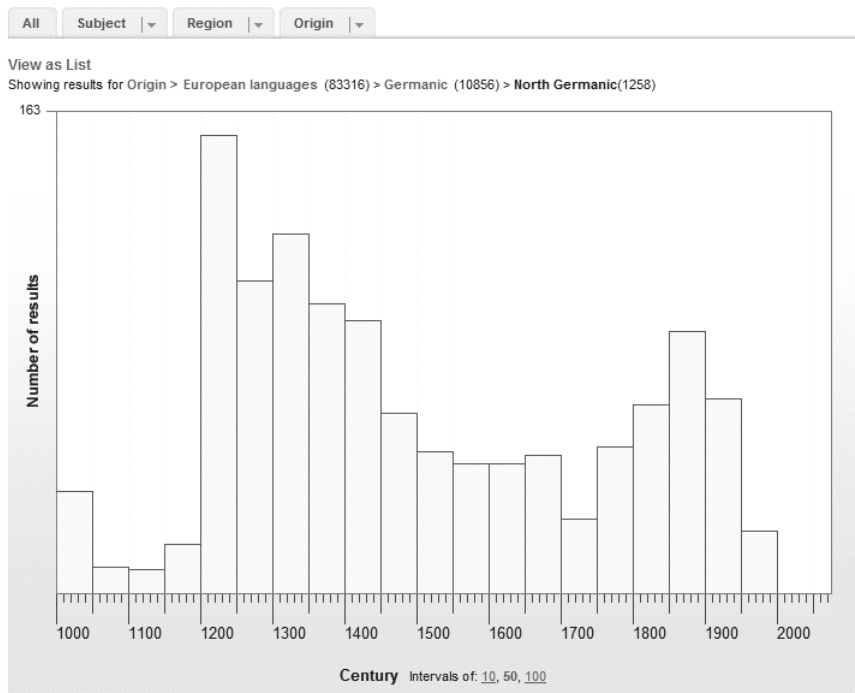
Editors from the original edition of the *OED* in the late nineteenth century would not have a great deal of difficulty accommodating themselves to work on the twenty-first century *OED*, once they had assimilated the enormous cultural changes since their time. The principal differences are indeed the cultural perspective, which includes a less Britain-centred editorial perspective, and the sense that the editor is not working on a word in isolation, but on a word which forms part of a much larger network of language. The computer makes it easier to see this extended view.

## A new mode for searching the OED

When the *OED* relaunched its online site in November 2010 it took advantage of the current search mode of using filters to narrow down results. This allowed the user to refine results in a way that was often possible on the former online system (available since 2000), but which involved very complex search routines. The new system had the benefit of simplicity.

Combined with filtered searches, the relaunched web site was able to represent results of searches not only as a list of entries (or other category) which matched the search request, but also by means of timelines which offered the user a visual representation of the results. This compressed view typically gives an instant, visual, impression of the outcome of a search, which can in turn be refined according to other criteria.

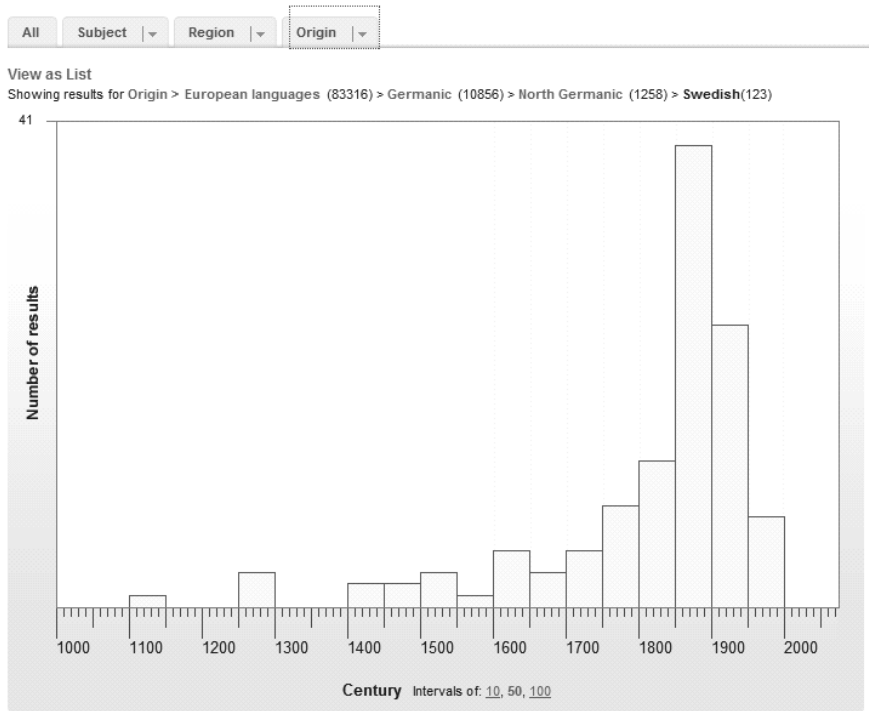
## Timelines



This timeline shows the user the dates at which words are recorded in the *OED* as entering English from the Scandinavian languages. There are a small number in the Old English period, then a significant mass from 1200 to 1450 AD, followed by a reduction but not complete absence in later centuries. The user is therefore asked to consider the underlying facts which cause this pattern to be created. Why, for example, are more words of Scandinavian origin recorded in English in the Middle Ages than in the Old English period when (from the history of Viking raids) we might expect to see a more dominant number. The answer lies elsewhere, in the mechanics of language transmission, in the power relationship of English with (Anglo-)French, etc.

If we then employ a filter, and reduce this Scandinavian word set down to the Swedish words recorded as entering English, we find a different 'shape' to the graph:

## Timelines



This shape shows that the heaviest period for the introduction of Swedish words (as opposed to those from the other Scandinavian languages) appears to be the late 19th century. Again, what are the reasons for this? How much do we need to discount the fact that this was the period just prior to the compilation of the First Edition of the *OED*? There are 123 words in this group, of which 66 are recorded as entering English between 1850 and 1950. What are these words? The 1850-1950 cohort includes a large number of minerals and chemicals (signifying the importance of Sweden in that period in the geological and chemical sciences), along with various 'cultural' terms which found their way into English for a number of reasons – words such as *Lapp* (Sw. lapp, same), *ombudsman*, *Riksdag*, *sloyd* (Sw. slöjd), *smorgasbord*, and *trolldom*. We can compare this set with the much smaller set (five) of Swedish words recorded as entering English in the period 1600-1650. What sort of words were these? They were: *barling* (Sw. bärling, a pole used in boat-building), *gantlope* (Sw. gatlopp), *rot* (Sw. rote), *skol* (Sw. skål), and *trow* (Sw. troll). At this period we see a different prolife in the English perception of Sweden: military and sea-faring terms, drinking, and

the hobgoblins of mythology. In different periods we see different influences on the interaction between peoples. The obscure Swedish loanword *gantlope* went on, after minor surgery, to have an illustrious career in the common or garden English expression ‘to run the gauntlet’. It would be extremely interesting to be able to view the comparable list of words borrowed into Swedish from English in those time periods.

The online site presents a different picture of the data every three months, when the *OED*’s database is updated with newly revised and freshly written entries. So although these timeline graphs show the state of play in mid 2011, the equivalent graphs for 2012 will be perceptively different, as a result of new material, new analysis, fresh discoveries.

## Authorial innovation

The filter and timeline system is not restricted to borrowed words. Another filter allows the user to see a chart of the thousand most frequently cited authors and sources in the dictionary. Again, this timeline needs interpretation: the output of a writer such as Jane Austen was (in terms of words) much lower than say that of Charles Dickens, so one would not expect to find Jane Austen as high on the chart as Dickens: in fact they stand at positions 13 and 251 out of 1,000 respectively, with Austen credited with the first recorded use of 44 words, and Dickens 258. The statistics on the chart change every three months, along with the rest of the dictionary data on the site, in line with the latest editorial alterations and additions.

500.	Richard Surflet	fl. 1600–1616	908	110	301
501.	Helkiah Crooke	1576–1648	907	160	423
502.	Mary Wortley Montagu	bap. 1689, d. 1762	903	33	124
503.	Roger Ascham	d. 1568	901	97	317
504.	Isaac D’Israeli	1766–1848	894	27	86
505.	William Withering	1741–1799	892	47	273
506.	Hugh Miller	1802–1856	892	45	189
507.	Herman Melville	1819–1891	890	60	251
508.	John Wilkins	1614–1672	889	132	352
509.	Newsweek	1933–	888	71	204
510.	George Turberville	c1543–c1597	884	151	360

Here we are at the mid-point of the chart. The columns represent, in order: sequence number, name of source, date(-range) of source, total number of quotations, earliest quotation for a word, earliest quotation for a sense.

From this it is possible to see, for example, that although Mary Wortley Montagu and Isaac D’Israeli are both cited around 900 times in the entire dic-

tionary, they both have a low ratio of first usages (word and sense). George Turberville (the sixteenth-century writer of animal husbandry), however, has a much higher recorded innovation ratio. The reader may ponder the lexicographical and real reasons for these differences.

One point that is a consistent feature of the revision is that the major, canonical authors of the language are gradually having their positions eroded. This is not to say that we think any less highly of them as literary figures, but that their hegemony is being rightly challenged by the mass of less literary sources to which editors can now turn for historical evidence of the language.

## James Joyce

James Augustine Aloysius Joyce (1882–1941), writer.

The 144th most frequently quoted source in the OED, with a total of 2344 quotations (about 0.07% of all OED quotations).

[See article in the Oxford Dictionary of National Biography](#)

Total number of quotations	2344	<a href="#">Find quotations</a>
Quotations providing first evidence of a word	102	<a href="#">Find quotations</a>
Quotations providing first evidence of a particular meaning	483	<a href="#">Find quotations</a>

Most quoted texts		Quotations: first in entry	Quotations: first in sense	Quotations: other
Name	Total Quotations	Percentage of quotations		
1. Ulysses	1671	71.0%		<a href="#">Find quotations in OED</a>
2. Portrait of Artist	174	7.0%		<a href="#">Find quotations in OED</a>
3. Finnegans Wake	169	7.0%		<a href="#">Find quotations in OED</a>
4. Let.	146	6.0%		<a href="#">Find quotations in OED</a>
5. Dubliners	123	5.0%		<a href="#">Find quotations in OED</a>

Here is James Joyce's 'page' on the OED Online web site. As of December 2011 it shows Joyce as the 144th most-cited 'source' in the *OED*. He has dropped one place since mid 2011, and I would expect the gradual slide to continue, as with almost all other celebrated writers.

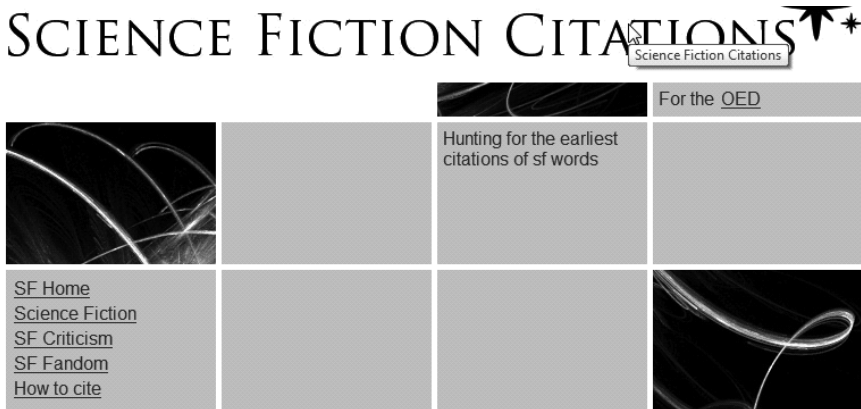
Joyce is currently shown as the first recorded reference for 483 words and meanings. During revision around 90 similar 'first usages' have been 'lost' to other sources, especially those newly available for searching on online historical databases. By contrast, new usages by the author find their way on to the database, buoying up but not maintaining his ranking.

As time passes, we shall eventually have a far better view of the types of words and meanings where Joyce (and other canonical authors) are truly innovative.

## Crowdsourcing

The *OED* has always crowdsourced, encouraging people throughout the world to increase our knowledge of the language by contributing directly to the dictionary. As we think about ways in which the *OED* will develop in the future we consider the possibility of setting up a wiki, to allow people to post their own earlier quotations, comments on etymology or definition, etc.

Early experiments with a web site designed for science-fiction enthusiasts, at which anyone can contribute new material on the subject to the *OED*, have proved extremely useful.



For several years people have been able to suggest new sci-fi usages, or to supplement existing ones, at <http://www.jessesword.com/sf/>. Here, for example, is the site's record for *meteor storm*:

<b>Definition</b>	a particularly intense meteor shower, esp. occurring when the earth intercepts a meteor stream soon after the originating comet has passed
<b>OED requirements</b>	antedating 1872
<b>Earliest cite</b>	Richard Proctor, "Meteors and Shooting-stars"
<b>Comment</b>	<p>Mike Christie submitted a 1948 citation from Mack Chapman Lea's "The Gorgons".</p> <p>Roberto Labanti submitted a 1938 cite from a newspaper report (The Hammond Times [Hammond, Indiana], April 20, 1938).</p> <p>Bill Mullins submitted an 1877 cite for "meteor-storm" from the "History of Kendall County, Illinois".</p> <p>Roberto Labanti submitted a cite from the Washington Post, November 19, 1928.</p> <p>Bill Mullins submitted a cite from the Worcester Spy, November 2, 1903.</p> <p>Bill Mullins submitted a cite from the Wheeling Register, December 20, 1896.</p> <p>Adam Buchbinder submitted a cite from Richard Proctor's article "Meteors and Shooting-stars" in the Eclectic Magazine for November 1872.</p> <p>Earliest cite in OED2: 1977. Antedated to 1948 in OED3 in March 2003.</p>
<b>Last modified</b>	7 May, 2009

The current online version of the entry for *meteor* shows the effect that work on this web site has had. The original version of *meteor storm* dated the expression

from 1977, as the page above states. The page then continues to narrate the contributions provided by users of the site, on the way to providing the current earliest example of the term in 1872. As usual, we know that around the next corner lies yet another earliest example – for this or any other word.

We are considering the possibility of increasing the number of these specialist sites, as well as offering a general facility to post new material on or in parallel with the main *OED* site.

## Extending the network

Traditionally dictionaries link between related entries by means of cross-references. Online this mechanism can be enhanced by simple techniques such as pop-up windows showing the meaning to which a cross-reference is given (and therefore making it unnecessary in all cases to click through to the entry):

- 1823 P. NICHOLSON *New Pract. Builder* 386 The Rod is from five to ten feet in length, and is used for measuring
- 1844 H. STEPHEN as to ascertain the depth of the
- 1898 H. GANNET and hinged so
- 1922 B. H. M. H. in use · the
- 2000 J. N. WILF measuring

### Cross Reference

#### Napier's bones, *n.*

Several narrow rods or strips of bone, ivory, wood, or other solid material, divided into compartments marked with certain digits, and formerly used to facilitate the operations of multiplication, division, and the extraction of roots according to a method invented by John Napier. The method was described by Napier in his book *Rabdologiae seu Numerationis per Virgulas Libri Duo* (1617).

(Hide quotations)

**b.** In *pl. Math.* = [NAPIER'S BONES](#) *n.* **Occas. in *sing.*:** each of the sticks making up a set of Napier's bones. **Now *hist.***

Thesaurus »  
Categories »

- 1618 T. BRETNOR *Neue Almanacke & Prognost. To Mathem.*, There is ·· an excellent treatise of Rabdologia, or Arithmeticall Rods, inuented and published by the learned Lord of Merchiston.
- 1667 W. LEYBOURN *Art of Numbring* i. 1 Rabdologie is the Art of Counting by Numbering Rods.
- 1709 G. VAUX tr. J. C. Sturm *Mathesis Juvenilis* I. v. 34 Some make these Rods in the form of Cylinders (keeping the same Order of the Numbers) but I find it more convenient and ready to make them of thin Columns of Paste-board.
- 1825 *Ann. Philos.* New Ser. 9 147 A box of rods, named the Rhabdological Abacus. ·· These rods were invented by Mr. Goodwyn for the purpose of facilitating the multiplication of long numbers.
- 1911 *Encycl. Brit.* XIX. 174/2 Each rod therefore contains on two of its faces multiples of digits which are complementary to those on the other two faces.
- 2004 B. BUNCH & A. HELLEMANS *Hist. Sci. & Technol.* 180/2 It can be used by moving several rods to show a pair of factors; the product then can be read from other numbers on the rods.

(Hide quotations)



Here we have an instance (at the entry for *ROD n.*) of a definition provided as only a cross-reference to another entry. By floating the cursor over the entry online, it is possible to read the intended definition without having to migrate to the relevant page.

## Links by means of a thesaurus

We are very fortunate that a new linking device, quite in keeping with the new idiom of linking around online web sites, has been provided for the *OED* by the compilation and publication of the *Historical Thesaurus of the OED*, a project initiated by Professor Samuels at the University of Glasgow and finally brought to fruition in 2009.

The project sought to re-order the semantic landscape from a sequence which was word-based and alphabetical to one which was concept-based, in the manner of a thesaurus. Since the project was dealing with the *OED*'s historical data, the objective could include the ordering of words chronologically by concept, so that all words in the dictionary with the general sense 'spoon' could be listed together, and the user could see when and how these usages entered and feel out of use in English.



This is a snapshot of the online database which lists the thesaurus structure around the term 'spoon' as an item of cutlery. By clicking on 'spoon...types of'

the user is taken to this list of types of spoon (illustrated here only up to the mid eighteenth century):

maidenhead	1495
<i>spec.</i> A representation of the head or head and shoulders of the Virgin Mary, forming an ornamental finish to the handle of a spoon; (occas.) the spoon itself. <i>Now hist.</i>	
slip	c1530
A spoon-handle having the top cut off obliquely; a spoon with a handle or stem of this form. (Cf. <i>SLIPPED adj.</i> 1.) <i>Now Hist.</i>	
Apostle spoon	1631
A silver spoon, the handle of which ends in a figure of an Apostle.	
tea-spoon	1686
A small spoon, usually of silver or silvered metal, of a size suitable for stirring tea or other beverage in a cup.	
hall-spoon	1688
a spoon made of hall-marked silver.	
pap spoon	1691
tablespoon	?1741
A large spoon (larger than a dessertspoon) for serving food. Also occas.: a (smaller) spoon used for eating.	

At present 'tea-spoon' is recorded half a century before 'tablespoon', but various terms which are much less familiar (and which we might not have thought of consulting), such as 'slip' and 'pap spoon', are shown too.

The thesaurus module is an extremely strong 'add-on' to the *OED*, available to all users. It provides a new way of looking at the vocabulary of English across the centuries. And yet again, the data is updated every quarter, so that new discoveries will cause the display of data to change and better reflect the reality of the language today and in the past.

## A new view of the language

What this new technology means is that we are able to revise and update the dictionary with a mass of detail unavailable to the original editors of the *OED*, and at the same time to present the information in new ways too. These new ways offer not just a look-up dictionary, in which finding the information looked up represents the end of the search. Rather, it presents the dictionary as a land of exploration, in which links are followed through a variety of different types of searches, and the network of language is much more successfully harnessed to provide answers not just to the little questions, but to larger questions to which the answers were previously entirely swamped by the mass of specific detail.

Beyond this, we are now in a much better position to review users' real uses for the dictionary, both as a guidepost in a system of online text and other data, but also as a repository for new information, provided perhaps through a wiki or other crowdsourcing medium, which will allow the dictionary to become more accurate, and more dynamically respond to the needs of the user.<sup>1</sup>

*John Simpson*  
Oxford English Dictionary.  
john.simpson@oup.com

---

<sup>1</sup> Illustrations of the *OED* both online and in other formats are published by kind permission of the Secretary to the Delegates of Oxford University Press.

# Strategier for funktionsbaseret betydningsordning

*Birger Andersen*

Traditional lexicography has established three general principles for sense ordering – historical, frequential and logical – but has not generally attempted to connect these principles to issues such as dictionary user, including user situation, and dictionary function. An analysis of a variety of monolingual and bilingual dictionaries shows that the logical and frequential principles are used in dictionaries from each type, and that the logical principle is used with widely varying degrees of ‘lumping’. This is attributed to the fact that the compilers of the dictionaries have been ‘forced’ to apply one of the general principles of sense ordering, since these dictionaries are designed as polyfunctional and often without any clear perception of user characteristics. Electronic dictionaries offer the possibility of dynamic data presentation in which sense ordering can be made sensitive to both dictionary function and dictionary user characteristics, and the article outlines proposals for functionally based sense ordering in a dictionary project involving the establishment of a database capable of generating three monofunctional English phrasal verb dictionaries.

*Nøgleord:* Historisk, frekventiel og logisk betydningsordning. ‘Lumping’. ‘Splitting’. Elektroniske ordbøger. Ordbogsfunktion. Ordbogsbruger. Funktionsbaseret betydningsordning.

I forbindelse med den ordbogsmæssige behandling af polyseme opslagsord er der i den leksikografiske litteratur gennem årene efterhånden skabt en tilsyneladende konsensus om, at der eksisterer tre generelle principper for betydningsordning:

1. Historisk (eller kronologisk) betydningsordning, som regel med den kronologisk ældste betydning som den første. Det er dog ofte blevet fremhævet, at dette princip kan være vanskeligt at arbejde med, da det i mange tilfælde er umuligt at afgøre den eksakte historiske rækkefølge af et givet ords betydninger. Uanset disse vanskeligheder er det naturligvis et princip, der som

- hovedregel forsøges anvendt i etymologiske ordbøger, men princippet er også blevet anvendt i traditionelle almene monolingvale ordbøger.
2. Frekventiel betydningsordning med den mest frekvente betydning som den første. Princippet anvendes som regel i forbindelse med det der kaldes 'splitting', dvs. skarpt adskilte betydninger uden markering af tilfælde af beslægtede betydninger. Det er blevet hævdet, at dette princip som hovedregel ses anvendt i f.eks. de engelske monolingvale learner-ordbøger, men som det senere vil blive påvist, kan der herske tvivl herom.
  3. Logisk betydningsordning, hvor der gøres forsøg på at identificere en række kernebetydninger, der følges af en række afledte betydninger. Der kan være tale om flere forskellige typer afledte betydninger, f.eks. fra det generelle til det specifikke, fra det konkrete til det abstrakte, fra det bogstavelige til det metaforiske, og så videre.

I den traditionelle leksikografi er det logiske og det frekventielle princip for betydningsordning af mange blevet anset for globalt anvendelige principper, uanset ordbogstype, og principperne er sjældent i den teoretiske diskussion blevet konkret koblet til aspekter såsom ordbogsbruger, ordbogsfunktion og brugersituation, men der er dog enkelte undtagelser.

Hornby (1965: 105) siger f.eks. om betydningsordning i monolingvale learner-ordbøger: "For those to whom English is a foreign language, those who are learning [...] English for present-day needs, the order in which semantic varieties are entered should perhaps be based on frequency rather than on historical principles. Such users of a dictionary are more likely to meet, and to need to use, words in their current senses. These are the senses, therefore, which should be entered first."

Og Kipfer (1983: 101) har følgende at sige om spørgsmålet: "There are arguments for and against each of these ways (de generelle principper for betydningsordning, *min tilføjelse*), and which of them the editor chooses depends on the ultimate purpose and use of the dictionary." Kipfer har dog ingen konkrete anvisninger på, hvilke kriterier der kan bringes i anvendelse for at tilpasse betydningsordningen til ordbogens formål eller brug.

Der er også eksempler på, at der argumenteres for et givet princip på baggrund af rent lingvistiske, og ikke leksikografiske, principper. Moerdijk (2003: 278) taler for en høj grad af logisk strukturering af ordbetydninger (også kaldet 'lumping') på følgende vis: "Because they want to account for too many of such various contextual semantic contributions in their dictionaries, most lexicographers are prone to split lexical meanings into too many different senses, whereas in fact they could and should, at least according to a number of semantic theorists, bring them together under fewer and more general meanings."

Det vil i det følgende kort blive illustreret, hvilke principper for betydningsordning der synes at være anvendt i en række forskellige ordbogstyper. Der er tale om tre ordbogstyper:

1. Almene engelske monolingvale ordbøger
2. Engelske monolingvale learner-ordbøger
3. Bilingvale engelsk-danske ordbøger

Illustrationerne vedrører opslagsordet *suit* som substantiv. Der er i oversigten over principper for betydningsordning, (Appendix 1), taget udgangspunkt i for det første det lineære princip med skarpt adskilte betydninger uden afledte betydninger, hvor rækkefølgen af betydningerne kan være frekventiel, men ikke nødvendigvis er det, og for det andet det hierarkiske princip med en række kernebetødninger med et antal afledte betydninger. Jeg skelner her mellem et lavt, et mellem, og et højt antal afledte betydninger, dvs grad af 'lumping'. Vurderingen hviler i dette tilfælde på et ret spinkelt grundlag, da opdelingen alene baserer sig på, hvordan de enkelte ordbøger behandler opslagsordet *suit*.

Det skal bemærkes, at rækkefølgen af kernebetødninger kan være frekventielt bestemt, hvilket er anført i skemaet i de tilfælde, hvor ordbogen eksplicit anfører, at dette er tilfældet. For en enkelt ordbog er der ingen eksplicit markering af, hvilket princip der er anvendt.

Der findes i Appendix 2 en gengivelse af, hvorledes udvalgte ordbøger behandler opslagsordet *suit* med udeladelse af irrelevante data.

Hovedparten af de almene monolingvale ordbøger benytter sig af det hierarkiske eller logiske princip, men med forskellig grad af 'lumping'. To ordbøger, Collins English Dictionary og Concise Oxford Dictionary of Current English, angiver, at man for kernebetødningerne har brugt det frekventielle princip.

Når det for Penguin Office Dictionary anføres, at der er tale om et højt antal afledte betydninger, skyldes det for det første, at den første kernebetødning ('tøj'-betødning) består af 3 underbetødninger, som i f.eks. BBC English Dictionary anføres som 3 separate betydninger, og at den femte kernebetødning består af 2 underbetødninger, der f.eks. i Collins English Dictionary anføres som 2 separate betydninger, nemlig betødning 7 og 9.

Den eneste almene monolingvale ordbog, der anvender det lineære princip er BBC English Dictionary, men ordbogen anfører intet om, hvorvidt de skarpt adskilte betydninger er ordnet efter frekvensprincippet.

For de engelske monolingvale learner-ordbøger er billedet en lille smule mere homogent. Langt hovedparten anvender et hierarkisk princip med et lavt eller middel antal afledte betydninger. Betydningsafledningerne vedrører alle *suit* i 'tøj'-betødningen. Historisk set kan man iagttage en udvikling henimod et

lavere antal afledte betydninger, altså en lavere grad af 'lumping', som f.eks. fra første til femte udgave af Longman, hvor betydning 1b i første udgave er skilt ud som separat betydning i femte udgave, og fra tredje til ottende udgave af Oxford, hvor det samme er sket. Kun Collins Cobuild Advanced Dictionary organiserer sine ordbetydninger efter et strengt lineært princip, dog uden eksPLICIT at angive, om princippet samtidig bygger på frekvens.

For de bilingvale engelsk-danske ordbøger er billedet igen ret broget fra Kjærulff-Nielsens hierarkiske opbygning med mange afledte betydninger, altså en ret høj grad af 'lumping', til den lineære, frekventielle opbygning i Politikens Engelsk-dansk ordbog med betydningsforklaringer.

Der er generelt set tale om et ret broget billede, hvilket efter min opfattelse skyldes, at vi i alle tilfælde har at gøre med trykte, polyfunktionelle ordbøger, med – også i flere tilfælde – en diffus brugergruppe. Ordbogsforfatterne har været tvunget til at vælge et af de generelle principper uden støtte i mulige overvejelser om specifikke funktioner for ordbogen, og i mange tilfælde også uden støtte i mulige overvejelser over brugerkaraktistika.

De dilemmaer, som forfattere til polyfunktionelle ordbøger med en diffus brugergruppe står over for, illustreres muligvis bedst ved et par citater fra den teoretiske litteratur.

Kipfer (1983: 107 f.) siger f.eks.: "There is no single system which would be both powerful and detailed enough to be used unequivocally and alone as the basis for ordering senses and which would command general authority and recognition [...] One problem is that all words cannot best be presented by any single method, that is it would probably be best if some words were presented in chronological order, others were presented in decreasing order of frequency, and still others presented by grouping basic meanings together into sub-categories [...]"

Samme opgivenhed sporer man hos Svensén (2009: 364), der siger: "Among the principles of arrangement discussed above (for betydningsordning, *min tilføjelse*), there is really none that by itself will deal with all problems; in practice, the lexicographer has to apply different principles in different situations, and sometimes perhaps even a combination of several of them."

Spørgsmålet om principper for betydningsordning kan imidlertid gøres yderst relevant i forbindelse med elektronisk baserede ordbøger, der giver mulighed for alle former for dynamisk datapræsentation. Lew (2009) diskuterer muligheden for variabel betydningsordning baseret på ordbogsfunktion og med inddragelse af brugerkaraktistika, hovedsageligt i elektronisk baserede monolingvale learner-ordbøger for 'advanced learners'.

Det skal indledningsvist siges, at hele diskussionen om betydningsordning hviler på den antagelse, at selv ret erfarne ordbogsbrugere generelt har en ten-

dens til at standse deres søgning ved den første betydning eller de allerførste betydninger i en given ordbogsartikel, specielt hvis der til den enkelte betydning er knyttet et stort antal data som f.eks. grammatiske data og eksempler. Dette er underbygget i en lang række undersøgelser af ordbogskonsultation i trykte ordbøger (se Lew 2004), og man må alt andet lige antage, at det samme gælder for elektroniske ordbøger. Det er derfor vigtigt at arrangere betydninger således, at ordbogsbrugeren får de betydninger først, der er relevante for det formål, hvor med ordbogen konsulteres.

For ordbogsfunktionen tekstreception argumenterer Lew for en betydningsordning ud fra frekvens med den mindst frekvente betydning først og den mest frekvente betydning sidst, da man må antage, at 'advanced learners' allerede er bekendt med de mest frekvente betydninger og derfor hovedsageligt i en receptionssammenhæng har brug for assistance til afkodning af de mindst frekvente betydninger.

For ordbogsfunktionen tekstproduktion foreslår Lew, at betydningerne ordnes efter faldende frekvens, da ordbogsbrugeren som regel i sin tekstproduktion vil anvende frekvente betydninger og i den sammenhæng har brug for grammatiske og stilistiske informationer om de højfrekvente betydninger.

Disse principper for betydningsordning kan gøres endnu mere specifikke for mere specialiserede ordbøger. Andersen (2010) beskriver et ordbogsprojekt, der indebærer udarbejdelsen af en database for engelske frasalverber, der skal være i stand til at generere tre elektroniske monofunktionelle ordbøger, en for tekstreception, en for tekstproduktion og en for oversættelse fra engelsk til dansk.

Produktionsordbogen har som primær målgruppe danske studerende, herunder engelskstuderende, på universitetsniveau, der under deres studium skal skrive essays, rapporter, og så videre, på engelsk. Receptionsordbogens primære målgruppe er også danske studerende på universitetsniveau, der under deres studium forventes at læse store mængder engelsksproget litteratur. Endelig har oversættelseordbogen to målgrupper: på den ene side professionelle oversættere og på den anden side engelskstuderende på universitetsniveau, der foretager oversættelser fra engelsk til dansk af forskellig art som del af deres studium. Med hensyn til brugerkaraktistika for de studerende kan det siges, at de generelt har et højt niveau med hensyn til engelsk sprogfærdighed.

Endelig skal det siges, at der i ordbøgerne er valgt det overordnede lineære princip for strukturering af betydninger, altså skarpt adskilte betydninger uden afledte betydninger. Efter min opfattelse giver det lineære princip en langt mere overskuelig opbygning af den enkelte ordbogsartikel end det hierarkiske princip.



Som sagt er der tale om specialiserede ordbøger i den forstand, at de som opslagsord har et meget specifikt udsnit af det engelske ordforråd, nemlig frasalverber – på engelsk 'phrasal verbs'. Det vil derfor være naturligt i forbindelse med beslutninger vedrørende betydningsrækkefølge at tage udgangspunkt i de specifikke vanskeligheder, som de volder ordbøgernes målgruppe set i forhold til den enkelte ordbogs funktion og målgruppens forudsætninger.

Mange frasalverber er karakteriseret ved at have en række – i nogle tilfælde ovenikøbet en meget lang række – betydninger gående fra konkrete betydninger, hvor betydningen er rimelig let at slutte sig til ud fra betydningen af frasalverbets enkelte elementer, til ret billedlige eller figurative betydninger, hvor det er stort set umuligt at slutte sig til betydningen ud fra betydningen af de enkelte elementer.

For receptionsordbogen vil det være optimalt at anføre de figurative betydninger først, da målgruppen for receptionsordbogen må antages at være så velbevandret i engelsk, at de enten er bekendte med de mest konkrete betydninger eller kan slutte sig til dem ud fra betydningen af de enkelte elementer af frasalverbet. Samme argumentation gælder for oversættelsesordbogen, specielt hvad angår målgruppen professionelle oversættere, der må anses for at have en overordentlig høj grad af sprogfærdighed i engelsk.

Med hensyn til produktionsordbogen er det antagelsen, at ordbogens brugere hovedsageligt vil slå de mest konkrete og gennemskuelige betydninger op med henblik på at finde grammatiske og stilistiske informationer om ordets anvendelse i disse betydninger. Antagelsen bygger på, at det er disse konkrete betydninger, ordbogsbrugerne hovedsageligt vil bruge i deres tekstproduktion, fordi de føler sig trygge ved at bruge dem, fordi de er rimeligt sikre på deres betydning. Når de skal udtrykke de figurative betydninger vil de have en tendens til at vælge enordsverber, der er synonyme med frasalverberne i disse betydninger.

Dette kan illustreres med nogle få frasalverber, der har flere end tre betydninger og samtidig udviser kombinationer af konkrete og billedlige/figurative betydninger. I alle tilfælde er betydninger listet fra det konkrete til det figurative, hvilket samtidig afspejler grad af betydningsmæssig gennemsigtighed.

---

## DIG IN

---

<i>Engelsk definition</i>	<i>Dansk oversættelse</i>
1. to mix a substance into soil by digging	grave ned
2. to put your hand into your pocket or bag	stikke ned i
3. to dig trenches for protection	grave sig ned
4. to refuse to do something	stå fast

- |                             |              |
|-----------------------------|--------------|
| 5. to start eating          | gå til fadet |
| 6. to begin to do something | gå i gang    |
- 

*dig in* har den meget konkrete betydning at grave et materiale ned i jorden. En mindre konkret betydning er at stikke hånden ned i sin lomme eller taske for at finde noget, men det fysiske element af bevægelse nedad er der stadig. Det er også tilfældet med soldater, der graver skyttegrave for at beskytte sig – at grave sig ned – og den fjerde betydning – at nægte at gøre noget/at stå fast – er formentlig afledt heraf. De to sidste betydninger – at begynde at spise noget/at gå til fadet – og gå i gang med noget – har betydningsmæssigt mistet meget af deres forbindelse til den konkrete betydning af *dig in*.

---

#### DROP OFF

<i>Engelsk definition</i>	<i>Dansk oversættelse</i>
1. to fall off	falde af
2. to take someone/something to a place by car	sætte af/aflevere
3. to become less, to be reduced	falde/aftage
4. to fall asleep	døse hen/falde i søvn

---

For *drop off* er der den konkrete betydning at falde af og der er vel en betydningsmæssig forbindelse herfra til den anden og tredje betydning – at sætte nogen af eller at aflevere noget og at falde eller aftage. Betydningen af 4 – at døse hen eller falde i søvn – er derimod vanskelig at slutte sig til ud fra den konkrete betydning af *drop off*.

---

#### GET AROUND

<i>Engelsk definition</i>	<i>Dansk oversættelse</i>
1. to travel around	komme (vidt) omkring
2. to avoid something	komme uden om/omgå
3. if news or information gets around it is told to many people	komme ud/komme frem
4. to persuade someone to do what you want	sno om sin lillefinger
5. to have sexual relationships with many people	være løs på tråden

---

*get around* har den konkrete betydning af gå eller rejse rundt og to mindre konkrete men dog beslægtede betydninger – at undgå eller omgå noget og om nyheder at komme eller slippe ud. De to betydninger 4 – at overtale nogen til at

gøre hvad man vil/at sno nogen om sin lillefinger – og 5 – at være promiskuøs/at være løs på tråden – kan kun vanskeligt siges at være beslægtede med den konkrete betydning.

---

**KNOCK OFF**


---

<i>Engelsk definition</i>	<i>Dansk oversættelse</i>
1. to make something come off	slå/banke/knække af
2. to cause someone to fall off something	vælte af
3. to reduce the price of something by a particular amount	slå af
4. to kill someone	gøre det af med
5. to defeat someone	besejre/slå
6. to copy something	kopiere
7. to stop doing something	holde op
8. to stop working (because it's time to go home)	holde fyraften
9. to produce something easily and quickly	få fra hånden
10. to steal or rob something	stjæle/negle/røve
11. to have sexual intercourse with someone	bolle med
12. to relax	tage en slapper
13. to drink something quickly	tylle i sig

---

Med hensyn til *knock off* er der to konkrete betydninger – 1 og 2. For mig at se er betydning 3 – at reducere en pris med noget – relateret til de konkrete betydninger, og det gælder muligvis også 4 – at slå nogen ihjel/at gøre det af med nogen – og i mindre grad 5 – at besejre nogen. Men de resterende 8 betydninger synes ikke at være relaterede til den konkrete betydning af *knock off*.

Hvis det skal forsøges at anvende en eller anden form for princip for ordning af de otte figurative betydninger af *knock off* må forslaget være, at hvad angår reception bør de mindst uformelle betydninger – og det vil sige 6, 8, 9, 10 og 12 – have forrang frem for de slangprægede betydninger 7, 11 og 13, da der er marginalt større mulighed for at disse betydninger vil optræde oftere i de tekster, som studenter forventes at læse. Det skal imidlertid understreges, at der er tale om uformelle betydninger som i hvert fald for akademiske tekster vil være ret infrekvente. Hvad angår oversættelse og produktion vil det være vanskeligt at opstille noget princip for rangordning af disse otte ret uformelle figurative betydninger.

## LITTERATUR

*Ordbøger*

- Allen, Robert, 2006: *The Penguin Office Dictionary*. London: Penguin Books.
- Axelsen, Jens, 2003: *Engelsk-Dansk Ordbog*, 13. udg. København: Gyldendal.
- Cambridge Advanced Learner's Dictionary*, 2003. Cambridge: Cambridge University Press.
- Cassell Giant Paperback Dictionary*, 1994. London: Cassell.
- Collins COBUILD Advanced Dictionary*, 6<sup>th</sup> ed., 2009. Boston MA: Heinle CENGAGE Learning.
- Collins English Dictionary*, 8<sup>th</sup> ed., 2006. Glasgow: HarperCollins Publishers.
- Hornby, A.S., 1974: *Oxford Advanced Learner's Dictionary of Current English*, 3<sup>rd</sup> ed. London: Oxford University Press.
- Hornby, A.S., 2010: *Oxford Advanced Learner's Dictionary of Current English*, 8<sup>th</sup> ed. Oxford: Oxford University Press.
- Kjørulff Nielsen, B., 1998: *Engelsk-Dansk Ordbog*, 6. udg. København: Gyldendal.
- Longman Dictionary of Contemporary English*, 1978. Harlow: Longman Group Limited.
- Longman Dictionary of Contemporary English*, 5<sup>th</sup> ed., 2009. Harlow: Pearson Longman.
- MACMILLAN English Dictionary for Advanced Learners*, 2<sup>nd</sup> ed., 2007. Oxford: Macmillan.
- Onions, C.T., 1966: *The Oxford Dictionary of English Etymology*. London: Oxford University Press.
- Politikens Engelskordbog Engelsk-Dansk/Dansk-Engelsk*, 2. udg., 2009. København: JP/Politikens Forlagshus A/S.
- Sykes, J.B., 1976: *The Concise Oxford Dictionary of Current English*. 6<sup>th</sup> ed. London: Oxford University Press.
- Zettersten, Arne & Lauridsen, Hanne, 1999: *Politikens Engelsk Dansk med betydningsforklaringer*, (2 bind). København: Politikens Forlag.

*Anden litteratur*

- Andersen, Birger, 2010: Etablering og strukturering af database til generering af tre elektroniske monofunktionelle 'phrasal verb' ordbøger. I: Harry Lönnroth & Kristina Nikula (eds.): *Nordiske studier i lexicografi 10. Rapport från Konferensen om lexicografi i Norden, Tammerfors 3-5 juni 2009*. Tammerfors. S. 56-71.
- Hornby, A.S., 1965: Some Problems of Lexicography. I: *English Language Teaching* 19. S. 104-110.
- Kipfer, Barara A., 1983: Methods of Ordering Senses Within Entries. I: R.R.K. Hartmann (ed.): *LEXeter '83. Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*. *Lexicographica Series Maior* 1. Tübingen: Niemeyer. S. 101-108.
- Lew, Robert, 2004: *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.

- Lew, Robert, 2009: Towards Variable Function-Dependent Sense Ordering in Future Dictionaries. I: Bergenholtz, Henning, Sandro Nielsen & Sven Tarp (eds.): *Lexicography at a Crossroads*. Berlin: Peter Lang. S. 237-264.
- Moerdijk, Fons, 2003: The codification of semantic information. I: Piet von Sterkenburg (ed.): *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing Company. S. 273-296.
- Svensén, Bo, 2009: *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Tono, Yukio, 1984: *On the Dictionary User's Reference Skills*. B.Ed. Thesis, Tokyo Gakuegi University, Tokyo.

*Birger Andersen*

Center for Leksikografi, Århus Universitet.

ba@asb.dk

## Appendix 1

## Principper for betydningsordning

	Lineær		Hierarkisk			Umarkeret princip
	Efter frekvens		Antal afledte betydninger			
	+	?	Lavt	Middel	Højt	
<b>Almene monolingviale engelske ordbøger</b>						
Penguin Office Dictionary					✓	
Cassell Giant Paperback Dictionary						✓
BBC English Dictionary		✓				
Collins English Dictionary			✓ (+ frekvens)			
Concise Oxford Dictionary of Current English				✓ (+ frekvens)		
<b>Monolingviale learner-ordbøger</b>						
Longman 1st ed				✓		
Longman 5th ed			✓ (+ frekvens)			
Oxford 3rd ed				✓		
Oxford 8th ed			✓			
Macmillan				✓		
Cambridge				✓		
Collins COBUILD		✓				
<b>Bilingviale ordbøger</b>						
Kjærulff-Nielsen Engelsk-dansk					✓	

Axelsen Engelsk-dansk		✓				
Politiken Engelsk-dansk med betydningforklaringer	✓					
Politiken Engelskordbog (skoleordbog)		✓				

## Appendix 2

### Den ordbogsmæssige behandling af *suit*

#### 1. Almene monolingvale engelske ordbøger

##### (a) Penguin Office Dictionary

**suit 1a** an outer costume of two or more matching articles of clothing, *esp* a jacket and trousers or a skirt

**b** a costume to be worn for a specified purpose. 2 all the playing cards in a pack bearing the same symbol, i.e. hearts, clubs, diamonds or spades. 3 a legal action; a lawsuit, 4 *informal* a business executive, *esp* one considered as faceless but influential. 5 *formal or archaic* a a petition or appeal b courtship.

##### (b) BBC English Dictionary

###### suit

1 [...] A man's suit consists of a matching jacket, trousers, and sometimes a waistcoat [...]

2 [...] A woman's suit consists of a matching jacket and skirt [...]

3 [...] A suit can also be a piece of clothing worn for a particular activity [...]

7 [...] In a court of law, a suit is a legal action taken by one person against another [...]

8 [...] A suit is also one of the four types of card in a set of playing cards [...]

##### (c) Collins English Dictionary

“As a general rule, where a headword has more than one sense, the first sense given is the most common in current usage.” (p X)

**suit** [...] 1 any set of clothes of the same or similar material designed to be worn together, now usually (for men) a jacket with matching trousers or (for women) a jacket with matching or contrasting skirt or trousers 2 (*in combination*) any

outfit worn for a specific purpose [...] 3 any set of items, such as the full complement of sails of a vessel or parts of personal armour 4 any of the four sets of 13 cards in a pack of playing cards [...] 5 a civil proceeding; lawsuit 6 the act or process of suing in a court of law 7 a petition or appeal made to a person of superior rank or status or the act of making such a petition 8 *slang* a business executive or white-collar manager 9 a man's courting of a woman

## 2. Monolingvale engelske learner-ordbøger

### (a) Longman Dictionary of Contemporary English (1<sup>st</sup> ed. 1978)

**suit** [...] 1 a a set of outer clothes which match, usu. including a short coat (JACKET) with trousers or skirt

b (*usu. in comb.*) a garment or set of garments for a special purpose [...] 2 a set (of armour) (in the phr. suit of armour/mail) 3 one of the 4 sets of cards used in games [...] 4 *fml* a request 5 *old use* the act of asking a woman to marry (esp. in the phr. **plead/press one's suit**) 6 LAWSUIT [...]

### (b) Longman Dictionary of Contemporary English (5<sup>th</sup> ed. 2009)

"The meanings of each are listed in order of frequency. The most common meaning is shown first" (p. xi)

#### **suit**

1. **CLOTHES** a set of clothes made of the same material, usually including a jacket with trousers or a skirt [...]
2. **bathing/jogging etc. suit** a piece of clothing or a set of clothes used for swimming, running, etc. [...]
3. **LAW** a problem or complaint that a person or company brings to a court of law to be settled [...]
4. **OFFICE WORKER** *informal* a man, especially a manager who works in an office and who has to wear a suit when he is at work [...]
5. **CARDS** one of the four types of cards in a set of playing cards

### (c) Oxford Advanced Learner's Dictionary of Current English (3<sup>rd</sup> ed. 1974)

**suit** [...] 1 [C] set of articles of outer clothing of the same material [...] *a man's* ~ jacket (waistcoat) and trousers; *a woman's* ~ coat and skirt; *a trouser* ~ woman's ~ of jacket and trousers; [...] 2 [C] (formal) request made to a superior, esp. to a ruler [...] 3 [C] (liter or old use) asking a woman's hand in marriage [...] 4 (**law**) ~ case in a law court, prosecution of a claim [...] 5 [C] any of the four sets of cards (spades, hearts, diamonds, clubs) used in many card games [...]



(d) Oxford Advanced Learner's Dictionary (8<sup>th</sup> ed. 2010)

“Meanings that are closely related share the same short cut” (p. x)

**suit**

1 a set of clothes made of the same cloth, including a jacket and trousers/pants or a skirt [...] 2 a set of clothing worn for a particular activity [...] 3 any of the four sets that form a PACK/DECK of cards [...]

4 = LAWSUIT [...] 5 [usually pl.] (*informal*) a person with an important job as a manager in a company or organization, especially one thought of as being mainly concerned with financial matters or having a lot of influence [...]

## (e) MACMILLAN English Dictionary for Advanced Learners

**suit**

1 a set of clothes made from the same cloth, usually a jacket with trousers or a skirt [...] 1a a type of clothing that you wear for a particular activity [...]

2 a claim or complaint that someone makes in a court of law [...]

3 one of four sets of PLAYING CARDS that together make a PACK [...]

4 *informal* someone who works in an office and wears a suit, especially a senior manager who you dislike because they think only about the financial aspects of a business

## (f) Collins COBUILD Advanced Dictionary

**suit**

1 [...] A man's **suit** consists of a jacket, trousers, and sometimes a waistcoat, all made from the same fabric [...] 2 [...] A woman's **suit** consists of a jacket and skirt, or sometimes trousers, made from the same fabric [...] 3 [...] A particular type of **suit** is a piece of clothing that you wear for a particular activity [...] 8 [...]

In a court of law, a **suit** is a case in which someone tries to get a legal decision against a person or company [...] 9 [...] A **suit** is one of the four types of card in a set of playing cards [...]

## 3. Bilingvale engelsk-danske ordbøger

## (a) Kjærulff-Nielsen Engelsk Dansk

**SUIT**

1. a) ( - of clothes) sæt tøj; [dame]spadseredragt; dragt [...]; habit; S mappedyr [...] b) sæt (fx a - of sails); c) (abt. hair; growth, crop): he had a big round head and a fine - of hair ... og pænt hår. 2. (card.) farve [...] 3. a) (hist.) følge; b) (jur.) =

*law-*; *c) in ~ with* i overensstemmelse med; passende til. 4. (*petition*) ansøgning.  
5. (*wooing*) bejlen [...]

### (b) Politikens Engelsk Dansk med betydningforklaringer

”De forskellige betydninger er opstillet efter hyppighed, således at den hyppigste betydning har nr. 1” (p.11)

#### suit

1. substantiv, tælleligt  
= **habit; jakkesæt** [...]
2. substantiv, tælleligt  
= **spadseredragt; dragt** [...]
3. substantiv, tælleligt: (nN)  
= **dragt** [...]
8. substantiv, tælleligt  
= **søgsmål; retssag; anklage; sagsanlæg** [...]
9. substantiv, tælleligt  
= **kulør; farve; kortfarve** [...]
10. Se også **bathing suit, birthday suit, boiler suit, romper suit, trouser suit.**

# Flødeboller, mælkebøtter og smørkringler

## Om leksikalsk-semantiske problemstillinger i arbejdet med Ømålsordbogen

*Lena Wienecke Andersen*

In my PhD project, I study how word meanings are ‘captured’ and defined in *Ømålsordbogen*, the Dictionary of the Danish insular dialects, and I study and try to develop new practices. Semantic relations is one of the central points in the project. In this article, I examine two problems concerning the extraction of semantic relations within a non-digitized collection of dialect words (already in or going to be in the dictionary). Since we want to mention semantically related words in the dictionary, we basically need to know what words we have in the collection of paper slips. As the words are historical and most of them unknown, general knowledge and introspection cannot be used. So, first, I will show one way to get an overview of the various words. Second, I outline the possibility to update the work with semantic relations – which so far has been done manually and non-digitized – by means of a Wordnet-inspired tool of modern standard Danish, *DanNet*. But it is argued that the geographical variation in the data has to be taken into consideration; it has to be incorporated that some of the synonyms in the data are actually equivalents. Overall, I argue for an onomasiological approach to the words when working with word meanings, even though the dictionary itself is alphabetically organized.

*Nøgleord:* DanNet, dialektologi, leksikalsk semantik, onomasiologi, seddelsamling, semantiske relationer, Ømålsordbogen

### 1. Indledning

Overordnet set kan man skelne mellem to tilgangsvinkler til et ordforråd: Den semasiologiske tilgang hvor ordene er det styrende princip, og den onomasiologiske tilgang hvor begreberne er det styrende (Bergenholtz et al 1997, Svensén 2004, Geeraerts 2010). Med udgangspunkt i mit ph.d.-projekt vil jeg i denne artikel (A) præsentere en fremgangsmåde til at få overblik over et geografisk varieret ordforråd og (B) diskutere hvordan man kan forholde sig til denne vari-

ation i arbejdet med leksikalsk-semantiske relationer. De to problemstillinger illustrerer nødvendigheden af den onomasiologiske tilgang, gerne kombineret med den semasiologiske.

## 2. Baggrund

Mit ph.d.-projekt der har arbejdstitlen: ”Fra virkelighedens verden til ordbogens – et studie i datastrukturering med henblik på betydningsdefinitioner i ordbøger”, placerer sig i krydsfeltet mellem dialektologi, leksikografi og leksikalsk semantik. Konkret leksikografisk og dialektologisk beskæftiger jeg mig med *Ømålsordbogen*, en dansk dialektordbog der beskriver dialekterne på Fyn, Sjælland, Lolland, Falster og de omliggende øer. Ordbogen er alfabetisk ordnet, og med de 10 bind der foreløbig er kommet, er redaktionen omtrent halvvejs. Konkret leksikalsk-semantisk skal jeg afprøve om strukturen i det digitaliserede begrebsnetværk over moderne standarddansk DanNet kan overføres til *Ømålsordbogens* materiale, og i så fald om DanNet kan bruges som analyseværktøj i forbindelse med de traditionelle dialekter.

## 3. Problemstilling A: overblik over ordforrådet

Som mange andre ordbøger vil vi i *Ømålsordbogen* gerne angive semantiske relationer, og her støder vi på første meget banale, men alvorligt mente problemstilling: Hvilke ord findes i det materiale vi arbejder med? Og det bringer præsentationens titel på bane:

I moderne standarddansk er en *flødebolle* en spiselig chokoladeovertrukket skumbolle (jævnfør norsk: kokosbolle, svensk: kokosboll), en *mælkebøtte* er en plante med gul blomst der af mange betragtes som ukrudt (norsk: løvetann, svensk: maskros), og en *kringle* er et stykke bagværk med remonce og for eksempel rosiner i (det samme i norsk, *kringla* i svensk).

I de klassiske dialekter på de danske øer var en *flødebolle* en serveringskande til fløde og en *mælkebøtte* en beholder til transport af mælk. *Smørkringlen* vender jeg tilbage til senere.

*Mælkebøtte* og *flødebolle* er eksempler på at selv om vi på *Ømålsordbogen* arbejder med dansk sprog, eller en del af det, så er der grænser for hvad man som redaktør kan bruge sin almene viden til. Vi sidder med ord fra de danske ødialekter, hovedsageligt fra perioden 1850-1920, og i det materiale findes der mange ord som vi enten slet ikke kender, eller som ikke betyder det samme længere. Redaktørens problem er egentlig ikke betydningsdefinitionen for det pågældende ord; det er faktisk mere reglen end undtagelsen at man skal lave

noget detektivarbejde når man redigerer en dialektordbog. Men hvordan henviser man på kryds og tværs mellem ordbogsartikler, herunder til artikler som endnu ikke er redigeret og måske først bliver det om mange år?

Man bliver nødt til at skaffe sig et overblik over ordforrådet; hvis man for eksempel skal henviser til synonymer, så skal man nemlig vide hvilke synonymer der findes. Vi kan ikke bruge introspektion, og eksisterende synonymordbøger dækker heller ikke. Hvad gør man så? Hvilke ord fandtes i de danske ømål? Hvilke ord forekommer i *Ømålsordbogens* samlinger?

#### 4. Ømålsordbogens seddelsamling

Som i mange andre ordbogsprojekter arbejder vi med sedler. Vi har rundt regnet tre millioner ikke-digitaliserede ordbogssedler. Hvor mange ord, lemmaer, det drejer sig om, er der ingen der ved. Men heldigvis er vores seddelsamling inddelt i emner; og inden for hvert af disse emner står sedlerne så alfabetisk. De cirka 60 emner afspejler den verden som dialekterne eksisterede i, nemlig landbokulturen som den så ud for godt 100 år siden. Der er for eksempel et emne der hedder *gård*, et der hedder *pløjning*, og et der hedder *mælk*. En sådan emneinddeling giver langt bedre mulighed for at få overblik over ordforrådet; inden for hvert enkelt emne vel at mærke.

Oprindeligt var det tanken at alle emner skulle gennemgås systematisk, altså genstand for genstand, proces for proces, før den alfabetiske ordbogsredaktion. Ved en samlet gennemgang af alle ord i emnet fik man dels overblik over ordforrådet, dels over de semantiske relationer heri. Som minimum oplyste man synonymer, men også andre relationer kunne registreres, for eksempel del-helhed og over-underbegreber. Filosofien var at man kun kan beskrive ordforrådet, for eksempel i en ordbog, når man kender genstandene og sammenhængene.

Hvad angår funktionsord, er en almindelig alfabetisk/semasiologisk gennemgang af materialet nok. Men hvad angår alle indholdsord, vil jeg sige at en systematisk/onomasiologisk analyse nærmest er nødvendig – i hvert fald for at få registreret alle de ukendte ord og lumske ord som *flødebolle* og *mælkebøtte* som der gemmer sig mange af i *Ømålsordbogens* materiale.

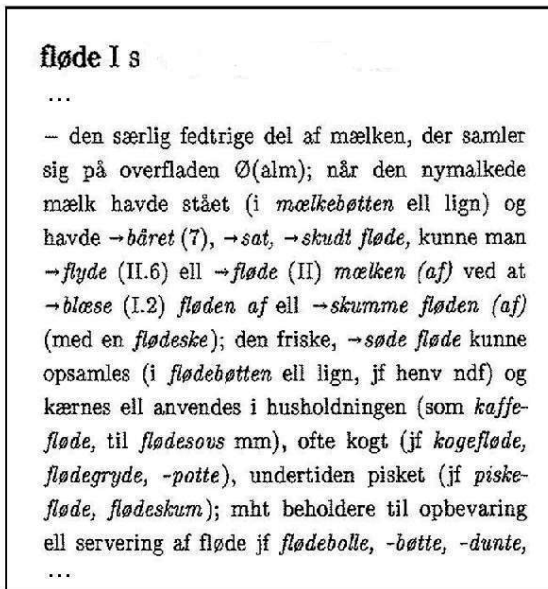
Når sådan et arbejde har fundet sted, er der ideelt set ingen tvivl om hvad der skal henvises til på kryds og tværs i ordbogen. Og det er vældig praktisk når nu ordbogen redigeres og udgives alfabetisk fra en ende af.

I praksis fik man desværre kun gennemgået omkring halvdelen af de cirka 60 emner. Derefter gik den alfabetiske ordbogsredaktion i gang, og man havde ikke resurser til at gøre begge dele.

## 5. Emnet *mælk*

Et af de emner der ikke nåede at blive gennemgået, er emnet *mælk*, og derfor skal jeg arbejde med det i mit ph.d.-projekt. Det omfatter ord for produkter som mælk, ost og smør, processer som at malke og at kærne, redskaber som en malkespand og en kærne, personer som malkepige og -komtesse og så videre. Kort sagt de ord som de dialekttalende brugte når de talte om emnet. Det drejer sig om i alt cirka 1.200 lemmaer fordelt på 10-13.000 sedler. Mælkeordene er hidtil kun blevet gennemgået alfabetisk, et for et. Det kan forsvares at gennemgå emnet systematisk nu selv om den alfabetiske ordbogsredaktion er lige knap halvvejs, idet de centrale begreber og ord inden for emnet er at finde i den endnu uredigerede del af alfabetet, for eksempel kærne, mælk, ost og smør. Umiddelbart er fløde det eneste centrale begreb der er redigeret.

Et uddrag af flødeartiklen i *Omålsordbogen* ses i figur 1:



Figur 1. Uddrag fra *Omålsordbogen*, bind 4.

I en ”almindelig” ordbog ville man måske have nøjedes med: ”den særlige fedtrige del af mælken der samler sig på overfladen”, men *Omålsordbogen* fortsætter. I artiklen forsøges nævnt alle de ord og begreber som fløde indgår i en relation med.

Meget kan gå galt her. For hvad gør man egentlig når emnet ikke er blevet gennemgået systematisk, og man har over 1.000 lemmaer fordelt på over 10.000 sedler?

Dels gennemser man sine sedler, i det her tilfælde flødesedler, og måske dukker der andre relevante ord og begreber op. Dels kan man gennemse den spørgeliste der har lagt grunden for mælkesamlingen (Møller 1938), og ud fra den kan man få ideer til hvor i alfabetet man kan lede efter relevante ord. I forhold til hvor mange ord der henvises til fra flødeartiklen, kunne jeg forestille mig at redaktøren også har gennemset hele seddelsamlingen af mælkeord. Endelig kan man selvfølgelig studere alle mulige andre kilder.

I mangel af bedre laver vi således en ad hoc-gennemgang af sedler, spørgelister og seddelsamlinger for at få overblik over hvilke relationer det aktuelle ord indgår i. Men det ville naturligvis være at foretrække hvis der forelå en fuldstændig systematisk gennemgang af hele emnet. Og dét er derfor en af opgaverne i mit ph.d.-projekt: At lave en systematisk gennemgang af mælkesamlingen.

## 6. Overblik over *mælk*

Det første jeg gjorde, var at indtaste alle mælkeordene i en database. Med en elektronisk base (modsat en seddelsamling på papir) kan de samme data sorteres på forskellige måder blot ved at trykke på en knap; når man vel at mærke har indtastet de oplysninger der skal sorteres efter. Som minimum indtastede jeg lemma, ordklasse, en grov semantisk kategorisering og ofte også stikord til en definition. Den grove semantiske kategorisering, som er baseret på intuition, har været nødvendig for mig for at få overblik; så jeg ikke skulle arbejde med 1.200 enheder svarende til de 1.200 lemmaer, men så jeg i stedet havde et overskueligt antal enheder at arbejde med.

I skrivende stund er jeg ved at gennemgå genstanden kærne systematisk. En lang række ord har fået mærkaten ”redskab/kærne”; det er de ord der betegner det redskab man brugte til at fremstille smør med. Muligheden for semantisk sortering ved hjælp af denne kategorisering har gjort at jeg hurtigere har kunnet afgrænse og overskue relevante ord end ellers.

Der er overordnet set to slags kærner: Kærner hvor beholderen står fast, og delene indeni er bevægelige, og kærner hvor beholderen er bevægelig, med faste dele indeni – og inden for disse typer er der igen flere typer og igen flere typer. Der er godt 100 betegnelser i materialet for helheden kærne og alle dens dele hvoraf nogle ses nedenfor:





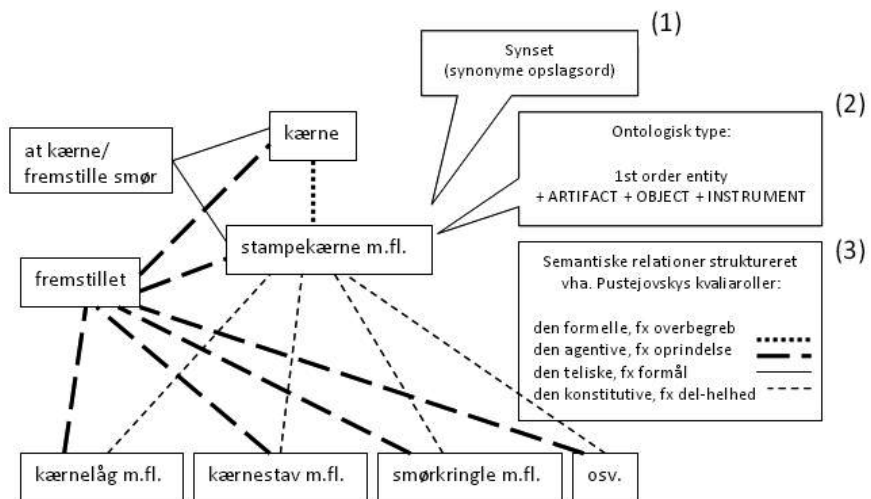
Næste skridt i mit ph.d.-projekt er at afprøve om et begrebsnetværk kan være et nyttigt redskab når man skal arbejde med de semantiske relationer i de danske ømål. Dette arbejde introduceres kort i det følgende.

## 7. Begrebsnetværk og semantiske relationer

Som nævnt indledningsvis vil jeg afprøve om DanNet kan bruges på mine data. Jeg vil ikke komme nærmere ind på baggrunden for DanNet og dets internationale forløber Wordnet, men blot henvise til DanNets hjemmeside samt Pedersen et al 2010.

DanNet er et digitaliseret begrebsnetværk for moderne standarddansk; et komplekst net der ideelt set skulle bestå af en masse knuder forbundet af en masse linjer; knuderne repræsenterer begreber og betegnelser for dem, linjerne repræsenterer forskellige semantiske relationer.

I figur 3 ses den første brøkdæl af et mælkenetværk, nemlig begrebet stampekerne og dets umiddelbare omgivelser i nettet. I det videre arbejde vil stampekerne skulle kobles til flere ord der igen er koblet til flere ord og så videre.



Figur 3. Stampekærne-synset i et DanNet-lignende netværk.

Udgangspunktet i figuren er begrebsknuden for stampekerne; i samme begrebsknude vil ordene stavkærne og stangkærne og så videre skulle placeres. I Word-

netterminologien er der tale om en gruppe synonyme, et synset, jævnfør (1) i figuren.

Næste skridt er at begrebsknuden skal kategoriseres. Efter et katalog af ontologiske typer skal alle begrebsknuder have en mærkat, jævnfør (2). I det her tilfælde er der tale om en konkret genstand, det vil sige en såkaldt "1st order entity", mere specifikt en menneskeskabt genstand beregnet som redskab. Derefter kobler man andre ord og synsets på ved hjælp af fire kategorier af semantiske relationer, jævnfør (3):

Overbegreb:	kærne
Oprindelse:	menneskeskabt, det vil sige fremstillet
Formål:	at kærne, at fremstille smør
Del-helhed:	stampekærnen består af en række dele i form af synsets for låget, staven, smørkringlen og så videre.

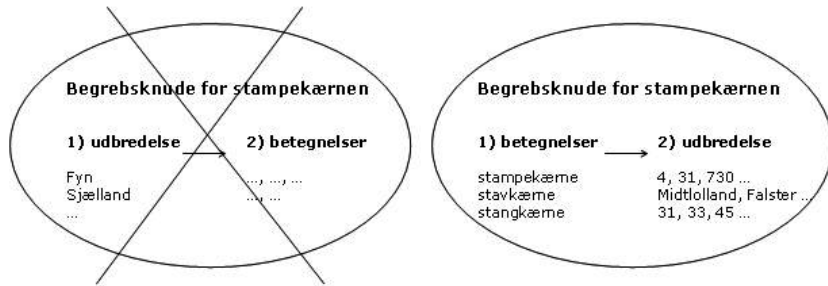
Dertil kommer at stampekærnen overbegreb såvel som dele også er fremstillede. Og ligeledes har overbegrebet kærne samme formål som stampekærnen.

Og sådan skal alle mælkeordene efterhånden placeres og forbindes på kryds og tværs i begrebsnetværket – når problemstillingen som præsenteres i det følgende, er blevet løst.

## 8. Problemstilling B: den geografiske variation

Betegnelserne for for eksempel *smørkringlen* er ikke bare almindelige synonyme der ubetinget kan anbringes i ét synset. I *Ømålsordbogen* kan betegnelser for den samme genstand være (1) "almindelige synonyme" hvor to ord mere eller mindre betyder det samme. Jeg har for eksempel en seddel hvor samme person har angivet at kunne sige både *trille* og *trisse* om træringen. De to betegnelser ville kunne placeres i et synset uden problemer. Men betegnelserne kan også være (2) "geografiske synonyme", eller ækvivalenter om man vil, der dækker over at man i én dialekt bruger et ord, i en anden dialekt bruger et andet ord for det samme. Denne geografiske synonymi skal inkorporeres i Wordnetmodellens synsets på en eller anden måde hvis den skal kunne bruges i arbejdet med *Ømålsordbogen*.

I figur 4 ses to begrebsknuder hvori der på to forskellige måder er taget højde for den geografiske variation. Det venstre eksempel er krydset over idet det ikke er en mulig løsning på grundlag af *Ømålsordbogens* materiale.



Figur 4. To eksempler på hvordan den geografiske udbredelse kan indarbejdes i en begrebsknode. Eksemplet til venstre er ikke muligt på grundlag af *Ømålsordbogens* materiale.

I begrebsknuden til venstre er tanken at hver dialekts betegnelse(r) holdes samlet sådan at man til slut ville kunne lave et net med alle begrebsknuder i hver enkelt dialekt, for eksempel fynsk. Dette forudsætter dog dels at man helt *principielt* mener at man kan afgrænse dialekterne skarpt; dels at det kan lade sig gøre i *praksis*.

I begrebsknuden til højre ses vidt forskellige angivelser af udbredelse angivet med topografiske numre (Skjelborg 1967) eller områdebetegnelse. Det topografiske nummer 31 omfatter et enkelt sogn, Vejby Sogn, mens det topografiske nummer 4 omfatter hele Nordsjælland. Ligeledes ses også de omfattende lokaliteter Midtjylland og Falster. På en stangkærneseddel ses igen sogn nummer 31. I Vejby kunne man altså både sige stampekerne og stangkærne om denne type kærne; her er tale om almindelig synonymi. I nogle tilfælde er de samme ord således almindelige synonymmer, i andre tilfælde er de samme ord geografiske synonymmer.

I Wordnettankegangen kan der laves et net per (national)sprog; således ville der (vel) principielt også kunne laves et net per dialekt. På grundlag af *Ømålsordbogens* materiale kan en skarp afgrænsning af dialekterne dog ikke lade sig gøre i praksis. På *Ømålsordbogens* sedler er angivet en lokalitet, som regel et sogn (for eksempel Vejby), men undertiden mere upræcist en egn eller et dialektområde (for eksempel Midtjylland, Falster). Det mest præcise vi kan gøre, er derfor at registrere de forskellige typer lokalitetsangivelser sådan som det ses i begrebsknuden til højre, altså angive udbredelsen så fint- eller grovmasket som den nu

engang er registreret.<sup>1</sup> Og derfor vil vi hellere tale om geografiske synonymer end om et antal dialekter med et antal ækvivalenter.

Om strukturen i begrebsknuden til højre kan løse mine problemer, kan jeg i skrivende stund ikke gennemskue, men jeg arbejder videre med problemstillingen i mit ph.d.-projekt.

## 9. Sammenfatning

Således kan jeg sammenfatte at med et materiale som *Ømålsordbogens* hvor ord som *mælkebøtter*, *flødeboller* og *smørkringler* betyder noget helt andet end man skulle tro – simpelthen fordi det meste af den verden materialet handler om, er mere eller mindre ukendt i vore dage – så vil en systematisk/onomasiologisk gennemgang af materialet være af væsentlig betydning når man sidder med den alfabetiske ordbogsredaktion.

Et begrebsnetværk i lighed med DanNet kan sandsynligvis være en hjælp i arbejdet med at få struktureret data yderligere og få overblik over dem. Men det kræver nogle justeringer. Det er – om noget – den geografiske variation vi arbejder med på *Ømålsordbogen* – så den *skal* indarbejdes i de værktøjer vi skal arbejde med.

## KILDER OG LITTERATUR

- Andersen, Lena Wienecke, under udarbejdelse: Fra virkelighedens verden til ordbogens. Et studie i datastrukturering med henblik på betydningsdefinitioner i ordbøger. Igangværende ph.d.-projekt på *Ømålsordbogen*, Afdeling for Dialektforskning, Nordisk Forskningsinstitut, Københavns Universitet.
- Bergenholtz, H., I. Cantell, R.V. Fjeld, D. Gundersen, J.H. Jónsson og B. Svendsén, 1997: *Nordisk Leksikografisk Ordbok*. Oslo.
- DanNet: <http://wordnet.dk> (maj 2011).
- Geeraerts, Dirk, 2010: *Theories of Lexical Semantics*. Oxford.
- Møller, Kristen, 1938: Mælken og dens Produkter. I: *Danske Folkemaal* 12. S. 1–16.
- Pedersen, B.S., A. Braasch, S. Nimb, J. Asmussen, N. Sørensen, H. Lorentzen og L. Trap-Jensen, 2010: *Lingvistiske specifikationer for DanNet Version 1.0.2*. <http://wordnet.dk> (maj 2011).
- Skjelborg, Åge, 1967 (revideret afskrift): *Topografisk ordning af Danmarks egne med alfabetisk nøgle*. København.

---

<sup>1</sup> At vi så i *Ømålsordbogen* vælger at generalisere til for eksempel S(alm) om ord der er almindeligt forekommende på Sjælland, L-Fa(spor opt) om ord der er sporadisk optegnet på Lolland og Falster og så videre, er en anden sag.

- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. Stockholm.
- Ømålsordbogen. En præsentation, udg. af Afdeling for Dialektforskning, Københavns Universitet, 2004. København.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer, red. af Afdeling for Dialektforskning, Københavns Universitet, udg. af Universitets-Jubilæets Danske Samfund, 1992ff, bind 1ff. København.
- Ømålsordbogens samlinger. Afdeling for Dialektforskning, Københavns Universitet. København.
- Ømålsordbogens tillægsbind, red. af Afdeling for Dialektforskning, Københavns Universitet, udg. af Universitets-Jubilæets Danske Samfund, 1992. København.

*Lena Wienecke Andersen*

Ømålsordbogen, Afdeling for Dialektforskning, Københavns Universitet.  
wienecke@hum.ku.dk

# Koordinering i behandlingen af faste ordforbindelser

*Torben Arboe*

The paper describes aspects of coordination in the lexicographical treatment of phrasal verbs and the corresponding adverbs in the *Jysk Ordbog* (JO), i.e. the Dictionary of the Jutland Dialects: which parts of the material are at best placed in the microstructure of the verbs (e.g. comprehensive verbs as *holde, gå* (= to hold, to go)), and which parts will be more adequately placed under the corresponding adverbials. Comparisons are made (section 2) with the treatment of the same verbs and adverbs in other large dictionaries of the Danish language, including *Ømålsordbogen*, i.e. the Dictionary of the Danish Insular Dialects. Examples are then given of the references from adverbs to other phrasal verbs (section 3), and some other problems in the microstructure of adverbs are addressed with regard to the rendering of phrasal verbs and other types of set phrases (section 4). To conclude, it is mentioned that a new IT platform which is being implemented for the *Jysk Ordbog* will make it easier to secure the large amount of references needed between the entries of phrasal verbs and the corresponding adverbs.

*Nøgleord:* Faste ordforbindelser, tryktabsforbindelser ("verbsamband"), mikrostruktur, dialektordbøger, rigsmålsordbøger.

Tidligere har jeg undersøgt behandlingen af faste ordforbindelser i Jysk Ordbog (JO) især i artikler på verber og substantiver med mange underbetydninger, altså i komplicerede mikrostrukturer, herunder indplacering af tryktabsforbindelser, også kaldet *partikelverber*, *verbsamband*, *phrasal verbs* mv. (Arboe 2008)<sup>1</sup>. Denne gang vender jeg perspektivet og undersøger behandlingen af faste forbindelser i artikler på præpositioner og adverbier, herunder opgaven med koordinering af oplysningerne mellem disse artikler og artikler på især substantiver og verber.

---

<sup>1</sup> Tryktabsforbindelser mv. indplaceres i et generelt skema over fraseologismetyper i Arboe 2011: 266, 268.

## 1. Praksis i Jysk Ordbog

Som eksempler brugte jeg tidligere mikrostrukturene i verberne *have*, *holde* og *hugge*, hvoraf her bringes en lidt ændret og forkortet udgave fra artiklen *2.holde* (se bilag 1)<sup>2</sup>.

Det ses, at der i *2.holde* afsnit F er repræsenteret ca. 20 præpositioner og adverbier: fra betydning 22 *holde 'ad* til betydning 43 *holde 'ved*. Nogle af dem optræder i tryktabsforbindelser med adskillige betydninger, fx 7 ved forbindelsen *holde 'af*, bl.a. betydningen ”synes om, kunne lide”, men også fx ”holde ud” (jeg kan ikke holde det 'af = ikke holde det ud). Det samme var i lidt mindre grad tilfældet ved verberne *have* og *hugge*: omkring 10 tryktabsforbindelser hver, fx *have 'for* og *have 'i* under verbet *4.have*, samt adskillige betydninger af både *hugge 'af* og *hugge 'for* under verbet *hugge* (Arboe 2008: 11, 18). Der henvises kun sjældent fra disse verber til artikler på de tilhørende, trykbærende præpositioner og adverbier. Det sker dog ved *2.holde* betydning 15 og 23, hvor der henvises til artiklen *ad*; her nævnes adskilligt specielt om præpositionen og adverbiet, hvilket jeg kommer jeg tilbage til nedenfor.

Andre præpositioner og adverbier trækker ligeledes interessante tråde til de verber, de indgår tryktabsforbindelse med, fx til verbet *gå*, som jeg derfor har taget med i strukturoversigterne denne gang (se igen bilag 1). Her har jeg understreget de præpositioner og adverbier, der hidtil er redigeret til JO. Disse er anført i bilag 2, fra *ad* til *ind*, med den redaktionelle struktur skitse-mæssigt angivet. Forkortelsen FF er anvendt for ”fast ordforbindelse” de steder, hvor der i JO anføres ”fast(e) forbindelse(r)”. Enkelte steder, hvor ordforbindelsen vurderes som mindre fast<sup>3</sup>, noteres i den redaktionelle tekst kun ”i forbindelsen”; dette markeres i bilag 2 med blot F (fx i *ad* betydning 10-12).

I bilag 1 gengives en del af strukturen i artiklen på verbet *gå*: lidt af afsnit A, desuden en større del af det afsluttende afsnit F vedr. tryktabsforbindelser. Det ses, at der opregnes ca. 35 tryktabsforbindelser, betydning 18-51, og at der er anført oversættelser af de fleste forbindelser fra betydning 18, *gå 'ad*, frem til betydning 34, *gå 'inde* (som ovenfor svarende til de hidtil redigerede afsnit af JO); de sidste ca. 15 tryktabsforbindelser fra *gå 'med* til *gå 'væk* er blot nævnt

<sup>2</sup> Uddragene fra JO og andre ordbøger er placeret som bilag, bl.a. fordi det dermed (via PDF med JO's font) kan sikres, at der bruges korrekte trykstreger foran adverbier mv. i de mange anførte tryktabsforbindelser fra JO. I den løbende teksts font findes trykstreg el. lign. desværre ikke, men må gengives med ' (apostrof), der i visse layoutprogrammer ganske misvisende omsættes til kursiv, når de efterfølgende tegn står i kursiv.

<sup>3</sup> Afgørelser omkring graden af ordforbindelsers ”fasthed” kan være temmelig vanskelige at træffe, som mange ordbogsredaktører har erfaret; jf. også diskussionerne i Steyer 2004.

summarisk. Under betydning 18 *gå 'ad* nævnes dels en betydning (a) ”slippe op”, dels en fast forbindelse (b): *hvad går der 'ad (ham/dig)*, der så jævnføres med det mere almindelige udtryk for det samme i betydning 19 (k), *hvad går der 'af ham/dig* = hvad er der i vejen med dig. Ser man i stedet på, hvad der nævnes under præpositionen *ad* (i bilag 2), fremgår det, at forbindelsen *gå 'ad hånden* anføres her, idet den særlige betydning vedr. brug af hestevogne mv., ”gå til venstre for kusken”, snarere ligger i præpositionen *ad* end i verbet *gå* og derfor behandles her (jf. betydning 1.1, FF efter anden redaktionelle ”planke” //). Forbindelsen kunne også være medtaget under subst. *hånd*; men dér nævnes den ikke blandt de faste forbindelser, heller ikke som en henvisning, lige som den ikke nævnes under *gå 18*. Her bør måske indføjes henvisninger, hvis der engang bliver mulighed for supplerung af det i forvejen meget omfattende henvisningsapparat i JO.

Lignende forbindelser med *ad* ses under betydning 4.1 (igen FF efter anden ”planke”): *holde 'ad sig*, om at styre hesteforspand for vogn mv. til venstre. Begge steder henvises til modsvarende forbindelser med præpositionen *fra*, nemlig *gå 'fra hånden* henholdsvis *holde 'fra sig*, om at gå til højre for kusken henholdsvis styre forspandet til højre, jf. *fra* betydning 3.3-3.4 (i bilag 2). Disse forbindelser med verbet *holde* nævnes ikke blandt tryktabsforbindelserne i 2. *holde* afsnit F (jf. bilag 1), derimod i afsnit D, betydning 15 med henvisninger til behandling i artiklen *ad 4.1*. I afsnit F ses desuden andre betydninger, nemlig dels i punkt 22 *holde 'ad* (”blive ved”), dels i punkt 23 en helt anden betydning af *holde 'ad sig*, nemlig ”indsmigre sig hos” (vi refererer ofte til betydninger blot med ordet ”punkt”). Denne betydning behandles i en andetsteds under *ad*, nemlig i punkt 1.2, hvor der anvendes formuleringen ”holde sig gode venner med”, der er stort set ækvivalent med den nævnte *fra 2. holde 23*. Denne positive betydning af *ad* i forbindelse med personer er i *ad 4.1* (bilag 2) også registreret for verbet *være* i forbindelsen: *der er noget ad nogen*, svarende til udtryk som ”ham er der noget ved, han duer til noget”. Her er valgt en formulering med præsens, da en infinitiv som ”være noget ad nogen” let vil forekomme for abstrakt eller ubestemt. Forbindelsen er på sin plads her under adverbiet *ad*; den næppe vil næppe blive nævnt under verbet *være*, da mange lignende konstruktioner så også skulle nævnes dér. Det samme gælder den faste forbindelse i slutningen af *ad* betydning 4.2: *være i' mod' ad* (= være genstridig), der udmærker sig ved at forbinde to trykstærke adverbier; den nævnes også i artiklen *i' mod*, betydning 6 (jf. bilag 2).

Et sidste forhold, der skal nævnes, er forbindelsen i betydning 10 *det bliver bedre ad*, der betyder ”det bliver bedre og bedre”, og tilsvarende for *værre*; dette har vi i JO også valgt at behandle her under adverbiet *ad* frem for i artiklen på adjektivet 1. *bedre* henholdsvis den kommende artikel på adjektivet *værre*.



## 2. Sammenligning med andre større ordbøger

Som vi har set ved *holde* og *gå* ovenfor, placerer JO ved de store verber tryktabsforbindelserne i et særligt afsnit med fortløbende betydningsnummerering. *Ordbog over det Danske Sprog* (ODS) følger samme princip med afsnitsopdeling og et betydningsnummer til hver fast forbindelse, se bilag 3.1 vedr. verbet *II. holde*. Det anføres, at kun de vigtigste tryktabsforbindelser medtages; alligevel når man op på mere end 20, nemlig fra betydning 27 *holde ad* til betydning 48 *holde ved*, foruden at der henvises til behandling af forbindelser som *holde borte*, *holde hårdt* under det pågældende adverbium eller adjektiv. Den særlige betydning af *holde ad sig* = ”dreje (hest) til venstre”, som vi betragtede ovenfor, er også i ODS behandlet under *ad*, hvortil der henvises sidst i betydning 27.2. I denne artikel er fremhævet, at præpositionen er trykbærende (hvad ODS ellers normalt ikke markerer), jf. bilag 3.2 vedr. *ad* betydning 1.1: ”holde ad / 'a / sig”, og som i JO sammenlignes med forbindelsen *holde fra sig*. I ODS-supplementet nogle årtier senere udbygges dokumentationen: der anføres et par litterære citater og indføjes en henvisning til behandlingen under verbet i ODS, se bilag 3.3. Desuden er der tilføjet en parallel fast forbindelse, nemlig *køre ad sig* = ”holde til venstre”, som blev behandlet under verbet *køre* i ODS. (Her kan jeg i parentes bemærke, at det er en stor fordel at redigere ”på nettet”: i JO kan vi straks indarbejde denne type henvisninger, som kommer til i en senere redaktionsfase, og dermed undgå oprettelse af supplementet).

*Ømålsordbogen* (ØMO) har ved tryktabsforbindelser valgt en anden praksis, nemlig at anføre dem under de enkelte betydninger af verbet, ikke i et særligt afsnit efter den grundlæggende opdeling i betydninger, jf. bilag 3.4. Her sætter man forbindelserne *holde ad sig* og *holde fra sig* ind i en større saglig sammenhæng ved at referere til synonyme som adjektiverne *nærmer* og *fjermer*; desuden anføres en udvidet fast forbindelse: *holde ad sig om*. Disse forbindelser er ikke direkte nævnt i artiklen *ad* i ØMO, men antagelig underforstået i formuleringerne i betydning 9 dér: *i forbindelse med andre verber, i forskellige faste forbindelser*.

*Den Danske Ordbog* (DDO) har som venteligt ikke en fast forbindelse *holde ad*; den findes antagelig ikke i ordbogens mediebaserede, moderne materiale fra slutningen af 1900-tallet. I artiklen på verbet *holde* anføres først en række betydninger, 1-14; derefter gives en række faste forbindelser uden nummerering, men med fed skrift som opslagsformen, fra *holde af* til *holde vejret*. For så vidt samme princip som ODS og JO: et antal faste forbindelser anføres alfabetisk til sidst, men som *underopslagsformer* eller *sublemmaer* i modsætning til angivelser med *betydningsnumre* i ODS og JO. Under *holde* har DDO 33 sublemmaer mod de nævnte ca. 20 i ODS så vel som i JO. Men de 33 forbindelser omfatter ud over

tryktabsforbindelser som *holde igen*, *holde inde* også adskillige faste forbindelser med *holde sig* mv. (se bilag 3.5), vilke behandles under verbets primære betydninger i de andre ordbøger. I DDO nævnes under *ad 2* ingen forbindelser med verbet *holde*, derimod fx *spørge ad*, *splitte ad*; disse omtales i JO under *ad* betydning 11-12 (jf. bilag 2).

Der er hermed givet et indtryk af forskellig praksis i behandlingen og placeringen af tryktabsforbindelser i større danske ordbøger; de forskelligartede praksisser er hver for sig fornuftige og velbegrundede i den helhedsbetragtning, den enkelte ordbog anlægger. Det vigtige er, at ordbogen sørger for relevante henvisninger fra adverbiet til verbet og omvendt.

### 3. Enkelte andre typer faste forbindelser

Faste forbindelser tilstræbes generelt behandlet under det vigtigste ord i forbindelsen, eller med en formulering fra JO's redaktionsregler – ganske vist vedrørende placeringen af ordsprog, men også gældende for andre typer ordforbindelser: de placeres under ”det betydningsfuldeste ord, dvs. det centrale, bærende ord, der *selvfølgelig* ofte kan være svært at udpege” (JO 1970: XXIII f., min udhævning). Her lægges altså op til, at der foretages en grundig afvejning og udøves det bedst mulige redaktionelle skøn. Det er naturligvis ofte substantiver, der udpeges som det centrale ord i faste forbindelser<sup>4</sup>; under *ad 5* henvises således til *færd 5.1* angående forbindelsen *ad færde* = på færde, jf. bilag 2. I udpræget grad ses i samme bilag henvisninger til substantiver i artiklen *2.ae* betydning 2 (= borte fra): til behandling under fx *1.by 2.2*, *drift 4* og til flere fremtidige artikler, fx *land x*, *sted x* (med *x* betegnes, at artiklen endnu ikke er redigeret, men indtastet i en database for ord til medtagelse i senere redaktion). I sammenhængen her, faste forbindelser med præpositioner og adverbier, er det dog oftest verber, der henvises til, fx *tage*, *løbe*, *slå*, *give* i artiklen *efter* (ligeledes bilag 2). Videre ses i artiklen *1.ind* betydning 3.1-3.4 henvisninger til både dialektale verber (*odde*, *fætte*) og verber kendt fra rigsmål, som fx *hugge* og *slå*; tryktabsforbindelsen *hugge 'ind* henvises til behandling under *hugge 14* (det vedrører kornhøst med le, hvor stråene mejes ind mod det uhøstede korn). Men der henvises også ofte *mere indirekte* til adskillige verber i artikler som adverbiet *af* betydning 20-21 (jf. bilag 2). I betydning 20 nævnes her bl.a. *måle 'af*, *skridte 'af*; i betydning 21 bl.a. *fodre 'af*, *pløje 'af*, *gøre 'af*, *hilse 'af*, hvor adverbiet endvidere kan foranstilles nogle af verberne i participium til sammensætninger som *afgjort*, *afløben* (markeret med “sammensætnings-prik” i JO).

<sup>4</sup> Jf. også Vikør 2008: 467f.

En type fraser, man vil forvente repræsenteret i ordbogsartikler på præpositioner og adverbier, er fraseokabeloner (dvs. ordforbindelser med tompladser<sup>5</sup>), og der ses da også adskillige i udvalget her. Det gælder forbindelsen *ad ... til* brugt om steder (jf. *ad* 4.1, bilag 2), fx *de skal ad byen til*, i nogle egne også omvendt: *de skal til byen ad*. Også præpositionen *efter* kan bruges på denne måde, fx *køre efter byen til*, nogle steder udvidet til *efter byen til-den*. Præpositionerne *fra* og *af* optræder sammen i forbindelsen *fra ... af*, ofte brugt om tid (jf. *af* betydning 3), fx *helt fra begyndelsen af*. En anden type fraseokabelon er *X for X*, der i artiklen 2. *for* betydning 10 ses med tidsangivelse: *år for år* og med en anden målangivelse: *fod for fod*. Endelig bemærkes typen *fra X til Y*, her med tidsudtryk som *fra dag til dag*, jf. *fra* betydning 6. Disse typer kendes også fra rigsmål; men JO er forpligtet til at vise, at de også findes i dialekterne. En særlig dialektal type findes med *imod*, nemlig tidsudtryk som: *om en time imod*, der har betydningen ”om cirka en time” (jf. *i' mod* 3, bilag 2).

#### 4. Andre redaktionelle forhold ved faste forbindelser

I princippet anvendes i JO i den redaktionelle tekst infinitiv i formuleringen af faste forbindelser, men enkelte gange er det nødvendigt at bruge en præsenskonstruktion med *der*, hvilket vi i afsnit 1 strejfede ved forbindelsen *der er noget ad nogen*, om en, der er dygtig, duer til noget. Det sker ligeledes ved udtryk som: *der går* (sygdom mv.) *i noget*, fx *miltbrand i kornet*, hvor man naturligvis kan bruge infinitiv, men hvor præsens er mere læservenligt (jf. *gå* 3, bilag 1). Ved *imellem* (bilag 2) bliver forholdene endnu vanskeligere: man kan i betydning 5.3 nok anføre infinitiv: *være i' mellem* (= være uenige), men ikke i betydning 5.4 gengive strukturerne her præcist ved brug af infinitiv alene: formuleringen *være (noget) i' mellem* får ikke personrelationerne med, som det tilstræbes med brug præsens: *der er noget i' mellem (nogen)* = der er uenighed mellem (nogen). Det samme gælder de faste forbindelser i 3. *i* betydning 1.2: en infinitivkonstruktion som fx *\*være en (+ subst.) i ham/hende* kan ikke bruges til angivelse af det tilsigtede; der kræves en præsensformulering som fx *der er en (+ subst.) i ham/hende* til at udtrykke, at ”han/hun er en (+ subst.)”, fx ”han er en slagsbroder”.

Det skal også bemærkes, at JO i behandlingen af de faste forbindelse enkelte gange lader det dialektale materiale afspejle sig i den redaktionelle tekst, som ellers formuleres på rigsdansk. Her vil jeg først nævne opslagsformerne *ae* og *frae*, der gælder særlige udtaleformer af *af* og *fra*, fx under 2. *ae* 2 (bilag 2) opfor-

<sup>5</sup> Jf. Farø & Lorentzen 2009: 77f., 94f.

dringen: *bliv æ havet i dag* = ”lad være at tage på havet i dag”. Denne form af adverbiet *af* findes også i rigsdansk talesprog, blot uden tilsvarende skriftlig form, mens JO har valgt at repræsentere den som en artikel; tilsvarende for *fræ*.

Dernæst drejer det sig om sideformer med forstavelsen *i-*, *e-* el. *æ-* til nogle af præpositionerne og adverbierne i bilag 2, fx *tage æ' bag* = være afvigende, unormal (jf. 4. bag 3). Der findes næppe en tilsvarende rigsmåls vending *\*tage bag*, så her må det være rigtigst at anføre den dialektale faste forbindelse. Tilsvarende gælder forbindelsen *være æ' for* = være foran.

Det sidste forhold gælder det helt generelle dialekttræk, at substantiver har foransat bestemt artikel i vestjysk og sønderjysk, i modsætning til østjysk og øvrigt dansk (og nordisk som helhed). Første eksempel, *de er af æ folk* (jf. 1. af 4, bilag 2) ville direkte oversat lyde: ”de er af folkene”; det betyder ”af familien, slægten”, hvilket man dog henvises til 1. folk 5 for at få oplyst. Dette udtryk kan ikke bruges på østjysk eller rigsmål med samme betydning. Det samme gælder forbindelsen *i æ dage* = for tiden (jf. 3. i betydning 4; direkte oversat: i dagene). Endnu fjernere fra rigsdansk ligger udtryk som *gå i æ bare ben* (jf. 3. i betydning 1), som direkte oversat til rigsdansk ville lyde ”\*gå i barbenene” eller ”\*gå i de bare ben”. Under alle omstændigheder ville en direkte omsætning lyde ubehjælpeligt på rigsdansk; vi har derfor bibeholdt det dialektale i formuleringen af den faste forbindelse i den redaktionelle tekst. Denne forbindelse er nævnt uden bestemt kendeord i artiklen 1. ben 1 i JO, men ikke som fast forbindelse; under adjektivet 2. bar er den ikke nævnt. Dermed fortjener den at blive nævnt under præpositionen 3. i.

Dette illustrerer problematikken med at vurdere, hvor meget man skal gøre ud af faste forbindelser i ordbøger: hvor etablerede skal de være for at fortjene særlig påpegning i en ordbogsartikel. Forbindelsen er fremhævet i én artikel vedr. et indgående ord (3. i), blot nævnt som eksempel i artiklen på et andet centralt ord (1. ben), og slet ikke nævnt i artiklen på det tredje centrale ord (2. bar), altså vurderet som af forskellig vigtighed i de tre artikler. Mange steder henviser JO, som vi har set eksempler på, fra ét ord i en fast forbindelse til behandling under et andet ord i forbindelsen (oftest substantiv og verbum); sådanne henvisninger er vigtige for en fuldstændig behandling af en given fast forbindelse i større ordbøger.

## 5. Afslutning

Om de faste forbindelser præsenteres som sublemmaer som i Den Danske Ordbog, eller som et antal særbetydninger til sidst i artiklen som ved verberne *holde* og *gå* i JO og ODS, eller under de enkelte betydninger som i Ømålsordbogen,

er måske ikke så afgørende for læseren, blot der er præcise henvisninger. Her kan den nye it-baserede redaktionsplatform, der er under udvikling på JO, antagelig i en senere fase give god mulighed for at sikre, at henvisningerne faktisk findes fra alle vigtige ord i en given fast forbindelse til det eller de ord, hvor forbindelsen behandles mest udførligt eller endegyldigt. It-platformen kan også medvirke til ensartethed i fastsættelsen af, hvad der skal regnes som faste ordforbindelser, og hvad der er mere tilfældige ordforbindelser, som ikke kræver udførlige henvisninger.

## FORKORTELSER

FF = fast forbindelse; F = (mindre fast) forbindelse; it = informationsteknologi.

## KILDER OG LITTERATUR

- Arboe, Torben, 2008: Fraseologiske enheders indplacering i Jysk Ordbog. I: Nordiske Studier i Leksikografi. 9. Red. af Ásta Svavarsdóttir (m.fl.). Reykjavik. S. 1-18.
- Arboe, Torben, 2011: Jysk fraseologi og idiomatik. I: Jysk, øsmål, rigsdansk mv.; Studier i dansk med sideblik til nordisk og tysk; Festskrift til Viggo Sørensen og Ove Rasmussen. Red. af Torben Arboe og Inger Schoonderbeek Hansen. Aarhus. S. 257-275.
- DDO = Den Danske Ordbog. 2003-2005. bd. 1-6. København.
- Farø, Ken & Henrik Lorentzen, 2009: De oversete og mishandlede ordforbindelser – hvilke, hvor og hvorfor? I: *LexicoNordica* 16. S. 75-101.
- JO = Jysk Ordbog. 1970-1979, hæfte 1-4; <http://www.jyskordbog.dk/> 2000ff. Peter Skautrup Centret for Jysk Dialektforskning, Aarhus Universitet.
- ODS = Ordbog over det Danske Sprog. 1914-1956. bd. 1-28. København.
- ODS-Supplementet = Ordbog over det Danske Sprog. Supplement. 1992-2005. bd. 1-5. København.
- Steyer, Kathrin (Hrsg.), 2004: *Wortverbindungen – mehr oder weniger fest*. Berlin / New York.
- Vikør, Lars S., 2008: Behandling af fraseologi i Norsk Ordbok. I: Nordiske Studier i Leksikografi. 9. Red. af Ásta Svavarsdóttir (m.fl.). Reykjavik. S. 465-476.
- ØMO = Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer. 1992ff. Afdeling for Dialektforskning, Københavns Universitet.

*Torben Arboe*

Jysk Ordbog, Peter Skautrup Centret for Jysk Dialektforskning,  
 Institut for Æstetik og Kommunikation, Aarhus Universitet.  
 jysta@hum.au.dk

**BILAG 1.**
**Faste ordforbindelsers placering i verbets mikrostruktur / (under)betydninger i Jysk Ordbog**

I afsnittene med tryktabsforbindelser er de præp./adv., der er anført i bilag 2, understreget

**gå** verb.

A. om bevidst (i alm. måltret) bevægelse.

3) ... || der går (sygdom, skadedyr etc.) i noget.

F. i tryktabsforb. med efterfølgende adv.

18) gå 'ad a) = slippe op, b) hvad går der 'ad (person) > 19k - 19) gå 'af a) = forlade noget; b) = stikke af; c) = abortere; d) = dø; e) = gå løs; f) = affyres; g) = forsvinde; h) = fortælles; i) = forløbe; j) = gå an; k) hvad går der 'af (person) - 20) gå 'an a)-c) - 21) gå der'hen, gå der'over - 22) gå 'efter = efterse - 23) gå 'for sig - 24) gå for'bi - 25) gå 'fra a) = løsne sig; b) = fratrækkes; c) = gå hurtigere end; d) = benægte; e) det går fra (én) = man glemmer det - 26) gå én noget 'fra = snyde - 27) gå 'frem a)-b) - 28) gå 'hen a)-d) - 29) gå 'hen og ... - 30) gå 'hjemme, syn.: gå 'inde - 32) gå i'mod a) = støde på; (overført: modarbejde; b) gå i møde - 33) gå 'ind a) = begynde; b) = slutte; c) gå 'ind på = acceptere - 34) gå 'inde [35]-51): gå 'med, 'ned, 'om, 'op, 'over, 'på, 'rundt, 'sammen, 'til, til'bage, 'ud, 'under, 'ved, 'ved sig, 'væk]

**holde** verb.

D. om styring, kontrol, ledelse.

15) || h. 'ad sig, h. 'fra sig = styre til venstre / højre, se *ad 4.1, fra 3.4* - h. det med én - 16) || h.

'lade-/'tømmevogn - 17) || h. 'halvt = investere fælles > 'halv 1.5 b - 18) || h. hus > hus

F) i tryktabsforb. m. efterflg. adv.

22) h. 'ad = blive ved - 23) h. (nogen) 'ad sig = indsmigre sig hos, se *ad 1.2* - 24) h. 'af, a) = synes om, b) c) d) e) f) = holde 'ud, g) = holde 'op - 25) h. 'an, a) b) - 26) h. 'bi > 'bi - 27) h. 'for, a) = standse foran, b) = være den, det kommer an på el. går ud over, c) = mene, synes, regne med - 28) h. 'frem, a) b) - 29) h. 'hen - 30) h. 'her - 31) h. 'hid - 32) h. i'gen - 33) h. 'i = klæbe til, holde fast - 34) h. 'inde, a) b) - 35) h. 'med - 36) h. 'op, a) = løfte, b) c) d) e) = standse, f) = sinke, udsætte - 37) h. sig 'op, a) b) - 38) h. 'på, a) b) - 39) h. 'sammen, a) b) - 40) h. 'til, a) b) = få bedækket (hundyr), c) d) = opholde sig - 41) h. sig 'til - 42) h. 'ud, a) = blive ved trods besvær, b) c) - 43) h. 'ved, a) = holde godt fast, b) c) = fortsætte

**BILAG 2.**
**Udsnit af artikelstrukturer i JO vedr. faste ordforbindelser (FF) og øvr. ordforbindelser (F)**
**ad** præp., adv.

A. som præp. (og hermed beslægtet adv.).

1) = ved, hos. - 1.1) om sted el. genstand || FF: falde/ligge/sidde 'ad = ... tæt ind til (om fx klædningsstykke) - være 'ad = være her, være til stede - være 'allesteder 'ad = være allevegne || FF (om trækdyr:): gå 'ad hånden = gå til venstre i spandet (idet en kusk alm. går på denne side, når han leder det); i modsætning til 'fra hånden (se *fra 3.3*) [jf. holde 'ad sig > 4.1] - (overf.): (være/skulle etc.) 'ad noget = i gang med noget - 1.2) om person el. familie || FF (med overf. betydning): (der er) noget ad nogen = noget (positivt) ved vedkommende - holde nogen 'ad sig = holde sig gode venner med vedkommende - (være) lig ad (= ligne, være ligesom) se *lig x*.

4) = til; (ud) i; imod; efter. - 4.1) om målet for bevægelse. || FF: ad ... til = rigsm. - til ... ad = d.s. - fra/til/tilbage ad = fra hhv. til (et underforstået sted, fx hjemmet) || FF: holde (etc.) 'ad sig = styre (vogn etc.) til venstre (i modsætning til holde -fra sig); skønt ulogisk bruges udtrykket også til og om kuskan, når han ved kørsel sidder til højre på bukken. - (kommando til trækdyr: (hold) 'ad dig! = (hold) til venstre (i modsætning til -fra dig). - stå/læg ad dig/sig! = gå til side. ... - 4.2) om målet for ytring, holdning, tanke el. interesse || FF: lægge 'ad = lægge kræfter i - lægge sit vid / sin hu ad (el. lign.) = være opmærksom på; gøre sig umage med - være 'i'mod 'ad = være kontrær - se ad (= se til, sørgе for) se *se x*

5) = gennem; hen ad; langs med || FF: ad døren = ud; af sted; hjemmefra - ad færde se *færd 5.1*. - (overført:): ad ét reb = i samme retning. - ad vejen = på færde; derudad. ...

B. som adv.

10) = stadig; efterhånden; især i F *blive bedre/værre ad* = blive bedre og bedre etc.)

11) = om; med hensyn til; især i F som *spørge ad, bede ad, vare ad* (= varsle, advare)

12) = fra hinanden; i stykker; i F som *skille(s) ad, splitte(s) ad, sprede ad* etc.

**'af** præp., adv.

A. som præp. (og hermed beslægtet adv.).

1) = (ned, op, ud, hjem etc.) fra. – 1.1) om bevægelse. ... – (som adv. udvidelse:) *fræ 'wòs ðá:ə* = fra os af (dvs. fra vores gård). // mulig som FF// ... – jf. *af 'uden* (se *uden x*).

3) = fra (om tid). – kun i flg. FF: fra ... **'af** – **af det første** = fra begyndelsen, fra først af – **af ny** (= fra ny af) se *ny x*. – af nylig se *nylig x*.

4) = stammende fra; boende i. ... – jf. *af æ folk* (= af familien), se *'folk 5*.

B. som adv.: undertiden svarende til forleddet *af-* i rigsmål el. andre dialekter (fx *\*jør aa = afgøre*); spec. i Sønderjy

20) = fra (til adskillelse); også i F som *dole/måle/skridte/trine 'af*

21) = færdig, til ende, op; fx i F som *fodre/malke/pløje/røgte 'af, gå 'af* (= ende, foregå), *gøre 'af* (= afgøre, afslutte, klare), *hilse/takke 'af* (= sige farvel), *lægge 'af* (= holde op med; nedlægge), *fly 'af* (= give en omgang prygl); desuden *gjort/kørt 'af* (= udaset), se *'afgjort, afkørt*; endelig *afhestet, afstyret, afløben* etc. (som udtryk for, at et husdyr er gjort drægtigt).

26) i FF med andet adv. (se dog *fra ... af* i betydning 3). – 26.1) **af og til** = nu og da. – 26.2) **af og i** (= som det kan blive, tilfældigt) – 26.3) **af og på** = op og ned; nogenlunde – 26.4) **op og af** = op og ned; nogenlunde.

**²æ** præp., adv. [ca. 1325: *auæ* (DgL.II.409)]; < *'af*, dannet i analogi til *oppe, nede, henne* etc., og ligesom disse brugt i forb. med tilstandsverber ... ]

2) = væk, borte fra (om afstand) ... bliv æ havet i dag || FF (med ligefrem el. overført betydning): *æ by(en)* se *'by 2.2*. – *æ drift* se *drift 4*. – *æ landet* se *land x*. – *æ lave* se *lave x*. – *æ led* se *led x*. – *æ sted* se *sted x*. – *æ vejen* se *vej x*. – *være 'æ vinden* = ikke fange vinden (om mølle) ... – *være/blive (etc.) 'æ med* = være/blive fri for, komme af med; miste, mangle; etc. ... han er æ med det ene øje

**'bag** adv. – med sideformen **i'bag** (el. med lignende tryksvag forstav.). [sideformen < gammeldansk præp. *a* (= på) + *'bag*]

A. uden efterflg. præp. el. adv.

1) = bagi, bagtil. || *ikke være æ'for* som andre *æ'bag* (el. lign.) = være afvigende, unormal

3) *tage æ'bag* (el. lign.) = være el. gøre noget for sent, forgæves; gå med uforrettet sag [formentlig < betydning 2 (= bagefter)]

**efter** præp., adv., konj. – med sideformen **after**.

A. som præp. (evt. brugt som adv.).

4) = i overensstemmelse med, på grundlag af (forlæg, norm, disposition, opfattelse etc.). || FF: *tage efter* (= tro på; tage varsel af) se *tage x*. – *sæde efter* (= regne med) se *sæde x*.

8) = (i)mod; til. – 8.1) om retning || FF: *efter ... til ...* efter byen til – *efter ... til-den* – *efter ... på*. – *til ... efter* – (overført:) *tæt efter* – *efter alle sider* se *side x*. – *løbe efter enden* se *løbe x*. – 8.2) om tid ... || *slå efter enden* se *slå x*. – 8.3) om egnethed. – kun i flg. FF: *være 'efter (det)* = passe, egne sig til

B. som ægte adv.

12) (i stedsbetydning:) *give 'efter* (for pres etc.) se *'give 14.4 a*.

16) = tilbage, retur; i forb. *give 'efter* (= eftergive) se *'give 14.4 b*.

**²for** præp., adv. – med sideformen **i'for** (el. med lignende tryksvag forstavelse). ... *e-*, *æ-*

A. med stedsbetydning el. betydning i umiddelbar forlængelse heraf.

5) = i nærværelse af, hos; over for, ... beregnet på; som medhjælp for; især i F som *synge for (nogen), tage hatten af for (nogen), sy for tækker* (se *sy x*), *drage for høle(en)* (se *³drage 7.1*) || FF: (hver) for *¹sig, for sig ¹selv* = rigsm. ... - *være æ¹for* = være i gang, i virksomhed

B. med tidsbetydning el. betydning i umiddelbar forlængelse heraf.

10) = pr. (tidsenhed), svarende til || (også) = *efter* (om rækkefølge); især i F som *år for år, fod for fod* etc. [spor. afhjemlet]

**fra** præp., adv., konj. – ... også (med tryksvag forstavelse) *æ-, a-, e-, i-*

A. med stedsbetydning.

3) som mere speciel retningsangivelse (i reglen med underforstået og mindre gennemskuelig styrelse). ... – 3.3) *gå ¹fra hånden* = gå til højre (i spandet, om trækdyr for plov el. vogn); i modsætning til *gå ¹ad, ¹til hånden* (se *ad 1.1, til x*) – 3.4) *holde ¹fra (dig/sig)* (el. lign.) = styre (vogn etc.) til højre; i modsætning til *holde ad sig* (se *ad 4.1*), *holde til sig* (se *til x*)

B. med tidsbetydning.

6) = siden; startende med. || (spec.:) *fra dag til dag* = dag efter dag (og lign. forb.)

**frae** præp., adv. – samme udtale som den trykstærke form af *-fra*, blot med længde i stedet for stød og i TonalOmr (K 1.9) med tonal acc.2.

1) = borte fra; løsrevet fra. ... *holde sig frae* (noget). || i overført betydning ... FF: *være ¹frae* (det) = være "ude" af en leg – *være ¹frae det/den* = være ude af sig selv (af skræk, vrede etc.)

**³i** præp. og adv.

A. med stedsbetydning.

1) om placering el. bevægelse. – 1.1) i alm. || (spec.:) *være ¹i* = være med (i leg mv.) – 1.2) i kropslig el. (overført:) sjælelig henseende || FF: *der er kalv* (el. tilsv.) *i den* = den er drægtig (om ko mv.) – *der er en (+ subst.) i ham/hende* = han/hun er en (+ subst.) ... fx *dreng / slagsbroder* – *¹af hånden og ¹i munden* (og lign.) – *af 1.1*, slutn. – 1.3) vedr. påklædning. || (spec.:) *gå i (æ) bare ben / fødder / hoved* (mv.) = gå barbenet / barfodet / barhovedet

B. med tidsbetydning.

4) om bestemt tidsrum el. tidspunkt || FF: *i dag, i går, i aften, i morges* (etc.) = (rigsm.) se under de enkelte subst. – *i tiden, i æ tid* = for tiden – *i æ dage* = for tiden ...

**i'mellem** præp., adv., konj. – med sideformen **mellem**.

A. som præp.

5) vedr. menneskelige relationer. ... 5.3) *komme/være i'mellem* = blive/være (stærkt) uenige – 5.4) *der kommer/er noget i'mellem (nogen)* (el. lign.) = der opstår/er uenighed, strid mellem nogen || (også med brug af personligt pronomem): *vi er/bliver noget imellem* = d.s. || (hertil:): *gå imellem bark og træ* (= sætte splid imellem personer) – *¹bark 1*.

**i'mod** præp., adv. – med sideformen **mod**.

B. vedr. tid mv.

3) = i retning af tidspunkt på dagen / året mv. – *om en time* (etc.) *i'mod* = om (ca.) en time

C. vedr. persons holdning/indstilling.

6) = i opposition til; negativ over for. ... || *være i'mod ¹ad* = være genstridig, kontrær.

**¹ind** adv.

A) som retningsadv.

3) vedr. andre genstande, ting mv. – 3.1) = hen mod ting mv. ... jf. *hugge ¹ind* (*-hugge 14*), *odde ¹ind* (*-odde x*), *slå ¹ind* (*-slå x*). – også: *lægge sig ¹ind* (om hest) – 3.3) = hen til række, (spec.:) hen mod midterlinje ... jf. *fætte ¹ind* – *³fætte*. – 3.4) = hen til opbevaringssted, samlingssted el. lign. || (hertil:): *få ¹ind* (= indsamle, indtjene) – *¹få 17*. – *købe ¹ind* – *købe x*.



Bilag 3.

3.1. Uddrag fra ODS II. holde

**D.** i faste forb. m. adv. og trykstærk præp. (kun de vigtigste anv. er medtaget; jf. ogs. u. bet. 23.1; om forb. som holde borte, fast, haardt, hjemme, se II. borte (1.6) osv.).

**27) holde ad.** (om holde ad for holde af se bet. 28.1). **27.1)** (nu kun dial.) til bet. 2.1 (og 3.1) i forb. holde en (noget) ad sig, (jf. ad A.1) søge at faa noget hen til sig; nærme til sig; m. h. t. person: knytte fast til sig; faa til at komme til sig. *Moth.H235.*

...

**246.** jf. *Gram.Nucleus.1157. 27.2)* (til bet. 23.1) i forb. holde ad sig (ogs. jf. ad A.6) af sig. *PAHeib.Sk.II.59. Tolderl.F.III.154), dreje til venstre* (se ad A.1.1).

3.2. Uddrag fra ODS ad

**ad,** præp. og adv.; som præp.: [a], kunstlet: [að]; som adv. [að]; ad. *Høysg.AG.24,*

...

*skr.VilhThoms.278* || om gaa ad glemme se Glemme; om op ad dage se Dag)

**A.** som præposition (med styrelse).

**1)** om bevægelse i retning henimod.

**1.1)** (dial. ell. foræld.). uden efterstillet adv.

...

familien B. paa vejen). *Bang.GH.201.* || **holde ad** [ʼa] sig, (dial.). dreje (hestene) til venstre, mods. holde fra sig (udtrykket, fordi kørsvendten, fx. ved pløjning, gaar til venstre for trækdyrene). *Moth.A17. Høysg.S.267. Esp.1. Thorsen.121. Feilb.I.3. ad dig!, hold hestene til venstre! Køværd. jf. adhaands. 1.2)* i nutidens rigsspr. som oftest

3.3. Uddrag fra ODS-Suppl. I. ad

**I. ad,** præp. og adv. + (o.1700 ogs. at. \*ned at Bakken. *Rose.Ovid.II.41.* jf. atskillig u. adskillig 2.2 S).

**A.** som præp.

**1)** || 95.5 holde ad sig, + (se ogs. (m. videre anv.) holde 27). Naar du holder ad dig paa hin Side den første Bro, har du den lige Landevej for dig til Schwerin. *Ing.VS.I. 154.* Ja, hold a' Dig (fodnote: Hold a' Dig = kjør til Venstre), saa jeg kan komme forbi. *J Helms.G.56. jf.: Adsigvendinger. PNSkovgaard.B.29.* køre ad sig, se II. køre 10.

3.4. Uddrag fra Æmålsordbogen holde II

-5.1) mht hest (ell køretøj) Ø(alm): hál he'sten

...

vF(Viss); i forb holde →hid (1.4), holde →hørs om at styre henh til venstre og til højre med et hesteforspand (idet kusen som regel går ell sidder til venstre), holde til →nærmer, holde til →fjermer (dvs nærmer og fjermer hest) ds, holde ad sig, holde fra sig ds nvS-

...

fåbij' vF(Ørst), nu skal han holde lidt fra sig "trække til højre i linerne" vF(Br), husk at holde ad dig, når du kører ud af ledstedet, at ikke baghjulet skal gå på L(Rad), å hálle hæstene a sø vF(Kavs); holde ad sig om, holde fra sig om om at køre henh venstre om og højre om

3.5. Uddrag fra Den Danske Ordbog – holde

<p><b>holde igen</b> forsøge at modvirke en stærk påvirkning, fx et træk; [NGN holder igen (på NGT)] □ så begyndte det</p> <p>...</p> <p><b>holde inde</b> afbryde en igangværende aktivitet, ofte tale; [NGN holder inde (med NGT/at+INF)] □ <i>Stedfar bad ham</i></p> <p>...</p>	<p><b>holde sig for god til</b> ikke nedværdige sig til; føle sig for god til □ <i>Astrologi</i></p> <p>...</p> <p><b>holde sig for ørerne (øjnene, næsen, ..)</b> lægge sin hånd el. sine hænder for ørerne (øjnene, næsen, ..), især for at undgå at høre, se, lugte osv.</p>
---	---

# *Norsk Ordbok* på nett: Utfordringar og løysingar

*Sturla Berg-Olsen & Åse Wetås*

Norsk Ordbok (abbreviated NO) is a scientific dictionary covering Norwegian in its written Nynorsk form and all spoken dialects. The dictionary is under production and will be finished in 12 volumes in 2014. The work on NO began in 1930, and the first volume was published in 1966. From 2002 the project organization NO 2014 has been working towards the completion of the dictionary, all the time on the premise that the final product will be a printed dictionary. Over the last few years a growing number of dictionaries have been published in an electronic format on the World Wide Web – either as a supplement to or an alternative to a printed edition. We see a development where more and more lexicographical products are exclusively published in an electronic form. This format renders possible a more user-friendly search functionality and more flexible ways of displaying dictionary entries. In addition, the format makes it fairly easy to link lexicographical resources with other existing, relevant electronic resources (cartographic data, source texts, encyclopaedic resources etc.). We at NO 2014 would also like to publish our dictionary electronically, but we have found ourselves faced with certain lexicographical challenges that have to be looked into in more detail. In this article we discuss some of the problems encountered in the work on the online version of NO, with particular emphasis on search mechanisms.

**NØKKEWORD:** e-leksikografi, eittspråklege ordbøker, søkjemekanismar, dialektordbøker

## 1. Innleiing

*Norsk Ordbok* (NO) er ei vitenskapleg dokumentasjonsordbok som dekkjer det nynorske skriftspråket og det norske folkemålet. Ordboka er under produksjon og skal vere ferdig i 12 band til 200-årsjubileet for den norske Grunnlova i 2014. Arbeidet med NO blei starta i 1930, og det første bandet av ordboksverket kom ut i 1966. Frå 2002 har prosjektorganisasjonen Norsk Ordbok 2014 (NO 2014) arbeidd med ferdigstillinga av ordboksverket, og det har heile tida vore føresetnaden at verket skulle gjevast ut på papir. Dei seinaste åra har fleire

og fleire ordbøker blitt publiserte elektronisk på nettet, anten som supplement til ei prenta utgåve eller som eit alternativ til prenting. Utviklinga går i retning av at stadig fleire leksikografiske arbeid utelukkande blir publiserte elektronisk. Dette opnar for meir brukarvenleg søkjefunksjonalitet og fleksibel artikkelvisning, og det elektroniske publiseringsformatet legg også til rette for ei samankopling av dei leksikografiske ressursane med andre tilgjengelege og relevante elektroniske ressursar (kartdata, grunnlagskjelder, ensyklopediske ressursar osv.). I NO 2014-organisasjonen ønskjer vi òg å publisere ordboksverket på nett, men dette publiseringsmediet utløyser ein del leksikografiske utfordringar som må diskutertast meir i detalj. I denne artikkelen kjem vi nærare inn på nokre av problemstillingane vi har drøfta i arbeidet med ei nettutgåve av ordboka, og då spesielt spørsmål som gjeld søkjemekanismar.

## 2. Bakgrunn<sup>1</sup>

NO dekkjer både det norske folkemålet (dvs. alle dei norske dialektane) og nynorsk skriftspråk frå 1860-talet og fram til i dag. Dette gjer at vi trygt kan ta utgangspunkt i at ordboksverket har eit samansett publikum. Mange potensielle brukarar vil primært vere interesserte i å hente ut informasjon om ord i eigen dialekt, medan andre kan vere på leit etter ulike nemningar for eit gjeve fenomen, eller etter eldre skriftspråklege belegg av eit gjeve ord. Eit mål vi har sett oss i arbeidet med ein nettversjon av ordboka, er at brukarane både skal få dei rette forventningane til kva dei kan finne i NO, og at dei samstundes får god rettleiing til å finne det dei leiter etter.

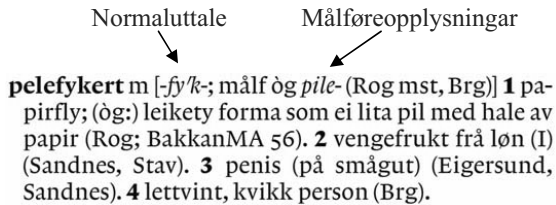
Artikkelmaterialet i NO bruker 1938-normalen for nynorsk som redaksjonell målform, med nokre mindre modifikasjonar som ikkje er så viktige i denne samanhengen. Dei formene frå 1938-normalen som er valde ut, ligg stort sett ikkje så langt frå ein moderne midtlinenynorsk, men normalen skil seg frå dagens nynorsk på nokre punkt. T.d. har substantivet *jente* oppslagsforma *gjente* i NO.<sup>2</sup> Ordboka tek utgangspunkt i at nynorsk skriftspråk og norske talemål er to sider av ei og same sak, eller siamesiske tvillingar, om ein vil. Det inneber at den redaksjonelle handsaminga av ord som berre er dokumentert nytta dialektalt, også kan få ei normert nynorsk form. Den nynorske forma slike ord får i NO, følgjer 1938-normalen. Døme på dette er belegga bak substantivet *myse-*

<sup>1</sup> Viss information i nedanstående figurer har gått förlorad på grund av svart-vitt tryck istället för färg (red:s anm.).

<sup>2</sup> Mellom dei sentrale systemdraga i NO-nynorsken er *a*-infinitiv, supinum på *-i* (*skriv*, *les*) og adjektivendinga *-ut* (*strip**ut*, *rugl**ut*).

*prim*, der mange har form /møsbrøm/, eller substantivet *matgjerd*, som finst i mange dialektar som /maje/ eller /majæ/.

For at tilhøvet mellom normert nynorsk oppslagsform og talemålsvariantane som ligg under skal komme til uttrykk, blir målføreformene gjevne att i ein eigen del av artikkelen. Her er det berre avvik ut over det som er heilt regelbunde, som blir gjevne att. Det inneber at vanlege, utbreidde former ofte ikkje blir gjorde greie for i artiklane. Til grunn for det ligg det ein føresetnad om at brukarane kjenner til regelbundne målføredrag (slik som lenisering av *p*, *t*, *k* til *b*, *d*, *g* i Sør-Rogaland og Agder, jf. figur 1).

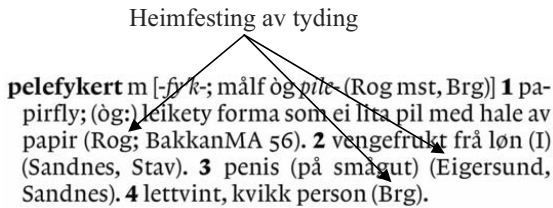


Figur 1. NO-artikkel (i papirversjon) med normaluttale og målføreopplysningar.

I artikkelen **pelefykert** kjem opplysningar om normaluttale av ordet først i den innleiande parentesen med skarpe klammer («hakeparentesen»). Grunnen til at det er ført den typen opplysningar i *pelefykert*-artikkelen, er at trykkplasseringa er avvikande frå den som er den vanlege i norsk. Etter opplysningane om normaluttale følgjer registrert målførevariasjon. Belegga frå Rogaland gjer greie for uttalen /pilyf:grt/. Dette området har lenisering, men i og med at dét målføredraget blir rekna for å vere sjølvgeve, blir det ikkje gjort greie for i NOs framstilling.

Derimot representerer ikkje *e > i* i førsteleddet av dette ordet eit regelbunde målføredrag, og målføreopplysninga *pile-* blir difor førd. Målføreformer som ikkje er sjølvgeve, blir førde inn i hakeparentesen i ordartiklane med geografisk plassering på heradsnivå. Heradssystemet reflekterer kommuneinndelinga i Noreg slik ho var før dei store kommunesamanslåingane på 1960-talet.

Vidare fører NO heimfesting av ordtydingar i tydingsdelen av ordartiklane. I den elektroniske redigeringsapplikasjonen blir heimfestingane førde i eit eige felt knytt til den aktuelle definisjonsteksten. I prenta artiklar kjem dette inn som opplysningar i parentes etter definisjonane (jf. figur 2 nedanfor). Dette er aktuelt i alle dei tilfella der ordtydinga ikkje er vanleg over heile landet. For ordtydingar som er ålment utbreidde, blir det ikkje ført heimfestingar.



Figur 2. Pilene viser heimfestingar av ordtyding.

NO i prenta versjon er tilgjengeleg for tingarar og gjennom forskings- og folkebibliotek som har verket i boksamlinga si. Når vi nå presenterer ordartiklane i ei nettbasert ordbok, er det rimeleg å vente at gruppa av potensielle brukarar av verket blir mykje større. Fordi ordboka i nettversjon møter ei stor og samansett brukargruppe, er det viktig å gjere artikkelpresentasjonen på nett så brukarvennleg som mogleg. Det aller viktigaste å sikre då, er at brukarane så langt som råd får treff som svarer til det dei faktisk ønskte å søkje på. For ei ordbok som presenterer analysar av både talemålsmateriale og materiale frå skriftspråkelege kjelder, er utfordringa endå større enn for dokumentasjonsordbøker som berre dekkjer den eine av desse kjeldetypane.

Det vi ønskjer for nettversjonen av NO, er at brukaren ikkje berre skal kunne søkje på oppslagsord, men også søkje i fritekst eller på geografisk stad ned til heradsnivå. For å sikre mest mogleg relevante treff, ønskjer vi vidare at brukaren også skal kunne søkje seg inn i ferdige ordartiklar gjennom dei koplingane som finst mellom ordartiklar og grunnlagsmateriale. Der er koplingspunktet NOs Metaordbok.

Metaordboka er ein elektronisk indeks der materiale frå ulike materialdatabasar er samla. Dette omfattar mellom anna NOs setelarkiv, Trønderarkivet, grunnmanuskriptet til NO, *Nynorskordboka* (NOB), *Bokmålsordboka* (BOB), Matias Skards *Nynorsk ordbok* (1954) og ei rekkje kjelder som ligg inne i Ordbokshotellet.<sup>3</sup> Belegg frå desse kjeldene er i Metaordboka samla under normerte oppslagsord og grovlemmatiserte.<sup>4</sup>

Det er Eining for digital dokumentasjon (EDD) ved Universitetet i Oslo som står bak NOs redigeringsapplikasjon, og EDD står også for utviklingsarbeidet bak nettutgåva av ordboka. Til første versjon av NO på nett har vi valt å prioritere funksjonalitet høgare enn design. Planen er å arbeide vidare med utsjå-

<sup>3</sup> Ordbokshotellet inneheld lokale ordsamlingar i elektronisk form, saman med ordartiklar frå ymse leksikografiske kjeldeverk, slik som *Nynorsk etymologisk ordbok* (Torp 1919) og *Norsk landbruksordbok* (Rommetveit 1979). For meir informasjon om Ordbokshotellet sjå Tvedt, Lien og Eide (2007).

naden på sidene etter at første versjon av eNO er lagt ut og testa. Etter testing vil vi kunne gjere endringar i måten vi presenterer og legg til rette for søk, for å gjere treffprosenten endå høgare og brukarane endå meir nøgde.

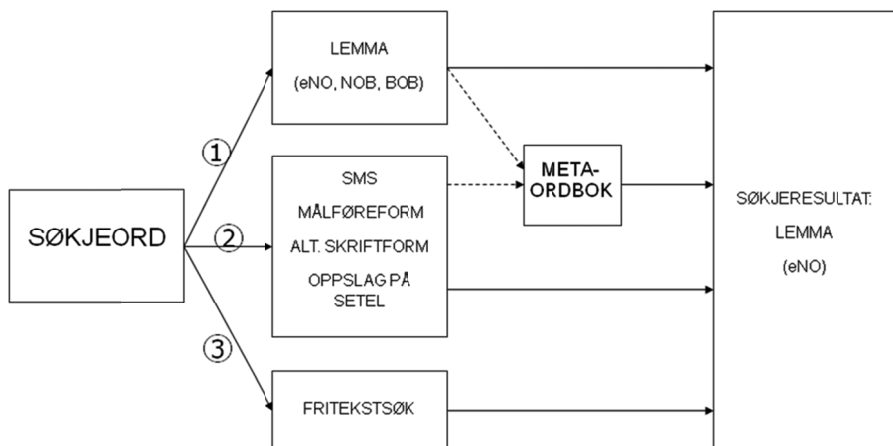
### 3. Søkjevegar ved enkelt søk i eNO

Dei fleste brukarane av eNO, eller i alle fall dei som har minst kjennskap til ordboka frå før, vil truleg bruke funksjonen «enkelt søk». Denne funksjonen vil ha eit søkjefelt utan valmoglegheiter. Vi har som mål å lage ein søkjemekanisme for enkle søk som gjev flest mogleg brukarar treff på det dei er ute etter, men dette lèt seg ikkje nødvendigvis gjere så lett. Dels kjem det av faktorar som alle ordbøker må ta høgd for – som at brukargruppa er heterogen og at nokre brukarar truleg vil vere ute etter informasjon som nett denne boka ikkje er designa for å gje (t.d. gjev ikkje eNO opplysningar om gjeldande offisiell rettskriving). Men vi støyter også på ein del utfordringar knytte til særdrag ved verket NO, som skissert ovanfor. Det gjeld til dømes diskrepansar på ein del punkt mellom oppslagsformene i NO (som følgjer normalen frå 1938) og dei formene som høyrer til dagens norm. Endå meir krevjande er tilhøvet mellom oppslagsformene i standardisert nynorsk, og målføreformene som ein gjeven artikkel representerer, og som ofte skil seg heller mykje frå den normaliserte forma. Ettersom NO – og dermed eNO – også er ei dialektordbok, vil ein tru at mange brukarar vil vere opptekne av å finne ord og uttrykk frå talemålet ulike stader i Noreg. Korleis kan vi utforme eit søkjesystem som fangar opp flest mogleg søk på målføreformer, leier brukaren til rett normalisert lemma og samstundes viser fram tilhøvet mellom den målføreforma det blei søkt på og lemmaet?

Ein stor føremon for søkjesystemet er det faktumet at alle innhaldsartiklar i NO er kopla til ein artikkel i Metaordboka. Slik kan Metaordboka tene som nav i søkjesystemet, fordi mange formalt ulike inngangar til eit lemma er samla her. Korleis dette vil fungere i praksis, er illustrert i nokre konkrete søkjescenarior i avsnitt 7. Figur 3 gjev eit oversyn over korleis søkjesystemet for enkle søk i eNO er tenkt å operere.

---

<sup>4</sup> Metaordboka skil ikkje mellom homografe lemma som høyrer til same ordklasse, eller same genus for substantiv. Såleis ligg til dømes belegg på homografane **I singel** «småkorna stein, grov grus» og **II singel** «1. singelplate; 2. spel med éin deltakar på kvar side» i same metaordboksartikkel, medan adjektivet **V singel** har eigen artikkel i Metaordboka. For meir om Metaordboka sjå Svoldal (2003) og Ore og Tvedt (2006).



Figur 3. Søkjevegar for enkle søk i eNO.

Nummera i figuren illustrerer rekkjefølgda søka opererer i. Når brukaren tastar inn eit søkjeord i rubrikken for enkelt søk, skjer følgjande:

- 1) Systemet søkjer etter ordet blant oppslagsformene i eNO, og samstundes også i oppslagsformene i *Nynorskordboka* (NOB) og *Bokmålsordboka* (BOB), som begge er kopla saman med eNO gjennom Metaordboka. Alle oppslagsord som svarer til søkjeordet i desse tre basane, gjev treff og blir viste i søkeresultatet. På denne måten vil ein brukar som søkjer på substantivet *jente*, få treff på ordartikkelen *gjente* i eNO (jf. det som blei sagt om denne forma i pkt. 2 ovanfor).
- 2) Om ein ikkje har fått treff, blir det søkt fleire andre stader i eNO-artiklane og i metaordboksartiklane. Det gjeld: a) Samansetjingsrekkjer inne i eNO-artiklar (forkorta SMS i figur 3). Desse er rekkjer med samansette ord som ikkje er definerte i NO, men berre lista opp under (ei tyding av) det grunnordet som utgjer førsteleddet i samansetjinga. Om søkjeordet er ei slik samansetjing, skal ein få opp den aktuelle grunnordsartikkelen der samansetjinga er førd opp. b) Målføreformer som er førde i eNO-artiklar. c) Alternative skriftformer som er førde i eNO-artiklar. Dette er former som ikkje fell saman med oppslagsforma, men som er eller har vore utbreidde i skrift. d) Oppslag på setlar i det elektroniske setelarkivet, som er kopla til dei grovlematiserte artiklane i Metaordboka. Oppslaget (eller overskrifta) på ein setel er nemleg ikkje alltid identisk med oppslagsforma i eNO (eller NOB eller BOB), og som vi skal sjå nedanfor, gjev det i nokre tilfelle høve til å fange opp søk som ikkje blir fanga opp på annan måte.

- 3) Om ein framleis ikkje har fått treff, blir det søkt i all tekst i eNO. Fritekst-søk vil til dømes kunne gje treff i unormaliserte sitat eller i oppslagsformer i eldre leksikografiske kjelder der ordet ligg føre.

## 4. Nokre moglege søkjescenario, problem og løysingar

I det følgjande skal vi gå gjennom nokre tenkte søkjescenario. Det er verdt å understreke at desse scenarioa er baserte på kva vi tenkjer oss at folk kan komme til å vilje søkje etter. Dei representerer med andre ord ikkje reelle søk. Når eNO er lansert for publikum, vil vi analysere korleis folk bruker søkjefunksjonen og om naudsynt justere søkjefunksjonen for å gje fleire relevante treff. Vi ser også for oss at informasjon frå søkjeloggar bør supplerast med andre undersøkingar som kan gje oss ein peikepinn ikkje berre om kva folk søkjer på og kva treff dei får, men også om kva dei faktisk var ute etter å finne då dei søkte.

### 4.1 Søkjescenario 1: «sjangle»

Brukaren tastar inn søkjeordet «sjangle» med føremålet å finne artikkelen for verbet som ein kan skrive slik i infinitiv på nynorsk og bokmål. Søkjeveg 1 i figur 3 fører for det første til treff på oppslagsforma *sjangle* i eNO, som er eit substantiv i hokjønn. Søket gjev også treff på verbet *sjangle* i NOB og BOB, som i Metaordboka er kopla til verbartikkelen **sjangla** i eNO (som nemnt i avsnitt 2 bruker NO *a*-infinitiv av verb, i tråd med 1938-normalen).

The screenshot shows two search results. The first result is for the word "sjangle" as a noun in the masculine gender. Below the word is a horizontal line and the text "innslag av striper i ty (Holla)". The second result is for the verb "sjangla" with the suffix "-a". Below this is a search box containing the text "gå ustøtt, rava (ll,3a), skjaga (1a), vingla; (ôg:) halda på å falla; vakla". Below the search box is a snippet of text: "han sjangla mest sanselaus ut på sleen då han skulde køyre heim att (Mehl.Hall 231)" and "han (fekk) eit drag over nakken so han sjangla burtetter golvet (NFL110d 41)".

Figur 4. Søkjetreff ved søk på «sjangle»: artiklane **sjangle** og **sjangla**.

Søkjerultatet (figur 4) inneheld to artiklar, mellom dei verbartikkelen som brukaren ønskte å finne. Vi ser altså at sjølv om infinitiven i eNO-artikkelen endar på *-a*, kan vi ved hjelp av Metaordboka og koplingane til NOB og BOB lett fange opp søk der *e*-infinitiv blir nytta. Dette gjeld berre i dei tilfella der



ordet det blir søkt på står i NOB og/eller BOB. Om det ikkje gjer det, må vi stole på andre metodar.

## 4.2 Søkjesenario 2: «skjangle»

I dette tilfellet søker brukaren etter verbet *sjangla* med ei ikkje-standardisert skriftform (*skj* i staden for *sj*) og med *e*-infinitiv. Søkjeveg 1 i figur 3 gjev ingen treff, ettersom *skjangle* ikkje er oppslagsord korkje i eNO, NOB eller BOB. Rett nok har eNO ein tilvisingsartikkel **skjangle** som viser til **sjangla**, men *a*-infinitiven i oppslagsordet hindrar søket frå å bli fanga opp på dette stadiet. Søkjeveg 2 fører heller ikkje til treff, men søkjeveg 3 (fritekstsøket) gjev treff i feltet for spesifikasjon av eldre kjelder:

**sjangla** verb -a 

---

Eldre kjelder  
H u *sjake*, A1 633 *skjangle*, A2 *skjangle* og u *skjaga*, R *skjangle*, R3

gå ustøtt, rava (ll,3a), skjaga (1a), vingla; (ög:) halda på å falla; vakla ✖

*han sjangla mest sanselau ut på sleen då han skulde køyre heim att* (Mehl.Hall 231)  
*han (fek) eit drag over nakken so han sjangla burtetter golvet* (NFL11Ød 41)

np ✖

*ein trøyt natranglar kom skjangle etter gata* (Forsb.DNR 71)

Figur 5. Søkjetreff ved søk på «skjangle»: artikkelen **sjangla**.

I feltet «Eldre kjelder» i artikkelen **sjangla** er det ført opp at kjelda A<sup>1</sup> (dvs. Aasen 1850) fører ordet på side 633 med forma *skjangle*. Via denne forma får brukaren opp nett denne artikkelen (og berre denne). Vi ser for øvrig at skriveforma med *skj*- også er attestert i eit av døma i artikkelen. Ein kan sjølv sagt hevde – med rette – at vi i dette tilfellet er heldige, ettersom det var langt frå gjeve at forma *skjangle* skulle dukke opp i artikkelen i det heile. Dette er likevel ein illustrasjon av at det gjeld å nyttiggjere seg informasjon i mange delar av artiklane ved søk i eNO.

## 4.3 Søkjesenario 3: «sjaundeferd»

Her søker brukaren på det samansette ordet *sjaundeferd*. Søkjeordet fell saman med den normaliserte forma av ordet. I og med at det ikkje finst ein artikkel med dette som oppslagsord korkje i eNO, NOB eller BOB, gjev ikkje søkjeveg

1 noko resultat. Derimot er ordet nemnt saman med andre samansetjingar i artikkelen **sjaunde-**, som er ein føreleddsartikkel som gjev opplysningar om i prinsippet alle ord som har dette føreleddet. Dette blir plukka opp av søkjeveg 2, og søkeresultatet må i dette tilfellet bli føreleddsartikkelen **sjaunde-**:

**sjaunde-**



til *sjaund*; såleis (omfram dei som fylgjer ndf på alfab plass) t d  
sjaundeday, sjaundeferd, sjaundeveitsle, sjaundeøl

Figur 6. Søkjetreff ved søk på «sjaundeferd»: føreleddsartikkelen **sjaunde-**.

I tilfelle som dette, der søkjeordet står inni ein artikkel, bør treffet uthevast eller markerast med ein særskild farge, slik at brukaren lett ser nett det ho var ute etter å finne.

#### 4.4 Søkjesenario 4: «sjondarbåt»

I dette scenarioet søker brukaren på «sjondarbåt», som er ei målføreform (attestert frå Hålandsdal i Hordaland) av det lemmaet som i NO er normert **sjaunde-båt**. I mange tilfelle vil det truleg vere naturleg for brukarane å søkje på dialektord i den forma dei er vande med å høyre og bruke, fordi dette ofte er ord som ein til vanleg ikkje ser i skrift. Her gjev søkjeveg 1 ingen resultat, men søkjeveg 2 vil gje treff på oppslaget på den eine setelen frå Hålandsdal der forma *sjondarbåt* er brukt. På papirsetelen er *sjondarbåt* førd som tentativ normalisert form av innskrivaren, og denne forma er førd vidare som oppslag då papirsetelen blei digitalisert, slik vi ser i figur 7, som gjev att eit utsnitt frå Metaordboka der denne eine setelen er markert:

Oppslag	Grammatikk	K-status	Antal
☐ sjaund*e*båt	m	ok	2
☑ sjondarbåt	subst mask ubf sg appell	ok	
☐ sjundar-båt	m	ok	

Figur 7. Metaordboksartikkelen **sjaundebåt**; setelen med oppslaget *sjondarbåt* er markert.

Via Metaordboka kjem ein då fram til artikkelen **sjaundebåt**, som representerer det lemmaet søkjaren var ute etter:

sjaunde-båt substantiv, hankjønn



båt brukt til likferd, sjaund (SHI R, Hålandsdal)

Figur 8. Søkjetreff ved søk på «sjondarbåt»: artikkelen *sjaunde*båt.

#### 4.5 Søkjesenario 5: «sennagress»

Dette scenarioet liknar mykje på det førre. Igjen søker brukaren på ei målføreform, «sennagress», som er attestert frå Vardø i Finnmark. Lemmaet det er snakk om, er i NO normalisert *sine-* eller *senne-gras*. NOB har oppslaget *sennegras*, medan BOB har *senne-gras* eller *-gress*. Som vi ser, fangar ingen av desse oppslaga opp den forma brukaren søker på. Men med søkjeveg 2 blir det igjen treff, her i form av eit oppslag i Ordbokshotellet, som er éin av ressursane som er samla i Metaordboka (sjå avsnitt 3). I figur 9 er det aktuelle oppslaget det fjerde nedanfrå. Det dreier seg om eit oppslag i Støme (1994), ei ordsamling frå Vardø.

Oppslag	Grammatikk	K-stat...	Antal
☐ sine*gras	n	ok	31
☐ senagras	subst nøyt ubf sg appell	ok	
☐ senegras	subst nøyt ubf sg appell	ok	
☐ senegras	subst nøyt ubf sg appell	ok	
☐ sen-gras	subst nøyt ubf sg appell	ok	
☐ sennegras	Ukjent	ok	
☐ sennelgras el. ~gress	Ukjent	ok	
☐ sinegras	n	ok	
☐ sinagras		ok	
☐ sennægras		ok	
☐ <b>sennagress</b>	n.	ok	
☐ senngræs	n.	ok	
☐ sennegras	subst nøyt ubf sg appell	ok	
☐ sinegras	n	ok	

Figur 9. Utdrag av metaordboksartikkelen *sinegras*; oppslaget *sennagress* er markert.

Også i dette tilfellet får altså brukaren treff på den artikkelen ho var ute etter, nemleg denne:

**sine-** eller **senne-gras** substantiv, inkjekjønn

1. **BOTANIKK**, plante(art) i storfamilien; *Carex vesicaria* (VSO; Snåsa, Velfjord, Rødøy, Skånland, Nordreisa; J.LidElvenNF 1020) ✕  
*gjetorde för ut yver verda som eld i sennegräs* (KohTB 13)
2. (myr)gras el stort som blir turka og brukt som isolasjon i kommagar og anna skoty (Forsand, Nesna, Korgen, Fauske, Gratangen) ✕  
*inni kommagan bruka di tørt, mjukt senngræs, og då e di godt beskjetta på føten* (Håloygm 1962,76)
- 3.
- a. eit slag gras som veks tett; (òg:) sine (II,1a) (Sande på Snm, Ulstein, No A2, Beiarn, Vardø)
- b. tettegras (Hetland)
- c. namn på plante(arte)n kornstorr (No R4)

Figur 10. Søkjetreff ved søk på «sennagress»: artikkelen **sine-** eller **sennegras**.**4.6 Søkjesenario 6: «sennagras»**

Til slutt vil vi sjå på eit meir problematisk, men ikkje urealistisk, søkjesenario. Brukaren søker på «sennagras». Dette er ikkje oppslagsform i eNO, NOB eller BOB, og ein finn det heller ikkje brukt som oppslag på setlar eller andre ressur sar kopla til den aktuelle artikkelen i Metaordboka. Heller ikkje eit fritekstsøk hjelper oss i dette tilfellet, og søket gjev dermed ingen treff. Det hjelper lite at forma *senna-* er førd opp som målføreform i føreleddsartikkelen **II sine-**, der nettopp opplysningar om samansetjingar av typen *sennagras* er samla:

**II sine-**

Målføreformar (målførevariasjon annan enn lik oppslagsord eller sjølvgeven)

*sine-*  
 Ulstein

*sene-*  
 No R4

*sena-*  
 Forsand, Sauda

*sen-*  
 No A2, Korgen, Beiarn

*senn-*  
 Fauske, Senja

*senna-*  
 Nesna, Gratangen, Vardø

til II sine

Figur 11. Føreleddsartikkelen **II sine-** med målføreformar.

Det er svært viktig kva informasjon brukaren får i dei tilfella der søket hennar ikkje gjev noko treff. Ho bør få informasjon om normeringsprinsippa som ligg bak forma på oppslagsorda i eNO (vi kan ikkje forvente at alle brukarane av eNO har eit reflektert forhold til tilhøvet mellom målføreformer og normaliserte standardformer). Samstundes bør ho få tips om meir avanserte søkje-strategiar som kanskje kan hjelpe henne å finne rett artikkel. I tilfellet med *sennagras* vil til dømes eit trunkert søk på «senn%øgras» føre fram.

## 8. Konklusjon

NO er ei ordbok der artiklane har ein kompleks og informasjonstett struktur. I tillegg føreset den prenta utgåva at brukarane har ein viss kompetanse i norsk målførevitskap og ein viss kjennskap til normutviklinga i nynorsk (jf. til dømes at ordboka ikkje fører sjølvgevene målføreformer, og at verket er redigert etter 1938-normalen). Formidlinga av dette ordboksmaterialet til eit breitt publikum er ei utfordring, og vi må gjere ein god del tilpassingar av informasjonsstrukturen og presentasjonen av artiklane i eNO. Søkjescenarioa vi har sett opp her, viser at den fullelektroniske ordboksbasen og dei elektroniske koplingane mellom artiklar og underliggjande grunnlagsmateriale likevel gjev oss eit godt utgangspunkt for å tilby brukarane god søkjefunksjonalitet og relevante treff. Redigeringsapplikasjonen er utvikla av informasjonsteknologar ved EDD, som også står for utviklinga av nettversjonen eNO. Fordi redaksjonen og EDD står i eit fortløpande samarbeids- og utviklingsforhold, kjenner både leksikografane og informatikarane form, innhald og søkjefunksjonalitet i dei mange samankopla basane og i ordboksapplikasjonen. Dette gjer utviklingsarbeidet lettare.

eNO blei lansert i ein testversjon med alfabetbolken *i-skodda* 15. mars 2012. Neste steg i arbeidet med eNO vil vere å teste denne førsteversjonen på eit større publikum og bruke responsen vi får attende til vidare arbeid med utsjånad og funksjonalitet i eNO.

## LITTERATUR

- Aasen, Ivar 1850: Ordbog over det norske Folkesprog. Kristiania.  
 BOB = Bokmålsordboka <<http://www.nob-ordbok.uio.no>> (september 2011)  
 eNO = Nettutgåve av Norsk Ordbok. Testversjon: *i-skodda*. <[no2014.uio.no](http://no2014.uio.no)> (april 2012)  
 Tvedt, Lars Jørgen, Elisabeth Lien og Øyvind Eide 2007: Ordbokshotellet – varig lagring og formidling av norske ordsamlingar. I: Torben Arboe (red.) Nordisk dialektologi og sociolingvistik: foredrag på 8. Nordiske Dialektologkonferanse, Århus

- 15.–18. august 2006. S. 379–388. <[www.edd.uio.no/artiklar/leksikografi/Ordbokshotellet.pdf](http://www.edd.uio.no/artiklar/leksikografi/Ordbokshotellet.pdf)> (september 2011)
- NO = Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet. 1966–. Oslo: Det Norske Samlaget.
- NOB = Nynorskordboka <<http://www.nob-ordbok.uio.no>> (september 2011)
- Ore, Christian-Emil Smith 2001: Metaordboken – et elektronisk rammeverk for Norsk Ordbok? I: Martin Gellerstam et al. (red.): Nordiska studier i leksikografi 5. Rapport från Konferens om Lexikografi i Norden, Göteborg 26.–29. maj 1999. S. 202–216.
- Ore, Christian-Emil og Lars Jørgen Tvedt 2006: Digital sats eller digital satsing? I: Henrik Lorentzen og Lars Trap-Jensen (red.): Nordiske studier i leksikografi 8. Rapport fra Conference om Leksikografi i Norden Sønderborg 24.–28. maj 2005. S. 315–322.
- Rømmetveit, Magne (red.) 1979: Norsk landbruksordbok. Band 1–2. Oslo.
- Skard, Matias 1954: Nynorsk ordbok for rettskriving og litteraturlæring. Femte utgåva. Oslo.
- Støme, Oddvar 1994: Vår eigen stemme. Tartu.
- Svardal, Terje 2003: Norsk Ordbok – om normering av Metaordboka. I: Zakaris Svabo Hansen og Anfinnur Johansen (red.): Nordiske Studier i Leksikografi 6. Rapport fra Conference om leksikografi i Norden Tórshavn 21.–25. august 2001. S. 351–356.
- Torp, Alf 1919: Nynorsk etymologisk ordbok. Kristiania.

*Sturla Berg-Olsen*

Norsk Ordbok 2014

[sturla.berg-olsen@iln.uio.no](mailto:sturla.berg-olsen@iln.uio.no)

*Åse Wetås*

Norsk Ordbok 2014

[ase.wetas@iln.uio.no](mailto:ase.wetas@iln.uio.no)

# Breaking away from tradition: Linking a database of inflection to an electronic dictionary

*Kristín Bjarnadóttir*

Icelandic inflection is complex and there is an abundance of variant forms, which are very often changeable according to time and style. The choice of variants within a paradigm in some cases depends on meaning and context. The tradition in Icelandic dictionaries is to give the principal parts of words, as indicators of inflectional classes. These, however, contain insufficient information for the remainder of the paradigm to be predictable. In an online dictionary this problem can be solved by a link to a database of inflection, such as the Database of Modern Icelandic Inflection (DMII). This is the method chosen in ISLEX, the Icelandic-Scandinavian online dictionary, and as a result the information provided far exceeds the tradition in Icelandic dictionaries. Both projects are under construction at The Árni Magnússon Institute for Icelandic Studies.

The breakaway from tradition also takes place in the actual production of the paradigms for the DMII. Icelandic textbooks and handbooks on morphology describe the system as a whole, i.e., they are constructed ‘top-down’, with a limited number of examples for each inflectional class. The production of DMII is, in contrast, ‘bottom-up’, as it is based on research on individual words, not inflectional classes. The aim is to provide enough information for the user to make an educated choice by himself.

## 1. Introduction

The topic of this paper is the link-up of two projects in progress at The Árni Magnússon Institute for Icelandic Studies in Reykjavík, i.e. the web-based Icelandic/Danish, Norwegian and Swedish dictionary ISLEX,<sup>1</sup> and The Database of Modern Icelandic Inflection, BÍN (Beygingarlýsing íslensks nútímamáls).<sup>2</sup>

---

<sup>1</sup> ISLEX. Halldóra Jónsdóttir, Þórdís Úlfarsdóttir (red.) The Árni Magnússon Institute for Icelandic Studies, Reykjavík. <http://islex.hi.is>; [islex.dk](http://islex.dk); [islex.no](http://islex.no); [islex.se](http://islex.se).

<sup>2</sup> BÍN = Beygingarlýsing íslensks nútímamáls. Kristín Bjarnadóttir (red.) The Árni Magnússon Institute for Icelandic Studies, Reykjavík. <http://bin.arnastofnun.is/>.

Instead of the traditional information on inflection in Icelandic dictionaries, each headword in ISLEX is connected to a full paradigm for the word in BÍN by a hyperlink. This provides the user with much more information on inflection than hitherto possible in Icelandic dictionaries. The information on inflection contained in BÍN can also be said to break away from the Icelandic tradition of treatment of the inflectional system by using a ‘bottom up’ approach to describe the system, instead of the ‘top-down’ approach traditionally used in grammatical description. The difference is that of a survey of the system as a whole, as demonstrated in textbooks, in contrast with showing the paradigms for every single word, showing the inflection “as is”, i.e. actual usage, as attempted in BÍN. It must be stressed, however, that work on BÍN is still an ongoing process, and a surprising aspect of the work has in fact been how deficient the existing research material is when it comes to describing the inflectional system of Icelandic in the necessary detail for this approach.

## 2. The two projects, ISLEX and BÍN

ISLEX is an Icelandic/Danish, Norwegian, Swedish online dictionary being prepared at The Árni Magnússon Institute for Icelandic Studies, in cooperation with DSL (Det Danske Sprog- og Litteraturselskab) in Copenhagen, The University of Bergen, and The University of Gothenburg (Halldóra Jónsdóttir & Þórdís Úlfarsdóttir, this issue). ISLEX contains 50,000 entries. The work commenced in 2006, and ISLEX is due to open on November 16th 2011.<sup>3</sup>

Work on The Database of Icelandic Inflection (BÍN) commenced in 2002 at The Institute of Lexicography (Orðabók Háskólans), now the Department of Lexicography of The Árni Magnússon Institute for Icelandic Studies. BÍN was initially created to serve two purposes, i.e. to produce data for use in language technology projects, and to make the results available to the general public online. From the outset, the intention was also to utilize the data in lexicography. Work on the project started in 2002, as a part of an language technology program launched by the Minister of Education, Science and Culture (Rögnvaldsson et al., 2009). From the beginning, the aim was to show “all” and only existing word forms from the modern language in a set of paradigms. At present BÍN contains 270,000 paradigms, 5.8 million inflectional forms.

---

<sup>3</sup> ISLEX was opened as planned. Work on Faroese is now in progress, and work on Finnish is expected to begin in 2012.



### 3. Information on Inflection in Icelandic Dictionaries and Grammars

The tradition in Icelandic dictionaries is to give partial information on inflection, i.e. the principle parts of words. These are not necessarily sufficient to predict all the remaining inflectional forms of the word. The principle parts are the entry forms, followed by the the genitive singular and the nominative plural for nouns, and the past tense and past participle for verbs. (For weak verbs the past tense form is in the singular, for strong verbs the past tense is singular and plural. The past participle is omitted when it is predictable from the past tense. In the case of umlaut, the subjunctive past tense is added.) The inflection of adjectives is only shown in irregular cases. The examples here are from *Íslensk orðabók* (Mörður Árnason, 2007):

<b>sokkur</b> kk -s, -ar	‘sock’ masc., gen.sg. <i>sokks</i> , nom.pl. <i>sokkar</i>
<b>þröskuldur</b> kk -s/-ar, -ar/-ir	‘threshold’ masc., gen.sing. <i>þröskulds/þröskuldar</i> , nom.pl. <i>þröskuldar/þröskuldir</i>
<b>labba</b> s. -aði	‘walk’ v., p.t. <i>labbaði</i>
<b>biðja</b> s. <i>bað</i> , <i>báðum</i> , <i>beðið</i> ; vh.pt. <i>beði</i>	‘pray’ v., p.t.sg. <i>bað</i> , p.t.pl. <i>báðum</i> , p.ptc. <i>beðið</i> ; subj.p.t. <i>beði</i>
<b>góður</b> l. (mst. <i>betri</i> , hst. <i>bestur</i> )	‘good’ adj., comp. <i>betri</i> , sup. <i>bestur</i>
<b>reiður</b> l.	‘angry’ adj. [No information on inflection.]

Table 1. Examples of information on inflection in *Íslensk orðabók* (2007).

Until the arrival of electronic editions, dictionaries have, by reasons of space, been constrained by their format to condense inflectional information to the barest minimum, as evidenced by the examples above. As the principle parts shown are not always sufficient to predict the whole paradigm of words, this sometimes leaves the user stranded, or at best with grammar books as the only solution. As the function of the grammar books is to give surveys of the system, not to give information on individual words, the user can still be stranded, i.e. if he happens to be searching for words that happen not to be among the relatively few examples selected for demonstration in the grammar books. Furthermore, some parts of the system are systematically missing from the literature, as will be demonstrated later.

## 4. The BÍN Paradigms

Morphologically, Icelandic is a rich language, with 16 inflectional forms to a noun (4 cases, singular and plural, +/-definite), 120 inflectional forms to an adjective (3 genders, 4 cases, singular and plural, +/-definite; positive, comparative, superlative), and 107 inflectional forms to a verb, (indicative/subjunctive, present/past, person, number, etc.). Inflectional form here signifies a word form with grammatical tag. These figures are for a full paradigm, excluding variants. These variants are the result of the fact that the inflectional endings that mark grammatical categories are not unique, and the result is a proliferation of dual forms. In such cases, the tradition in Icelandic grammar books is to say that a word can belong to two inflectional classes. In BÍN, however, each headword or lemma is shown in full in one paradigm, i.e. a word belongs to one inflectional class only. Each inflectional class therefore includes all variants, and the result is a proliferation of inflectional classes, or over 600, as of summer 2011. The inflectional classes are only a tool for the production of the paradigms, and the users of BÍN and ISLEX do not have access to the classification. The only visible access is the individual paradigm, represented in the traditional manner, as shown in the paradigm for *sokkur*.<sup>4</sup>

### sokkur

Karlkynsnafnorð

Eintala			Fleirtala		
	án greinis	með greini		án greinis	með greini
Nf.	sokkur	sokkurinn	Nf.	sokkar	sokkarnir
Pf.	sokk	sokkinn	Pf.	sokka	sokkana
Pgf.	sokk	sokknum	Pgf.	sokkum	sokkunum
Ef.	sokks	sokksins	Ef.	sokka	sokkana

<sup>4</sup> The metalanguage for BÍN is as yet only in Icelandic. *Karlkynsnafnorð*: n.masc., *eintala*: sg., *fleirtala*: pl., *án greinis*: indefinite, *með greini*: definite, *Nf.*: nominative, *Pf.*: accusative, *Pgf.*: dative, *Ef.*: genitive.

The link-up between ISLEX and BÍN is done simply by placing hyperlinks in each entry in ISLEX, giving the users of ISLEX the same access to BÍN as other users of the website.<sup>5</sup>

## 5. Variants in BÍN

In BÍN, an attempt is made to order variants according to use, i.e. by placing the most frequent, the most acceptable one, etc. before those of lesser importance. This is by no means easy to do, as there are many criteria to be taken into account. For technical reasons, there is, however, no way of marking the variants inside the paradigm in any manner, such as by font or colour. BÍN was initially created mainly for language technology use, where the database has to be as inclusive as possible, but the ordering of variants is not of prime importance, at least not in the language technology projects contemplated so far. The reason for inclusion is obvious; if a word form is not in the database, it will not be included in the output used for a search engine, etc. Speakers of the language, on the other hand, want information on the use of the variants, not just a list of them. Therefore, the online version of BÍN contains notes on the choice of variants, usage, etc., which are included at the top of the paradigm. These notes are intended to help the user pick the right variant, or at least to keep him from using highly restricted ones, stylistically or otherwise, as seen in an English translation of the note on the noun *rödd* below:

*Note:* Sometimes an obsolete inflectional form, *röddu*, appears in the dative singular in texts:

*Þeir brópuðu hárrí röddu.*  
'They shouted in a loud voice.'

The same word form may very rarely appear in the accusative singular:

*Þeir heyra ekki röddu hans.*  
'They do not hear his voice.'

*Table 2. The note on the use of variants in the noun rödd 'voice'*

The notes in BÍN are gradually being expanded, in part in response to online users demanding help with variants. Although the subject matter of BÍN is Ice-

---

<sup>5</sup> The data from BÍN is also used in the ISLEX search engine, giving users access to entries through the use of any inflectional form, as in *kött*, *ketti*, *kattar*, *kötturinn*, *köttinn*, *kettinum*, *kattarins*, *kettir*, *köttum*, *katta*, *kettirnir*, *kettina*, *köttunum*, *kattanna* → *köttur* 'cat'.

landic inflection in modern Icelandic, it has proved necessary to include notes on obsolete forms as these are very much a part of today's language, in idioms, fixed expressions and quotations, etc.

## 6. The Unpredictable Forms

In the instance of *sokkur*, the traditional dictionary representation of the inflection would be *-s*, *-ar*, as referred to above. This information, however, does not give the necessary information on the dative form, which is not predictable from the principle parts or from the rules found in grammar books. In fact it is not predictable in any manner, as the comparison of phonetically similar words below shows.

Masculine nouns, singular, indefinite				
nom.	<i>sokkur</i>	<i>flokkur</i>	<i>lokkur</i>	<i>kokkur</i>
acc.	<i>sokk</i>	<i>flokk</i>	<i>lokk</i>	<i>kokk</i>
<b>dat.</b>	<b><i>sokk</i></b>	<b><i>flokki</i></b>	<b><i>lokk/lokki</i></b>	<b><i>kokk</i></b>
gen.	<i>sokks</i>	<i>flokks</i>	<i>lokks</i>	<i>kokks</i>
'sock'	'flock, party'	'lock' (of hair)		'chef'

Table 3. *The unpredictable dative of 'sokkur'.*

The definite forms of the dative singular of these words are *sokknum*, *flokknum*, *lokknum*, *kokkinum/kokknum*, showing that the definite forms can not be predicted from the indefinite forms either, complicating the issue even further.

The dative singular of masculine nouns is notoriously unpredictable, and acknowledged in the grammatical literature as such (Friðrik Magnússon, 1984). Even so, people have a tendency to generalize, and to assume that words such as *sokkur* should have *-i* in the dative, as the grammar books suggest *-i* in the dative singular for masculine nouns ending in *-ur* in the nominative singular, apart from a few listed exceptions. This is of course the result of the fact that the grammars are surveys, intended to demonstrate how the system works, not exact manuals on the behaviour of individual words. The problem is that people will regard the rules in the grammar books as absolutes, and stick to them, at least consciously, as when answering questions on usage, even though they may never actually use the inflectional forms they consider or mark as 'correct'. This is the case with the dative of *sokkur* 'sock' cited above. The dative form *sokk* is attested in the sources at the Árni Magnússon Institute for Icelandic Studies, i.e. in the examples in the archives and text collections of 65 million running words, as well as in the available sources from Old Icelandic. The usage seems to

be unequivocal, as in: “Ég fór úr blautum sokk/\*sokki” ‘I took off a wet sock’. Search on the web gives the same result, although one example was found of the dative *sokki*.

To reach conclusions such as the one on the dative of *sokkur*, every possible source is consulted in the course of the production of the paradigms for BÍN, as far as time allows. The main sources are archives and text collections of the Árni Magnússon Institute for Icelandic Studies, and other available text collections, such as those at the University Library. The database is constantly revised.<sup>6</sup>

## 7. New Information on the Inflectional System

According to the literature, the case of the dative singular of neuter nouns should be simpler than that of the corresponding masculine ones, as the dative singular marker *-i* is supposed to be (almost) universal in the neuter. The only exceptions listed in a recent survey on Icelandic morphology are *tré* ‘tree’, *fé* ‘money; livestock’, and *hné/kné* ‘knee’ (Guðrún Kvaran, 2005). The *-i* is in fact the dative singular marker for single syllable and affixed words from the inherited Icelandic vocabulary, as for compounds formed from these. The research for BÍN shows, however, that neuter loanwords, especially multisyllable words, very often do not have an ending in the dative singular. Words of this type are not mentioned at all in most grammars, although some of them are by no means new in the language, such as the words *fenikel*, *kanel*, *vítríól*, *stúdíum*, *examen*, *flaskó* and *ópíum* which are attested in the archives of the Árni Magnússon Institute for Icelandic Studies in citations from the turn of the 18th century.<sup>7</sup>

The neuter loanwords cannot be left out of BÍN, of course, although they find no place in the grammars. BÍN would be of less use in language technology if certain inflectional classes were left out, even if they exhibit traits not wholly in compliance with the historical facts on the inflectional system. Bisyllabic or multisyllabic single morpheme stems are fairly alien to the language, but their existence is a fact in the loanwords. Some of the 18th century loanwords mentioned above do in fact have Icelandic equivalents that have replaced them in the modern language. A case in point is *examen*, which has been superseded by *próf*, whereas the word *ópíum* is in full use, without a successful attempt at sub-

<sup>6</sup> Users of the online version of BÍN contribute generously with comments on revision, additions and corrections. There were 398,768 visits in the period May 2010 to May 2011, 326,559 of those from Iceland.

<sup>7</sup> Ritmálssafn Orðabókar Háskólans:  
[http://arnastofnun.is/page/arnastofnun\\_gagnasafn\\_ritmal](http://arnastofnun.is/page/arnastofnun_gagnasafn_ritmal).

stitution by a native word. The number of these nouns in use is large enough not to make it possible to ignore them altogether and therefore they have to be a part of BÍN. That, however, makes it necessary to engage in research needed to find out how they behave in context.

The result is not as neat as that of the grammatical surveys, and here is where the difference between those and BÍN comes into focus. The surveys are ‘top-down’, i.e. they describe the system as a whole, choosing examples to demonstrate inflectional classes and exceptions from those. In BÍN, each paradigm stands on its own, and the inflectional classes are no more than a convenient way of producing the paradigms. When a new variant necessitates a new inflectional class, one is simply added. The system is truly ‘bottom up’. Instead of a simple rule in the textbooks, stating that the dative ending of neuter nouns is *-i*, with the exceptions *tré*, *fé*, *hné/kné*, the facts on the neuter loanwords in BÍN are as follows:

The dative of neuter loanwords in BÍN: *-i* or *-0*?

- Multisyllable words can have *-0* or *-i*, sometimes depending on stem structure: *ópíum -0*, *statif -i*
- Very many loanwords have variants: *bíó -0/-i*, *glúten -i/-0*.
- Some subgroups can never have the ending *-i*. This is true of the names of countries: *Íran*, *Ísrael*, *Mexíkó*:

*Ég var að koma frá Mexíkól\** *Mexíkói* ‘I just came from Mexico’

In contrast, names of countries that are a part of the inherited vocabulary, and compound names with a native last part do have an *-i* in the dative:

*Ég var að koma frá Noregi* ‘I just came from Norway’

*Ég var að koma frá Finnlandi* ‘I just came from Finland’

The case of the neuter loanwords is used here for demonstration, as it is an instance of a systematic gap in the treatment of Icelandic inflection in the literature. More often than not, the lack is simply that of omission of information on individual words, if indeed it can be called omission, as the intention in the textbooks was probably never that of complete coverage. The problem is that the nature of a survey, which is at the core of the grammatical descriptions, is forgotten, and the rules intended for description are taken as absolutes.

The difference between the approaches to inflection between the textbooks and BÍN is as follows:

- The generalized rules in grammars and textbooks do not give enough information for individual paradigms. In BÍN, every single inflectional form has to appear.

- The scope of the textbooks is very narrow; they share most of the few examples they show. The vocabulary in BÍN is extensive, running into hundreds of thousands of headwords.
- There is a strong historical bias in the textbooks, giving obsolete and archaic forms their space, to the exclusion of newer words or variants, especially ones found to be unacceptable to purists. The aim in BÍN is to reflect actual use in the modern language, including loanwords and slang, irrespective of acceptability. Obsolete and archaic forms are only included if they commonly appear in Modern Icelandic.

As the data and research on individual words is surprisingly scarce in the literature, the amount of research required for BÍN is considerable. In an online project this poses less of a problem than previously, although the users may have to get used to the fact that the description of the inflectional system is not something to be determined once and for all.

## 8. Conclusion

The work on linking ISLEX and BÍN is still at an experimental stage, but the hyperlinks are in place. ISLEX is expected to serve both native speakers of Icelandic and others needing an Icelandic-Scandinavian dictionary, whereas the twofold purpose of BÍN was centered primarily on language technology, and then, secondarily, on native speakers of Icelandic. BÍN is evolving into a multi-purpose database, and as such it is difficult to serve all needs at once, the LT groups, native users, researchers, and students, at home and abroad. There is only one version of the database; it has to be as inclusive as it can be.

For BÍN to be as useful as possible to the users of ISLEX, the user interface needs to be translated to the languages used there. The website should also contain a manual and a thorough introduction to the methods used in the production of BÍN. These are minimum requirements, and they should not be too difficult to reach. The fact that non-native speakers might be better off with less information on variants, leaving out obscure or very specific instances, is harder to handle. In the original version of BÍN, variants were not linearly ordered. Now, an attempt has been made to order variants according to 'acceptability', and the notes accompanying the paradigms are intended to help the users cope. These notes, however, are probably of limited use to learners of Icelandic, as they assume native or near-native competence.

The linking of ISLEX and BÍN is therefore a break with traditions in Icelandic lexicography and inflectional description in two ways. Firstly, it gives the

users of ISLEX access to all the known facts on the inflection of individual words, perhaps sometimes to a point of superfluity, as in the instance of proliferating and/or obscure variants. Secondly, the attempt at describing the inflectional system ‘bottom up’ should lead to the discovery of the actual facts of language use, as in the case of the elusive dative ending in the multisyllable loanwords discussed above. That would at least give an alternative to the rule systems of the textbooks, which is sometimes based on partial data. As access to available data grows, with better tools of analysis, access to larger text collections and tagged corpora, etc., the picture should become clearer. BÍN is only the first step.

## REFERENCES

- Árnason, Mördur (red.), 2007: Íslensk orðabók [Dictionary of Icelandic], 4. útg., Edda útgáfa hf., Reykjavík.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir, 2012: ISLEX – en flersproget nordisk ord-bog. I: Nordiska studier i lexikografi 11. Lund. (Skrifter utgivna av Nordiska föreningen för lexikografi 12.)
- Kvaran, Guðrún, 2005: Orð. Handbók um beygingar- og orðmyndunarfræði. Íslensk tunga, II. bindi. Almenna bókafélagið, Reykjavík.
- Magnússon, Friðrik, 1984: Ein lítil beygingarending. Mímir, Blað stúdenta í íslenskum fræðum 32:33–43.
- Rögnvaldsson, E., H. Loftsson, K. Bjarnadóttir, S. Helgadóttir, A. B. Nikulásdóttir, M. Whelpton and A. K. Ingason, 2009: Icelandic Language Resources and Technology: Status and Prospects. Rickard Domeij, Kimmo Koskeniemi, Steven Krauwer, Bente Maegaard, Eiríkur Rögnvaldsson and Koenraad de Smedt (eds.): Proceedings of the NODALIDA 2009 workshop Nordic Perspectives on the CLARIN Infrastructure of Language Resources. S. 27–32. Northern European Association for Language Technology (NEALT), Tartu University Library.
- BÍN = Beygingarlýsing íslensks nútímamáls. Kristín Bjarnadóttir (red). The Árni Magnússon Institute for Icelandic Studies 2002-, Reykjavík. <http://bin.arnastofnun.is/>
- ISLEX. Halldóra Jónsdóttir, Þórdís Úlfarsdóttir (red.) The Árni Magnússon Institute for Icelandic Studies, Reykjavík. <http://islex.hi.is>; [islex.dk](http://islex.dk); [islex.no](http://islex.no); [islex.se](http://islex.se).
- Ritmálssafn Orðabókar Háskólans (The Written Language Archive): The Árni Magnússon Institute for Icelandic Studies, Reykjavík. [http://arnastofnun.is/page/arnastofnun\\_gagnasafn\\_ritmal](http://arnastofnun.is/page/arnastofnun_gagnasafn_ritmal)

*Kristín Bjarnadóttir*

The Árni Magnússon Institute for Icelandic Studies.

[kristinb@hi.is](mailto:kristinb@hi.is)



# Nye kilder til dansk ordbogshistorie

*Simon Skovgaard Boeck*

This paper gives a very brief survey of the history of Danish lexicography in the 15th and 16th centuries, focusing on thematic school dictionaries, many of which are only known through book catalogues and the like. The books themselves have been lost, torn by generations of school children. The paper focuses on a Danish version of the influential *Nomenclator omnium rerum propria nomina variis lingvis explicata indicans* by the Dutch scholar Hadrianus Junius. The Danish version (*Epitome*) was compiled by Hans Poulsen Resen – whose linguistic and pedagogical interests are known from other sources. Unlike Junius's learned work, *Epitome* was intended for school children, and the number and quality of information in the original is therefore dramatically reduced to Latin and Danish equivalents. Also, the possible relationship between *Epitome* and some later school dictionaries is examined.

*Nøgleord:* ordbogshistorie, tematiske ordbøger, skoleordbøger, Hadrianus Junius, Hans Poulsen Resen, Daniel Matras.

## Baggrund

Der er som bekendt overleveret seks danske ordbøger fra 1500-tallet (jf. Andersson m.fl. 1997):

Christiern Pedersen: *Vocabularium ad usum dacorum*, 1510, nye udgaver 1514 og 1518

Henrik Smith: *Hortulus synonymorum*, 1520

Jon Tursten: *Vocabularius rerum*, 1561, ny udgave 1579

Henrik Smith: *Libellus vocum Latinarum*, 1563

P[oul] N[ielsen] H[ingelberg]: *Vocabulorum variorum expositio*, 1576

Mads Pors: *De nomenclaturis Romanis*, 1594.

Disse i alt ni udgaver har imidlertid ikke kunnet mætte markedet, og det vides at sprogmanden Jakob Madsen Aarhus (1538-1586) har arbejdet på endnu en ordbog. Mere end et manuskript blev det ikke til, og dette brændte ved Københavns brand i 1728, således at bogen nu kun kendes gennem en notits i et katalog over Anders Lemvigs bibliotek. Lemvig var Aarhus' svoger, og han overtog

dennes samlinger efter hans død. Manuskriptet beskrives som "Dictionarij Latino-danici D. Jacobi Matthiæ" og det specificeres bestående af syv alfabetisk ordnede dele, hvoraf ganske vist første del "A, B, C" manglede på katalogiseringsstidspunktet (Katalog i Rigsarkivet, dat. 1603, s. 121r; jf. Birket Smith 1882: 103).

Peter Skautrup beklager tabet af Aarhus' ordbog: "Thi i denne måtte vi – forudsat at værket ikke har været en mindre original, ny udgave af et af de da foreliggende vocabularier – i al fald have kunnet hente rigere sprogstof, end det vi nu kan samle sammen i De literis [Madsen Aarhus' sprogvidenskabelige hovedværk]" (Skautrup 1931: 69). Men Skautrups forbehold er nok overflødig. Som sagt har Aarhus' ordbog tydeligvis haft alfabetisk makrostruktur; og den eneste latin-danske forgænger med denne struktur af Christiern Pedersens pionerværk. Denne ordbog fra begyndelsen af århundredet kan vist ikke siges at have mistet sin værdi fuldstændigt på Aarhus' tid; Calepinus' berømte *Dictionarium* som Pedersens er en dansk version af, oplevede genoptryk langt op i tiden (ikke kun i de reviderede, polyglotte udgaver værket er mest kendt som, også i den oprindelige skabning som monolingval latinordbog, Labarre 1975). Desuden ligger Calepinus til grund for Poul Jensen Coldings store *Etymologicum Latinum* 1622. Men Pedersens bearbejdning lader alligevel en del tilbage at ønske, fx har han flere steder misforstået Calepinus, og hans danske modsvarigheder er af og til oversættelser af de latinske forklaringer, ikke egentlige danske ækvivalenter (jf. Haastrup 1988). Det er nok mest oplagt at Aarhus' håndskrevne lister er et originalt arbejde – i det omfang leksikografiske arbejder, der jo altid har indebåret en udnyttelse af forgængere, kan betegnes som originale. Hvilke ordbøger han har udnyttet og hvordan han har bearbejdet dem, kan man kun spekulere om, snarest har han trukket på kontinentale værker. Men i det omtalte bibliotekskatalog er også optaget to af de ovenfor nævnte danske ordbøger, vel at mærke ikke Pedersens, men Smiths fra 1563 og Hingelbergs fra 1576. Det er muligt at allerede Aarhus har ejet disse bøger, men det kan ikke vides; i kataloget er mellem dem optaget to bøger udkommet efter hans død. Det drejer sig om to udgaver af samme ordbog:

Brevis Nomenclator Hadriani Junij, Haf. 1596

Brevis Nomenclator Had. Junij. Haf. 1601

(RA-katalog, s. 93v).

Med katalogets "Brevis" betegnes at der er tale om uddrag, nemlig af Hadrianus Junius' *Nomenclator omnium rerum propria nomina variis lingvis explicata indicans* – en af de mest indflydelsesrige ordbøger i 1500-tallet. Ingen af de to udgaver kendes i dag, de er sandsynligvis blevet slidt op i skolerne.

I katalogets første post er noget slettet der *kan* være ”Resenij”, og det passer med hvad vi ellers ved om bogen, nemlig at Hans Poulsen Resen (1561-1638) netop i 1596 (3. maj) fik kongeligt privilegium for seks år på udgivelsen af *Brevis nomenclator Hadriani Junii cum utili appendice pro pueris* (KancBrev). Udga- ven af Hadrianus Junius’ ordbog skal dels ses i lyset af Resens pædagogiske interesser, dels i lyset af de sprogvitenskabelige interesser han har næret; et håndskrevet ”Lexicon etymologicum Lingvæ Danicæ” er gået tabt, og han har stillet materiale til rådighed for Poul Jensen Coldings ordbogsarbejde (jf. ndf.).

Helt tabt er Resens bearbejdning ikke. I hvert fald et eksemplar (men måske heller ikke flere) af en udgave fra 1624 er kendt og beror i Det Kongelige Bibliotek. Dens titelblad lyder:

Epitome Brevis Nomenclatoris Hadr. Junii, &c. Cum utili Appendicula pro Puerulis. F A B. Minora facilius minores percipiunt. En tibi, parve puer, breviter, repetenda subinde, Nomina cum verbis, particulasque damus. M. DC. XXIV. Cum Gratia & Privilegio. HAFNIE, Apud Salomonem Sartorium.

## Hvordan forholder Epitome sig til Hadrianus Junius’ værk?

Junius’ ordbog oplevede en meget stor udbredelse med nye oplag og reviderede versioner på flere sprog (Se den udførlige oversigt i Claes’ bibliografi (Claes 1977) og den argumenterede fremstilling hos Müller 2001 kap. IV). Det er derfor langt fra sikkert at det er førsteudgaven, Resen har lagt til grund, selvom eksemplarer af denne kendes i Danmark, fx fra Lemvigs katalog. Men uanset hvilken konkret bog Resen har haft for sig, er der ingen tvivl om at Junius’ ordbog er det grundlæggende forlæg. Der er derfor god grund til at opholde sig lidt ved denne.

Hadrianus Junius selv havde berøring med Danmark, uden det dog har influeret på den danske bearbejdning. I midten af 1560’erne var han kortvarigt i København som medicinsk professor og livlæge for Frederik II (Rørdam 1869-1872: 567ff.). Junius befandt sig ikke godt i København og rejste i 1565 tilbage til Haarlem for at blive stadslæge og latinskolerektor. Kort efter – i 1567 – udkom altså hans *Nomenclator*, og han har muligvis arbejdet på dette værk i København. Junius døde i 1575.

Bogen var ikke hans første leksikografiske bedrift: allerede i 1548 havde han i Basel udsendt et *Lexicon Graeco-latinum* (ny udgave: Basel 1557). Og disse to hellige sprog indgår også i *Nomenclator*, latin som kildesprog. Men de suppleres af ækvivalenter på tysk, nederlandsk, fransk, italiensk, spansk og engelsk, dog således at ikke alle sprog er repræsenteret i samtlige artikler; sjældnest er engelsk.

Rækkefølgen af sprog, og omfanget af deres brug afspejler nok i nogen grad magt- og handelsforhold i samtiden. Nederlandske (hos Hadrianus angivet ved siglen *B.* 'belgisk') synes at være anvendt som målsprog i alle artikler. I denne polyglotte forsamling lyser det tredje hellige sprog, hebraisk, ved sit fravær. Men det hænger sikkert sammen med interessen for de levende sprog der også kommer også til udtryk i en anekdote der fortælles om Junius:

Johannes Sambucus (...) ondernam eene reis naar Holland, alleen met doel om *Junius* te zien en met hem in kennis te geraken; toen hij deswege te Haarlem gekomen zich aan het huis van Junius vervoegde, en op zijne vraag naar dezen ten antwoord ontving, dat hij niet te huis was, maar zich op eene plaats bevond, waar de handwerkslieden des avonds plegen bij een te komen, werd zijne verontwaardiging daarover zoo zeer opgewekt, dat hij besloot den volgenden dag, zonder *Junius* zelfs gezien te hebben, de terugreis weder aan te nemen. En waarom bezocht deze nu die bijeenkomsten der werklieden? Omdat hij geene moeite spaarde voor de zamenstelling van zijne *Nomenclator*, zoo dat hij zich niet schaamde met kunstenaars en handwerkslieden een gesprek aan te knoopen, ten einde de ware namen van zaken en werktuigen, in hun bedrijf gebruikelijk, te leeren kennen (van der Aa 1860: 241).

For en moderne bevidsthed er historien måske ikke videre morsom, men når den ses i forhold til renæssancens stadige orientering mod og seen op til antikken i alt fra lærde til tekniske forhold, er den tankevækkende. Junius udførte – hvis anekdoten taler sandt – feltarbejde blandt meddelere, mens hans samtidige fandt disse i gamle skrifter. Jakob Madsen Aarhus giver udtryk for samme indstilling til fagterminologien i en retorisk lærebog fra 1589 (Skautrup 1931: 97).

Efter førsteudgaven 1567 fulgte meget hurtigt bearbejdnings på en lang række sprog. I sin oversigt over tyske ordbøger i 1500-tallet har Peter O. Müller kaldt *Nomenclator* "das einflußreichste Werk innerhalb der Sachgruppenlexikographie des 16. Jahrhunderts" (Müller 2001: 349), og de tematiske ordbøger (sagordbøger) havde netop en blomstringstid i 1500-tallet. Den danske bearbejdning af værket understreger denne udbredelse. Men den viser også at udbredelsen ikke kun har haft form af ajourførte og supplerende optryk, men også af beskæring og tilpasning; bogen er blevet omplantet fra et lærd europæisk miljø til danske skolestuer. Af den mere berømte nederlandske filolog Justus Lipsius er Hadrianus Junius ellers blevet kaldt den lærdeste nederlænder siden Erasmus (Brugmans 1927: 693), og hans ordbog lever op til denne betegnelse – som allerede de mange sprog antyder. Der er tale om et stort værk, delt i to bøger, i alt omfattende 85 kapitler. Ordbogsdelen fylder 570 sider i den lille tykke bog, og hertil kommer mange og udførlige udenværker bl.a. en oversigt over citerede forfattere, en over diskuterede steder hos en række af disse, en kapiteloversigt og en liste med forkortelser og et register (66 sider á tre spalter).

Epitome er i 1624-udgaven en lille bog på 96 upaginerede sider; der er meget få udenværker, men man bemærker et dansksproget kapitelregister bagerst. I 1500-tallets tematiske ordbøger gives også gerne en kapiteloversigt, men der udelukkende på latin. Det bemærkes også at Junius' omfattende latinske ordregister – en læserservice i meget store tematiske ordbøger – ikke findes hos Resen.

Selve ordbogsdelen er inddelt i 44 kapitler. Allerede fra første kapitel viser Epitome sin forbindelse til Hadrianus Junius: Det første kapitel er ”Om Bøger og det som brugis til Bøger”. Mens hovedparten af de tematiske ordbøger indledes med Gud, er det et nybrud ved Junius' *Nomenclator* at den tager udgangspunkt i bøgernes verden, som grundlag for elevernes læring. Junius' struktur udkonkurrerer aldrig *De Deo*-traditionen, men Junius' enorme indflydelse sætter sig også spor i andre ordbøger; i Danmark bl.a. i Daniel Matras' ordbøger (jf. ndf.). Hverken i Resens eller Matras' versioner overtages Junius' pædagogisk-filosofiske argumentation for strukturen.

Det er altså en omfattende reduktion Resen har foretaget, og den har ikke blot ramt antallet af sider og kapitler, de enkelte oplysninger i hver artikel er også blevet begrænset til et minimum: latinske og danske opslagsord. Den første artikel kan illustrere forskellen:

**L**IBER dictus à libro, qui cortex est arborum interior tenuiorque, cui olim animi concepta inuare solebant. **C**ODEX ab arboris trunco, unde efformabantur tabulæ ceratæ. Codicē plurium tabularum coniunctionē Nonius exponit, βίβλος, λίσος, γράμμα Plutar. AL. Buch. BEL. Boeck. GAL. Livre. ITAL. LIBRO. ANGL. à boec.

Liber Boog

Det er åbenlyst at Resen ikke har fundet Junius' mange andre oplysningstyper (etymologi, ækvivalens på andre sprog og klassikerhenvisninger) relevante for en skolebog, og heri har han nok valgt ret.

Artikelrækkefølgen er naturligvis påvirket af de mange udeladelser og af at Resen har grupperet synonymer. Men selv med forbehold herfor, må det konkluderes at Epitomes rækkefølge ikke svarer til *Nomenclators*. I det første kapi-

tel er de enige om de første fire artikler: *Liber*, *Libellus* (hertil har Nomenclator en række forbindelser der er optaget som artikler), *Volumen* og *Volumen papyraceum*. Men efter dette ord – der betegner mængdeenheden bog (et antal ark) – har Epitome ikke ulogisk ”*Plagula Arck Papiir*”, et ord der ikke optræder i Nomenclators bog-kapitel. Noget lignende gør sig gældende med ”*Epitome Extract*” som den danske ordbog dårligt kunne undvære. Man må nok konkludere at den overordnede plan med kapitler er fælles for de to bøger, men at den konkrete udfyldning afviger i detaljen. Samme resultat giver en stikprøvesammenligning af kapitlerne om bonderedskaber, våben og metaller og sten. Det er muligt at uoverensstemmelser med hensyn til rækkefølge og artikelforråd afspejler at Resen har benyttet et senere oplag end førsteudgaven (som min kollokation bygger på), men det er lige så sandsynligt at de – der er få og små – er fremkommet i forbindelse med hans arbejde.

At Resen har udført en egentlig redaktion, bliver også klart når man betragter et appendiks på otte kapitler der ikke findes hos Junius. Tillægget rummer bl.a kapitler med sted-, personnavne, adjektiver og verber. I modsætning til hoveddelen er de fleste af disse kapitler alfabetisk ordnede, mens afsnittet om talord har sin egen naturlige orden, og det korte om pronominer halvvejs grammatisk orden.

Både på mikro- og makroniveau har Resen altså bearbejdet Junius' tekst. Og det er et tilbagestående spørgsmål hvorfra det supplerede materiale (og navnlig det danske ordstof) kom.

## Epitomes indflydelse

### Latinskolerne

Vi ved at eleverne i Aalborg skole i 1587 måtte afskrive Hadrianus Junius' ordbog efter diktat – fordi skolen ikke havde råd til at give dem et eksemplar hver. Omvendt ved vi at i hvert fald Vordingborg skole i 1617 havde økonomi til det (Jensen 1982: 92). Sandsynligvis er der i begge tilfælde tale om uddrag af Junius' ordbog, men det er muligt at det netop er Resens initiativ der har muliggjort at Vordingborgskolen har haft råd; Aalborgskolen kan måske have været henvist til dyrere udenlandske udgaver.

### Ordbogstraditionen

Den mest lærde af de bevarede danske ordbøger fra 1500-tallet, Mads Pors' fra 1594, udmærker sig ved en interesse for modersmålet der særligt kommer til

udtryk i et dansk ordregister. Der er meget store forskelle mellem Epitome og Pors' værk, men på netop dette punkt er der berøringspunkter. I Pors' ordregister og Epitomes dansksprogede kapiteloversigt ligger kimen til tematiske ordbøger der kan anvendes med mere end ét kildesprog (jf. Hannesdóttir 1998: 56). Epitome indeholder heller ingen eksempler på at det andet af to latinske synonymmer blot gengives med *item* på dansk – et ord der jo er meningsløst som dansk kildesprogsindgang, men meget hyppigt i de ældre ordbøger. Det er altså muligt at Resens bearbejdning er blevet brugt med læseretningen højre-til-venstre som dansk-latinsk ordbog selvom typografien ikke har understøttet en sådan brug.

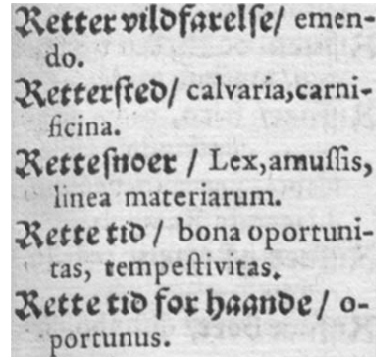
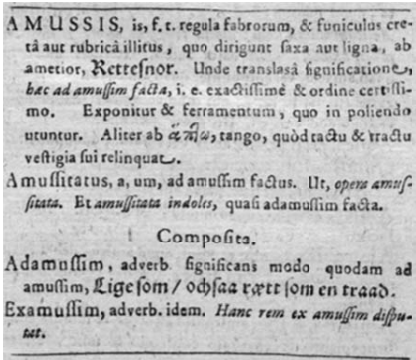
Den 1. juli 1619 skrev Christian IV til Fyns biskop, rektoren og konrektorren i Odense om grundlæggelsen af nye akademier og stiftsskoler og en deraf opstået mangel på skolebøger. De tre blev – i samarbejde med bl.a professor Jens Dinesen Jersin – pålagt at udarbejde en lang række skolebøger (KancBrev). Det umiddelbare resultat heraf blev Jersins latinske grammatik (se om denne og Resens reaktion Jensen 1982).<sup>1</sup> Det oplagt at sætte to privilegier fra 26. februar 1620 i forbindelse med dette pædagogiske tiltag; denne dag fik nemlig ikke blot rektoren i Haderslev Johannes Egnotus et 10-årigt privilegium på et *Dictionarium Latino-Danicum* der synes tabt, Poul Jensen Colding fik også 15 år på det der to år efter udkom med titlen *Etymologicum Latinum*. I det hele taget kan man konstatere en opblomstring af ordbogsudgivelser i 1620'erne, målt i antal, kvalitet og diversitet.

Hovedparten af 1500-tallets ordbøger er billigt trykte sager til brug i skolerne, med den yngste, Mads Pors', som undtagelsen der bekræfter reglen. Pors' bog er et værk på højt akademisk niveau og med internationale præentioner; bogen blev trykt i Frankfurt og har latinske, danske og græske modsvarigheder og klassikerhenvisninger i hver artikel. Denne lærde linje videreføres siden af Poul Jensen Colding hvis *Etymologicum Latinum* (Rostock 1622) er den indtil da største ordbog på dansk grund. Linjen løber videre gennem 1600-tallet med flere meget lærde ordbøger der inddrager stadigt flere sprog. Men ved siden af disse ordbøger har der stadig været behov for simple ordbøger til den elementære latinundervisning. Colding selv leverede en vending af sit værk, og gav resultatet titlen *Dictionarium Herlovianum* (Kbh. 1626). Allerede af titlerne

---

<sup>1</sup> Jens Dinesen Jersins lille latinske grammatik, *Epitome Grammaticæ Latinæ* udmærker sig som den første latingrammatik på dansk, og ikke blot forklares de latinske former på dansk, modersmålet inddrages også illustrativt: "Nomen kalder du saadant et Ord/ som du siger een eller eet for i Dansken *ut Deus* En Gud/ *Panis* et Brød/ *Bonus* En Goed" (s. A1r).

fremgår den store forskel mellem de to (*Herlovanum* sigter til skolen Herlufsholm hvor Colding var forstander 1622-1631). Begge var nok henvendt til skolerens latinundervisning. Men den første er et lærd, europæisk orienteret værk der grupperer sammensætninger og afledninger under et grundord; *Dictionarium Herlovanum* er en hjælp til begyndere. Et eksempel viser forskellen:



Som nævnt er der en direkte forbindelse mellem Resen og Colding. Det fremgår af fortalen til 1622-ordbogen hvor Colding takker Resen for lån af materiale:

Men en anden ting der i høj grad hjalp, var et initiativ fra den altid årvågne biskop over Sjællands stift hr. dr. Hans Resen, der, som den støtte han er for skolerne, og som den store elsker og dyrker af modersmålet han er, havde påbegyndt og næsten færdiggjort tre temmelig store bind af denne slags. Dem gav han mig mulighed for at bruge så jeg når der var mulighed for det, kunne supplere mit eget materiale fra dem, hvilket jeg åbent erkender at jeg har gjort, og jeg skal her offentligt udtrykke min gæld til ham. Hvis jeg med mit værk gør mig fortjent til fædrelandets tak, er det mit ønske og min agt at den skal deles med ham (Colding 1622: (:):2r; Peter Zeebergs oversættelse, renæssancesprog.dk).

Ikke langt fra Coldings Herlufsholm, på ridderakademiet i Sorø, tiltrådte i 1624 franskmænd Daniel Matras (som den første) stillingen som professor i fransk og italiensk.<sup>2</sup> Til den ende har han udgivet en række bøger som hjælpemidler for den studerende ungdom. Kendtest er måske hans ordsprogssamling, men også grammatikker over fransk og italiensk og tre ordbøger er overleveret fra

<sup>2</sup> Daniel Matras og hans forfatterskab er behandlet af Iver Kjær i efterskriften til udgaven af Matras' ordsprogssamling, Kjær 1981. Biografi og bibliografi findes desuden i Matras 1950.



hans hånd. Disse tre er: en lille fransk-dansk (1628), en lille fransk-latinsk (1635), og en stor fransk-tysk-italiensk-dansk ordbog (1643). Også disse ordbøger er blevet hårdt slidt, og i hvert fald det sidste årstal gælder ikke en førsteudgave. Allerede før 1625 synes Matras at have udgivet en latinsk-italiensk-fransk-tysk ordbog, en sådan henviser han til i en fransk-italiensk grammatik fra dette år, og øjensynligt ligger den til grund for 1643-ordbogen (jf. Kjær 1981: 283).

De to ordbøger med dansk målsprog er (som ordsprogsamlingen) tematisk ordnede,<sup>3</sup> men det er navnlig den lille fra 1628 der påkalder sig opmærksomhed i forbindelse med Hadrianus Junius. Den store indskriver sig i *De Deo*-traditionen og har først et kapitel om Gud, den lille indledes derimod af et kapitel *De Livres, & de leurs appartenances*. Allerede heri mærker man Junius' indflydelse, og når man hertil bemærker at den lille bog er trykt hos Salomon Sartor, er det oplagt at forvente at Resens oversættelse har dannet baggrund for Matras' danske modsvarigheder. I fortalen til ordbogen skriver Matras at han har haft brug af en ung vens hjælp til disse modsvarigheder, selv har franskmanden på dette tidspunkt ikke lært nok dansk til at kunne (jf. Kjær 1981: 283). Det er vel ikke urimeligt at forestille sig at denne ven har skelet til *Epitome*.

Ordbogens makrostruktur stemmer ganske godt overens med *Epitomes*; der er enkelte forskelle der som regel består i at *Epitome* behandler flere emner under ét kapitel, mens Matras fordeler materialet på flere. Denne overensstemmelse rækker til *Epitomes* appendiks med bl.a. funktionsord og *proprier* (fra kap. 37). Appendikset er alfabetiseret efter latinen og det er muligvis dette forhold der har fået Matras til at opgive rækkefølgen for at behandle disse ordgrupper efter sit eget system (og alfabetiseret efter de franske opslagsord).

Matras medtager et langt større ordforråd end *Epitome*, og selvom det er fristende at se Sartor som bindeleddet mellem de to bøger, er det muligt at Matras har ekstraheret sine franske opslagsord fra en uforkortet udgave af *Nomenclator*. I betragtning af Matras' ufuldstændige danskundersøgelser kan kildeforholdet til hans ordbøger fortjene en grundigere undersøgelse end her kan gives. Foreløbig kan et uddrag af kapitlet om mennesket og dets legeme illustrere hans plusstof (fra *Nomenclator* medtages kun de latinske opslagsord):

<sup>3</sup> Heller ikke den fransk-latinske ordbog har alfabetisk makrostruktur. Bogens franske titel lyder in extenso: *Dictionnaire François-Latin, Distribué en xv. Observations, Esquelles sont representez tous les Mots François qui descendent de la langue Latine*. Den første observation lyder: "A, finale in Dictionibus Latinis, vertitur in E, in Gallicis" (s. 1) og hertil føjes en alfabetisk og meget nøgen ordliste med franske og latinske versioner af samme ord.

Nomenclator 1567	Epitome 1624	Matras 1628
<i>Cor</i> <i>Præcordia</i>	<i>Cor</i> Hierte	<i>Le coeur, ou cueur.</i> Hierte. <i>Le vous remercie de bon coeur.</i> Jeg tacker eder storlig/ hiertelig. <i>Apprendre par coeur.</i> At lære vden ad. <i>*Il le dit de bouche, Mais le coeur n'y touche.</i> Hand siger det vel med Munden/ men hand tencker dog andet i hiertet. <i>Au fin coeur d'hyver.</i> Mit i Vinteren. <i>Au coeur de la France.</i> Mit i Franckerig. <i>Le choeur d'une Eglise.</i> Koor vdi Kircken. <i>Enfants de choeur.</i> Capellebørn/ Børn vdi Sangeriet.
<i>Pulmo, pulmones</i> <i>Septum transuersum</i>	<i>Pulmo</i> Lunge  <i>Anhelitus</i> Aande	<i>Vn poulmon.</i> En Lunge  <i>L'haleine.</i> Aande. <i>Vn homme qui a courte haliene.</i> En som haf- fuer en stacket Aande. <i>L'haleine luy pût.</i> Hans Aande lufter ilde.
<i>Mamma, mamilla, vber</i> <i>Sinus mammarum, vber</i>	<i>Mamma</i> Patte	<i>Vne mammelle.</i> Patte/ Spenne.  <i>Les tetons d'une fille.</i> En Jomfrues Patter.
<i>Papilla</i>	<i>Papilla</i> Pattevorte	<i>Le bout de la mamele, ou le tetin.</i> Patte vorte.
<i>Iris</i>		
<i>Ventriculus</i> <i>Ventriculi os</i> <i>Ventriculi fundus</i>	<i>Ventriculus</i> Mave	<i>L'Estomach, &amp; le petit ventre.</i> Maffue. <i>Vn estomach d'austruche qui digere le fer.</i> En haard oc sterck Maffue/ som kand fortære Jern som en Struß. <i>Vn bon estomach qui digere bien les viandes.</i> En god oc sund Maffue/ som vel kand for- døye spisen.

Det er tydeligvis navnlig fraseologi, Matras har suppleret med; hans interesse for ordsprog er som sagt bevidnet fra andre skolebøger (jf. Kjær 1981).

De mest indflydelsesrige ordbøger i 1600-tallet skyldes vel nok Stephan Johannes Stephanius (1599-1650), og også de indskriver sig i den tematiske tradition. Stephanius kom som barn til Sorø skole i 1608 – for tidligt til at have kendt Matras. Og han dimitterede fra Herlufsholm 1615 – syv år før den første af Coldings ordbøger. Hans leksikografiske produktion omfatter to værker: en mindre, basal ordbog og en større, mere avanceret i to dele (nomina og verber)

der konsekvent har klassikercitater; den lille rummer blot latinske og danske modsvarigheder. De to værker bærer næsten samme titel og det er vanskeligt at holde dem ude af hinanden. Ved siden af ordbøgerne omfatter hans sprogpædagogiske forfatterskab samtaleøvelser, også til den videregående latinundervisning. Hans sprogpædagogiske indsats viste sig meget indflydelsesrig, og hans system brugtes langt op i 1700-tallet.

Stephanius' ordbøger er *De Deo*-ordbøger, og hermed kan Epitomes indflydelse siges at være slut. En indireket indflydelse er dog mulig. En notits i Peder Syvs *Danske Boglade* vidner om at Viborgrektoren Villads Nielsen Brøns har samlet kritiske kommentarer til Junius' værk (det fremgår ikke om det gælder en dansk bearbejdelse), og at disse har været udnyttet af Stephanius (Suhm 1780: 96).

## LITTERATUR

### *Håndskrift*

Rigsarkivet. 265 Københavns Universitet, Universitetsbiblioteket. Kataloger over Universitetsbibliotekets samlinger 1603-1603 (Katalog over Anders Lemvigs bibliotek dateret 1603).

### *Tryk*

van der Aa, A.J., 1860: Biographisch Woordenboek der Nederlanden. Udg. af K.J.R. van Harderwijk. Bd. 9. Haarlem.

Andersson, Henrik, Ebba Hjorth og Merete K. Jørgensen, 1997: Man skal ikke give bagerbørn hvedebrød. En artikel om leksikografisk brødarbejde. I: Fl. Lundgreen-Nielsen mfl. (red.), *Ord, Sprog oc artige Dict. Et overblik og 28 indblik 1500-1700*. Festskrift til Poul Lindegård Hjorth, Kbh. s. 101-114.

Birket Smith, Sophus, 1882: Om Kjøbenhavns Universitetsbibliothek før 1728, især dets Håndskriftsamlinger. Kbh.

Brugmans, H., 1927: Junius (Hadrianus) . I: *Nieuw Nederlandsch Biografisch Woordenboek*. 7. del. Leiden, sp. 692-694.

Claes, Franz, 1977: *Bibliographisches Verzeichnis der deutschen Vokabulare und Wörterbücher, gedruckt bis 1600*. Hildesheim.

Hannesdóttir, Anna Helga, 1998: *Lexikografihistorisk spegel. Den enspråkiga svenska leksikografins utveckling ur den tvåspråkiga*. Göteborg.

Hastrup, Niels, 1988: When 'interpretamenta' are translated instead of 'lemmata' or 'Around' af Vicious Circle in *Vocalubarium[!] ad usum Dacorum (1510)* by Christiern Pedersen. I: Graham D. Caie og Holger Nørgaard (red.), *A literary Miscellany Presented to Eric Jacobsen*. Kbh. s. 353-366.

Jensen, Kristian, 1982: *Latinskolens dannelse. Latinundervisningens indhold og formål fra reformationen til enevælden*. Kbh.

Junius, Hadrianus, 1567: *Nomenclator omnium rerum*. Antwerpen (faksimileudgave: Georg Olms Verlag Hildesheim. New York 1976). [Nomenclator]

- KancBrev = Kancelliets Brevbøger vedrørende Danmarks indre Forhold. Kbh. 1887 ff.  
 Kjær, Iver, 1981 (udg.): Daniel Matras: Proverbes. Kbh. (= Danmarks gamle Ordsprog V).
- Labarre, Albert, 1975: Bibliographie du Dictionarium d'Ambrogio Calepino (1502-1779). Baden-Baden.
- Matras, Christian, 1950: Ættmenn uttanlendis, i: Útiseti, Árbók Føroyingafelags, s. 78-99.
- Müller, Peter O., 2001: Deutsche Lexikographie des 16. Jahrhunderts. Konzeptionen und Funktionen frühneuzeitlicher Wörterbücher. Tübingen.
- Resen, Peder Hansen, 1685: Petri Johannis Resenii Bibliotheca Regiæ Academiae Hafniensi donata cui præfixa est ejusdem Resenii vita. Kbh.
- Rørdam, Holger Fr., 1869-1872: Københavns Universitets historie i Kong Frederik den Andens tid (1559-1588) [=Københavns Universitets historie fra 1537 til 1621, bd. 2] Kbh.
- Skautrup, Peter, 1931: Jacob Madsens dansk. I: Christen Møller og Peter Skautrup (udg.), Jacobi Matthiæ Arhusiensis (Jacob Madsen Aarhus). De literis libri duo. bd. 2, Århus, s. 69-98.
- [Suhm, P.F.] 1780: Bidrage til den Danske Bog-Historie, af Hr. Peder Syvs Danske Boglade, i: Samlinger til den Danske Historie. 1. bind, 2. hæfte. København, s. 89-104.

*Simon Skovgaard Boeck*

Nordisk Forskningsinstitut og Det Danske Sprog- og Litteraturselskab  
 ss@dsl.dk

# Nordisk barneordbok på nett

*Torbjørn Breivik*

The Nordic Dictionary for Children is an electronic dictionary consisting of approximately 3,400 entries in each of the three Scandinavian languages, Danish, Norwegian (Bokmål and Nynorsk) and Swedish. Each entry has a short written text (e.g. definition, example). All entries including the short text will be recorded so that the users can listen to them. When one is working with the neighbouring languages in Scandinavia the pronunciation often causes interference problems which blur the understanding. The dictionary is intended for children from 12 years upwards, it is for free use on Internet and as an app for the iPhone. It is intended as a resource in the teaching of neighbouring languages in school, which is part of the curriculum in Scandinavia from undergraduate level onwards. Users knowing Faroese, Icelandic, Greenlandic, Finnish or Northern Sami may use their own language for searching, but the hits will be in one of the Scandinavian languages. The project is part of the campaign to maintain inter-Nordic language understanding among children and young people. It has Nordic funding and is edited by three seniors from the Language Councils in Scandinavia. The dictionary will be launched in August 2012.

*Nøkkelord:* Elektronisk, skandinavisk, barneordbok med lyttfunksjon.

Nordisk barneordbok på nett vil være en ressurs i nabospråksundervisninga i skolen i Danmark, Norge og Sverige. I fase to utvides boka til brukere fra de andre språkområdene i Norden. Ordboka vil ha flerspråklig søking: brukerne kan søke fra eget språk og velge om de vil se hva ordet heter på ett eller flere av de tre basisspråkene (dansk, norsk (bokmål og nynorsk) og svensk). Målgruppa er primært barn fra ca. 12 år og oppover. Ordboka skal inneholde et grunnleggende ordforråd med tillegg av hverdagsord for målgruppa. Oppslagsordene og en kort forklaring skal bli mulig både å lese og å lytte til, og boka vil kunne brukes på flere plattformer. Prosjektet er knyttet til Nettverket for språknemndene og inngår som en del av den nordiske språkkampanjen.

## Hvorfor en nettordbok for barn i Norden?

Gjennom Nordisk språkkampanje 2011 understreker Nordisk ministerråd at de ønsker å synliggjøre nordiske språk og nordisk kultur med fokus på språkforståelse. Ministerrådet ønsker at barn og unge skal bli oppmerksomme på fordelene med nordisk språkforståelse, og de støtter prosjekter og tiltak som kan gi lærere noen redskaper til å nå dette målet. Samarbeidsministrene som vedtok å sette i gang den nordiske språkkampanjen, uttalte at ”vi må gi de unge lyst til å lære de skandinaviske språkene fordi det åpner en helt ny verden for dem” (Nordisk ministerråd).

Ideen om en nettordbok for barn som skal lære seg å forstå andre nordiske språk, dukket opp gjennom diskusjoner om hva som kunne gjøres for å realisere den nordiske språkdeklarasjonens målsetting om nordisk språkforståelse. Undersøkelsen ”Håller språket ihop Norden?” (Delsing & Lundin Åkesson 2005) viser at det står dårlig til med ungdommers nabospråksforståelse i Danmark, Norge og Sverige. Undersøkelsen sammenligner resultatene med en nesten tretti år gammel undersøkelse (Maurud 1976) og finner at det har gått nedover med nabospråksforståelsen blant ungdom. Undersøkelsen fra 2005 viser imidlertid at det er håp: de ungdommene som antas å ha mest kontakt med nabolandene, forstår litt mer enn de som ikke har den typen kontakt.

I Danmark, Norge og Sverige er det nordiske og nabospråksforståelse en del av læreplanene for grunnskolen. Fra elevene er 11–12 år, skal de bli kjent med nabospråkene. Kravene til lesing og forståelse øker med klassetrinnene til at norske elever i videregående skole (gymnasiet) skal kunne gjøre rede for nordisk språkpolitikk – intet mindre! Noen av oss fant ut at vi kunne senke ambisjonsnivået litt, se hva dagens ungdom bruker av hjelpemidler i skolearbeidet, og dermed var ideen om en nettbasert ordbok født.

Det finnes mye litteratur og mange opplegg for nabospråksundervisning tilgjengelig for lærere (f.eks. på nettsidene til Nordisk språkkoordinasjon). Bøynen for de fleste lærere er tid, og det å finne materialet som er der ute. En undersøkelse som Språkrådet i Norge gjennomførte høsten 2010 blant norsklærere i ungdomsskolen og videregående skole, viser at kompetansen i nabospråksundervisning varierer. Holdningen til emnet er ofte negativ hos dem som har dårligst kompetanse i det (Språkrådet i Norge: [www.sprakradet.no/Tema/Skole/](http://www.sprakradet.no/Tema/Skole/)).

Med dagens teknologi tilgjengelig vil en trykt ordbok neppe bli tatt i bruk av barn og unge. Bokhylla er ikke det første stedet ungdom går til for å lete når det er noe de vil vite – de går til nettet, og de surfer gjerne via mobilen. En ordbok må derfor være tilgjengelig på disse mediene om den skal bli brukt. I tillegg bør den ha et brukervennlig grensesnitt og noen fasiliteter som gjør den ekstra attraktiv.

## Innhold

Oppslagsordene i ordboka er fra basisordforrådet og fra emneområder vi antar barn er opptatt av. Ordene får korte forklaringer, og forklaringene skrives etter co-build-systemet, som betyr at oppslagsordet gjentas i den korte forklaringen som følger. Det som særlig vil skille denne nettordboka fra andre, er at vi legger til lyd: alle oppslagordene og forklaringene blir lest inn. Brukerne kan da lytte til uttalen og forklaringen av ordet samtidig som de leser den på skjermen. Oversettelse av innholdet til de andre språkene i Norden gir mulighet for at brukerne kan søke på eget språk og se og høre hva ordet heter på dansk, norsk (bokmål og nynorsk) og svensk. Ord og forklaringer vil bli lest inn på disse tre språkene.

Basisordforrådet er ikke nødvendigvis forskjellig i dansk, norsk og svensk. Det er flust av fellesord som skiller seg lite når det gjelder skrivemåte, men kan gjøre det når det gjelder bøyingsmåter og noen ganger når det gjelder betydning. Vi har lagt vekt på å få med ord fra temaområder som samfunnsliv, kultur, skolehverdag, fritid, natur, slang, medie verden samt en del faste ord og vendinger. Barnelitteraturen har f.eks. ulike tradisjoner når det gjelder oversettelse av fiktive navn, og et utvalg av disse er med, f.eks. *Donald Duck* (*Kalle Anka*, *Anders And*), *Peter Pan*, *Spiderman* og *Harry Potter*. Målet er å vise brukerne hvor like språkene er på noen områder, men også at de kan være svært ulike på andre. De vanligste falske vennene må selvsagt også være med. For en norsk ungdom betyr *rolig* at det er *stille*, men en svensk vil oppfatte det som moro. *Artig* på dansk er å være høflig, mens det på norsk betyr å ha det moro. Et *marssvin* på norsk er utlukkende et lite kjæledyr, men for en dansk ungdom er det navnet både på et kjæledyr (overensstemmende) og på en liten hval (*nise* på norsk).

Det byr også på utfordringer å innpasse og forklare ord som bare finnes i ett eller to av språkene og mangler tilsvarende i det tredje: *bedstefar* (da), *bestefar* (no) og ?? i svensk. Ulike ord med samme referanse, er noe enklere: *amt* (da), *fylke* (no) og *län* (sv) eller *løbehjul* (da), *sparkesykke* (no) og *sparkcykel* (sv), *agern* (da), *eikenøtt* (no) og *ekollon* (sv). Mange ord er (nesten) like og har samme referanse: *adresse*, *bord*, *stol*, *skole*. De korte forklaringene som kreves på grunn av kostnadene med innlesingen, byr på utfordringer for redaksjonen som skal formulere dem. Jo lengre en forklaring er, jo lengre tid bruker vi i studio og det medfører økte kostnader for prosjektet.

Noen grupper av ord krever en særlig gjennomgang. Det er ord som har med kultur og religion å gjøre. Innvandrere har med seg sin egen kultur og religion når de kommer, og det må ordboka ta et visst hensyn til. Velger vi å ha med *biskop* og *prest*, må vi også ha med tilsvarende titler/posisjoner i de større religi-

onene som nå er representert i våre respektive land. Ord og uttrykk som gjelder skolehverdagen må være riktige og gjenspeile dagens ordbruk. Et eksempel er *gymnasium* som har gått ut av bruk i Norge. Dagens betegnelse på skolen som kommer etter ungdomsskolen, er *videregående skole*.

## Opplysninger om enkeltord

Brukerne vil bare få tilgang til å se selve ordet og forklaringen på det. I basen vil det i tillegg ligge opplysninger om ordklasse og bøyning. Når det gjelder bøyning, legger vi inn hovedformene og de vanligste alternativene til bøyning om det finnes noen. For norsk finnes det både på bokmål og nynorsk alternative skrivemåter og bøyingsformer for mange ord, og her er det gjort et skjønnsomt utvalg. Disse opplysningene legger vi inn av to grunner: ordboka kan bygges ut seinere og hensynet til søkemulighetene. Hvis brukerne søker på en bøyingsform fordi de ikke kjenner grunnformen, skal de få treff på søket sitt, men de vil få fram grunnformen av oppslagsordet med forklaring. Denne løsningen gir større fleksibilitet og øker bruksmulighetene for ordboka. Om noen seinere skulle ønske å utvide ordboka slik at også bøyingsformer og ordklasse er aktuelt å vise for brukerne, ligger informasjonen der.

## Trinnvis utvikling av ordboka

I første fase av prosjektet vil innholdet bli på dansk, norsk (bokmål og nynorsk) og svensk, og disse språkene utgjør kjernen i ordboka. I fase to av prosjektet vil innholdet bli sendt til oversetting til de andre nordiske språkene: grønlandsk, islandsk, færøysk, finsk og samisk (nordsamisk). Vårt prosjekt dekker ikke en fullstendig mangespråklig ordbok, men ved å oversette ordene og forklaringene til alle språkene, gir vi brukerne søkemulighet inn mot kjernematerialet. Når det gjelder disse oversettelsene, er det ikke nødvendig å ta hensyn til hvor mye plass forklaringene tar. Disse delene skal i første omgang ikke leses inn, og man kan da tillate seg å bruke litt mer plass til å forklare hva ordet eller uttrykket betyr. Denne delen er for brukere som ikke har norsk, dansk eller svensk som sitt førstespråk, men som ønsker å vite hva et dansk, norsk eller svensk ord betyr, og hvordan det høres ut. Forklaringene skal være til støtte for brukeren for å forstå betydningen av det danske, norske (bokmål eller nynorsk) eller svenske ordet. Arbeidsmåten tar høyde for seinere utbygging til en fullstendig mangespråklig ordbok. Antall søk og hvilke ord det søkes på, vil bli registrert og redaksjonen vil gjøre en vurdering med jevne mellomrom om vi skal legge inn



flere ord. Det er også mulig å tenke seg at brukerne kan melde fra om ord de savner.

Det nordiske samarbeidet er basert på at vi bruker dansk, norsk eller svensk i møter oss imellom, og det er forståelsen av disse språkene som i første omgang har vært vårt anliggende med prosjektet. Teknisk tilrettelegges ordboka for utvidelse til full flerspråklig ordbok, men økonomien setter stopper for det i denne omgang. Det vil imidlertid bli lagt inn pekere til andre ordbøker og nyttige ressurser fra inngangssida for ordboka.

I trinn tre ser vi for oss å utvikle egne undervisningsopplegg knyttet til ordboka, og tips for lærere om måter å bruke ordboka på både i morsmålsundervisningen og i andre fag, f.eks. når temaet er skole, politisk styresett, nordisk natur osv. Vi ser også for oss at det kan bli mulig å legge inn bilder og illustrasjoner koplet til ord og uttrykk for å hjelpe på forståelsen. Det var ønskelig å ha visuelle uttrykk med i første fase, men det er kostbart og av den grunn flyttet til trinn tre. Det er også mulig å tenke seg ordboka som en wiki-løsning der brukerne selv bidrar til utvikling av ordboka, noe som vil kreve en redaksjon som tar hånd om og vurderer forslag som kommer inn og sørger for å tilpasse dem formatet i basen.

## Prosjektgruppen

Prosjektgruppen består av en person fra hvert av de skandinaviske landene: Pia Jarvad er hovedredaktør og har ansvar for det danske innholdet. Birgitta Lindgren er redaktør og ansvarlig for det svenske innholdet. Torbjørg Breivik er prosjektleder, redaktør og ansvarlig for de to norske versjonene av innholdet. Jakop Halskov fra Dansk Sprognævn/Nordisk Sprogkoordination har bidratt med teknisk tilrettelegging av arbeidsverktøyet. I tillegg har Bodil Aurstad fra Nordisk Sprogkoordination bidratt med sin erfaring fra ordboksarbeid.

## Finansiering av prosjektet

Prosjektet har fått midler fra Nordpluss språk, fra Nordisk kulturfond, fra den nordiske språkkampanjen, både den nordiske og den norske delen, fra nettverket for språknemndene og fra noen av språknemndene. I tillegg bidrar de tre redaktørene med betydelig egeninnsats. Prosjektet har en total ramme på ca. DKK 1 mill. Språkrådet i Norge er juridisk eier av ordboka på vegne av Nettverket for språknemndene i Norden.

## Status pr. oktober 2011

Planleggingen av ordboka startet for alvor i februar 2010 etter at det på nettverksmøtet for språkrådene i august året før var blitt foreslått at man burde undersøke muligheten for å lage en gratis nettordbok for barn. Høsten 2010 ble prosjektet presentert for nettverkets medlemmer. Nettverket sluttet seg til forslaget, og Nordisk språkkordinasjon påtok seg å være administrativ tilrettelegger. Søknader om midler ble sendt i flere omganger, og med positiv tilbakemelding fra Nordpluss språk om 40 000 euro til prosjektet, gikk prosjektgruppa i gang i januar 2011.

For å komme raskt i gang og få et visst basisordforråd vedtok prosjektgruppa å ta utgangspunkt i en eksisterende ordbok for barn og bygge ut den med et utvalg av ord som dekker de temaene redaktørene mente burde være med i en ordbok for skoleelever. Redaktørene har nå gått gjennom ca. 2/3 av materialet, og har lagt en plan for det resterende som tilsier at ordboka skal kunne lanseres rundt årsskiftet. Det gjenstår korrektur og noe justering av innholdet og forklaringene. Når dette er ferdig fra redaktørene, blir innholdet levert til produksjon av lyd (innlesing) og produksjon av databasen.

Lyden vil bli produsert ved at profesjonelle innlesere (ungdommer) leser inn hvert ord med forklaring, og de innleste filene blir så koplet til selve databasen. Syntetisk tale til opplesingen ble vurdert fordi det er billigere enn å bruke mennesker til innlesing, men resultatet ble for dårlig. Hensikten med ordboka er å bidra til å øke forståelsen av de andre nordiske språkene, og da må lyd kvaliteten være så god at det ikke hersker tvil om hva som blir lest. Den tekniske løsningen i databasen er tilrettelagt for å gi rask respons på brukernes søking.

## Samarbeid med andre prosjekter

På vegne av språknemndene søkte prosjektgruppa om midler fra Nordpluss språk og fikk tilbakemelding om at vi burde vurdere mulighetene for samarbeid med Foreningen Nordens Forbund i Finland, som hadde søkt om midler til et nordisk parlørprosjekt. Parløren vil dekke hverdagssituasjoner for barn og unge som besøker et av nabolandene, og ordforrådet vil være dekket i ordboka.

I Norge er Språkrådet sammen med Foreningen !les og Foreningen Norden utpekt som bidragsterne til Nordisk språkkampanje 2011. Den norske tXtaksjonen går ut på å få barn og unge til å lese og lytte til tekster på svensk og dansk. Ordforrådet i tekstene skal være dekket i ordboka. Foreningen Norden i Norge har fått penger til et prosjekt der forfattere og litteraturformidlere fra Sverige og Danmark besøker ungdomsskoler og videregående skoler i Norge. Med besøket følger et undervisningsopplegg tilpasset læreplanens kompetanse-

mål, og ordboka vil være en støtte i undervisningen og nyttig oppslagsverk når barn og unge blir inspirert til å lese mer litteratur på nabospråkene.

## LITTERATURLISTE

- Delsing, Lars-Olof & Lundin Åkesson, Katarina, 2005: Håller språket ihop Norden? En forskningsrapport om ungdomars forståelse av danska, svenska og norska. Tema-Nord 2005:573, Nordiska Kulturfonden, København.
- Maurud, Ø., 1976: Nabospråksforståelse i Skandinavia. En undersøkelse av gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige. Nordisk utredningsserie 13. Nordiska rådet, Stockholm.

### *Aktuelle lenker*

Dansk Sprognævn: <http://www.dsn.dk>

Foreningen!les: <http://www.foreningenles.no>

Foreningen Norden Norge: <http://www.norden.no/forfatterbesok2011.html>

Nordisk kulturfond: <http://www.nordiskkulturfond.org/>

Nordisk ministerråd: <http://www.norden.org>

Nordisk språkkampanje 2011: <http://www.norden.org/no/aktuelt/nyheter/spraakkampanje-skal-styrke-det-nordiske-kulturfellesskapet>

Nordisk språkkordinasjon: <http://nordisksprogkoordination.org/>

Språkrådet i Norge: <http://www.sprakradet.no>

Språkrådet i Sverige: <http://www.sprakradet.se>

*Torbjørge Breivik*

Språkrådet, Norge

[torbjorg.breivik@sprakradet.no](mailto:torbjorg.breivik@sprakradet.no)

# Hur blir en traditionell, tryckt ordbok en webbordbok?

*Ilse Cantell & Caroline Sandström*

The history of *Ordbok över Finlands svenska folkmål* (Dictionary of Swedish Dialects in Finland) began in the 1920s when the collecting of dialect data was initiated. The editing process started at the beginning of the 1960s, and four volumes were published between 1982 and 2007.

When the digitizing process began in 2010 the first two volumes (*A–Hu*), which only existed on paper, had to be scanned. They were converted into XML format using the structure indicators in the dictionary articles. The manuscripts for the next two volumes (*Hy–Och*) were compiled in a word processing program. Today the dictionary is compiled in a dictionary editing program. In the future, new articles will primarily be published online. A pilot consisting of the interval (*I–K*) will appear in 2013.

In our presentation we discuss difficulties which arise when a traditional printed dictionary is transformed into an online publication. We focus on features that can be used to parse the dictionary text into structured text files. Important questions are e.g.: What structure indicators are typographical? When do we have to pay attention to the content of the text to get proper tagging?

Our target is an electronic dictionary with flexible search possibilities that fully use all elements of the articles. We aim at serving both members of the ordinary dialect-speaking public and researchers in linguistics.

## 1. Projektet *Ordbok över Finlands svenska folkmål*

Bakgrunden till *Ordbok över Finlands svenska folkmål* (FO) sträcker sig ända till 1920–30-talen<sup>1</sup>. Vid denna tidpunkt hade dialekterna redan dokumenterats i några decennier och upplysningar om de svenska dialekterna i Finland hade publicerats i bl.a. Rietz (1862–67), Vendell (1904–07) och Wessman (1925–

---

<sup>1</sup> Ahlbäck (1976: VIII–XIX) och Slotte (1981: 45–64) redogör för bakgrunden till FO, källorna och redigeringsprinciperna för ordboken. Olav Ahlbäck var huvudredaktör (1960–1981) och han efterträddes av Peter Slotte (1982–2009). Sedan 2009 är Caroline Sandström huvudredaktör för FO.

32). År 1927 inrättades Folkmålskommissionen<sup>2</sup> som beviljades statliga anslag och i samband med detta väcktes tanken om att öka insamlingen med sikte på en heltäckande dialektordbok. Ett omfattande, systematiskt insamlingsarbete av dialektala ord och uttryck inleddes i samarbete med Svenska litteratursällskapet i Finland (SLS). Under slutet av 1940-talet påbörjades arbetet på ett ordregister över det insamlade dialektmaterialet kompletterat med andra källor. Resultatet blev ett register som i dag består av omkring en miljon ordsedlar.

Redigeringsarbetet på dialektordboken inleddes i början av 1960-talet. Ordbokens redaktion överfördes till Forskningscentralen för de inhemska språken<sup>3</sup> då forskningscentralen grundades. Första häftet av FO kom ut 1976, följt av ytterligare fem häften med tre till fyra års mellanrum. År 1982 publicerades det första hela bandet (*A–E*) och det andra bandet (*F–Hu*) stod färdigt 1991. Utgivningen i häften frångicks under början av 1990-talet och tredje (2000) och fjärde banden (2007) – intervallet (*Hy–Kjys*) respektive (*Kjysa–Och*) – har utkommit som hela volymer. Från och med fjärde bandet är SLS förläggare för ordboken.

FO är en typisk *flergenerationsordbok* (Vikør 1999) med en lång tillkomst-historia. Kännetecknande för dessa ordböcker är att varierande redigerings- och trycktekniker har avlöst varandra under årens lopp, vilket bl.a. kan ha konsekvenser vid överföringen av ordböckerna till digitala databaser.

De första häftena av FO redigerades manuellt och sattes vid tryckningen med datorstyrd fotosättning. Under åren 1993–2005 redigerades FO med ett ordbehandlingsprogram och 2006 övergick man till ett redigeringsprogram. Publiceringen av nytt material sker under 2010–20-talen primärt på webben, men enligt överenskommelse med förlaget SLS kommer FO också i framtiden att tryckas. Den färdiga ordboken beräknas bestå av omkring 125 000 artiklar fördelade på sju band.

Utgångsläget för den digitalisering som behövs för webbordboken är varierande för de olika avsnitten av ordboken. De två första banden har skannats och genomgått textigenkänning (OCR) innan de kunnat konverteras till XML-kod varefter de ännu måste kontrolleras i flera omgångar vid överföringsprocessen. Tredje och fjärde banden redigerades i ett ordbehandlingsprogram och har konverterats och kontrollerats vid överföringen till XML-kod. Eftersom den pågå-

<sup>2</sup> Folkmålskommissionen, vars officiella namn är Kommissionen för undersökning av svenska folkmål och svensk folkkultur i Finland, var en fri sammanslutning av företrädare från olika vetenskapliga kulturinstitutioner.

<sup>3</sup> Forskningscentralen för de inhemska språken, sedan 1.1.2012 Institutet för de inhemska språken, grundades 1976 och är ett statligt finansierat språkinstitut underställt Undervisnings- och kulturministeriet.

ende redigeringen av nya artiklar sker med ett redigeringsprogram i XML-kod kan de nyaste artiklarna från femte bandet framåt tillsammans med artiklarna i de fyra första banden sammanslås till en XML-kodad databas som sedan konverteras till det format användargränssnittet på webben kräver.

## 2. Varför en webbordbok?

Förlagsbranschen både i Finland och på andra håll i Europa har under de senaste åren drabbats av nedskärningar som i synnerhet påverkat utgivningen av stora referensverk och ordböcker (se Laureys 2012). Dessa tendenser märks också i Institutet för de inhemska språkens utgivningspolitik; de egna ekonomiska resurserna för tryckning av pappersordböcker är begränsade och under de senaste åren har det blivit allt svårare att få utomstående finansiering för tryckning.

De stora pågående ordboksprojekten vid institutet har under 2000-talet genomgått flera utvärderingsprocesser. Trots de statliga sparkraven vill institutet hålla fast vid kärnverksamheten i lexikografi och språkvård och även i framtiden satsa på utgivning av kulturbärande historiska ordböcker. De kommer i fortsättningen att publiceras på webben. Framför allt för forskare ger webbordböcker möjlighet till nya ingångar till språkliga data. Webbpubliceringen medför också att dessa ordböcker når ut till nya användare. För FO:s del innebär de nya kraven och övergången till webbpublicering att också de tidigare utgivna banden av ordboken bör göras tillgängliga på webben så att det kompletta ordboksverket kan komma såväl forskare som allmänheten till godo.

## 3. FO:s strukturer och konverteringarna

Man kan uppfatta strukturen i en ordbok på många olika sätt. Den enda struktur som är konkret synlig i en tryckt ordbok är typografin och layouten. Men bakom den finns lexikografens tanke om hela ordbokens och de enskilda artiklarnas struktur.

Om det vore så att det fanns en direkt motsvarighet mellan varje lexikografiskt strukturelement<sup>4</sup> och en typografisk lösning, så skulle konverteringen vara enkel. Men så är inte fallet, utan konverteringen måste bygga på en analys av sambandet mellan text och typografi.

---

<sup>4</sup> Ibland används här också för korthetens skull termen *element* om inte risk för sammanblandning med begreppet 'element' i XML-strukturen föreligger.

### 3.1 Den typografiska strukturen

Då FO ska ges ut som webbordbok, ska texten i ordbokens fyra första band överföras till en kodning enligt märkspråket XML. Vid konverteringen finns det inte något annat att utgå ifrån än texten och typografin. Texten i den tryckta FO erbjuder tre typer av fästpunkter som kan användas vid konverteringen till XML-kod. Det kan vara rent layoutmässiga drag, typografiska drag, förekomsten av vissa textsträngar eller av tecken som bara förekommer i vissa delar av ordboksstrukturen. Allt detta ska återges i webbordboken, men ibland kan det inte genomföras på samma sätt som vid tryckningen. Fonter kan t.ex. bete sig annorlunda på skärmen jämfört med hur de ser ut i tryck.

I en webbordbok ter sig frågan om utrymme helt annorlunda än i en tryckt ordbok. På en datorskärm finns det inte lika mycket plats som på en tryckt sida. Å andra sidan kan själva materialet som blir tillgängligt på webben, dvs. databasen, vara nästan oändligt stort.

I FO:s makrostruktur markeras början och slutet av artiklarna av hängande indrag och radbyten. Som typografiska medel att skilja åt elementen i FO:s mikrostruktur har man i den tryckta ordboken utöver rak brödtext använt olika snittvarianter: fet och kursiverad stil, spärning och olika teckengrader.

Dessutom innehåller FO textsträngar som dels står som strukturella markörer mellan olika element i mikrostrukturen, dels bara kan ingå i vissa element. Det kan vara fråga om skiljetecken (semikolon, komma, tankstreck o.d.), siffror och ensamma högerparenteser som i FO används som markörer för de olika betydelsemomenten.

Tecken från landsmålsalfabetet<sup>5</sup> förekommer endast i elementet <uttal>, dvs. i uttalsvarianterna av uppslagsorden och i elementet <fras>, exempelfraserna. Sådana textsträngar kan kodas antingen som elementet <uttal> eller elementet <fras>. Båda dessa har dessutom ett attribut vars två olika värden ”fin” och ”grov” anger vilken ljudbeteckning som använts i källan. Apostrof som anger muljering och omvänd apostrof för partiell tonlöshet är också tecken som endast finns i elementen <uttal> och <fras>.

Spärrad stil har i den tryckta ordboken använts för hänvisningar men också i andra funktioner då man velat ge emfas åt ett ord eller en fras. Med hjälp av hänvisningsmarkörer (=, *Jfr/jfr* eller *se*) ska det vara möjligt att göra skillnad på

<sup>5</sup> Tre tecken som används i ljudbeteckningen i FO och som ingår i landsmålsalfabetet men inte i IPA har på initiativ av FO:s redaktion införts i Unicode-standardens version 5.1.0 i mars 2008. Dessa är tecknet för stockholms-e, tecknet för tonlöst l och o-tecknet med ring nertill. De ingår nu i fältet Latin Extended-C och deras koder är U+2C78 (Latin small letter e with tail), U+2C79 (Latin small letter turned r with tail) och U+2C7A (Latin small letter o with low ring inside).

explicita hänvisningar som automatiskt kan länkas ordboksinternt och ord och fraser som getts emfas genom samma typografiska lösning. Enskilda ord i spärrad stil är implicita hänvisningar, de kan ibland stå i böjd form och kan då inte länkas automatiskt. (Se exempel i artikeln *hydare* nedan.) Fraser i spärrad stil kan ofta inte i sin helhet tolkas som hänvisningar:

lyfta v. [- -] lyfta på (el. uppå) hatten el. mössan [- -]

Fraser av denna typ föregås vanligen av texten *i uttr.* (= i uttrycket). Frasen kan ses som en uppslagsform till efterföljande dialektala exempelfraser och ska då inte länkas till en annan artikel.

### 3.2 Den lexikografiska strukturen

Den lexikografiska strukturen, framför allt makro- och mikrostrukturen ska vid konverteringen överföras till XML-kod. Enskilda artiklar och nischer med sammansättningar ska analyseras fram och ordboksartiklarnas objektspråkliga och metaspråkliga strukturelement ska få rätt kod.

Artiklarna i FO kan vara antingen enkla artiklar eller nischer bestående av ett simplexord och därtillhörande sammansättningar. Sammansättningarna är dessutom ordnade i grupper beroende på förledens form, t.ex.:

lucka s.

A: luck-byxor

-kapp

-karm

-karvel

-nar

-ring

-skalk

B: lucka-byx

-byxor

C: lucku-byx

-byxor

-ränna

-sträckare

Vid XML-kodningen av elementen i FO:s makrostruktur gäller det att finna gränserna mellan artiklarna, att föra ihop sammansättningarna med simplexordet och att gruppera sammansättningarna enligt förledstyperna som i ordboken markerats med stor bokstav i fetstil.



Artiklarna (för exempel se artikeln *hydare* nedan) består av en formell del där uppslagsordet, dess dialektala uttalsvarianter och grammatiska upplysningar ingår. Betydelsebeskrivningen ges genast efter den formella delen i korta artiklar. I momentindelade artiklar kommer betydelsebeskrivningarna efter respektive siffra. Siffran står i fetstil och efterföljs av en ensam högerparentes. Utbredningen ges också för de olika betydelserna och i samband med betydelserna ges de eventuella exempelfraserna.

**hydare** *hy:dar* Kn, *hy:dar(e)* NL m.3. – 1) rackare, hudavdragare. NL. – 2) man som efter säljägarnas hemkomst från f ä r d e n deltog i arbetet med att h y d a sälarna. Kn.

Förkortningarna som används vid de grammatiska upplysningarna och markörerna för den geografiska utbredningen är konventionaliserade och därför lätta att få rätt kodade vid konverteringen. Dessutom innehåller artiklarna i FO interna och externa hänvisningar, etymologiska uppgifter med angivande av ord i ett långivande språk, t.ex. finska, ryska eller amerikaengelska, och i vissa fall också angivelser av en källa.

### 3.3 XML-kodad strukturerad text

För att ett stort material ska kunna överföras till en databas för utgivning både som tryckt ordbok och som elektronisk ordbok ska materialet kodas. Numera används märkspråket XML.

Den typografiska strukturen är utgångsläget då en traditionell, tryckt ordbok vars manus tagits fram antingen manuellt eller med ett ordbehandlingsprogram ska överföras till det XML-format som är målet för processen. Nedan står ett exempel på en färdig, överförd artikel, en mycket kort sådan. Artikeln *lullog* har överförts från typografiskt formaterad text till FO:s XML-kodade text. XML-kodningens hierarki kan här ses med hjälp av de olika indragningarna. De andra elementen är inkapslingar i det överordnade elementet för artikeln, <art>.

<pre> &lt;art&gt;   &lt;uppsl&gt;lullog&lt;/uppsl&gt;   &lt;text&gt;     &lt;uttal skrift="fin"&gt;lullo(gär) &lt;/uttal&gt;     &lt;gramm&gt;adj.&lt;/gramm&gt;     gungande. vacklande, ostadig.     &lt;ort&gt;Na&lt;/ort&gt;.   &lt;/text&gt; &lt;/art&gt; </pre>	<p><b>lullog</b> <i>lullo(gär)</i> adj. gungande, vacklande, ostadig. Na.</p>
---	---

Figur 1 nedan är ett försök att sammanföra den typografiska och den lexikografiska strukturen, dvs. i detta fall ordbokens mikrostruktur samt strukturen i den XML-kodade texten till en helhetsbild. Taggarna med namn i fetstil får illustrera den lexikografiska strukturen, medan taggarna i vanlig stil representerar elementen i ett XML-kodat FO-dokument. I de två sista kolumnerna till höger står element i den typografiska strukturen vid sina respektive lexikografiska strukturelement och textsträngar som bara kan stå på bestämda ställen i strukturen.

			Typografi	Text
<Mikrostruktur>				
<uppsl>				
</uppsl>				
<homograf>				
</homograf>				
<text>				
	<uttal>		Kursiv/stor rak stil	tecken från lands- målsalfabetet
	</uttal>			
	<ort>		Rak stil	Lista
	</ort>			
	<gramm>		Rak stil	Lista
	</gramm>			
	<ort>		Rak stil	Lista
	</ort>			
	<mom>		Fet stil	tankstreck, siffra, högerpar.
	<fras>		Kursiv/stor rak stil	tecken från lands- målsalfabetet
		<rak_text>	Rak stil	
		</rak_text>		
		<fras_forkl>	Fet stil	
		</fras_forkl>		
		<ort>	Rak stil	Lista
		</ort>		
		<hanv>	Spärrad stil	
			(inkapslingar)	
		</hanv>		
	</fras>			
	</mom>			
</text>				
</Mikrostruktur>				

Figur 1 Den typografiska och lexikografiska strukturen i FO.

#### 4. Planering av webbordboken

Inför planeringen av *Webb-FO*<sup>6</sup> har redaktionen haft stöd av språkteknolog- och lexikografkolleger vid Institutet för de inhemska språken. Arbetet har inletts med ett pilotprojekt där funktionerna testas på ett kortare avsnitt ur ordboken.

Kriterierna för en bra webbordbok är varierande. Pastor och Alcina (2010) har jämfört ett antal elektroniska ordböcker utgående från olika kriterier. Kriterier som de diskuterar är ordbokens databas som resurs och andra eventuella resurser som kan nås t.ex. via hyperlänkar. Vidare behandlar de sökningar och hur sökresultatet presenteras. Nedan i avsnitten 4.1 och 4.2 diskuterar vi hur *Webb-FO* kunde byggas upp utifrån dessa kriterier.

FO:s redaktion har bekantat sig med några webbordböcker och tänkt ut en rad prioriterade egenskaper då *Webb-FO* tas fram. Problem som iakttagits hos gratisordböcker på webben är att de innehåller så mycket reklam att det ibland är svårt att skilja mellan innehåll och reklam. Sökfunktionerna i det brokiga sortimentet är sällan särskilt genomtänkta eller sofistikerade.

En god förebild bland webbordböckerna är *EDD Online (English Dialect Dictionary)*. Den föreligger tills vidare bara i en betaversion som forskare kan komma åt att testa. Alla planerade funktioner är inte ännu färdiga, men de finns redan med som knappar eller kryssbara rutor i användargränssnittet. *EDD Online* har en del sådana egenskaper som kunde tillämpas för *Webb-FO*. Den är jämförbar med FO dels därför att den är en dialektordbok och innehåller liknande lexikografiska kategorier som FO, dels därför att den genomgår en motsvarande digitaliseringsprocess.

Samtidigt med utgivningen av *Webb-FO* pågår webbpublicering av den finska dialektordboken *Suomen murteiden sanakirja (SMS)*<sup>7</sup> vid Institutet för de inhemska språken. Artikelstrukturen i SMS och FO är rätt olika, så det är bara en del av arbetet med SMS som kan utnyttjas för FO. Bägge ordböckerna kommer att nås via materialbanken *Kaino* (<http://kaino.kotus.fi/>) och därför koordineras projekten. En annan förebild för FO är webbordboken över karelskan *Karjalan kielen verkkosanakirja*<sup>8</sup> med sina mångsidiga sökmöjligheter. Den publicerades redan i oktober 2009 på *Kaino*.

<sup>6</sup> Arbetsnamn för FO:s webbversion.

<sup>7</sup> Ett pilotavsnitt av SMS publicerades på webben i februari 2012.

<sup>8</sup> En recension av den karelska webbordboken finns i LexicoNordica 17 (Cantell 2010).

#### 4.1 Ordboksresurser

Vi har ovan i avsnitten 3.2 och 3.3 beskrivit databasen som resurs för webbordboken. Framför allt är det viktigt att FO:s databas och kodning ska utnyttjas till fullo. Om exaktare kodning behövs för någon funktion som annars inte är möjlig, är redaktionen beredd att satsa på detta.

Hänvisningarna i ordboken ska vara länkade så att man genom att klicka kommer fram till det ställe som hänvisningen avser. Dessutom kunde det skapas länkar vid hänvisningar till åtminstone två andra ordböcker som finns eller kommer att finnas tillgängliga på webben, nämligen till SAOB och SMS. I framtiden planerar Institutet för de inhemska språken att utveckla en ordboksportal och det vore fint att där kunna länka direkt mellan husets olika ordböcker. Externa länkar är sårbara, eftersom de inte nödvändigtvis är bestående; t.ex. databasen för SAOB kan i framtiden bli omplacerad.

#### 4.2 Sökningar

Webbordboken ska vara mångsidigt sökbar. Det ska vara möjligt att göra (1) enkla sökningar i lemmalistan, (2) avancerade sökningar med olika begränsningar och filter som är relevanta med hänsyn till innehållet och (3) fritextsökningar i hela databasen.

Avancerade sökningar i *Webb-FO* ska kunna begränsas till olika element i artiklarna. Förutom bland uppslagsorden – som kan vara simplexord och/eller sammansättningar – ska det vara möjligt att söka bland uttalsvarianterna och exempelfraserna och dessutom i brödtexten vars viktigaste del är betydelseförklaringarna. Betydelseförklaringarna är dock inte taggade, vilket betyder att sökningen här ger en hel del extra brus.

De avancerade sökningarna ska också kunna begränsas med hjälp av sökfiltter enligt beläggort (socken) eller dialektområde, uttalsvariant (t.ex. enligt kvalitet och kvantitet), ordklass, böjningskategori (deklination och konjugation), etymologi (t.ex. finska, ryska, amerikaengelska som långivande språk) samt enligt olika kombinationer av dessa.

*Webb-FO* ska t.ex. kunna konsulteras för följande sökningar:

- Vilka svaga maskuliner förekommer i dialekten i Tenala? (svar t.ex.: *buske* 1, *hägne*; *hyvel*)
- Vilka sammansatta adjektiv finns vars efterled slutar på *-og*? (svar t.ex.: *hårdmjölkog*, *häll-fattog*, *hämnd-girog*)
- Vilka substantiv ur gruppen femininer f.1/f.2 kan i vissa dialekter ha maskulin böjning m.1/m.2? (svar t.ex.: *al* 1, *dörr*, *jord*, *lagg* 1, *lagg* 2, *not* 1)

- Vilka homografer av ord som slutar på *-a* förekommer som både substantiv (svf.) och verb? (svar t.ex.: *balga*, *håsa*, *mara* med homografer)
- Vilka finska lån finns belagda i norra svenska Östbotten? (svar t.ex.: *kytta* s.1, *käpel* 1, *käpel* 2, *lamp* 1)

Exaktare kodning av innehållet i databasen behövs om *Webb-FO* ska kunna erbjuda sina användare möjlighet att söka synonymer eller antonymer eller s.k. indirekta etymologier, dvs. etymologiska angivelser som finns i den artikel som det hänvisas till men som också avser det ord som det hänvisas ifrån. Sökningar enligt bynamn eller önamn kan inte heller göras utan att exaktare kodning införs.

För att webbordboken ska kunna användas för sökningar som kan ge svar på frågor om t.ex. terminologi, utbredning och vissa typer av förleder eller efterleder i sammansättningar krävs likaså exaktare kodning. Sådana frågor är t.ex. följande: (a) Vilka sällfångstermer förekommer i dialekten på Bergö? (svar t.ex. *krabb* s.2, bet. 2, *kryp-stång*) eller (b) Vilka sammansättningar finns med adjektivisk förled som slutar på *-og*? (svar t.ex.: *hårog-buse*, *hårog-harpa*, *kålog-bisse*, *kåtog-nackad*).

Då artiklar som ingår i sökresultatet visas borde det vara möjligt att tekniskt komprimera i artikelfönstret för en översikt av betydelserna. Också för SMS har en komprimeringsfunktion utarbetats. Artiklarna i SMS visas antingen i sin helhet eller komprimerade så att endast dispositionen, dvs. momenten och de inledande förklaringarna, syns eller i en vy där exemplen uteblir.

## 5. Sammanfattning

Stora vetenskapliga ordböcker som är fritt tillgängliga på webben är ett steg i utvecklingen mot ett mer demokratiskt kunskapssamhälle. En webbordbok har i jämförelse med en tryckt ordbok en helt annan potential att nå nya användargrupper. Det ligger dock ett drygt arbete bakom en webbutgivning.

Då man planerar en helt ny elektronisk ordbok, kan man börja med att planera vilka sökningar som ska vara möjliga och sedan fatta beslut om databasens innehåll och kodning enligt önskemålen. Då man digitaliserar en existerande ordbok, utgår man från det innehåll som finns, tänker ut vad som kunde vara önskvärt att söka i detta innehåll och försöker sedan åstadkomma en kodning som gör alla de önskade sökningarna möjliga. Detta kan vara en stor utmaning, då ordboken inte ursprungligen planerades och byggdes upp med tanke på en webbutgivning.

Webbordboken planeras så att det blir möjligt att genom sökningar och relevanta begränsningar och sökfilter få fram det centrala i ordbokens innehåll. För FO:s del är exponeringen av variationen i de svenska dialektområdena i Finland viktig. Ordboken innehåller en mångfald av dialektala uttalsformer och belysande exempelfraser; ett digert dialektmaterial som redaktionen för FO vill göra tillgängligt för alla användare.

## LITTERATUR OCH KÄLLOR:

- Ahlbäck, Olav, 1976: Inledning. Ordbok över Finlands svenska folkmål, utg. av Forskningscentralen för de inhemska språken. Helsingfors. S. VII–XXVIII.
- Cantell, Ilse, 2010: Ordbok över karelskan på Internet. I: *LexicoNordica* 17. S. 277–288.
- FO = Ordbok över Finlands svenska folkmål, utg. av Forskningscentralen för de inhemska språken, 1976 ff. Helsingfors.
- Laureys, Godelieve, 2012: Mot en hållbar lexikografi. En balansgång mellan återanvändning och förnyelse. I: *Nordiska studier i lexikografi* 11. (Skrifter utgivna av Nordiska föreningen för lexikografi 12.) Lund
- Pastor, Verónica och Alcina, Amparo, 2010: Search Techniques in Electronic Dictionaries: a Classification for translators. I: *International Journal of Lexicography*, Vol. 23 No. 3. S. 307–354.
- Rietz, Johan Ernst, 1862–1867: *Svenskt dialektlexikon*. Ordbok öfver svenska allmogespråket. Lund.
- SAOB = Ordbok över svenska språket, utg. av Svenska Akademien, 1893 ff. Lund.
- Slotte, Peter, 1981: Ordbok över Finlands svenska folkmål – problem och metoder. I: *De finlandssvenska dialekterna i forskning och funktion*, red. Bengt Loman. Åbo. S. 45–64.
- SMS = Suomen murteiden sanakirja, utg. av Forskningscentralen för de inhemska språken, 1985 ff. Helsingfors.
- Vendell, Herman, 1904–1907: Ordbok över de östsvenska dialekterna. Helsingfors.
- Wessman, V. E. V., 1925–1932: *Samling av ord ur östsvenska folkmål*. Tillägg till H. Vendells Ordbok över de östsvenska dialekterna. Helsingfors.
- Vikør, Lars S., 1999: Fleirgenerasjonsordbøker og tida. I: *Nordiska studier i lexikografi* 4. Rapport från Konferensen om lexikografi i Norden Esbo 21–24 maj 1997, red. Peter Slotte et al. Helsingfors. S. 395–405.

## Internethänvisningar

- Aineistopalvelu Kaino (Materialbanken Kaino): <http://kaino.kotus.fi/> (september 2011)
- EDD Online: [http://www.uibk.ac.at/anglistik/projects/speed/startseite\\_edd\\_online.html](http://www.uibk.ac.at/anglistik/projects/speed/startseite_edd_online.html) (april 2011)
- Karjalan kielen verkkosanakirja (Webbordbok över karelskan): [http://kaino.kotus.fi/cgi-bin/kks/kks\\_etusivu.cgi](http://kaino.kotus.fi/cgi-bin/kks/kks_etusivu.cgi) (september 2011)

Suomen murteiden verkkosanakirja (Webbordbok över finska dialekter): <http://freax.kotus.fi/sms/?p=main> (april 2012)

*Ilse Cantell*

Ordbok över Finlands svenska folkmål  
ilse.cantell@kotus.fi

*Caroline Sandström*

Ordbok över Finlands svenska folkmål  
caroline.sandstrom@kotus.fi

# Maintaining consistency of monolingual verb entries with interannotator agreement

*Silvie Cinková, Lenka Smejkalová, Anna Vernerová,  
Jonáš Thál & Martin Holub*

We study the verb *throw* from the point of the inter-annotator agreement in the Word Sense Disambiguation (WSD) discipline. The most frequently used verbs are often polysemous. This poses a challenge when creating manually annotated data for machine learning, since polysemy threatens inter-annotator agreement. We argue that the classical WSD setup (selecting just one matching sense from a list) is inadequate for semantically complex verbs.

*Keywords:* Corpus Pattern Analysis, interannotator agreement, English verbs, vagueness, polysemy, word senses

## 1. Introduction

This study<sup>1</sup> approaches grouping readings (lexical units, senses) from the perspective of Natural Language Processing (NLP). A classical task in NLP is *Word Sense Disambiguation (WSD)*. Automatic WSD is mostly performed with a method of (statistical) *machine learning*. Machine learning is used in complex tasks for which humans are not able to create explicit rules. The computer learns to mimic human judgment from a *gold standard data* set with examples of

---

<sup>1</sup> This work has been using language resources developed and/or stored and/or distributed by the LINDAT-Clarin project of the Ministry of Education of the Czech Republic (project LM2010013). It has been supported by the Czech Science Foundation projects P103/12/G084 and P406/2010/0875 as well as by the Czech Ministry of Education project MŠMT ČR LC536 and EuroMatrixPlus (FP7-IST-5-034434-IP, 7E09003). We also thank Patrick W. Hanks, Pavel Rychlý and Adam Rambousek for a comprehensive CPA guidance, the permission to experiment on a copy of PDEV, as well as for the technical assistance we have been receiving.



good solutions of the given problem. It is supposed to apply the same strategies on unknown data, so that humans perceive its decisions as acceptable.

The gold standard data for WSD usually consists of a text corpus that is interlinked with a machine-readable lexicon. Originally, digitized dictionaries for human users were used. Later on, specialized lexical resources for NLP were developed (e.g. Miller 1995 and Fellbaum (ed.) 1998, Ruppenhofer et al. 2002, Palmer et al. 2005). Yet despite all efforts put into WSD in the last two decades, the issue still remains unsolved.

The typical setup of a WSD manual annotation task is providing human annotators with a list of senses for each word. Each annotator is allowed to choose exactly one sense per word occurrence. *Interannotator agreement* (IAA) is measured several times or continuously during the task by having several annotators process the same data. Manual annotation, as well as IAA measurement, is a standard method in computational linguistics.

Semantic tasks turn out to be difficult for humans. The annotators tend to disagree with each other and thus IAA drops in semantic tasks. This is an important issue, since the computer will never make better human-like judgments than humans themselves do. If annotators in a WSD task provide each a different answer for a task, the computer has no chance to learn any strategy for assigning word senses either. Finding an adequate way to create gold standard data is therefore vital for a satisfactory WSD or any other method of semantic interpretation.

## 2. The different types of polysemy

WSD works with the tacit assumption that word senses are mutually exclusive; i.e. that the words are ambiguous. We will argue that the classical WSD setup is bound to fail with words that denote vague concepts and show that the relation between two intuitively distinct readings of a verb is quite often that of vagueness rather than that of ambiguity, depending on the context. The following sections will explain the notions *ambiguity* and *vagueness* (Van Deemter 2010).

### 2.1 Ambiguity

When a word or an expression is *ambiguous*, it denotes several concepts that are cognitively distant. When using a clearly ambiguous word such as the noun *spring*, the speaker can impossibly mean more than one of its readings at the same time (*source of water, season, coil*) or “something in between” or even “not care which the recipient chooses, all being equally adequate for conveying the

message”. If in this case the recipient is unsure about the correct interpretation, it is the speaker’s fault.

## 2.2 Vagueness

Opposite of ambiguity is *vagueness* (van Deemter 2010). A concept is *vague*, when it allows for borderline cases. For instance, there is a general understanding of the adjective *tall*, but, confronted with a row of men of different height, people will disagree on individual men whether they are tall or not. In the language, a word denoting a concept is vague when there are no clear borders between its readings and there are sensible contexts in which the speaker can mean both at the same time, or something in between, or where this information may remain underspecified without compromising the message.

Let us have the noun *glass* as an example of a vague concept. MEDO (Rundell et al. 2002) gives it three contemporary readings: hard clear substance, small container for drinking and (a summarizing term for) attractive artifacts made of glass (*a collection of Italian glass*).

These readings have many common features: fragility, smooth surface, translucency, etc. and differ mainly in accentuating either the raw material or the (number of) artifacts made of it. In many cases, an underspecifying context can be perceived as ambiguous, since it would matter whether the speaker means the raw material or the artifact(s). Despite the three readings being intuitively easy to keep apart, many everyday contexts operate with a much coarser granularity of the semantic grouping, and this is where we can speak about vagueness. For instance, *glass polisher* is meant for glass in any form, and neither the speaker nor the recipient would care whether the material or artifacts made of it are meant. Cruse (1986) calls the relation among different readings of a word based on emphasizing one aspect of the more or less common denotate *semantic modulation*. For the purpose of this paper, we regard a concept as a *vague concept* when it is expressed by a set of mutually modulated readings of one word. The actual uses of that word in contexts where more readings can be considered can be regarded as *vague contexts*.

## 3. Verbs as vague concepts

Dividing verbs into readings is generally considered to be more controversial than doing the same with nouns, which we claim has to do with the inherent vagueness of most verbs. While nouns typically denote entities (event nouns disregarded), verbs typically denote events that entities undergo, as well as rela-

tions between entities. It is astonishing that our languages contain so many fewer verbs than nouns. In addition, their distribution is not even. Out of over 6,000 verb types in the English BNC corpus, the first most frequent 1,000 verbs cover more than 90% of all verb occurrences in the corpus. We apparently need only 1,000 words to describe almost all kinds of events and relations between all thinkable entities! It will be argued that this enormous descriptive power makes verbs, at least the most frequent lexical verbs, vague concepts.

If different people were to create the same verb entry independently of each other by making semantic clusters from random concordances, the number of clusters and the distribution of the concordances would be individual. Our intuitive criterion is namely the perceived similarity of the respective events, which is individual.

The similarity perception is partly affected by the morphosyntactic relations between the verb and its immediate context: e.g. any other instance of *throw* will be perceived as different from *throw up* (i.e. *vomit*), due to the particle alone. When the morphosyntactic realization of a predicate is the same, the inherent semantic characteristics of the participating entities play an important role in judging to what extent two events denoted by the same verb are similar: when *a boy throws stones*, is it the same event as when *somebody is thrown into the air and carried down the road by the motorbike or by an explosion*? Some similarity judgments are likely to be shared across the language community (*horse throwing a man* is different from *dawn throwing sunlight*), but some will inevitably be clustered differently by different speakers (and even differently by the same speaker at different times), depending on the granularity with which they regard the respective concordances as instances of *throwing*.

The difference in plausible contextual synonyms of *throw* between the respective clusters shows that the clusters reflect our perception of them as different sorts of events:

*A boy was throwing/hurling/tossing/blowing stones.*

*Fast driving on gravel roads throws up rocks which can scar the car. /My wheels spit gravel and I long for a bigger windshield.*

*... fragmental material thrown into the air by explosive volcanic activity. and The volcano was throwing/spewing stones and lava.*

*Dawn threw/cast sunlight across the ruins of the old city.*

To show the semantic differences between different cases of *throwing*, the contextual synonyms must have more collocational restrictions than the original verb. E.g. *spew* has more specific requirements on the agent as well as on the patient than *throw* has. The “spewer” must act as a sort of container and have a mouth or a mouth-like aperture, out of which the material is forced. Almost

anything in the world can *throw* almost anything, but only a subset of them *spew* when they *throw*.

Verbs with collocational restrictions, such as *spew*, explicitly express those semantic features of their arguments that are relevant to the event they denote. Using them, the speaker explicitly shares his world knowledge with the recipient. The use of frequent lexical verbs like *throw*, on the contrary, seems to shift the entire burden of world knowledge onto the recipient. For instance, to decode the message of the sentence *He threw the bread to the birds*, the recipient has to know that the implicit but probably the most *relevant* message concerns feeding the birds or discarding the bread rather than the explicitly mentioned propelling of bread crumbs, unlike e.g. *throwing darts*, although the hand-propelling motion is present in both. There is no need to decide which it is, because the relation between these readings is not ambiguity, but vagueness. In this way, using a semantically complex (vague) verb enables us to conflate a number of different aspects of the same event into one predicate, while the relevance of the different readings is assessed on the basis of the world knowledge (by both the speaker and the recipient).

Whenever the relations between readings in verb entries are those of vagueness rather than ambiguity, then the classical setup of the word-sense disambiguation task is completely out of place and the problem must be approached in a different way.

## 4. Vagueness of *throw* in semantic definitions and in the data

### 4.1 *throw* in OntoNotes

The inventory of wordsenses in the English OntoNotes 4.0 (Weischedel et al. 2010), an extensive linguistic resource, has arisen by merging the very granular WordNet senses (difficult for annotation) to achieve the right granularity with which two annotators reach a 90% agreement. The senses are:

1. Propel or hurl something/someone with force; toss or put forward an idea or gesture.
2. Discard, dispose, expel, or get rid of.
3. Manipulate or move something in order to operate.
4. Cast, emit, or radiate (including metaphoric expressions as *cast light*, *understanding*, etc.)
5. Confuse or bewilder
6. Hold an event (including *throw a fit*)
7. Form or shape, as pottery

8. Lose, as a game, intentionally
9. Give up, quit (including *throw in*, *throw in the towel*)
10. throw up: Vomit
11. throw in: Add as an extra or gratuity
12. None of the above

The verb *throw* has 11 senses (plus one “trash bin”) now and the interannotator agreement 79%, still below the desired limit. Further merging can be expected in the next release.

Some senses are easy to match to concordances, because they are distinguished by particles (*throw in*, *throw up*), and they are also semantically distinct from the physical throwing as propelling an object with force or acting as the propelling force. Senses 8 and 9 are similar (both mean a sort of *giving in*), but one is realized by a particle verb or an easily identifiable idiom (*towel*), while the other combines almost exclusively with synonyms of *match*, *game* and *elections*. So does Sense 6 (with the typical collocate being *party*). An additional annotator convention specifies that *fit* (i.e. also *tantrum* and *wobbly*) also belongs here, since the definition alone does not imply it. Sense 3 is also easy to identify, as the list of its typical collocates is short and homogeneous: *switch*, *handle*, *engine*, *car*, *back gear*. Sense 5 is associated with the preposition *into* and a synonym of *confusion* or *disarray*, or the idiom *out/off kilter* – also quite a homogeneous group. Sense 7 (*forming pottery*) is unmistakable in real contexts, too. So is 10 (vomiting).

The official release does unfortunately not show the disagreeing annotator judgments but only the adjudicated result with an indication that two annotators disagreed. Nevertheless, three senses stand out as the most likely causes of annotator disagreements: 1, 2 and 4. They are associated with semantically very heterogeneous events, as the annotated corpus concordances show.

Sense 1 encompasses very manifold events, for instance:

- The delinquents threw a brick through the school window.*  
*He threw sixes on both die.*  
*She threw herself enthusiastically into the project.*  
*The earthquake threw them onto the floor.*  
*He threw in a couple of wisecracks during her speech.*  
*She hangs out while the boys fight and throw money around.*

Sense 4 (cast, emit or radiate) is more homogeneously determined by the semantic features of the patient, which is supposed to be a liquid, vapor, light or sound.

In the example sentences of Sense 2 in the lexicon (discard, dispose, expel, or get rid of, including *throw away* and *throw off*), all occurrences of *throw* combine with the particles *away*, *out* or *off*, except the idiomatic *throw something into the wind*, but the particle is not prescribed as obligatory. This suggests that e.g. *throw the paper straight into the bin* can be both classified as 1 and as 2, causing a disagreement, since both the respective senses are broad and often combine.

Surprisingly, the huge OntoNotes 4.0 corpora contain only 91 sense-annotated occurrences of *throw*.<sup>2</sup> Of these 91 sentences, 29 double-annotated occurrences contained disagreements to be adjudicated. The most frequent adjudications are in Senses 1 (15) and 2 (9). The other readings had only marginal adjudications (Sense 4 had two. Senses 5, 6, and 10 had one each). Senses 3, 7, 8, 9 and 11 did not occur in the annotated data at all. We have therefore no idea what agreement the annotators would reach on them.

We have browsed the 24 occurrences of *throw* in OntoNotes 4.0 that were adjudicated to 1 or 2. We were not able to see whether the disagreement was caused by an annotator error or whether the lexicon entry was adjusted according to this annotation and the annotation was not revised afterwards. The patient was in most cases money, in the sense of wasting or spending it in a hope to solve a problem (4 cases). Two cases denoted discarding something (but bare *throw* was used without any particle), and the other two denoted adding to consideration. The annotators also disagreed on throwing stones etc. as missiles (2 cases). The remaining 17 occurrences were probably disagreements between 1 and 2, finally adjudicated to either 1 or 2.

We can only guess that the commonest disagreement combinations were 1, 2 and 2, 1 in Senses 1 and 2, but even so we can observe that Senses 1 and 2 are problematic for annotation, even though they are the most generic and also most frequently matched senses in the list.

Brown (Brown et al. 2010) performed IAA experiments with WordNet vs. OntoNotes sense-annotated data. They showed that IAA is harmed by the high granularity of senses rather than by the sheer number of senses (when they are not very fine-grained). Our observation does not falsify the experimental outcome; nevertheless, it suggests that most problems within the coarse-grained annotation arise in the coarsest senses anyway.

---

<sup>2</sup> However, the low number is understandable, given that a word is massively annotated only when ITA has reached 90% on a 50-sentence random sample and the latest ITA measured on *throw* was only 79% (cf. Weischedel et al. 2010:13).

## 4.2 *throw* according to Corpus Pattern Analysis

Corpus Pattern Analysis (CPA) is a corpus-supported manual method of creating dictionary entries for verbs (Hanks, in press, Hanks and Pustejovsky 2005.) The lexicographer takes a random corpus sample of 250-1000 concordances and sorts them into meaningful clusters. In addition, he considers the automatic collocation analysis provided by the Sketch Engine (Kilgarriff et al., 2004). Then he describes the prototypical morphosyntactic structure of each cluster and the prototypical lexical realization of relevant collocates. This part of a dictionary reading is called *pattern*. He also makes a paraphrase of the pattern in form of a finite clause (*implicature*). Only frequent cases get a pattern of their own. Rare and unintelligible cases are marked as unclassifiable. Borderline cases are associated to patterns, when possible, but have additional markup (*exploitations*). Many marked borderline cases in a pattern result in pattern revision (mostly in the creation of an additional pattern).

We have used the *throw*-entry in Hanks' Pattern Dictionary of English Verbs (PDEV, <http://corpora.fi.muni.cz/cpa/>), which had a promising 1000-concordance sample annotated by the lexicographer. It turned out, however, that the sample was not finished. Therefore we finished the annotation on our own<sup>3</sup>. The annotation led to an addition of new patterns and to a slight revision of Hanks' original patterns. The 1000 concordances yielded about 70 patterns in total. Most patterns were naturally phraseological units as *throw the baby out with the bathwater*, *throw a spanner into the works*, and *throw the towel/sponge in*, as well as *throw up*.

No effort was put into making the implicatures mutually exclusive. Some of our implicatures were clearly special cases of others. E.g. a very distinct pattern of *dead bodies being thrown* would fit into the less specific *garbage discarding* pattern. The granularity of the implicatures of the respective patterns was based entirely on the data. With the instruction to associate a concordance to the most specific cluster possible, each concordance had to match exactly one implicature – or none.

The original pattern inventory also contained a pattern for propelling physical objects towards a target and a pattern for discarding physical objects (spending assets and aiming gestures/words etc. had already been separated). The (single) annotation of the 1000 concordances uncovered a fuzzy border between propelling and discarding, which is likely to have caused disagreements

---

<sup>3</sup> A sample of 30 PDEV verbs was IAA-tested and revised. It is publicly available as VPS-30-En at <http://hdl.handle.net/11858/00-097C-0000-0005-BF95-B> in the LINDAT-CLARIN repository.

in the OntoNotes 4.0 annotation. Nevertheless, one can still think of many contexts in which the discarding implicature is clearly the dominant one. Therefore the self-standing *discard*-pattern for physical objects was kept and just extended with more lexical suggestions (*in the bin, in the garbage, etc.*) to encourage annotators to rely on their world knowledge, and the *target* condition was removed from the *propel* pattern. Another pattern was created for throwing physical objects at a target as *missiles*. Restructuring the patterns this way brought more homogeneous groups of concordances.

### 4.3 A case study of *throw* in a vague context

However, minute pattern splitting is a burden in WSD tasks, as Brown (Brown et al. 2010) has demonstrated. Any new batch of corpus data to analyze is bound to throw up concordances that do not quite match any current pattern, which in turn is bound to threaten the IAA.

The following example will show that, no matter the granularity of dictionary readings, frequent lexical verbs tend to denote several aspects of an event at once. Finding and encoding their typical combinations is very difficult and probably domain-, culture- and application dependent (cf. Kilgarriff 1997), such that petrifying them into hard-wired mutually exclusive readings would be counter-productive.

Let us have a look at a BNC sentence that describes an event by conflating several world-knowledge-based implicatures into the verb *throw*:

*Osbern has his father killed by a lowly mob and thrown to birds and wild animals.*

Five of our revised patterns are relevant for this particular event:

1. Human uses *hands* to *propel* a physical *object* in a direction for a short distance  
*Ex.: Tourists are encouraged to throw coins into the fountains for good fortune.*
2. Human violently *pushes* or shoves or kicks another *human* so that the other human loses control over his movements and *falls*  
*Ex.: He threw her to the ground/against the cupboard...*
3. Human *discards* or gets rid of an artifact or *stuff*  
*Ex.: He threw the paper straight into the bin/threw it away, threw it out.*
4. Human (*murderer*) disposes of or *discards* or *hides* the *body* of his *victim* to some place  
*Ex.: Their corpses were thrown down a well.*
5. Human *feeds an animal/animals* with a physical object or a substance  
*Ex.: It was like throwing a piece of meat to sharks.*



Now, if the annotator had to pick one reading only, she would have to decide which throwing is the most relevant one. Interannotator disagreement is inevitable here, as everyone can interpret the nuances of a text differently.

## Discussion

Liberman (2009: 2) claims that:

[...] linguistic annotation tasks are not really classification problems, but rather translation problems. We don't normally assume that there is only one correct translation of a Chinese sentence into English; nor do we try to make this true by constructing elaborate translation guidelines to cover every relevant contingency, though in principle we could.

Recent experiments (Rumshisky et al. 2009, Erk 2010) as well as empirical observations of experienced lexicographers (Krishnamurthy and Nicholls 2000, Hampton, 2007) suggest that WSD in its classical setup is extremely difficult for humans. On the other hand, Rumshisky showed in a 2009 experiment with the Mechanical Turk (Rumshisky et al. 2009) that even linguistic non-experts reach a good agreement on deciding whether two occurrences of a word (verb) are used in the same sense. In addition, their groupings turned out to be quite similar to a sense grouping made by a professional lexicographer, only coarser. Erk (Erk 2010) reported a similar result with concordances compared to WordNet (Miller 1995 and Fellbaum 1998) senses. Wilks and Ide (Wilks and Ide 2006) question the usefulness of WSD being performed as a separate task in NLP applications in general.

## Conclusion

Experiments with the traditional as well as the more recent approaches to human WSD annotation prove that humans are unable to pick “the right” sense from a predefined list to decode a word in context. This has two implications:

- 1) “picking the right sense” from a lexicon is not exactly the way in which humans use dictionaries to look up an unfamiliar word, although we clearly make use of dictionaries when decoding texts with unfamiliar words;
- 2) if the computer is supposed to mimic the human way of using dictionaries for text analysis, the task of word sense disambiguation (or text understanding in general, cf. Wilks and Ide 2007) has to be constituted in a different way.

## REFERENCES

- Brown, Susan et al., 2010: Number or Nuance: Which Factors Restrict Reliable Word Sense Annotation? In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valetta, Malta.
- Cruse, D.A., 1986: *Lexical Semantics*. Cambridge University Press.
- Davies, Mark, 2008: *The Corpus of Contemporary American English: 425 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.
- Van Deemter, Kees, 2010: *Not exactly. In praise of vagueness*. Oxford University Press.
- Erk, Katrin, 2010: What Is Word Meaning, Really? (And How Can Distributional Models Help Us Describe It?) In: Proceedings of the 2010 Workshop on Geometrical Models of Natural Language Semantics, ACL 2010, pp. 17–26, Uppsala, Sweden
- Fellbaum, Christiane (ed.), 1998: *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Hampton, J. A., 2007: Typicality, graded membership, and vagueness. *Cognitive Science*, 31, pp. 355–384.
- Hanks, Patrick Wyndham. in press. *Lexical Analysis. Norms and Exploitations*. MIT Press.
- Hanks, Patrick, Pustejovsky, James, 2005: A Pattern Dictionary for Natural Language Processing. In: *Revue française de linguistique appliquée* 10 (2).
- Kilgarriff, Adam, Rychly, Pavel, Smrz, Pavel, Tugwell, David, 2004: The Sketch Engine. In *EURALEX 2004, Lorient, France*; Pp 105-116. <http://www.sketchengine.co.uk>
- Kilgarriff, Adam, 1997: "I Don't Believe in Word Senses". *Computers and the Humanities*, 13 (2).
- Krishnamurthy, R., Nicholls, D., 2000: Peeling an onion: the lexicographers' experience of manual sense-tagging. *Computers and the Humanities*, 34(1–2).
- Liberman, Mark, 2009: The Annotation Conundrum. In: Proceedings of the EACL 2009 Workshop on the Interaction between Linguistics and Computational Linguistics: Virtuous, Vicious or Vacuous? Athens, Greece.
- Miller, George A., 1995: WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38:11, pp. 39–41.
- Palmer, M., Kingsbury, P., Gildea, D., 2005: The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics* 31:1, pp. 71–106.
- Rumshisky, Anna et al., 2009: The Holy Grail of Sense Definition: Creating a Sense-Disambiguated Corpus from Scratch. In: *Fifth International Workshop on Generative Approaches to the Lexicon (GL 2009)*. Pisa, Italy.
- Rundell, M. et al., 2002: *Macmillan English Dictionary for Advanced Learners. International Student Edition*. MacMillan Education. Oxford. (online access: [http://online.macmillandictionary.com/mc\\_au2/macmil.htm](http://online.macmillandictionary.com/mc_au2/macmil.htm))
- Ruppenhofer, Josef, Collin F. Baker and Charles J. Fillmore, 2002: The FrameNet Database and Software Tools. In: Braasch, Anna and Claus Povlsen (eds.), *Proceedings of the Tenth Euralex International Congress*. Copenhagen, Denmark. Vol. I, pp. 371–375.

Weischedel, A. et al., 2010: OntoNotes Release 4.0. Linguistic Data Consortium, Philadelphia

Wilks, Yorick, Ide, Nancy, 2007: Making Sense about Sense. Word Sense Disambiguation, Algorithms and Applications (pp. 47–73). Springer.

*Silvie Cinková*

Charles University in Prague, Faculty of Mathematics and Physics - Institute of Formal and Applied Linguistics, Malostranské nám. 25, 118 00 Praha 1, Czech Republic.

cinkova@ufal.mff.cuni.cz

*Lenka Smejkalová*

Charles University in Prague.

smejkalova@ufal.mff.cuni.cz

*Anna Vernerová*

Charles University in Prague.

vernerova@ufal.mff.cuni.cz

*Jonáš Thál*

jonlatash@gmail.com

*Martin Holub*

Charles University in Prague.

holub@ufal.mff.cuni.cz

# Ellips – ett beskrivningsproblem i en historisk ordbok

*Birgit Eaker*

We often use ellipsis in spoken and written language. We shorten and leave out where the context will show what we mean. In time, elliptical uses may become established and form new words.

A common type of shortening in Swedish is the exclusion of one of the elements of a compound word so that only one of the elements remains, the first element (*spark* from *sparkstötting*) or the last (*konvalj* from *liljekonvalj*).

In a historical dictionary, like the SAOB, ellipsis can sometimes cause problems. Compounds are presented under the word which forms the first element, the headword. However, the elliptical form of a compound may well be identical to one of its original elements, and it can sometimes be difficult to separate the ellipsis from the element. Also, new compounds can be created with an elliptical form as one of the elements, and in practice it may be impossible to know whether a compound is a primary formation or contains an ellipsis. Furthermore, it can be difficult to judge whether a particular meaning of a simplex form is the result of an ellipsis of a compound, or whether it is a case of change of meaning in the original word.

To follow a logical or chronological model can consequently meet with difficulties when ellipsis is involved. In my paper I will discuss the presentation of ellipsis in the SAOB.

NYCKELORD: lexikografi, ellips, etymologi, betydelser

## Inledning

Ordet *ellips* har fler än en betydelse och jag vill inledningsvis redogöra för hur det används här. Själva ordet kommer från latinets *ellipsis*, av grekiska *ἐλλειψις*, av ett verb med betydelsen 'utelämnas, fattas'. Inom språkvetenskapen används *ellips* dels för handlingen att bilda en ofullständig fras eller ett ofullständigt ord genom att man utelämnar ett led i frasen eller ordet och dels för den förkortade fras- eller ordbildningen som blir resultatet. I denna artikel ska jag diskutera problem som kan uppstå för lexikografen till följd av elliptisk ordbildning uti-

från en undersökning av hur ellips har behandlats i SAOB. Min utgångspunkt är den som redaktör i SAOB.

Syftet med min undersökning var att se hur ellips har behandlats i SAOB, om behandlingen har förändrats över tid och i vilken mån elliptisk ordbildning har kommenterats i artiklarna. För att få svar på dessa frågor gjorde jag en sökning i webversionen av SAOB på *ellip\** och fick 1897 träffar enligt sökning i frekvens (3 januari 2011), härunder orden *ellips*, *elliptisk*, *elliptiskt* osv. Bland dessa finns naturligtvis några där ordet förekommer i språkprov som illustrerar någon annan företeelse (t.ex. *Cometernes verkliga rörelser sker i ellipser*) men de sorterades lätt bort. För att också få en uppfattning om i vilken mån ellips faktiskt kommenteras i ordboken har jag slumpmässigt slagit upp ett antal ord som kan anses vara elliptiska bildningar. Jag har också gjort kompletterande websökningar på bl. a. *kortord\** och *kortform\**. Jag ska inte gå närmare in på själva undersökningen här men använder den som underlag för diskussion.

Ellips är ett vanligt förekommande fenomen i vårt språk. Ja, vi kan knappast föra ett samtal utan att använda ellips. Vi utesluter ständigt sådant som är självklart i sammanhanget. Men termen *ellips* förutsätter att det finns – eller åtminstone någon gång har funnits – en föreställning om en språklig helhet, det gäller både tillfällig ellips i samtalsituationen och när det är fråga om ordbildning.

Förkortning av ord kan ske på olika sätt. Man kan behålla sista delen av ett ord, som i *bil* av *automobil* och *buss* av *omnibus*, eller första delen som i *bio* av *biograf* och *leg* av *legitimation*. En dubbel sammansättning kan förkortas genom ellips av mellanledet som i *livbolag* av *livsförsäkringsbolag*. Sådana former kallas ibland klammerformer. En annan typ är teleskopord, t.ex. *motell* av *motorhotell*, *genrep* av *generalrepetition*, där det ofta är fråga om kontraktion av flera led. Man kan nöja sig med förledets initial som i *ubåt* för *undervattensbåt* eller kombinera initial och kortform som i *kpist* för *kulsprutepistol* eller bara använda vissa initialer eller bokstäver som *tv* för *television* eller *televisionsapparat* eller *SAOB* för ”Svenska Akademiens ordbok”. Den sista typen kallas ofta initialord. Man kan också bilda kortformer genom att både ta bort och lägga till. Med ändelsen *-is* bildar vi kortformer som *dagis* av *daghem* och *skådis* av *skådespelare* och med *-are* kortformer som *deckare* av *detektiv* eller *detektivroman*. I SAOB saknas ibland förväntade kortformer beroende på att de vid de aktuella bandens redigering inte var etablerade i språket.

Kortformerna är inte alltid inhemska bildningar. Förkortningen av *automobil* till *bil* skedde först i danskan. *Omnibus* har förkortats till *buss* efter påverkan från engelskans kortform *bus* och *ubåt* har vi efter tyskans *U-Boot*, kortform av *Unterseeboot*.

Vissa handböcker använder termen *stympning* för ordbildningar som *bil* och *buss* och begränsar termen *ellips* till att gälla sammansatta ord där ena ledet

stryks (Thorell 1981: 159). Andra använder genomgående termen stympning i stället för ellips när det är fråga om ordbildning (Söderbergh 1968: 177) eller växlar mellan termerna (Wessén 1992: 21, 22). Termen stympning används inte i SAOB numera men har använts ett fåtal gånger i början av alfabetet.

Vad vi än kallar fenomenen är det otvivelaktigt så att vi förkortar där det av sammanhanget klart framgår vad vi menar och utesluter led i både satser och ord. Så småningom kan sådana förkortningar bli lexikaliserade och bilda nya ord och uttryck.

Den typ av kortform som jag vill diskutera närmare i den här artikeln är den vanligaste och den som vanligen kallas ellips i SAOB, nämligen när ett av leden i en sammansättning har uteslutits.

Först bara några ord om SAOB. Det är en historisk ordbok med uppgift att beskriva det svenska språket och dess utveckling från ca 1521 fram till våra dagar. Som flera andra historiska ordböcker följer vi – så gott det nu går – en logisk-kronologisk beskrivningsmodell. Den är kronologisk så till vida att vi börjar vår beskrivning med den betydelse som vi anser vara närmast ordets ursprungliga betydelse och därefter försöker beskriva betydelsernas utveckling i logisk (eller genealogisk) ordning. Som flera andra ordböcker av den här typen redovisar vi sammansättningar under sitt huvudord, dvs. under det ord som utgör första sammansättningsledet och som vi menar att sammansättningen är bildad till.

## Tillfällig ellips av sammansättningsled

I satssammanhang är det mycket vanligt att första ledet i en sammansättning utesluts. Ofta är det då fråga om konkreta föremål. Några exempel:

En stigbygel används av ryttare som stöd för foten vid uppsittning och ridning. Ordet *stigbygel* är bildat efter tyskans *steigbügel*, som är en sammansättning av *steigen*, 'stiga' och *bügel*, 'bygel'. Båda leden har alltså direkta motsvarigheter i svenskan. När det klart framgår av sammanhanget vad vi menar kan vi utesluta det första ledet. Några rader ur en reklamtext: *Jag är helnöjd med flex-ride stigbygel... När jag först såg bygel så ...* [osv.]

Det vanliga redskapet för virkning har en liten krok i änden som man fångar upp garnet med. Den svenska benämningen är *virknål*, en sammansättning av *virka* och *nål*. När sammanhanget är klart räcker det med *nål*. Ur en virkbeskrivning: *För in virknålen i maskan längst till vänster och om garnet som dras igenom till 2 m på nålen.*

*Taktpinne* är benämning på den stav eller pinne en dirigent använder för att markera takt och dirigera med. Ordet är en sammansättning av *takt* och *pinne*.

Om sammanhanget är klart kan det räcka med *pinne*. *Stycket var slut, dirigenten la ner pinnen.*

I de här tre exemplen är ursprungsordet sannolikt en sammansättning och användning av enbart efterledet en ellips.

Men det är inte alltid givet att det är på det viset. Vi kan också för att förtydliga lägga till ett extra led till ett ord. Tänk er att vi sitter vid en falnande brasa. Någon säger. *Lägg på en pinne till. Ta en björkpinne – gran sprakar så mycket.* Här är det inte *björkpinne* som är ursprunget, utan *pinne* är grundordet i sin egentliga betydelse och *björk* är ett förtydligande led.

Vid redigeringen av ordböcker utgår vi vanligen från korpus av något slag, och vare sig de är digitala eller utgörs av handskrivna lappar så är de utsnitt eller excerpter ur språket. Redaktören får inte alltid så stora utsnitt som skulle behövas för att kunna avgöra om en användning är elliptisk. I SAOB kan detta uttryckas som i BYGEL: ”1) stycke metall l. läder o. d. som bildar en hel l. half ring o. d. samt tjänar till handtag l. fäste l. skydd l. stöd o. d. för ngt; *ofta elliptiskt för ngn ssg, ss. STIG-, VAR-BYGEL m. fl.*” [min kursivering]. Eller som i definitionen av PINNE: ”1) stavformigt, jämförelsevis smalt o. kort föremål av trä ... (*ofta i l. elliptiskt för ssg med förled som anger l. antyder användningen ...*)” [min kursivering].

För att vi ska förstå när orden *bygel*, *nål* och *pinne* betyder ’stigbygel’, ’virknål’ respektive ’taktpinne’ behövs ett sammanhang där betydelsen klart framgår. Sådana användningar kan kallas kontextberoende. I allmänhet anser man inte att sådana användningar måste redovisas i ordböcker. Men om det är fråga om att själva sammansättningsledet har fått en speciell betydelse, som t.ex. *nål* i *virknål* brukar den redovisas i SAOB. I artikeln NÅL, moment 1, finns därför ett undermoment: ”e) om nål med krökt spets, använd vid ... virkning o. d. ...; jfr ... VIRK-NÅL m. fl.” Här har redaktören inte explicit uttryckt att användningen är elliptisk.

## Lexikaliserad ellips

I svenskan finns också många ord där förkortningen har blivit permanent, ellipserna har lexikaliserats.

Ordet *liljekonvalj* eller *liljekonvalje* är bildat efter nylatinets *lilium convalium*, eg. ’dalarnas lilja’. Växten kan också benämnas *konvalj* eller *konvalje*. Processen är densamma som för *stigbygel* > *bygel*, men här har ellipsen resulterat i en etablerad eller lexikaliserad kortform, i det här fallet t.o.m. två variantformer, varav den utan *-e* är vanligare numera. Till skillnad från ordet *bygel* är ordet *konvalj* helt entydigt och kan bara användas om blomman liljekonvalj (*Convallaria*

majalis L.), medan *bygel* kan ha flera betydelser. Både *konvalj* och *liljekonvalj* är egna uppslagsord i moderna ordböcker, t.ex. i Svenska Akademiens ordlista (SAOL) och Svensk ordbok (SO). I SAOB har *konvalj(e)*, däremot behandlats som sidouppslagsformer till *liljekonvalje*, vilket snarast visar att kortformen vid tryckningen av *konvalj(e)* 1936 ännu inte hade nått den status formen har idag.

*Sparkstöttingen* är sannolikt en nordisk uppfinning. Ordet är en sammansättning av verbet *sparka* och *stötting*, en mindre kälke. Dessa enkla men praktiska fortskaffningsmedel för vinterföre glider på två medar. Man står på den ena meden och skjuter ifrån med den andra foten, sparkar sig fram. I SAOB redovisas ordet bland sammansättningarna under verbet *sparka*. Nuförtiden är nog den elliptiska formen *spark* minst lika vanlig. Här är det alltså efterledet som har fallit bort och förledet som har bibehållits. I det här fallet ser ellipsen likadan ut som substantivet *spark*, sparkande, stöt med foten. Ellipsen har resulterat i att vi har fått två ord *spark* med olika etymologi, ett som är bildat till verbet *sparka* och ett som är kortform av *sparkstötting*. Med den logisk-kronologiska beskrivningsmodellen som utgångspunkt skiljer SAOB på homonymer och har följaktligen redovisat två SPARK, dels sbst.<sup>1</sup>, sparkande, stöt med foten, dels sbst.<sup>2</sup>, sparkstötting.

Jag minns också från min barndom att mina föräldrar ibland kunde säga *stötting*. *Ta stöttingen, Birgit*. Det är givetvis också en ellips, av samma typ som *nål* för *virknål*. Den har upptagits i SAOB under ordet STÖTTING, sbst.<sup>3</sup>, som betydelse 2: ”om sparkstötting; särsk. i l. elliptiskt för ssgn SPARK-STÖTTING.”

Systembolaget, det statliga aktiebolag som i Sverige handhar försäljning av alkoholhaltiga drycker, kan få tjäna som ytterligare ett exempel på att båda leden i en ursprunglig sammansättning kan användas för en och samma sak. Ordet *systembolag* är en sammansättning av substantiven *system* och (*aktie*)*bolag*. Det används i två betydelser, dels om aktiebolaget, dels om butik som tillhör aktiebolaget. Sammansättningen är upptagen under sitt huvudord, SYSTEM, sbst.<sup>1</sup>, och där förd till betydelse 5 ”om sammanfattningen av allt som hör till viss företeelse l. verksamhet l. visst företag o. d. ...”

Både *bolaget* och *systemet* är lexikaliserade ellipser för *systembolaget*. Valet av kortform kan vara en fråga om ålder. *Systemet* är sannolikt vanligare bland yngre. Regional eller social variation kan kanske också spela in. Det anses möjligen lite finare att säga *systemet* än att säga *bolaget*. Mot det antagandet kan dock anföras att systembolagets tidskrift heter *Bolaget*. Orden används oftast i bestämd form men obestämd form förekommer, som i denna fråga i ett shoppingcenter: *Finns det något system här?*



## Diskussion

I en ordbok som försöker följa en logisk-kronologisk modell skapar ellipser ibland problem. Innan vi kan få ellipser som *bygel* av stigbygel, *pinne* av takt-pinne, *spark* av sparkstötting osv. måste ju sammansättningarna *stigbygel*, *takt-pinne* och *sparkstötting* ha funnits. Ellipserna kan inte existera innan sammansättningen – eller åtminstone en föreställning om den – gör det.

När det är första ledet i sammansättningen som har fallit bort (som hos *bygel* av *stigbygel*) är redovisningen egentligen inte så problematisk, även om den inte är helt logisk. I SAOB tas den typen av ellips vanligen upp i det huvudord som utgör efterledet (se t.ex. BYGEL 1, PINNE 1, STÖTTING, sbst.<sup>3</sup> 2).

Men det är också mycket vanligt att efterledet i en sammansättning utesluts så att bara förledet blir kvar, som hos *spark* och *system*. På liknande sätt har *läsk* bildats av läskedryck, *kyl* av kylskåp, *frys* av frysbox eller frysskåp, *polo* av polotröja, *knyt* av knytkalas, osv. Liksom i fråga om ellips av förled är grundordet inte alltid självklart. Den elliptiska ordformen har ofta flera betydelser och kan ha uppkommit ur olika sammansättningar. Ordet *flyg*, bildat genom ellips, var nytt när det beskrevs i SAOB på följande sätt: 1) elliptiskt för FLYG-VÅSEN, resp. -VAPEN. 2) elliptiskt för FLYG-MASKIN l. dyl. Första belägg på betydelserna är nästan samtida, 1924 respektive 1925.

Denna typ av ellips kan vara krångligare att redovisa. I SAOB anses det inte att betydelsen 'systembolag' ska redovisas i det ursprungliga huvudordet SYSTEM, sbst.<sup>1</sup>. Betydelsen 'systembolag' kommer ju från sammansättningen *systembolag* där ledet *bolag* har fallit bort. Om betydelsen 'systembolag' skulle redovisas i ordet SYSTEM, sbst.<sup>1</sup>, skulle sammansättningen som utgör ursprunget till betydelsen placeras under sin avkomma, dvs. ursprunget skulle komma efter resultatet. Det blir en logisk och kronologisk kullerbytta. I SAOB tas därför den elliptiska formen upp som ett eget huvudord, SYSTEM, sbst.<sup>2</sup>, systembolag; systembolagsbutik. (Jfr SPARK ovan.)

SAOB har varit under arbete under en mycket lång period och min undersökning visar att behandlingen av ellips inte är helt konsekvent genom tiden. Ett exempel på A- och ett på T- kan få illustrera detta.

*Astrakan* är benämning på flera sorters sommaräpplen, särskilt vit respektive röd astrakan. Ordet är upptaget som ASTRAKAN, sbst.<sup>2</sup> (sbst.<sup>1</sup> är benämning på ett slags skinn eller tyg) i den del av SAOB som trycktes 1902. Etymologin i SAOB är fantasieggande. För det första menar man att ordet är elliptiskt för sammansättningen *astrakanäpple*, och det är säkerligen på det viset. Man antar vidare att ellipsen har bildats i svenskan och spritt sig vidare härifrån. Det är också troligt. I grunden ligger ett ortnamn, den ryska staden Astrachan (förr Astrakan) nära Kaspiska havet, och man förmedlar i SAOB ett antagande att

äpplena har spritts via kärnor från svenska krigsfångar från södra Ryssland på Karl den tolfte tid. Det verkar något fantastiskt. Men hur det än är med den saken är ordet elliptiskt, eftersom Astrakan är ett ortnamn och vi använder det om äpplen – det är inte en rimlig eller logisk betydelseutveckling. Kronologin stöder antagandet om ellipsis. Sammansättningen är tidigare belagd än simplex, år 1762 respektive 1794. Ännu tidigare omtalar Linné ”Astracanske Äpplen” (Hortus upsaliensis, s. 130, 1748, SAOBs) och ”klara Astracaniska äpplen” (Sk. resa 1751, s. X, SAOBs). Sannolikt är det tack vare Linné som benämningarna har spritts. Jag kan i sammanhanget inte låta bli att citera ett språkprov ur OED från 1822: “The Red Astracan Apple, imported from Sweden, but now growing here, is one of the very best early apples.” (*London Jrnal. Arts & Sci.* 4 47.)

Ordet *astrakan* om äpplen kan också ha sammansättningar och *astrakan-äpple* är en fullt möjlig sammansättning med betydelsen ’äpple av sorten astrakan’. *Det där astrakanäpplet var verkligen gott.* Ordformen *astrakanäpple* kan alltså antingen vara ursprunget till den elliptiska bildningen *astrakan*, eller en ny sammansättning, ett äpple av sorten astrakan. Vad gör en redaktör som försöker följa en logisk-kronologisk modell med den här sammansättningen? I SAOB behandlas inte ortnamn, om de inte har fått appellativ status. Det hade annars varit möjligt att redovisa ordet både som sammansättning till ortnamnet Astrakan och som sammansättning till den elliptiska bildningen *astrakan*. Men hur ska lexikografen kunna veta var gränsen går – om ens användaren vet det? I det här fallet har redaktören valt att sätta sammansättningen *astrakanäpple* under den elliptiska bildningen *astrakan* och komplettera med en etymologisk upplysning: ”ssgn eg. bildad till ortnamnet Astrakan”.

Numera skulle en redaktör i SAOB sannolikt i stället låta *astrakanäpple* få bli ett eget huvudord. Så har man gjort på senare tid i liknande fall. En parallell kan vara *tsetse* i betydelsen ’tsetseflugor’. Ordet *tsetse* är en ellipsis av *tsetseflugor* och i SAOB har både ellipsen *tsetse* och den ursprungliga sammansättningen *tsetseflugor* redovisats som egna huvudord.

Men mycket ofta är lexikografens verklighet inte så tydlig. I samtal och texter kortas konkreta sammansättningar ofta när betydelsen framgår av sammanhanget, och t.o.m. etablerade ellipser kan kompletteras till sammansättningsform efter behag eller för större tydlighet. Ofta kan redaktören helt enkelt inte avgöra om ett ord är en ellipsis av en sammansättning eller en särskild betydelseutveckling hos ett simplex.

Ta t.ex. ordet *tvilling* om en kristall som består av två med varandra symmetriskt sammanväxta kristaller. Samma betydelse har sammansättningen *tvillingkristall*. Vad är ursprungligare – ordet *tvilling* i en bildlig användning om en sådan typ av kristall eller sammansättningen *tvillingkristall*? Redaktören har ansett det troligast att simplex i bildlig användning är det primära och har redo-

visat betydelsen i ordet *tvilling*, som moment 2 b. Ordet *tvillingkristall* redovisas på ett sätt som anger att denna sammansättning är bildad till *tvilling* i bet. 2 b. Mycket talar för att redaktören har gjort en riktig bedömning, såväl kronologin, dvs. tiden för första belägg, som kristallens utseende och i det här fallet också källan. Första belägg på *tvilling* i betydelse 2 b är från 1829, medan första belägg på *tvillingkristall* är något senare, från 1836. De tidigaste språkproven är excerpter ur den banbrytande forskaren Jöns Jakob Berzelius skrifter och det är väl troligt att sammansättningen har bildats genom att *kristall* har lagts till som förklarande efterled. Men redaktören har också pekat på möjligheten att betydelse 2 b hos *tvilling* kan ha uppkommit genom ellips och har satt den möjligheten inom parentes: ”om (l. elliptiskt för) tvillingkristall”.

Här blir det lite komplicerat. Om *tvilling* i den här betydelsen faktiskt är en ellips av ordet *tvillingkristall* har vi ju att göra med en ny ordbildning som egentligen är senare än den sammansättning som den kommer från (även om vi inte råkar ha skriftligt belägg för det) och därför inte kan presenteras före den – dvs. om man ska följa en strikt logisk-kronologisk modell.

Men som verkligheten ser ut för oss ordboksredaktörer kan vi inte följa en idealisk beskrivningsmodell till punkt och pricka. Det gäller inte minst när det är fråga om att avgöra om en viss betydelse hos ett simplex har utvecklats inom ordet eller beror på ellips av ett sammansättningsled, dvs. om den är resultat av betydelseutveckling eller ordbildning. Därtill är vår kännedom om ordens tillkommelse och de sammanhang där de har använts alltför begränsad. Vår strävan ska givetvis vara att reda ut förhållandena så långt som det är möjligt med en rimlig arbetsinsats. Men ibland kan vi helt enkelt inte avgöra hur utvecklingen har varit. Frågan blir då om det alltid är så viktigt att följa modellen? Vi vet ju att datorer kräver strikt logisk ordning för att kunna processa information. Men vi människor är inte på samma sätt beroende av att allt är konsekvent och logiskt ordnat. Vi kan förstå och uppfatta att det kan finnas oklarheter och olika tolkningar. Jag tror för min del att våra läsare uppskattar att vi talar om att bilden inte alltid är kristallklar och redovisar olika möjligheter, t.ex. påpekar när en viss betydelse skulle kunna vara resultat av ellips.

## KÄLLOR OCH LITTERATUR

OED = Oxford English Dictionary. Webversionen: <http://www.oed.com> (september 2011)

SAOB = Ordbok över svenska språket, utg. av Svenska Akademien, 1893 ff. Lund.  
Webversion av SAOB. <http://g3.spraakdata.gu.se/saob/> (januari 2011)

SAOBs = Samlingar till Ordbok över svenska språket. Lund.

SAOL = Svenska Akademiens ordlista över svenska språket. 13 uppl. Stockholm 2006.

- SO = Svensk ordbok utgiven av Svenska Akademien. Stockholm 2009.  
Söderbergh, Ragnhild, 1968: Svensk ordbildning. Skrifter utgivna av Nämnden för svensk språkvård 34. Stockholm.  
Thorell, Olof, 1981: Svensk ordbildningslära. Stockholm.  
Wessén, Elias, 1992, Svensk språkhistoria 2. Ordbildningslära. Ny utg. av 5 uppl. 1971. Nytryck i Nordiska Språk. Stockholm.

*Birgit Eaker*

Svenska Akademiens ordboksredaktion, Lund  
birgit.eaker@svenskaakademien.se

# Fra alfabet til begrep: Bokmålsordboka og NorNet

*Ruth Vatvedt Fjeld, Rune Lain Knudsen & Julie M. Torjusen*

This article discusses the use of a monolingual dictionary (*Bokmålsordboka*) as basis for a wordnet for Norwegian (NorNet). The first part of the article describes the method for constructing the wordnet, and discusses the structure of NorNet with regard to synsets, unique beginners and top ontology. The second part of the article discusses some of the problems that arise due to inconsistencies in the dictionaries, and how these are solved in NorNet. The selection of entries for the wordnet is also discussed. Finally, the article describes plans and future developments. ction of entries to the wordnet is also discussed. Finally, the article accounts for plans and future developments.

NØKKEWORD: ordnett, merge-tilnærming, ontologi

## 1. Ordnett generert fra en ordbok – og ordbok rettet ved hjelp av et generert ordnett

Et ordnett er en datamaskinleselig beskrivelse av de semantiske relasjonene mellom et utvalg ord. Et tidlig moderne ordnett er Princeton WordNet (PWN) for amerikansk-engelsk, seinere er EuroWordNet (EWN) utviklet, der mange europeiske språk er koblet sammen via engelsk, og det er et mål at også det danske DanNet skal gå inn i EWN. I Sverige holder man på å utvikle ordnettet Swesaurus, i Finland har man oversatt PWN til finsk, og på Island er man også i gang med å lage et ordnett. Denne artikkelen skal handle om arbeidet med å utvikle et norsk ordnett kalt NorNet. Ordnettet utvikles ved Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo, avdeling for bokmålsleksikografi.

I Nygaards hovedoppgave *Frå ordbok til ordnett* (Nygaard 2006, Fjeld & Nygaard 2009) og ble semantiske relasjoner mellom substantiver automatisk generert ved hjelp av enkle regler som analyserte definisjonene i *Bokmålsordboka* (BOB) (Wangensteen 2005). Tabell 1 viser noen eksempler.

Regel	Om definisjonen består av ett enkeltstående substantiv eller flere enkeltstående substantiv i lemmaform adskilt med komma, er disse substantivene synonym til definiendum.
Eksempel	<b>vidde</b> [...] <u>område</u> , <u>areal</u>
Resultat	<b>vidde</b> SYNONYM_FOR <b>område</b> <b>vidde</b> SYNONYM_FOR <b>areal</b>
Regel	Dersom definisjonen har andre elementer enn enkeltstående substantiv, er det første substantivet i lemmaform overbegrep (hyperonym) til definiendum, med mindre dette substantivet står i en stoppliste (en liste med ord som ikke skal være underbegreper (hyponym)).
Eksempel	<b>bistro</b> m1 (fr 'vertshus(holder)') liten <u>restaurant</u>
Resultat	<b>restaurant</b> HYPERNYM_FOR <b>bistro</b>

Tabell 1: Eksempler fra det opprinnelige regelsettet og informasjonen som blir trukket ved hjelp av dette. Ord som fanges opp av regelen og får tildelt relasjoner er understreket.

NorNet ble automatisk generert i 2008, med utgangspunkt i et sett av slike regler som i Tabell 1 og definisjonene i BOB. Foreløpig inneholder ordnettet 37787 substantiv, 15516 overbegreps-/underbegrepsrelasjoner og 7172 synonymirelasjoner. Siden ordnettet er automatisk generert, inneholder det en del feil og mangler, og gjennomgår derfor en grundig redigeringsprosess. Når det arbeidet er ferdig, vil vi ha et godt grunnlag for videre utvidelse av ordnettet, og vi kan begynne å teste ut nye metoder for automatisk uttrekning av semantiske relasjoner. Vi vil også få et håndfast mål på hvor godt reglene som genererte ordnettet, hentet ut de ønskede relasjonene fra substantivsdefinisjoner.

Alle lemmaene i ordnettet er knyttet til ordboka gjennom et felles ID-nummer. Redigeringsarbeidet gir også som resultat at vi kan forbedre ordboka. Mange av feilene som forekommer i det uredigerte ordnettet, skyldes inkonsekvenser i definisjonene i BOB. Ordboka er redigert med semasiologisk tilnærming og alfabetet som sorteringsprinsipp. Resultatene fra redigeringa av NorNet vil dermed kunne brukes til å gjøre ordboka mer konsekvent og systematisk. Dette kommer vi tilbake til i punkt 5. Arbeidet med redigeringa har også medført en rekke observasjoner av hvordan algoritmene og implementasjonen av dem har fungert.

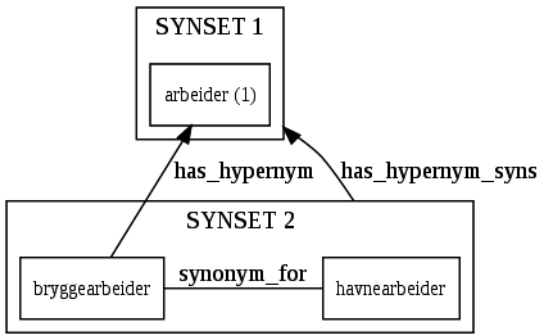
Foreløpig er ca. 25 % av under-/overbegreps- og synonymirelasjonene mellom delbetydningene rettet. Rettinga har i hovedsak blitt utført av Julie M. Torjusen. I første omgang har hun systematisk kontrollert den automatiske tilde-  
linga av relasjoner mellom delbetydninger som ble konstruert ved hjelp av det opprinnelige regelsettet og slettet ukorrekte relasjoner som programmet lagde. En vanlig feil er at reglene ikke har skilt mellom homonymer, slik at for eksem-

pel adjektivet *kort* har blitt tolka som substantivet *kort*. Det har ført til at ord som *andakt* med definisjonen ”kort tale med utgangspunkt i en bibeltekst” har fått *kort* i betydninga ”lite rektangulært blad av papir, papp el. plast som brukes til forskjellige bestemte formål” som overbegrep. Til nå har det blitt utført i alt 3846 rettinger. 1788 av disse er korrigeringer av genererte relasjoner, de resterende 2058 rettingene er innsetting av manglende relasjoner. Dette arbeidet har vist at samsvaret mellom de automatisk genererte relasjonene og den manuelle redigeringa er på under 70 %, altså at godt over en fjerdedel av de automatiske uttrekkingene må rettes på et eller annet nivå. I tillegg indikerer antall innsettinger av manglende relasjoner en noe lav dekningsgrad. Anslaget for samsvar kan virke noe lavt, men siden det gjenstår ca. 75 % som vi ikke har undersøkt, er dette foreløpig ikke godt nok grunnlag for å vurdere algoritmens kvalitet. Det er grunn til å tro at resten av de automatisk genererte relasjonene vil ha omtrent like mange feil, men vi kan ikke si noe sikkert om dette før vi har gransket en større del av materialet.

## 2. Synsetta i NorNet

Synonyme delbetydninger har en symmetrisk relasjon SYNONYM\_FOR mellom seg (se figur 1). Den samlede mengden av synonymer for et begrep utgjør ett synsett. En delbetydning kan ha overbegrepsrelasjon til en annen delbetydning, slik at den ene delbetydninga er overbegrep for den andre. Dette danner grunnlaget for to lag med overbegrepsrelasjoner: mellom delbetydninger isolert og mellom synsettene de er en del av. Overbegrepsrelasjonene mellom synsetta vil utledes fra overbegrepsrelasjonene funnet for de enkelte delbetydningene i et synsett. Overbegrepsrelasjonen for et synsett vil velges ut ifra hvilken delbetydning som oftest er angitt som overbegrep for de synonyme delbetydningene. En delbetydning som ikke har fått tildelt noen egne overbegrepsrelasjoner, vil dermed kunne få dette ved hjelp av synsettet den er en del av. Denne formen for synsettklassifisering vil vise nyanser mellom synonymer som får ulike overbegreper, samt kartlegge inkonsistens og konflikter i overbegreps hierarkiene. Et eksempel på overbegrepsrelasjoner mellom delbetydninger i forhold til overbegrepsrelasjoner mellom synsett er gitt i figur 1.

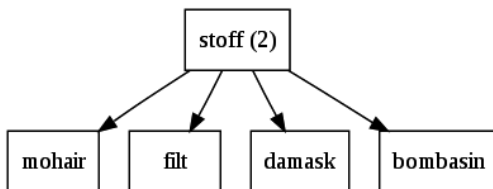
I synsett 2 er det bare *bryggearbeider* som har fått tildelt overbegrepsrelasjon. Basert på dette defineres synsett 1 som overbegrep for synsett 2. Underbegrepet *havnearbeider* i synsett 2 blir inkludert i overbegreps hierarkiet fordi den er del av samme synsett som *bryggearbeider*.



Figur 1: Eksempel på relasjoner mellom delbetydninger i forhold til relasjoner mellom synsets.

### 3. Unike begynnere

Unike begynnere er betegnelsen for de lemmaene som står øverst i ordnettets overbegreps-hierarki, dvs. lemma som ikke selv har noen overbegreper (Miller 1998:28). Dette er illustrert i figur 2. NorNet har til nå i alt 22887 substantiv som ikke er tildelt noen overbegreper i ordnettets. Av disse begrepene har 20810 heller ikke noe underbegreper, altså delbetydninger som ikke har fått tildelt overbegreps-/underbegrepsrelasjoner og dermed er helt isolerte i forhold til ordnettets. Disse klassifiseres ikke som unike begynnere før de har fått tildelt én eller flere relasjoner. Av disse har imidlertid 5043 fått synonymirelasjoner og får derfor enten overbegrep ved hjelp av ett eller flere synsett-hierarkier, eller de blir en del av et synsett som er klassifisert som en unik begyner.



Figur 2: Eksempel på en unik begyner. Enkelte underbegreper er utelatt på grunn av plass-hensyn.

Når vi trekker fra antall lemmaer som verken har overbegreps- eller underbegrepsrelasjoner individuelt, står vi igjen med 2077 unike begynnere med ett eller flere underbegreper.



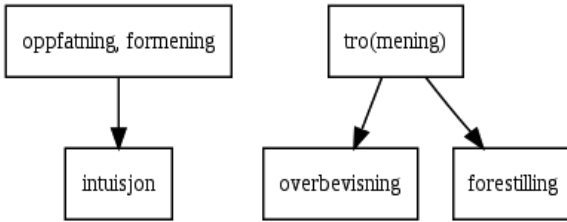
Lemmaer med høyest antall overbegreper er listet opp i Tabell 2.

Hyp.fr	Lemma.
1114	person
76	grunnstoff
72	stoff (1)
64	lyd
56	område
53	tid, ord
44	måte
42	stoff (2)
41	bevegelse, gjenstand
38	tilstand
36	myntenhet
34	enhet
33	musikkstykke
31	rekke
30	handling

Tabell 2: Liste over unike begynnere med 30 eller flere underbegreper. De med likt antall underbegreper står på samme rad adskilt med komma

Her ser vi at lemmaet *person* skiller seg ut med hele 1114 underbegrepsrelasjoner. Dette ser ut til å samsvare med en undersøkelse av samme lemma i DanNet (Pedersen og Braasch 2009), der 4246 lemmaer har *person* som nærmeste overbegrep. Siden underbegrepsfrekvensen til *person* er såpass høy, og siden begrepet har vært gjenstand for studier i DanNet, er personbetegnelser et semantisk felt vi ønsker å fokusere ekstra på, særlig for å kunne sammenligne med beskrivelsen i DanNet. Blant annet vil vi undersøke i hvor stor grad vi kan redusere antall umiddelbare underbegreper ved å innordne dem i et mer detaljert hierarki.

Isolerte begrepsøyer er klynger av begreper som er knyttet sammen ved hjelp av semantiske relasjoner seg imellom, men som mangler relasjoner som knytter de opp mot resten av ordnettet, jf. figur 3.

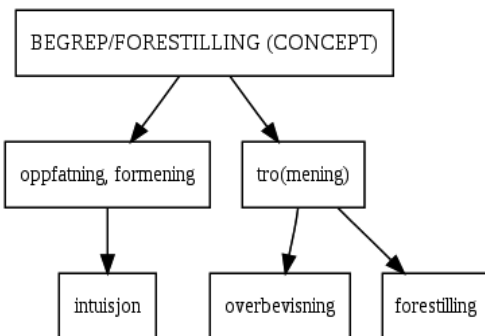


Figur 3: Eksempel på isolerte begrepsøyer/unike begynnere

Dette gjør disse klyngene usynlige i mange søk, siden stien av relasjoner fra ett begrep til neste blir avbrutt. Arbeidet med å redusere antall unike begynnere ved å få dem ordnet i et bedre hierarki er derfor av høy prioritet. Et lavt antall unike begynnere er en god indikasjon på et lavt antall isolerte begrepsøyer, og dermed et “tettere” nett av begreper som gir mer komplett informasjon ved søk i ordnettet.

#### 4. Toppontologi

En toppontologi for ordnett forsøker å beskrive overbegreper som samler så mange underbegreper som mulig under seg i en hierarkisk struktur. Å bygge opp en dekkende og veldefinert toppontologi er viktig for å redusere antall isolerte begrepsøyer i ordnettet. EWN har utviklet en toppontologi som er basert på sammenligninger av de ulike toppontologiene i alle EWN-ordnetta. Vi mener at denne toppontologien er et godt utgangspunkt for å strukturere NorNet. Med en slik felles toppontologi blir det enklere å slå sammen, sammenligne og kommunisere mellom ordnett for forskjellige språk. Vi har derfor tatt utgangspunkt i EWNs toppontologi og bygger en kjerneontologi for NorNet som er mest mulig i samsvar med den, jf. figur 4.

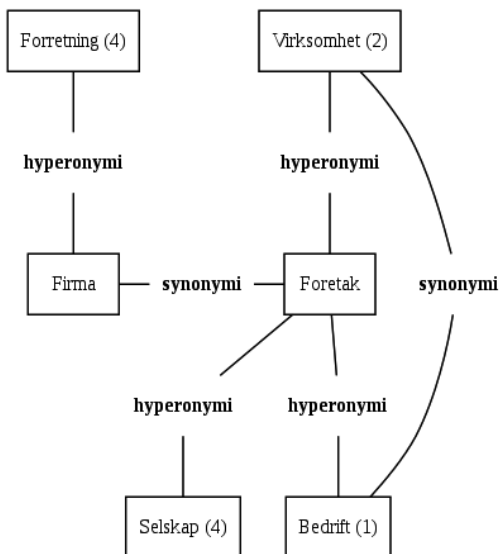


Figur 4: Eksempelen fra figur 3 knyttet sammen ved hjelp av et begrep fra en toppontologi.

## 5. Redigeringsarbeidet

Som nevnt i punkt 1 er NorNet basert på BOB. Det er en enspråklig allmennordbok som først og fremst er ment for menneskelige brukere, og det har ført til enkelte problemer når definisjonene i BOB har blitt lest av et dataprogram. I dette punktet skal vi gå gjennom noen av disse problemene og hvordan vi har prøvd å løse dem. Vi skal også sammenlikne problemene vi har hatt i arbeidet med NorNet med problemer som har dukka opp i arbeidet med DanNet (Pedersen m.fl 2009). Det gjelder inkonsekvens i valg av overbegrep, ettordsdefinisjoner som ikke er synonymmer og ad hoc-sammensetninger. Vi skal også diskutere lemmaseleksjon i ordnettet (jf. punkt 6).

Et av problemene når ordnett konstrueres ut fra definisjonene i en ordbok, er at ord som tilhører samme semantiske felt, ofte kan ha ulike overbegreper i definisjonene. Dette ser vi f.eks. ved synsettet {firma; selskap (4); bedrift (1); virksomhet (2); forretning (4); foretak}<sup>1</sup>. Ifølge definisjonene i BOB er *firma* underbegrep til *forretning* og synonymt med *foretak*, mens *selskap* er underbegrep til *foretak*. Det samme gjelder *bedrift*, men *bedrift* er i tillegg synonymt med *virksomhet*. *Virksomhet* er i tillegg overbegrep til *foretak* (se figur 5).



Figur 5: Synsettet {firma; selskap (4); bedrift (1); virksomhet (2); forretning (4); foretak}

<sup>1</sup> Vi bruker notasjonen {} for synsett (Miller 1998).

Selv om en slik betydningsbeskrivelse kan gjenspeile vanlige språkbrukeres forståelse og bruk av disse ordene og dermed fungere godt i en tradisjonell ordbok, er dette er en kaotisk løsning for et ordnett. Ved liknende tilfeller i DanNet har man valgt å slå sammen alle begrepene i ett synsett (Pedersen m.fl 2009:276). En slik løsning har vi også valgt for NorNet. Dermed vil f.eks. et søk på *firma* også gi treff på de andre lemmaene i synsettet.

En annen type inkonsekvens i definisjonene i BOB finner vi ved ettordsdefinisjoner, som ifølge vårt regelsett blir tolket som synonymdefinisjoner. Men i en del tilfeller viser det seg at det ikke stemmer, som i betydning 3 av *bue*, jf. figur 6:

<u>boge</u>	I bue el. I boge <u>m1</u> (da. <i>bue</i> , norr <i>bogi</i> , besl med <i>III bøye</i> )
<u>bue</u>	
	1 krumning, kurve <i>sirkelb-</i> , <i>ellipseb-</i>
	2 hvelving i byggverk <i>rundb-</i> , <i>spissb-</i> I <i>romanske og gotiske b-r</i>
	3 skytevåpen <i>pil og b-</i> / overf. <i>Spenne b-n for høyt sikte</i> for høyt, overvurdere seg selv
	4 del av strykeinstrument <i>fiolinb-</i>

Figur 6: Definisjon og betydningsinndeling av bue/boge i Bokmålsordboka

Logisk sett er *bue* underbegrep til skytevåpen, ikke synonym. Dette må rettes manuelt i NorNet. I denne retteprosessen oppdager vi altså en del oppslag i ordboka som bør endres og gjøres mer konsistente. I dette eksempelet ser vi at betydning 4 viser en meronymi-relasjon ("del av strykeinstrument"). Det burde betydning 3 også gjøre, for en bue er ikke et skytevåpen uten en pil. Viss denne definisjonen blir endra til "del av skytevåpen", vil den ikke lenger være en ettordsdefinisjon og ifølge vårt regelsett ikke et synonym.

BOB inneholder mange ad hoc-sammensetninger, dvs. sammensetninger som brukes som overbegreper i definisjonene, men som ikke er egne oppslagsord i ordboka. I og med at disse sammensetningene ikke fins i ordboka som egne lemmaer, har ikke programmet funnet noe overbegrep for dem. Dette ser vi i definisjonene til lemmaene *biedermeier*, *buttskjøt* og *bøffelgress*, jf. figur 7.

<b>biedermeier</b>	<b>biedermeier</b> <i>-en</i> (utt bidermeier el. <i>-maier</i> ; ty. 'spissborger', omdannet av <i>Biedermann</i> og <i>Bummelmaier</i> , to ty. besteborgertyper) møbelstil fra første halvdel av 188-tallet
<b>buttskjøt</b>	<b>buttskjøt</b> sveiseskjøt mellom jernplater som ligger butt i butt (I, 1)
<b>bøffelgras</b>	<i>-gras</i> el. <i>-gress</i> småvokst præriegress som lager tette tuer el. <i>Matter</i> , <i>Buchlöe</i>
<b>bøffleggess</b>	<i>dactyloides</i>

Figur 7: Eksempler på lemmaer uten overbegrep

Verken *møbelstil*, *sveiseskjøt* eller *præriegress* er egne oppslagsord i BOB. Dette skaper et hierarkisk problem, fordi vi får unike begynnere som ikke hører til høyt oppe i hierarkiet. Dette kan løses ved å gi lemmaene overbegrep etter det siste leddet i sammensetninga, som for de to siste eksemplene er identiske med sisteleddet i lemmaet. Ved *biedermeier* blir *stil* overbegrep, og dermed får man ikke noe mellomnivå mellom *biedermeier* og *stil*, men man slipper å få *biedermeier* som unik begyner.

Lorentzen og Nimb drøfter ad hoc-sammensetninger i ordbøker som danner basis for ordnett i sin artikkel om forholdet mellom tradisjonelle ordbøker og ordnett. De foreslår to mulige løsninger: å opprette sammensetningene som egne begreper i ordnettet, eller å erstatte dem med ord som allerede fins i ordnettet (Lorentzen og Nimb 2010:332). Når det er snakk om ad hoc-sammensetninger, er det trulig mindre hensiktsmessig å opprette dem som egne begreper i et ordnett. De står ikke på egne bein i samme grad som ord som har gjennomgått en utvelgingsprosess for lemmastatus i en ordbok. Dette kan eventuelt undersøkes ved hjelp av korpus.

## 6. Lemmaseleksjon

Lemmaseleksjonen i et ordnett kan gi litt andre konsekvenser enn lemmaseleksjonen i en vanlig ordbok fordi det er en viss forskjell på lemmaenes synlighet i dem. I en ordbok (særlig elektronisk) er det mange ord man sjelden vil se i bruk, med mindre man søker spesifikt etter dem. I et ordnett vil man derimot få vist mange ord man ikke har søkt etter, fordi de står som over- eller underbegreper til et ord man har søkt etter. I NorNet er lemmaseleksjonen basert på definisjonene av substantivene i BOB, og ofte er det ikke nærmeste overbegrep som står først i definisjonen, men et overbegrep lenger oppe i det ontologiske hierarkiet. Det fører i en del tilfeller til at enkelte lemmaer får veldig mange underbegreper (jf. tabell 2). Leksikografisk Bokmålskorpus (LBK) er et balansert, lingvistisk annotert korpus for moderne bokmål, som er en viktig ressurs for lemmaselek-

sjonen til NorNet. Et mye brukt overbegrep i BOB er for eksempel *evne* (1), som bare under bokstaven a har underbegrepene *absorberingsevne*, *absorpsjonsevne*, *abstraksjonsevne*, *akselerasjonsevne*, *anskuelsesevne*, *arbeidsevne* og *avdråttsevne*. Mange av disse lemmaene er marginale i faktisk språkbruk, ved søk i LBK finner man kun ett treff på *absorberingsevne*, ett treff på *absorpsjonsevne*, fire treff på *abstraksjonsevne*, to treff på *akselerasjonsevne* og tolv treff på *arbeidsevne*. De øvrige lemmaene får man ingen treff på. Hvorvidt så sjeldne ord skal tas opp som lemmaer i et ordnett, kommer an på hva ordnettet skal brukes til. Ved oppbygging av et nytt ordnett, mener vi det er viktigere å ta med mer sentrale deler av ordtilfanget først, så ord med svært lav eller ingen forekomst i LBK blir merket for undertrykkelse ved visning, men lemmatiseres i NorNet som kandidater for et fulldekkende ordnett i framtida. På denne måten bruker vi LBK som datagrunnlag for en statistisk analyse av delbetydningene for å anslå NorNets dekningsgrad og relevans for moderne bokmål.

## 7. Metodevalg

Eksempelene i punkt 5 viser at det er umulig å lage et ordnett med utgangspunkt i en ordbok uten å bruke tid på manuell retting. Et alternativ kan være å oversette et ordnett laget for et annet språk, noe som er ansett som en raskere og billigere metode, og som samtidig gjør at ordnett for forskjellige språk blir kompatible (Pedersen m.fl. 2009:271). Men ved å oversette fremmedspråklige ordnett risikerer man at strukturer som fins i målspåket går tapt, fordi de ikke fins i kildepråket. Viss vi skulle oversatt DanNet til norsk, ville det bl.a. bli problemer med de mange sammensetningene i norsk og dansk som har forskjellig bruk og betydning. I DanNet er *brusebad* et *bad*, mens en *dusj* er en *sprut* i NorNet, og *killling* er et *pattedyr* i DanNet, mens en *kattunge* er en *unge* (1). I NorNet kan *flom* bety en stor mengde, som i *tåreflom* og *ordflom*, men denne betydninga fins ikke i DanNets søkefunksjon [andredord.dk](http://andredord.dk). Det viser seg altså ved sammensetninger at dansk og norsk har forskjellige begrepsstrukturer, som man ikke vil få fram ved å oversette det danske ordnettet til norsk. Man må belage seg på mye manuelt ekstraarbeid også om man velger å oversette ordnett. Basert på observasjonene vi selv har gjort fra arbeidet med NorNet og hva andre som har laget ordnett har rapportert, mener vi at behovet for etterarbeid er høyt uansett hvilken metode man velger. Vi har derfor valgt å fokusere på metodeutvikling, da vi mener at dette vil være verdifullt ikke bare for utviklingen av NorNet, men også for utvikling av relaterte leksikalske ressurser.

## 8. Videreutvikling av NorNet

Planen fremover er å inkludere flere ordklasser og relasjonstyper, både mellom delbetydninger internt innenfor hver ordklasse og der avledninger og sammensetninger under samme begrepsfelt koples sammen på tvers av ordklassene. Vi vil eksperimentere med forskjellige teknikker for uttrekking av semantisk informasjon, både fra ordboksdata og naturlig tekst, og dokumentere resultatene. Eksperimentene vil basere seg på regelbaserte metoder, statistiske modeller og kombinasjoner av disse.

Definisjonsspråket i tradisjonelle ordbøker er dels formelt, dels naturlig, og det er derfor interessant å analysere – spesielt med regelbaserte metoder. Ved å forsøke å utvikle en generell metode som i stor grad kan automatiseres, vil det åpne for nye måter å hente ut informasjon fra liknende, strengt strukturerte teksttyper på, også om de i utgangspunktet ikke ble laget med maskinlesbarhet i tankene. På bakgrunn av disse observasjonene er analyse og klassifisering av definisjonene i Bokmålsordboka under utvikling. Denne analysen baserer seg på ordklassemønstrene for definisjonene og vil kunne si noe om konsistens ved hjelp av statistikk over unike eller sjeldne definisjonsmønstre, avvik fra de vanligste mønstrene mm. Resultatet av denne analysen danner også grunnlaget for utvikling av flere regelbaserte metoder for uttrekking av semantisk informasjon som vil bli brukt i NorNet.

## LITTERATUR

- Andre ord, ved Anders Johanssen. <http://andreord.dk/>
- BOB = Wangensteen, Boye, 2005: Bokmålsordboka, Kunnskapsforlaget, Oslo
- Fjeld, Ruth Vatvedt & Lars Nygaard (2009): NorNet – a monolingual wordnet of modern Norwegian. I: B.S Pedersen, A. Braasch, S. Nimb & R. V. Fjeld (red.): Proceedings of the NODALIDA 2009 workshop, NEALT, s. 13–16
- LBK=Leksikografisk Bokmålskorpus (<http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/index.html>)
- Lorentzen, Henrik & Sanni Nimb, 2010: Fra ordbog til wordnet. Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse? I: H. Lönnroth og K. Nikula (red.): Nordiske studier i leksikografi 10, s. 329–344
- Miller, George A. 1998: Nouns in WordNet. I: C. Fellbaum (red.): WordNet – An electronic database, MIT Press, Cambridge, s. 23–46
- Nygaard, Lars, 2006: Frå ordbok til ordnett, upublisert hovedoppgave, UiO
- Pedersen, Bolette Sandford & Anna Brasch, 2009: What do we need to know about humans? Viby, Mnemo, s.158–165
- Pedersen, Bolette Sandford, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen & Henrik Lorentzen, 2009: DanNet: the challenge of compiling a

wordnet for Danish by reusing a monolingual dictionary. I: Language Resources and Evaluation 43, s. 269–299.

*Ruth E. Vatvedt Fjeld*

Institutt for lingvistiske og nordiske studier, Oslo universitet  
r.e.v.fjeld@iln.uio.no

*Rune Lain Knutsen*

runelk@ifi.uio.no

*Julie M. Torjusen*

julieimt@student.ilos.uio.no



# Ømålsordbogens beskrivelse af verbalbøjning

*Asgerd Gudiksen*

The article presents and discusses the lexicographical description of conjugation in *Ømålsordbogen* (The Dictionary of the Insular Dialects). *Ømålsordbogen* describes the traditional dialects with 1850–1945 as its core period. The source materials are mainly records collected from about 1920 to about 1940, often unique information not to be found elsewhere. However, only a minor number of verbs have been recorded systematically. Consequently, the morphological data of a verb are often incomplete and therefore difficult to interpret and describe. Another challenge for the dialect lexicographer is to account for the inflectional variation that may be both geographical and chronological, without making the microstructure too difficult to access.

*Nøgleord:* dialektleksikografi, bøjningsbeskrivelse, mikrostruktur

## Indledning

Når man slår et verbum op i *Ømålsordbogen*, er det lidt uforudsigeligt hvad man finder om bøjning. De korteste bøjningsafsnit består af tre ord. De længste er på tre spalter.

Hvordan bøjningsbeskrivelsen er udformet, er bestemt af i hvert fald tre faktorer: hvor indviklet bøjningen er, hvor meget bøjningen varierer geografisk og endelig hvor dækkende materialet er. Der har været meget andet man skulle optegne end verbalbøjning, og det er kun en mindre del af verberne der er blevet systematisk optegnet.

Baggrunden for mit foredrag er at jeg er i gang med en undersøgelse af den svage verbalbøjning i de danske øsmål, dvs. dialekterne på Sjælland, Møn, Lolland-Falster og Fyn med omliggende øer. Undersøgelsen er primært dialektologisk, men eftersom jeg har taget udgangspunkt i de 10 bind der er udkommet af *Ømålsordbogen*, fra *a* til *kursmed*, har jeg fået nogle erfaringer som bruger og er kommet til at reflektere over vores praksis.

Man kan udskille tre typer af bøjningsbeskrivelse i Ømålsordbogen, som jeg skal vise fordele og ulemper ved, men inden jeg går i gang med at diskutere bøjningsbeskrivelserne, vil jeg først præsentere ordbogen ganske kort og derefter sige noget om vilkårene for vores arbejde med bøjningerne, med andre ord noget om bøjningsklasser og endelser i ømålene og noget om ordbogens kilder.

Ømålsordbogen beskriver de traditionelle dialekter med 1850-1920 som kerneperiode for lyd- og bøjningsbeskrivelsen (og med 2. verdenskrig som øvre grænse for ordforrådet). Der er tale om en indgående beskrivelse ikke kun af sproget, men også af den bondekultur som ordene indgår i. Målgruppen er bred: sprogforskere af forskellig slags, etnologer, museumsfolk og ikke-professionelle brugere med interesse for dialekter og bondekultur. Det medfører at der vil være oplysningerne i ordbogen som en del af vores brugere ikke vil være interesserede i, bl.a. de meget udførlige beskrivelser af udtale og bøjning. Det er samtidig oplysninger vi nødig vil skære ned på, for det er ofte oplysninger man ikke finder andre steder. Ømålsordbogen bygger nemlig i høj grad på materiale indsamlet igennem mange år af ordbogens ansatte og lokale medarbejdere, dels i form af optegnelser på papir, dels i form af lydoptagelser (jf. Ømålsordbogen. Tillægsbind og Gudiksen og Hovmark 1999). Den ideelle løsning ville være flere udgaver af ordbogen, men den er ind til videre ikke inden for rækkevidde. Den løsning vi har valgt, er at udforme lyd- og bøjningsafsnittene med tanke på sprogforskerne, men forsøge at opdele ordbogsartiklerne tydeligt sådan at brugere der er ikke interesserede i bøjningen, forhåbentlig ikke farer vild i artiklen uden at nå frem til det de søger.

## Redegørelse for endelser og bøjningsklasser

I Ømålsordbogen opererer vi med tre svage bøjningsklasser: ede-bøjning, te-bøjning og de-bøjning, mens man i beskrivelser af rigsmålet kan nøjes med to: ede-bøjning og te-bøjning.

Betegnelserne ede-bøjning, te-bøjning og de-bøjning i Ømålsordbogen skal betragtes som formler, som abstrakte kategorier der dækker over forskellige lyd-lige realisationer i ømålene. te-bøjningen er den bøjning der er lettest at identificere. Den er overalt repræsenteret ved en endelse der korresponderer entydigt med skriftens endelser *-te* i præteritum og *-t* i perfektum participium. Udtalen er dog normalt med [d], fx [glæmdə, glæm'd] 'glemte, glemt'. Forklaringen er at fonemerne /t/ og /d/ er faldet sammen i dansk i indlyd og udlyd (undtagen bornholmsk). Der er fri variation mellem /t/ og /d/, men [d] er den normale manifestation.

Hvad ede-bøjningen angår, går der et skel mellem de dialekter der har fonemet spirantisk *d* ("blødt *d*") ligesom dansk rigsmål, og dem der ikke har det. Dialekterne med spirantisk *d* omfatter først og fremmest de sjællandske dialekter, desuden vestmønske og sydfalsterske. I dialekterne med spirantisk *d* ligger bøjningsendelserne nær ved rigsmålets og skriftens endelser, fx bøjes verbet *drille* således på sjællandsk:

[*drelə, drelɪ, dreləðə, dreləð*] (Sjælland)

I de øvrige øsmål der ikke har ikke spirantisk *d*, realiseres ede-bøjningen på en måde der afviger markant fra skriften og rigsmålet, fx bøjes *drille* således på østfynsk:

[*drelə, drelɪ, drelə, drelə*] (Østfyn)

Som det ses, er der ikke forskel på præteritum, perfektum participium og infinitiv, det hedder [*drelə*] i alle tre former. Kun præsens har sin egen form.

Dialekterne på Fyn med omliggende øer afviger desuden fra dialekterne på Sjælland, Møn og Lolland-Falster ved at det er relevant at operere med endnu en bøjningsklasse, nemlig de-bøjning. Præteritum har samme form som i ede-bøjningen, men de-bøjningen afviger fra ede-bøjningen ved at participiet er enstavet, hvilket i øst- og vestfynsk, der befinder sig inden for stødområdet, kommer til udtryk ved at participiet har stød, fx bøjes *dele* således på østfynsk:

[*dielə, dielɪ, dielə, di'el*] (Østfyn)

Den enstavete participium har de-bøjningen tilfælles med te-bøjningen. Og forklaringen er at de historisk set er varianter af samme bøjningsklasse, anden svage bøjningsklasse, selv om te-bøjning og de-bøjning fremtræder meget forskelligt i nyere tids dialekter.

I gammeldansk havde 2. svage bøjningsklasse fire varianter afhængigt af verbalstammens udlydende konsonant (jf. Horneman Hansen 2001: 39ff.). Systemet så i præteritum sådan ud:

1. *-te* efter ustemt konsonant, fx *lekte* 'legede' (af *leke*), *døpte* 'døbte' (af *døpe*).
2. *-de* efter stemt konsonant undtagen *r*, fx *delde* 'delte' (af *dele*) (hvor *d* må antages at betegne stemt *d*).
3. *-dde* efter postvokalisk *-th*, fx *fødde* 'fødte' (af *føthe*).
4. *-the* efter *r* og efter vokal, fx *hørthe* 'hørte' (af *høre*), *nathe* 'nåede' (af *na*) (hvor *the* betegner spirantisk *d*).

Lydudviklinger og analogisk omdannelse har medført at dette system er gået i opløsning.

Stemt *d* forsvandt ud af dansk, og siden blev den fonematiske forskel mellem *d* og *t* som sagt ophævet i indlyd og udlyd. Det har jo så medført at endelse 1 og 2 ovenfor er faldet sammen til én endelse skrevet *-te* og normalt udtalt [-*də*]. Endvidere udviklede gammeldansk *dd* sig i det meste af ømålsområdet til [*d*], dvs. at endelse 3 er faldet sammen med 1 og 2 de fleste steder. Hvad angår den 4. endelse, svinder spiranten allerede i gammeldansk, og verber af denne type er ofte gået over til *te*-bøjning eller *ede*-bøjning.

De verber hvis bøjning er kategoriseret som *de*-bøjning, er først og fremmest verber hvis stammer i skriftsproget ender på *-l*, *-nd* eller *-r* som fx *dele*, *kende* og *høre*. Gammeldansk *-ld*, *-nd* og *-rth* er nemlig blevet til *-l*, *-n* og *-r* ved lydret assimilation. Præteritummer som fx [*dielə*] 'delte' og [*hyɜrə*] 'hørte' repræsenterer altså den lydrette udvikling af de gammeldanske former, mens [*dielɔ*] og [*hyɜrdə*] er eksempler på morfologisk omdannelse. De assimilerede former er bevaret i et vist omfang i fynsk (og for øvrigt også i jysk og ikke mindst i bornholmsk, jf. Horneman Hansen: 98 ff. og 124 f.). I sjællandsk, lolland-falstersk og i rigsmålet har denne type verber derimod normalt fået *te*-bøjning ved morfologisk omdannelse bortset fra nogle få relikter, fx *gjorde* præteritum af *gøre*.

Kategorien *de*-bøjning omfatter endvidere nogle verber hvis stamme ender på vokal, fx *fri* med præteritum og participium [*friːə*, *friː*] – igen former der ikke har noget konsonantisk element i bøjningen, men som repræsenterer en lydret udvikling af de gammeldanske former. Endelig omfatter *de*-bøjningen enkelte eksempler på verber med morfologisk omdannelse, fx *gemme* der lydret skal have *te*-bøjning, men stedvis har *de*-bøjning med præteritum og participium: [*gæːmə*, *gæːm*].

Størstedelen af de verber der er beskrevet i Ømålsordbogen, går efter *ede*-bøjningen i hele området. Der er desuden også mange verber der har *te*-bøjning overalt. Men bøjningsvariation er ikke sjælden. Der er bl.a. mange verber der varierer dialektgeografisk mellem *ede*-bøjning, *te*-bøjning og evt. *de*-bøjning. Der er også eksempler på at oprindeligt stærke verber stedvis er overgået til svag bøjning, fx *grine* der forekommer både med den oprindelige stærke bøjning og med *te*-bøjning, *de*-bøjning og *ede*-bøjning. Og denne variation skal beskrives i ordbogen. Hertil kommer variation af de enkelte bøjningsformers udtale der også kan høre hjemme i bøjningsbeskrivelsen. Ved verbet *bide* er der for det første variationen mellem former med og uden spirantisk *d*, for det andet variation i udtalen af rodvokalen i præteritum. Rodvokalen kan udtales med langt *e* som

i rigsmålet, altså: [be'ð] eller [be'd], men kan også være repræsenteret af en diftong med varierende udtale, fx [bi'ɜð], [be'ɜð] eller [bæj'ð].

## Om kilderne til bøjningsbeskrivelsen

Der er to hovedkilder til oplysningerne om verbalbøjning i Ømålsordbogens samlinger.

Den første hovedkilde er trykte og utrykte grammatiker og ordbøger af dialekttalende over deres egen dialekt. De vigtigste trykte beskrivelser er: Thorsen 1894 (fra Nordvestsjælland), Christensen 1936 og Jørgensen 1943 (fra Lolland), Jacobsen 1885 (fra Vestfyn) og Gotfredsen 1955 og 1972-74 (fra Langeland). Mange af disse kilder giver en grundig beskrivelse af verbalbøjning, men den dialektgeografiske fordeling er spredt og tilfældig.

Den anden hovedkilde til verbalbøjning er de såkaldte ordlisteoptegnelser. Det er optegnelser efter en spørgeliste der har til formål at få oplysninger om udtale og bøjning (jf. Gudiksen og Hovmark 1999: 17). Den indeholder 82 svage verber hvis bøjning der skal gøres rede for, og 85 stærke. Ordlisten har været brugt til at optegne efter i stort set hele ømålsområdet, men det er altså kun forholdsvis et begrænset antal verber der har været optegnet systematisk.

Der ud over er der ganske vist hundredvis af andre optegnelser der også indeholder verber, men ikke altid i lydskrift. Desuden medfører visse sammenfald i præteritum mellem bøjningsklasserne at enkeltstående præteritummer ikke altid kan henføres sikkert til bøjningsklassen. Jeg nævnte før at præteritum har samme form i de-bøjning og ede-bøjning. Et andet eksempel er verber som *bytte*, *kaste*, *koste* og *hente*. I rigsmålet har verber af denne type ede-bøjning, i ømålene er der variation mellem ede-bøjning og te-bøjning (dog med ede-bøjningen som den mest almindelige). I de dialekter som ikke har spirantisk *d*, er præteritum af ede-bøjning og te-bøjning imidlertid faldet sammen, så det altså ikke kan afgøres hvilken bøjning en enkeltstående præteritum repræsenterer. På sjællandsk, vestmønsk og sydfalstersk (som alle har spirantisk *d*) er det i princippet muligt at skelne mellem de to bøjninger:

dialekt uden spirant *d*:  
ede-bøjning: [byd̥ə, byd̥ə]  
te-bøjning: [byd̥ə, byd]

dialekt med spirantisk *d*:  
ede-bøjning: [byd̥əðə, byd̥əð]  
te-bøjning: [byd̥ə, byd]

Der er imidlertid den komplicerende faktor at sjællandsk, vestmønsk og sydfalstersk har variation mellem endelsen med [ð] og en reduceret endelse uden [ð]. Også i disse dialekter kan det altså være usikkert om en optegnet præteritumsform som [*bydø*] repræsenterer te-bøjning eller ede-bøjning.

Alt det betyder at det undertiden ikke fremgår hvordan et verbum bøjes i hele dets forekomstråde, eventuelt slet ikke. De verber der ikke er forsynet med bøjningsangivelser fordi der ikke har været materiale til det, er dog i regelen verber der kun er optegnet i få sogne. Det kan fx være ord for at bære ad på en uheldig måde dannet ved kreativ omdannelse af typen *ffuske*, *ffunte*, *ffumpe* eller verber afledt af et substantiv, fx *flotte* 'forsyne et fiskegarn med flåd'.

## Bøjningsbeskrivelsen i Ømålsordbogen

Ømålsordbogen har tre typer af bøjningsbeskrivelse. Den første type består i at der kun er anført hvilke(n) bøjningsklasse(r) verbet tilhører. Den anden type er en udvidet udgave af den første. Den tredje type omfatter bøjningsrækker ofte suppleret af et "endvidere-felt", som jeg skal komme tilbage til. Desuden er der som sagt den mulighed at der ikke er anført oplysninger om bøjning overhovedet, og den forekommer faktisk ret hyppigt, fx er 73 af de 241 verber under bogstav *b* uden bøjningsoplysninger. De resterende 168 verber fordeler sig sådan:

Type 1: kun angivelse af bøjningsklasse	142 verber
Type 2: angivelse af bøjningsklasse med eksempler	20 verber
Type 3: bøjningsrækker	6 verber

Type 1 bruges især ved verber der udelukkende har ede-bøjning i ømålene, fx har 126 af de nævnte 142 verber ede-bøjning i hele området, men skabelonen er også brugt ved beskrivelsen af verber som fx *fyre* der normalt har ede-bøjning, men dog enkelte steder optræder med te-bøjning, nemlig i to sogne på T(åsinge) og i Marstal Landsogn på Æ(rø), jf figur 1.

### fyre I v

(også *fyre* II): [*fy.ɹ*], sjæld [*fyj.ɹ*] Ø(vist alm, tilsv Thorsen.SS:16, jf FØS:269, FØF:237), m acc 1 T, Æ(alm); [*fu.ɹ*] nS(Gill). – Bøjn: ede-bøjn, præterit opt m te-bøjn [*fy.ɹdø*] T(Bjb), [*²fyrdø*] Æ(MarstLd), part m te-bøjn [*fy.ɹd*] T(Bregn).

Figur 1. Lyd- og bøjningsafsnit til *fyre* I.

Den kan endvidere være brugt ved verber der som *hvæse* har te-bøjning i de meste af området, men stedvis er gået over til ede-bøjning, jf. figur 2.

hvæse II v

[..]

– Bøjn: te-bøjn Ø(vist alm, Thorsen.SS:107, LollO, PJac.MF:168), ede-bøjn M(Bor), Fa(Sklb), L(To LollO), vF(Kavs), sF(Gudb), ældre prææs *hvæs* [..]

Figur 2. Bøjningsafsnit til *hvæse* II.

Man kan bemærke at te-bøjning beskrives som ”vist alm” – en modifikation man ofte møder i Ømålsordbogen. *Vist* afspejler at der er optegnelmæssige huller i materialet, men ikke større end at redaktøren har turdet generalisere – med forbehold.

Fordelen ved type 1 er naturligvis at bøjningsbeskrivelsen er meget let læst, ulempen er at det ikke antydes at realisationen kan afvige ganske meget fra skriften. Beskrivelse af verberne i selve ordbogen suppleres dog af Tillægsbindet, hvor termerne ede-, te- og de-bøjning forklares, desværre lidt uformidlet. Der mangler nogle typeeksempler og en dialektgeografiske oversigt over endelsernes realisation, jf. fx forklaringen af ”ede-bøjn”:

ede-bøjn: Den svage verbalbøjning, hvis præteritums- og participiums-endelse korresponderer med skriftsprogets endelser -ede og -et. Se fx nogle af bøjningsrækkerne under *bygge*, *bande* III, *bøje* II, *behøve*, *age*, *flå* II (Ømålsordbogen. Tillægsbind: 24).

Der henvises jo ganske vist til bøjningsrækker man kan kigge på, men eftersom de pågældende verber har både ede-bøjning og te-bøjning, og rækkefølgen af bøjningsrækkerne i disse verber tillige er opstillet geografisk, så er det ikke helt let at overskue hvordan ede-bøjningen realiseres.

Et eksempel på type 2 er verbet *gemme*, der ”vist” kan have te-bøjning i hele området (se figur 3). De former der er anført i artiklen *gemme*, er dels de-bøjningsformer, der medtages fordi denne bøjningsklasse ikke findes i rigsmålet, dels uregelmæssige te-bøjningformer. Begge dele er typisk. Der kan dog også i et eller andet omfang være bragt regelmæssige te-bøjningformer. Derimod er der aldrig eksempler på ede-bøjnings-former, selv om formerne uden [ð] også er klart rigsmålsafvigende.

gemme II v

[..]

– Bøjn: te-bøjn Ø(vist alm); m vekslende vokallængde: *gæ·mə*, *gæ·m.ɪ*, *gæmdə*, *gæm't* nvS(Sejr Thorsen.SS:102 og 108), [*gæ̃·mə*, *gæ̃·m.ɪ*, *gæ̃mdə*, *gæ̃'md*] øF, vF(alm), u stød sF(VSk), tilsv m [ø] øF(alm), sF(Gudb); ede-præt L(To LollO); de-bøjn m præt og part [*gæ̃·mə*, *gæ̃'m*] vF(Kavs, Br) – inf i medial form [..]

Figur 3. Bøjningsafsnit til *gemme* II.

Det er i reglen let at overskue den dialektgeografiske fordeling af bøjningsklasserne, men det kan måske være lidt problematisk at læseren normalt kun møder eksempler på det afvigende eller specielle.

flække III v

S [*flæg*, *flæg.ɪ*, *flæwd*, *flæwd* ell *flæw'd*] nS, øS-(alm, jf FØS:404f), nvS(Udby, Gra, Sod, tilsv Skib Kryger.Gr:389, DyrI.S),

[*flæg*, *flæg.ɪ*, *flægd*, *flægd*] nS(Seng, KarlsI), nvS(Ug, Rørby, Hjb, Sejr, tilsv Thorsen.SS:111), svS(Hrlm, Ørs), sS(alm).

Endv: præs [*flæ'g.ɪ*] S(spor opt i omr m kortvokalstød).

L [*flæg*, *flæg.ɪ*, *flægd*, *flægd* og tilsv m *æg*] M, nFa, L(alm),

[*flæg*, *flægr*, *flæjd*, *flæjd*] sFa(alm).

F [*flæg*, *flæg.ɪ*, *flæg*, *flæg*] F, T, Lgl, Æ(spor opt, m acc 2 Marst, m v-stød nvF),

[*flæg*, *flæg.ɪ*, *flæjd*, *flæjd*] Lgl(alm).

Endv: ede-bøjn S, M, L-Fa(spor opt, nyere) – st bøjn: præt *flak* nS(Hels), svS(Tjæ, Fodby), L-(MLoll), vF(Kavs), sF(Eg, SFyn), part *flukket* L(Bennike.S), *flukken* sF(Eg, SFyn), jf *flukken* adj – medial form: inf og præs [*flægs*] nvS(Odd, Kværk), svS(Fodby), Fa(Sklb), L(Rad), præt [*flæwds*] nvS(Kværk), [*flægs*] nvS(Hjb), skrevet *flakts* M(StegeLd), L(Rad).

Figur 4. Lyd- og bøjningsafsnit til *flække* III.



Type 3, bøjningsrækker, kan være meget omfangsrige bøjningsbeskrivelser. Rekorden er vist verbet *gøre* med tre spalter; *flække* (figur 4) på ca. en halv ordbogsspalte er en af de korte. Ved denne skabelon er udtaleafsnit og bøjningsafsnit integreret. Bøjningsrækkerne er ordnet efter hovedområder: S for Sjælland, L for Møn og Lolland-Falster og F for Fyn og øerne syd for (Langeland, Tåsinge og Ærø). I reglen er der desuden et eller flere felter indledt med ”endv” (dvs. endvidere). I artiklen *flække* III afslutter feltet hele lyd- og bøjningsafsnittet, men der kan også være et endv-felt efter hvert af de tre områder (S, L og F) plus et felt i slutningen af lyd- og bøjningsafsnittet. Endv-feltet indeholder ofte supplerende lydformer i infinitiv, alternative udtaler af former fra bøjningsrækkerne eller oplysninger om særlige præsensformer, men der kan også som i artiklen *flække* være andre bøjningsklasser end dem der er opstillet i bøjningsrækkerne. Hvis man vil have svar på spørgsmål som: hvordan kan dette verbum bøjes i sjællandsk?, eller: hvilken udbredelse har de enkelte bøjningsklasser i Ømålsområdet?, må man derfor ofte stykke oplysningerne sammen to-tre steder fra i bøjningsafsnittet. Desuden må man være i stand til at identificere hvilken bøjningsklasse en given bøjningsrække tilhører for at nå frem til det rigtige resultat. Til gengæld får den læser der kan følge med, jo grundig besked og får også set eksempler på både regelmæssig og uregelmæssig bøjning. Det er også en fordel at lyd- og bøjningsafsnittet er klart adskilt fra den øvrige artikel så den læser der ikke har brug for oplysningerne, let kan springe dem over.

## Afslutning

Ømålsordbogens beskrivelse af verbalbøjningen er ofte knap, måske sommetider for knap, undertiden udførlig, sommetider måske for udførlig. Under alle omstændigheder kræver den en del af læseren. Vores målgruppe er dog som nævnt forholdsvis snæver, når det gælder lyd- og bøjningsafsnit, nemlig sprogforskere, især sproghistorikere og dialektforskere, der behersker et skandinavisk sprog. Opgaven er altså først og fremmest at sikre at denne type brugere nogenlunde ubesværet kan få fat i de oplysninger de søger. Der er efter min mening to punkter hvor vi især kan – og bør – sætte ind. For det første bør teksten i Til-lægsbindet forbedres, så det fremgår eksplicit hvad der karakteriserer de enkelte bøjningsrækker, for det andet bør vi gøre type 3, bøjningsrækketypen, mere tilgængelig ved at markere hvilken bøjningsklasse de enkelte bøjningsrækker tilhører. Desuden bør vi nok fravælge type 3 til fordel for type 2 hvis materialets sammensætning vil medføre at ”endvidere-feltet” svulmer op.

Når Ømålsordbogen en gang ude i fremtiden bliver tilgængelig også som netordbog, vil det forhåbentlig blive muligt at komme brugerne mere direkte i

møde med til- og fravalg, men indtil videre kræver vi altså undertiden lidt af vores brugere og håber at de trods alt finder det frem som de er interesserede i.

## LITTERATUR

- Christensen, Balthasar, 1936: Lollandsk Grammatik – Radstedmaalet. København.
- Gotfredsen, L., 1955: Langelandsk Grammatik i Hovedlinier. København
- Gotfredsen, L., 1972-1974: Langelandsk Ordbog. København.
- Gudiksen, Asgerd og Henrik Hovmark, 2009: Måske husker De noget alle andre har glemt. I: Asgerd Gudiksen m.fl. (red): Dialektforskning i 100 år. S. 13-64. København.
- Horneman Hansen, Lise, 2001: Jysk -de-bøjning. En undersøgelse af svag præteritumsbøjning. Uppsala.
- Jacobsen, Peder, 1885: Mundarten i Flemløse Sogn. – Vestfynsk. Udgivet af Poul Andersen 1933. København.
- Jørgensen, J., 1943: Lollandsk Ordbog (Thoreby-maalet). København.
- Thorsen, P.K., 1894: Sprogarten på Sejerø. København.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Møn, Lolland-Falster, Fyn og omliggende øer, 1992 ff. København.
- Ømålsordbogen. Tillægsbind. Indledning, nøgler, forkortelseslister, kort, 1992. København.

*Asgerd Gudiksen*

Ømålsordbogen, Nordisk Forskningsinstitut, Afdeling for Dialektforskning, Københavns Universitet  
gudik@hum.ku.dk

# Ivar Aasen og svensk leksikografi

*Kjetil Gundersen & Dagfinn Worren*

In the language project of Ivar Aasen (1813–1896) an important aspect was to demonstrate that Norwegian was a national language in its own right, alongside Danish and Swedish. In his lexicographical work Aasen used a Danish model, but he emphasized that, from a linguistic point of view, Norwegian actually had just as much in common with Swedish. In this article we present a study that demonstrates the dimension of the specific common vocabulary between the two languages. The method used for this purpose is to extract references to Swedish sources in Aasen's 1873 edition of his dictionary. The study reveals that 7.5% of the vocabulary is common, and that Aasen mainly refers to two Swedish sources (Dalín 1855 and Rietz 1867). Not surprisingly, a significant part of this common vocabulary mostly occurs geographically along the Swedish border. The conclusion is that Aasen's ideas regarding a common vocabulary for Norwegian and Swedish are to a certain extent verified. But more importantly, the still rather sparse occurrence justifies the establishment of Norwegian as a distinct language of its own.

*Nøkkelord:* Dalin, fellesord, heimfeste, heimfesting, Rietz, Aasen

## 1. Norsk og svensk

Eit av måla med språkprosjektet til Ivar Aasen (1813–1896) var å syne at norsk var eit særlege språk til liks med dei to andre skandinaviske språka, dansk og svensk. I det språklege jamføringsarbeidet som han gjorde, vart det då eit viktig poeng å syne at norsk hadde ordtilfang som ikkje fanst i det danske skriftmålet. Her var han oppteken av at det i den naturlege geografiske fellesskapen med Sverige også kunne liggje ein leksikalsk fellesskap (Aasen 1965: 344). Målet med denne artikkelen er å dokumentere korleis svensk språk og leksikografi kom til uttrykk i prosjektet hans, men utan å jamføre med dansk. Det tillèt ikkje ramma for artikkelen.

Først skal bakgrunnen og føresetnaden for Aasens målreisingsprosjekt skisserast opp. Med det utgangspunktet avgrensar vi oss til å syne korleis svenske kjelder og svensk språk kjem til uttrykk i hovudverket hans, *Norsk Ordbog* (1873). Særskilt skal det ordtilfanget som Aasen dokumenterer som sams for

svensk og norsk i denne første fullnormerte nasjonale ordboka for norsk, dragast fram og analyserast i lys av desse kjeldene med vekt på opphav og geografisk utbreiing.

## 2. Bakgrunn

Sidan 1500-talet hadde Noreg hatt sams styre med Danmark og dermed også sams akademisk kultur, irekna skriftspråk. Noreg fekk si eiga grunnlov i 1814, som også la rammene for Noreg som nasjonalstat. I alle fall for det kulturelle Noreg framstod det då som eit problem at skriftspråket var dansk. Allereie som 23-åring peikte Aasen på denne nasjonale mangelen i artikkelen ”Om vort Skriftsprog” frå 1836. Her gjorde han framlegg om at dei norske dialektane måtte beskrivast systematisk som grunnlag for eit nasjonalt skriftspråk lagt fram i ei ordbok og ein grammatikk (Aasen 1912: III,7).

## 3. Føresetnader i samtida

Det var to hovuddrag i samtida som gjorde eit slikt språkprosjekt mogleg å gjennomføre, nemleg romantikken og den historisk samanliknande språkvitskapen. Den romantiske idérørsla opna for mangfaldet i det folkelege kulturlivet og framheva det historisk røtte folkemålet som eit av dei fremste nasjonale uttrykka (jf. t.d. Torp og Vikør 1993: 139).

Den historisk samanliknande språkvitskapen førte systematikk inn i språkgranskinga. Tenkjemåten i denne språkteorien kravde både systematisk kunnskap om fleire samtidsspråk og tilsvarende kunnskap for eldre språksteg og evne til å sjå parallellutviklingar over tid (Hovdhaugen et al. 2000: 158). I tiåra før Aasen-prosjektet var dansken Rasmus Rask (1787–1832) utan tvil den lysande lingvistiske stjerna i Norden. Saman med tyskarane Jacob Grimm (1785–1863) og Franz Bopp (1791–1867) var Rask ein av grunnleggjarane av denne nye teorien. Også svensken Johan Ihre (1707–1780) var ein særmerkt føregangsmann på same feltet. Rask hadde særleg utvikla teorien på dei nordiske språka. Han kom jamvel fram til at det gamle nordiske språket var ”mere grammatikalsk fuldkomment end Tysken” (Rask 1811: vi). Det vart rekna som eit kvalitetsmerke. Aasen var ikkje upåverka av slike synsmåtar (Venås 1994: 387). Det måtte styrkje han i trua på prosjektet sitt, for dei norske dialektane var openbert formrike.

## 4. Oppdrag med dilemma

Dei nye europeiske ideane om språk utløyste også økonomisk støtte til eit norsk målreisingsprosjekt. Frå 1842 og resten av livet fekk Aasen høve til å setje språkvisjonen sin ut i livet, i første omgang som oppdrag frå Det kongelige Norske Videnskabers Selskab i Trondheim.

Oppdraget var ikkje utan dilemma og motstand. For det første utstyrte Vitskapselskapet Aasen med hjelpemiddel som var tunge hegemoniske verk frå dansk språkvitskap. Det galdt både Molbechs *Dansk Ordbog* (1833) og Rasks *Vejledning til det Islandske eller gamle Nordiske Sprog* (1811) (Walton 1996: 78). Etter den danske litteraturforskarer Hans Hauges beskriving av dei norske akademiske tilhøva var Aasen i ein slags postkolonial situasjon (Hauge 2001: 371). Det var nettopp denne stoda Aasen skulle frigjere det kulturelle Noreg frå. Løysinga på eit slikt postkolonialt dilemma har vore å vidareutvikle saksfeltet på eigne premissar (op. cit. 376). Det var nettopp det Aasen gjorde med å bruke Molbechs ordbok som modell i målreisingsprosjektet sitt samstundes som ordboka ville bli avløyst som autoritet for skriftspråket i Noreg (Worren 2006: 392).

For det andre var det sterke motkrefter i Noreg mot Aasens prosjekt, fordi det ville bryte opp den etablerte kulturelle fellesskapen med Danmark. I ein offentleg språkpolemikkk vart Aasen konfrontert nettopp med Jacob Grimms syn på utviklinga for dei europeiske språka. I *Geschichte der deutschen Sprache* (1848) ser Grimm det som sannsynleg at dansk om nokre hundre år vil gå opp i svensk. Dermed skulle det vere heilt urimeleg å etablere eit eige nasjonalspråk i Noreg, hevda dei norske kritikarane. Aasens kommentar var at vitskapsmannen Grimm hadde han den største respekt for. Derimot var det ingen grunn til å leggje vekt på synsaren Grimm. Aasen avslutta med at han trudde at det ville falle naturlegare med tre skandinaviske språk i staden for Grimms eine (Aasen 1912: 177 ff.). Her har ettertida gjeve Aasen rett. Til Grimms orsaking skal det seiast at han ikkje kjende Tore Jansons *Språkens historia* (2010) og hans framstilling av korleis politiske realitetar lagar andre språklege utviklingar enn det språkvitskaplege vurderingar kunne tilseie.

## 5. Den svenske lina

### 5.1 Introduksjon

Den svenske lina i Aasens prosjekt byggjer særleg på tre hovudkjelder, Dalin (1855), Rietz (1867) og Rydquist (1883).

Fullføringa av dei første ordbøkene i morsmålsleksikografien i Noreg og Sverige fall i tid nokså nær saman. Ivar Aasens første ordbok, *Ordbog over det norske Folkesprog*, låg føre i 1850. Anders Fredrik Dalin (1806–1873) gav ut *Ordbok öfver svenska språket* frå 1850 til 1855. Til 1850-ordboka si fekk Aasen såleis ikkje nytte av Dalin. Men frå han byrja med arbeidet på 1873-ordboka i 1865, var ”Dalins Ordbog” det svenske leksikografiske referanseverket. ”De svenske Ord ere anførte efter Dalins Ordbog” heiter det i kjeldelista (s. XV).

Om lag samstundes kom det også to andre verk som skildra det svenske språket, men med andre innfallsvinklar enn Dalins ordbok. Det eine var Johan Rietz (1815–1868) si dialektordbok, *Svenskt dialekt-lexikon: ordbok öfver svenska allmoge-språket*, som var ferdig i 1867. Det andre var Johan Rydqvist (1800–1877) sitt omfattande verk om svensk språkutvikling, *Svenska språkets lagar. Kritisk afhandling*, som kom ut i perioden 1850 til 1883. Desse to verka sørgde for at Aasen hadde eit solid grunnlag for jamføringa med svensk også dialektalt og historisk.

## 5.2 Omtale av svenske kjelder i Norsk Ordbog

Når Aasen viser til svenske språktilhøve i Norsk Ordbog, dreier det seg som regel om tilvisingar til eit svensk ord som liknar eller er identisk med det norske oppslagsordet, som i oppslaget *corp*:

Korp, m. Ravn (= Ramn). Østl. Sv. korp. Jf. Krump. (Aasen 1873)

Men dette vil ikkje seie at Aasen freista å gjennomføre ei uttømmende og fullt ut systematisk jamføring av norsk og svensk ordtilfang. Til dømes er vanlege norsk-svenske ord som *katt*, *bein* og *hus* handsama utan tilvising til dei tilsvarende svenske formene, sjølv om Aasen hadde tilgjenge til desse orda i svenske ordbøker. Aasen greier heller ikkje ut om kva kriterium han gjekk etter når han valde ut ord som skulle jamførast. Ei aning om korleis Aasen prioriterte, får ein likevel i ein passasje i fortalen i Norsk Ordbog der han kort kjem inn på praksisen med å jamføre norske oppslagsord med ord i nærskylde språk: ”Ved adskillige mindre bekjendte Ord [i Norsk Ordbog] er en Sammenstilling med svenske, tyske eller engelske Dialekter tilføiet, for saavidt som der var Anledning dertil” (Aasen 1873: XII). Det kan sjå ut som Aasen ønskte å løfte fram felles ordtilfang i tilfelle der det ikkje er opplagt, og at han var oppteken av å syne slektskap mellom *målføre*. Det siste bør ein vel her forstå som dei munnlege folkemåla. Såleis verkar det rimeleg at når Aasen trekkjer fram Rietz’ dialektleksikon i fortalen i Norsk Ordog, så lovprisar han verket for å ha samla kunnskap om ordtilfanget i svensk talemål:

Det er ikke liden Lettelse for os at forefinde en saa righoldig Ordsamling som Rietz's svenske Dialekt-Lexikon, hvori flere ældre Glossarier ere opptagne eller benyttede, saa at man der med Lethed kan faae en Oversigt af det store Forraad af svenske Ord, som ikke ere komne til Brug i Bogsproget (Aasen 1873: VI).

### 5.3 Omfang i Norsk Ordbog

Det har ikkje vore kjent kor mykje ordtilfang Aasen dokumenterer som sams for norsk og svensk i Norsk Ordbog, og såleis heller ikkje kva som kjenneteiknar dette ordtilfanget. Dette er bakgrunnen for at vi vil gje eit oversyn over oppslag i Norsk Ordbog som inneheld tilvisingar til svensk, og så presentere ein analyse av dette ordtilfanget med tanke på opphav og utbreiing. Arbeidet med å få oversyn over omfanget av tilvisingar tok til med å overføre ein elektronisk versjon av ordboka til ein database der det er mogleg å sortere oppslaga etter tekstinnhald.<sup>1</sup> Dei svenske tilvisingane vart så identifiserte ved å søkje fram oppslag som inneheld forkortingar, ord(kombinasjonar) og forfattartilvisingar som Aasen nytta i samband med jamføringane. Såleis søkte vi etter teiknrekkjer som *sv.*, *i Sverige*, *svenske*, *Dalin*, *Rietz* og *Rydgvist*.

Desse søka genererte til saman 2772 unike oppslag, av totalt 36902 i heile Norsk Ordbog.<sup>2</sup> Tabell 1 syner førekomsten av oppslag for kvar av teiknrekkjene vi søkte etter.<sup>3</sup>

Vi fann tre oppslag der teiknrekkjene *i Sverige* eller *svensk(e)* er nytta i andre samanhengar enn til å gje opplysingar om svensk språk. I oppslaga *Jamt* og *Lapplending* er *i Sverige* del av definisjonsteksten, medan *svensk* er eige oppslagsord. Når desse oppslaga er trekte frå, er det 2769 oppslag med tilvisingar som har å gjere med svensk språk, noko som utgjør om lag 7,5 % av oppslaga i ordboka. Sagt på ein annan måte: På ei gjennomsnittleg side i Norsk Ordbog er det om lag tre oppslag med tilvisingar til svensk.

<sup>1</sup> Denne versjonen var tekstfila som vart nytta som grunnlag for den førebels siste trykte utgåva av Norsk Ordbog som vart utgjeven i 2003 (Kruken og Aarset 2003).

<sup>2</sup> Dette talet er noko lågare enn talet på oppførte ord, som er om lag 40 000 ifølgje omslagsteksten til 2003-utgåva av Norsk Ordbog. Skilnaden kjem av at ein del av orda i Norsk Ordbog, som regel avleiingar eller samansetjingar, ikkje er skilde ut som eigne oppslag, men er gjorde greie for under oppslag for andre ord.

<sup>3</sup> Summen av førekomstane for dei einskilde teiknrekkjene er høgare enn talet på oppslag som inneheld minst éin førekomst av ei slik teiknrekkje. Dette kjem av at det same oppslaget kan innehalde fleire enn ei av teiknrekkjene. Til dømes finn ein ofte *Rietz* i same oppslag som avstytinga *sv.*

Type oppslag	Førekomst
Alle i Norsk Ordbog	36902
Som inneheld <i>sv.</i>	2431 (448 av desse med formuleringa <i>sv. Dial.</i> )
Som inneheld <i>Rietz</i>	236
Som inneheld <i>svensk(e)</i>	333
Som inneheld <i>i Sverige</i>	47
Som inneheld <i>Rydgqvist</i>	15
Som inneheld <i>Dalin</i>	3

Tabell 1. *Tilvisingar til svensk i Norsk Ordbog*

Sidan dei fleste tilvisingane står utan kjelde, er det ikkje mogleg å fastslå kva kjelde som ligg til grunn for kvar einiskild tilvising. Ein kan rett nok rekne med at Dalin er kjelda til dei fleste av dei kjeldelause tilvisingane, sidan Aasen fører opp Dalins ordbok som referanseverket for svensk i kjeldelista. Men det finst unntak. Særleg ser det ut til at dei 448 tilfella der *sv.* blir følgt av avstytinga *Dial.* (for *dialekt*), byggjer på andre kjelder, og sidan det er snakk om svenske målføreord, er Rietz den opplagde kandidaten. Vi kontrollerte 20 av dei kjeldelause oppslaga opp mot ordbøkene til Dalin og Rietz, og i alle oppslag med *sv. Dial.* fann vi det svenske ordet att i Rietz, men ikkje i Dalin. Biletet som dannar seg, samsvarar med det Aasen sjølv teiknar i kjeldelista og i fortalen i Norsk Ordbog: Dalin er hovudreferansen til svensk standardmål, medan dialektleksikonet til Rietz er ei nyttig utfylling fordi det dekkjer svensk ordtilfang som ikkje blir handsama hos Dalin. Den tredje kjelda vi har vore inne på, Rydgqvists *Svenska språkets lagar. Kritisk afhandling*, ser ut til å vere nytta i berre nokre få tilfelle. I dei tilfella Aasen viser eksplisitt til Rydgqvist, er det typisk at det handlar om etymologi, som i oppslaget *galen*, der vi får opplyst at adjektivet er danna til fortidspartisippet av verbet *gala*:

galen, adj. 1) gal, afsindig. G. N. galinn. Vel egentl. forhexet (ved Tryllesang), og saaledes Particip af gala (jf. Rydgqvist, Sv. Spr. L. 1, 172). (Aasen 1873)

Når vi skulle analysere materialet vidare, måtte vi vurdere i kva grad dei identifiserte oppslaga representerer sams norsk-svensk ordtilfang. Det syntte seg då at dei aller fleste tilvisingane til svensk er jamføringar med openbert nærskylde ord, som i oppslaget *corp* ovanfor. Av og til jamfører rett nok Aasen med eit ord som ikkje like opplagt er skyldt med det norske, og stundom med eit samansett ord som berre delvis fell saman med det norske. Vi vurderte det likevel ikkje



som fruktbart innanfor rammene av denne artikkelen å gjere ein grundig analyse av kvar og eitt oppslag for eventuelt å luke ut somme slike oppslag. Hovudbiletet er i alle fall at den store hopen av jamføringar med svensk dokumenterer likskapar og skyldskap mellom norsk og svensk ordtilfang. Dei einaste oppslaga vi heldt utanfor vidare analyse, var eitt der opplysinga dreier seg om at oppslagsordet *ikkje* blir brukt i svensk (*hulder*), og to der det blir vist til svenske ord som openbert ikkje er skylde (norsk *amt* og *brosme* blir jamførte med svensk *län* og *lubb*). Dermed hadde vi eit materiale på 2766 oppslag som vi rekna som Aasens norsk-svenske ordtilfang, og som vi har nytta i den følgjande analysen.

#### 5.4 Innslag av norrøne arveord i Aasens norsk-svenske ordtilfang

Ei viktig side ved den norske og svenske språkfelleskapen i tillegg til den sjølv-sagde geografiske nærleiken var for Aasen at språka over lang tid hadde utvikla seg side om side utan at eitt av dei hadde vore underlagt det andre (Venås 1996: 435 f.). Historisk sett blir då dei norrøne arveorda naturlege jamføringspunkt. Såleis er det interessant å ta for seg dei norsk-svenske orda som Aasen viser til i Norsk Ordbog, og undersøkje kor stor del av dei som blir jamførte med ei norrøn opphavsform. Eit døme på eit slikt oppslag har vi i *fjos* (der *G. N.* står for *gammalnorsk*, som var omgrepet Aasen nytta om norrønt):

*Fjos* (oo), n. Fæhuus, Kostald, eller Faarestald. [...] *G. N.* *fjos*, n. I Sverige: *fjus*, *fjøs* og *fl.* (Rietz 133). (Aasen 1873)

Empiri på kor stort innslaget av norrøne arveord er i fellesorda, fann vi ved å jamføre førekomsten av norrøne tilvisingar i ordboka totalt med førekomsten av slike tilvisingar i oppslag med fellesord (tabell 2).

	Oppslag med tilvising til norrønt i heile Norsk Ordbog	Oppslag med tilvising til norrønt blant oppslag som har tilvising til svensk
I tal	5371 (av 36902)	1320 (av 2766)
I %	14,6 %	47,7 %

Tabell 2. *Oppslag med tilvising til norrønt i heile Norsk Ordbog og blant oppslag som inneheld tilvising til svensk*

I ordboka er det 5371, eller 14,6 %, av oppslaga som inneheld ei tilvising til norrønt. Blant oppslaga med fellesord er innslaget av slike tilvisingar forholdsvis mykje høgare: Her blir 1320, eller nær halvparten, av orda førte attende til norrønt. Kor godt desse tala speglar det faktiske innslaget av arveord blant norsk-

svenske fellesord, er usikkert sidan det er uvisse knytt til Aasens utval av både arveord og norsk-svenske fellesord. Det norsk-svenske fellesordet *gråbein* står til dømes utan tilvising til norrønt i Norsk Ordbog, sjølv om det er med i Johan Fritznors *Ordbog over det gamle norske Sprog* frå 1867. Uansett syner tala ein tendens som det ville vere interessant å få stadfest eller tilbakevist i ein meir omfattande analyse.

### 5.5 Geografisk utbreiing av norsk-svenske fellesord i Noreg

Skal ein undersøkje kor i Noreg Aasens norsk-svenske ord har størst utbreiing, er det naturleg å gå ut frå ein hypotese om at kjerneområdet for fellesorda ligg nær Sverige, altså i dei austlege delane av Noreg. Spørsmålet blir såleis om eit fellesord som *gyrje* er typisk i Norsk Ordbog, eit ord Aasen heimfester til tre område med grense mot Sverige – Smålenene (i dag Austfold), Romerike og Solør:

Gyrja (y'), f. 1) Mudder, Slam, Dyndjord. "Gjørju", Smaal. Rom. Solør. I svenske Dial. gjørja, gerja [...] (Aasen 1873)

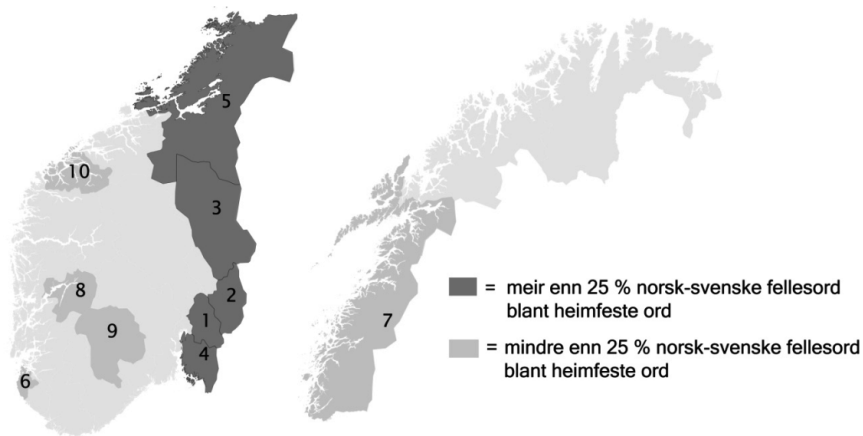
For å få ein peikepinn på om fellesorda har eit særleg nedslagsfelt lengst aust i Noreg, undersøkte vi førekomsten av tilvisingar til svensk i oppslag der Aasen heimfester oppslagsordet til område nær Sverige (Nordland, Trøndelag, Østerdalen, Solør, Ringerike, Austfold) og jamførte tala for desse områda med tilsvarende tal frå eit knippe område i andre delar av landet (Sunnmøre, Hardanger, Jæren, Telemark). Funna i denne undersøkinga må takast med atterhald om at ei heimfesting ikkje alltid tyder at eit ord er typisk for heimfestingsområdet. Ei heimfesting kan til dømes vere oppført fordi det aktuelle området har avvikande uttale av eit ord. Heimfesting signaliserer med andre ord at eit ord blir brukt i eit visst område, men ikkje at ordet *ikkje* blir brukt andre stader. Resultata er illustrerte i tabell 3 og figur 1.

Heimfesting	Førekomstar totalt i Norsk Ordbog	Samførekomstar med tilvising til svensk	
		I tal	I %
1 Rom. (Romerike)	231	87	37,7 %
2 Solør	255	89	34,9 %
3 Østerd. (Østerdalen)	598	193	32,3 %
4 Smaal. (Austfold)	989	262	26,5 %
5 Trondh. (Trøndelag)	1180	310	26,3 %

6 Jæd. (Jæren)	402	94	23,4 %
7 Nordl. (Nordland)	2063	436	21,1 %
8 Hard. (Hardanger)	1019	168	16,5 %
9 Tel. (Telemark) <sup>a</sup>	3811	624	16,4 %
10 Sdm. (Sunnmøre)	2987	460	15,4 %

a) Når Aasen heimfeste ord til Telemark, meinte han truleg dei øvre delane av fylket. Difor er den sørlegaste delen av fylket, landskapet Grenland, ikkje markert på kartet.

Tabell 3. *Førekomstar av oppslag med eit utval norske heimfestingar og av oppslag med både den norske heimfestinga og tilvising til svensk*



Figur 1. *Konsentrasjon av ord som Aasen peiker ut som norsk-svenske, blant heimfeste ord i eit utval heimfestingsområde i Norsk Ordbog (Kartgrunnlag: Statens kartverk)*

I tabellen er heimfestingane plasserte i fallande rekkjefølgje etter kor stor prosentdel av dei som førekjem saman med tilvising til svensk. Kolonne nummer to frå venstre syner totalførekomsten av kvar av heimfestingane, medan kolonnane til høgre syner talet på samførekomstar med tilvising til svensk i høvesvis absolute tal og i prosent av totalførekomsten. På kartet representerer mørkegrått dei fem heimfestingsområda med høgast konsentrasjon av fellesord i utvalet vårt, medan lysegrått representerer område med lågare konsentrasjon av slike ord.

Resultatet er i tråd med hypotesen: Dei fem heimfestingsområda med høgast konsentrasjon av tilvisingar til svensk er alle grenseområde, og berre grenseområdet Nordland, som hamnar under Jæren, vik av frå trenden. Mest slåande er den høge konsentrasjonen av tilvisingar blant orda med tilknytning til Østerdalen, Solør og Romerike. Meir enn tredjeparten av orda som blir heim-

feste til denne delen av landet, blir også jamførte med svenske ord. Dette tyder på at målføra i desse områda har eit vesentleg innslag av fellesord med svensk. På andre sida skal ein ikkje underslå at område i andre delar av landet også har sams ordtilfang med svensk. Det syner seg til dømes at dei høgfrekvente heimfestingane for Telemark (3811 oppslag) og Sunnmøre (2987 oppslag) dukkar opp i høvesvis 22,6 % og 16,6 % av oppslaga med tilvising til svensk. Naturlegvis må tala i denne undersøkinga takast med atterhald når det gjeld kva dei seier om faktisk utbreiing av norsk-svenske fellesord i norske målføre. For det første er det som nemnt ikkje eintydig kva ei norsk heimfesting seier om utbreiinga av eit ord, og for det andre er tala ufullstendige for område utanom svenskegrensa. Likevel syner undersøkinga tendensar som kan vere interessante å følgje opp.

## 6. Sluttord

I denne artikkelen har vi synt korleis Aasen i 1873-ordboka med eit geografisk utgangspunkt dokumenterte eit felles norsk-svensk ordtilfang. Det var med på å gje særmerke til og å godtgjere eit eige norsk språk i jamføring med dansk.

Men vi har også sett at han bygde på eit leksikografisk hovudverk i dansk, Molbechs ordbok. I fullføringa av prosjektet til Aasen var Dalins ordbok ei hovudkjelde. På Aasens tid var dette dei opplagde vala både som mønster og kjelder. Det var kvalitativt gode val, ikkje minst med tanke på tydingskildringa som er hovudsaka i morsmålsleksikografi. Både Molbech og Dalin har fått ettermæle som pålitelege og presise definisjonsutformarar. Lindegård Hjorth (1983: 30) går god for Molbech og Sven-Göran Malmgren (1998) har gått god for Dalin. Ei tilsvarende omdømming får Aasen av Alf Hellevik (1968: 202). Samanfattande kan vi seie at Molbech, Dalin og Aasen danna den første meisterklassa skandinavisk leksikografi.

## LITTERATUR

- Dalin, A. F., 1850–1855: Ordbok öfver svenska språket. Stockholm: Författarens förlag. Også tilgjengeleg på nett hos Språkbanken: <http://spraakbanken.gu.se/dalin/index2.html> (mai 2011).
- Fritzner, Johan, 1867: Ordbog over det gamle norsk Sprog. Kristiania: Feilberg og Landmark. Tilgjengeleg på nett hos Nasjonalbiblioteket: <http://www.nb.no> (september 2011).

- Grimm, Jacob, 1848: *Geschichte der deutschen Sprache*. Leipzig: Weidemannschen Buchhandlung.
- Hauge, Hans, 2001: Postkolonialisme. I: Johannes Fibiger et al. (red.), *Litteraturens tilgange – metodiske angrebspunkter*. København: Gads Forlag. S. 369–395.
- Hellevik, Alf, 1968: *Ords-kattar*. I: *Det Norske Samlaget. 1868–1968*. Oslo: Det Norske Samlaget. S. 201–214.
- Hovdhaugen, Even et al., 2000: *The History of Linguistics in the Nordic Countries*. Helsinki: Societas Scientiarum Fennica.
- Janson, Tore, 2010: *Språkens historia*. Stockholm: Norstedts.
- Kruken, Kristoffer og Terje Aarset, 2003: *Ivar Aasen. Norsk Ordbog med dansk Forklaring*. Ny utgåve. Oslo: Det Norske Samlaget.
- Lindgård Hjort, Paul, 1983: *Danske ordbøgers historie*. København: Reizel.
- Malmgren, Sven-Göran, 1998: Almqvist, Dalin och den svenska definitionsordbokens födelse. I: Gertrud Pettersson (red.), *Studier i svensk språkhistoria*. Lund: Lund University Press. S. 195–213.
- Molbech, Christian, 1833: *Dansk Ordbog indeholdende det danske Sprogs Stammeord, tilligemed afledede og sammensatte Ord, efter den nuværende Sprogbrug forklarede i deres forskellige Betydninger, og ved Talemaader og Exempler oplyste*. Kiøbenhavn: Gyldendalske Boghandlings Forlag.
- Rask, Rasmus Kristian, 1811: *Vejledning til det Islandske eller gamle Nordiske Sprog*. Kjøbenhavn: Schuboths Forlag.
- Rietz, Johan Ernst, 1867: *Svenskt dialekt-lexikon*. Ordbok öfver svenska allmoge-språket. Malmö: B.A. Cronholm. Tilgjengeleg på nett hos Projekt Runeberg: <http://runeberg.org/dialekt/> (mai 2011).
- Rydqvist, Johan Er., 1850–1883: *Svenska språkets lagar: kritisk afhandling*. I–VI. Stockholm.
- Torp, Arne og Lars S. Vikør, 1993: *Hovuddrag i norsk språkhistorie*. Oslo: Ad Notam Gyldendal.
- Venås, Kjell, 1994: Ivar Aasen om norsk i høve til andre nordiske mål. I: Jens Allwood et al. (red.), *Gothenburg Papers in Theoretical Linguistics 69 II*. Göteborg. S. 385–397.
- 1996: *Då tida var fullkomen*. Ivar Aasen. Oslo: Novus forlag.
- Walton, Stephen J., 1996: *Ivar Aasens kropp*. Oslo: Det Norske Samlaget.
- Worren, Dagfinn, 2006: Molbech som mønster for Aasen. I: Henrik Lorentzen og Lars Trap-Jensen (red.), *Nordiske studier i leksikografi 8. Rapport fra Konferanse om leksikografi i Norden*. Sønderborg 24.–28. mai 2005. København. S. 391–406.
- Aasen, Ivar, 1850: *Ordbog over det norske Folkesprog*. Kristiania: Carl C. Werner & Comp.
- 1873: *Norsk Ordbog med dansk Forklaring*. Christiania: P.T. Mallings Boghandel.
- 1912: *Skrifter i Samling. III*. Kristiania og Kjøbenhavn. Gyldendalske Boghandel. Nordisk Forlag.
- 1965 (1. utg. 1864): *Norsk Grammatik*. Oslo: Universitetsforlaget.

*Kjetil Gundersen*

Norsk Ordbok 2014, Institutt for lingvistiske og nordiske studium, Universitetet i Oslo  
kjetil.gundersen@iln.uio.no

*Dagfinn Worren*

Norsk Ordbok 2014, Institutt for lingvistiske og nordiske studium, Universitetet i Oslo  
dagfinn.worren@iln.uio.no

# Ordbog over Dansk Talesprog (ODT) – værkstedspræsentation af en ordbogsportal

*Carsten Hansen & Martin H. Hansen*

The article presents the initial investigating stages of the development of a dictionary of spoken Danish, based on a corpus consisting mainly of sociolinguistic interviews at the LANCHART Centre in Copenhagen. The data are stratified and represent different segments of Danish speakers with regard to geography, gender, age and social class, recorded over several decades.

The purpose of the dictionary is to present spoken Danish to the interested layperson as well as the research-oriented user, and the dictionary will be designed as an online portal with these recipients in mind.

The article presents several examples of work in progress: establishing the editing principles of linguistic fact boxes which will give a more detailed presentation of a specific topic; a sketch of the portal design; a pilot project on interjections with an, as yet, unfinished dictionary entry on the lemma *av* ('ouch'); and finally, an investigation of laughter as part of spoken language which also demonstrates how the tagging of a sociolinguistic corpus can be put to use in a dictionary.

*Nøgleord:* Leksikografi, Talesprogs-korpus, Pragmatik, Konversationsanalyse

## ODT og Sprogforandringscentret

På basis af et talesprogs-korpus bestående af primært sociolingvistiske interviews optaget i perioden 1978-2010 og som i den ortografiske transskription tæller op mod 7 mio. løbende ord, arbejder vi på Danmarks Grundforskningsfonds Center for Sociolingvistiske Sprogforandringsstudier, i daglig tale Sprogforandringscentret, på en ordbogsportal kaldet *Ordbog over Dansk Talesprog* (ODT). Centrets baggrund og formål blev ved åbningen i 2005 formuleret sådan af centerleder Frans Gregersen:

Vi vil opbygge en ny sprogforandringsteori baseret på empirisk forskning i sprogforandring i virkelig tid inden for det danske sprogsamfund. Danmark er som sprogsamfund et meget velegnet laboratorium for en undersøgelse af hvordan sprog forandrer sig og

hvorfor. Landet er overskuelig stort, det er veldokumenteret både i sproglig, i historisk og sociologisk henseende, og der er gennemført en række undersøgelser af dansk tale-sprog i dets variationer, geografisk og socialt. Alle disse karakteristika gør det oplagt at svare på grundlæggende spørgsmål om hvordan sproglig variation og sprogforandring hænger sammen.

Vi ved at sproget bruges forskelligt af de forskellige aldersgrupper, af de to eller flere køn og af socialt og geografisk forskellige personer, men hvordan hænger det sammen med de langtrækkende tendenser i det danske sprogsamfunds forandring fra et dialekt-samfund med få store byer omkring århundredskiftet fra 1800 til 1900, og til et gennemurbaniseret informationssamfund med mange minoritetssprog omkring årtusindskiftet?

Vi har på centret gentaget tidligere undersøgelser med så vidt muligt samme informanter, vi har suppleret med informanter der har samme baggrund, og endelig har og vil vi fortsat afprøve en række hypoteser og fordomme på dette materiale. Data er derfor omhyggelige genindspilninger af de informanter som tidligere medvirkede i syv forskellige dialektologiske og sociolingvistiske undersøgelser fra forskellige dele af Danmark, nemlig



- Vinderup, Jylland: båndoptagelser udført i 1973, 1978 og igen i 2006-2008
- Vissenbjerg, Fyn: optagelser i 1982-83 og igen i 1998-99
- Odder, Jylland: optagelser i 1986-89 og igen i 2005-2008
- BySoc (Projekt Bysociolingvistik), København: optagelser i 1986-89 og igen i 2005-2009
- Næstved, Sjælland: optagelser i 1986-90, 1993 og igen i 2005-2007
- Køge, Sjælland: optagelser i 1989-98 og igen i 2006-2007
- DASVA (Dansk og Svensk i Akkomodation), København: optagelser i 2001-2002, endnu ikke gentaget.

Figur 1. Undersøgelsesområder

For detaljer om de enkelte undersøgelser design, se oversigten på Sprogforandringscentrets hjemmeside: <http://dgcss.hum.ku.dk/undersoegelsesomraader/>

Gentagelsen har givet os data om sprogforandringer *i virkelig tid*, dvs. de samme taleres forandrede – eller konstante – tale efter 30, 20 og 5 år, henholdsvis.



Normalt arbejder man med aldersfordelte informantsamlinger så man slutter fra forskellene mellem aldersgrupperne: Sådan som de unge taler, bliver sproget i fremtiden når de gamle er forsvundet. Det kalder man at studere sprogforandringer i forestillet tid (*apparent time*). Det vi gør, er altså at studere sprogforandring i virkelig tid (*real time*), heraf projektets engelske kælenavn: LANCHART, Language Change in Real Time.<sup>1</sup>

Siden centrets åbning er nye undersøgelser kommet til, således at der nu som nævnt foreligger et transskriberet korpus med op mod 7 mio. løbende ord. Der ligger 1587 timer, 33 minutter og 25 sekunders lyd i Sprogforandringscentrets korpus, og der er brugt mere end 17.000 timer på transskription og opmærkning.

## Ånden i ODT

ODT's virkefelt er dobbelt. For det for første bedriver vi i redaktionen egen *forskning*, dels i talesprog på linje med Sprogforandringscentrets øvrige aktiviteter og dels i det uopdyrkede felt talesprogsleksikografi. På den måde *producerer* redaktionen, tilknyttede specialeskrivere og ph.d.-studerende ny viden. Den anden side af vores virkefelt er *formidling*. Her ser vi det som vores opgave ikke blot at frugtbargøre og eksponere centrets og redaktionens egne forskningsaktiviteter, men opgaven består også i at formidle talesprogsforskning og viden om talesprog indvundet uden for centret. Imidlertid skal der ikke herske nogen tvivl om at den bærende ide i vores arbejde er etableringen af talesprogsportalen, dvs. praktiseret talesprogsleksikografi.

ODT's brugersegment er også dobbelt. *Primæradressaten* er den sprogligt/lingvistisk dannede alménbruger der interesserer sig for talesprog, *sekundæradressaten* er den forskningsorienterede bruger. Begge adressater skal opleve portalen som tilgængelig (dvs. med hurtig accesstid), segmentdifferentieret (dvs. relevant), servicerende, reklamefri og gratis. Det skal i det hele taget være en positiv oplevelse at frekventere portalen.

ODT er sig sin rolle som dataredaktør bevidst. Ud over de redigerede ordbogsartikler vil ordbogsportalen indeholde en række bokse med oversigts- hhv. dybdegående artikler om talesprogsfænomener dokumenteret i korpus og relateret til ordbogsartiklerne. Brugeren kan i den endelige version gå bag om de redigerede fremstillinger såvel i ordbogsartikler som i bokse ved at foretage direkte søgninger i korpus, men grundlæggende dokumenteres søgningerne i artiklens dokumentationsdel.

---

<sup>1</sup> Den oprindelige ordlyd er tilpasset nærværende tekst.

ODT insisterer på at demonstrere at en ordbog der konsekvent er baseret på talesprogsdata, på afgørende punkter adskiller sig fra den skriftsprogsbaserede ordbog. Det drejer sig for det første i hvert fald om lemmabestanden – også når den skriftsprogsbaserede ordbog medtager pseudotalesproglige kilder som fx fiktionsdialoger. Det drejer sig dernæst om morfologiske paradigmer, syntaktiske mønstre, betydningsvarianter og om de pragmatiske og diskursive niveauer.

## Bokse

Sprogbokse skal formidle en (kortfattet) karakteristik af det talesproglige fænomen. Når ODT's interne redaktører og de eksterne boksforfattere skal redigere en boks, er de forpligtede på *boksmanualen*:

### Fremstillingen skal

- karakterisere det talesproglige fænomen
- funderes i belæg fra korpus
- supplerende funderes kontrastivt i forhold til skriftsprog
- adresseres dobbelt.

### Fremstillingen kan

- indeholde supplerende karakteristik på baggrund af andet end korpus. Det angives eksplicit når karakteristikken baseres på andet end centrets korpus
- inddrage forskning i fænomenet.

Det talesproglige fænomen kan være stort set alle aspekter af talesprog. Det kan være et leksem: *ah, hov, jamen, nå, undskyld*, det kan være en funktionsgruppe: *begejstring*, eller det kan være noget andet talesprogsrelevant: *latter, pauser* el. *-agtig*. I redaktionen er vi løbende på jagt efter boksemner og trækker gerne på kræfter uden for redaktionen. Blandt Sprogforandringscentrets forskere har vi sikret os bokse om følgende emner i tilknytning til deres igangværende forskning: *Epistemiske adverbier* (Tanya Karoli Christensen), *at-bortfald* (Kasper Boye), *Genstandsfald* (Jeffrey Parrott), *Generiske pronominer* (Torben Juel Jensen). Vi låner også gerne specialviden fra den relevante faglitteratur, fx *plus at* (Nørgård-Sørensen 2001), *"ja men nej"- og "nej men ja"-svar* (Thøgersen & Nielsen 2009) eller *her'ere og der'ere* (Therkelsen 2002).

## ODT som ordbogsportal

ODT er designet som en netbaseret ordbogsportal med mulighed for kontrastive parallelsøgninger i et skriftsprogskorpus og paralleloplag i skriftsprogsbaserede ordbogsressourcer, dvs. at ODT er et "sted på internettet som fungerer som indgang til søgning efter bestemte oplysninger, idet det indeholder en stor mængde links til forskellige hjemmesider som er opdelt efter emne eller vedrø-

rer et bestemt emne” (Den Danske Ordbog). Dette design afspejler sig i nedenstående udkast til portalens velkomstsider:

## Velkommen til Ordbog over Dansk Talesprog (ODT)

Her kan du få mere at vide om ODT

Her kan du få mere at vide om DGCSS' korpus

Her kan du få mere at vide om undersøgelsesområderne der ligger til grund for DGCSS' korpus

Her kan du få mere at vide om DGCSS

Her kan du få mere at vide om andre danske talesprogskorpora

Her kan du få mere at vide om engelske talesprogskorpora

Her kan du få mere at vide om norske talesprogskorpora

Her kan du få mere at vide om svenske talesprogskorpora

Her kan du få mere at vide om [xxx] talesprogskorpora

Kontakt ODT

Du er besøgende nr. 1236

Indtast søgeord:

ha

*Boks:* Kort og godt om *ha* og *latter*

*Links:*

- Her kan du få mere at vide om *latter* i andre ordbøger, korpora og fremstillinger
- *ha*-konkordans
- *ha*-statistik

## Pilotprojekt om interjektioner

ODT-projektet befinder sig i en indledende eksplorativ fase, både teoretisk og metodisk. I loyalitet med vores talesprogskorpus og det sociolingvistiske interview har vi valgt at gennemføre et pilotprojekt med de entydigt samtalepecifikke tokens der i korpus er annoteret som interjektioner. For at tage højde for eventuelle fejl i den automatiske PoS-opmærkning supplerer vi centrets egen inventarliste over interjektioner med Den Danske Ordbogs interjektionsliste. Vi går i denne eksplorative fase noget lemfældigt til værks mht. den ellers nødvendige kategoriale skelnen mellem interjektioner, onomatopoieta og modalpartikler. Schwitalla (2003) og Fiehler (2005) åbner op for en funktionel sammenfatning af disse grupper i termen *Gesprächspartikeln*.

Vi arbejder med to forskellige procedurer der supplerer hinanden. I den første procedure gennemgår vi de to interjektionslister semasiologisk og funktionsbestemmer hver kandidat: Hvilken slags job udfører kandidaten i samtalen? Fra Adolphs (2008) har vi hentet termen *funktionsprofil*, nemlig det sæt af forskel-

lige men supplerende funktioner der i deres hele udgør leksemets funktionsprofil. Fra Schwitalla (2003: 157) og Fiehler m.fl. (2004: 204ff) har vi hentet inspiration til en egen funktionsgruppetypologi, vel vidende at Schwitallas og Fiehler m.fl.s lister er orienteret mod sekventiel kategorisering, det vil sige mod turtagning i konversationsanalytisk forstand, i modsætning til vores der på nuværende stadium er mere indholdsorienterede.

I den anden procedure supplerer vi de forskellige funktionsgrupper med kandidater fra det semasiologiske analysearbejde. Det tjekkes om funktionsprofilerne kan bidrage med én eller flere funktioner til det samlede funktionsgruppelinventar. Tentative eksempler på vores funktionsgrupper er: *Beklagelse, bekræftelse, bekymring, erkendelse, hilsen, overraskelse (glædelig), overraskelse (forskrækket), selvrettelse, tøven, skepsis, væmmelse*.

Nedenfor er gengivet første version af en artikel over interjektionen *av*, skrevet af Mette Skovmand Juul, en af ODT's studenterredaktører. Artiklen er anonymiseret idet problematikken om informanternes anonymitet i forhold til ODT-portalen endnu ikke er afklaret.

**Lemma: *av***

*av*<sup>1</sup> (overraskelse)

Antal belæg: 40

Betydninger:

1. *av*: bruges for at udtrykke overraskelse eller anden (ofte ubehagelig) følelse: *Av det dur ikke* (Vinderup 2006)
2. *av* for satan/søren/pokker/dælen, *av* min arm: bruges som mildt kraftudtryk for at udtrykke ærgrelse eller overraskelse: *Av for satan en historie altså* (BySoc 2007)

Funktionsgruppe: overraskelse/forskrækkelse, beklagelse

*av*<sup>2</sup> (minimalrespons)

Antal belæg: 27

Betydninger:

1. *av*: ord man siger mens man lytter til en andens tale: *Ja... ja... av... ... laver I andet* (Odder 1988)

Funktionsgruppe: minimalrespons

*av*<sup>3</sup> (medgiven)

Antal belæg: 1

Betydninger:

1. *av* ja: give nogen ret i noget, medgive, indrømme: *Det er det jo det... av ja det er varmt* (Næstved 2006)

Funktionsgruppe: bekræftelse

## Værkstedsrapportering om latter

I vores interjektionsliste finder man også strengen *ha*. I det transskriberede og annoterede korpus er alle lydsekvenser der høres som latter, opmærket på samme måde, nemlig som *ha*. Uanset om der bliver leet højt og længe, om der bliver leet kort og forsigtigt, eller om der faktisk bliver sagt 'ha', står der *ha* i korpus. Det har den fordel at vi hurtigt kan identificere al latter i korpuset, men det har til gengæld den ulempe at det er vanskeligt at skelne mellem forskellige former for latter. Der er i alt cirka 60.000 forekomster af *ha*. Redaktionens opgave består nu i en teoretisk og metodisk afklaring af hvordan vi håndterer latter.

### Det teoretiske afsæt

En traditionel opfattelse består i at opfatte latter som en kausal reaktion på en stimulus, typisk humor. Både Carroll (2001) og Vøge & Wagner (2010) peger på et generelt problem i denne opfattelse, nemlig vanskeligheden ved at forklare hvorfor en bestemt stimulus nogle gange udløser latter og andre gange ikke. Den konversationsanalytiske forskning (CA) antyder en løsning på dette problem idet latter faktisk følger et forudsigeligt mønster når vi betragter den som interaktion. Desuden tyder CA-resultater på at latter alligevel er en slags sprog: Ganske vist udfører latter ikke en semantisk reference til samtaleens indhold, men den udfører en pragmatisk funktion i samtalsituationen. Latter er med andre ord en systematisk produceret, socialt organiseret aktivitet der tilpasser sig eller påvirker den løbende samtale – den er interaktionelt funderet (Jefferson m.fl. 1987).

### Fra kvalitativ til kvantitativ analyse

CA har anvendt en kvalitativ metodisk tilgang til at analysere latter: Hvem ler først? Hvilken funktion har det at nogen ler først? Ler man med hinanden eller ad hinanden? Fortæller man noget sjovt eller alvorligt mens man ler – og hvilken effekt har dette for hvordan samtaleren bør reagere? Et meget stort antal isolerede analyser har givet stof til udvikling af en latterens typologi. Men man har derimod ikke noget kvantitativt overblik over de forskellige lattertypers udbredelse i forskellige kontekster, og dette manglende overblik indebærer også

at man ikke har mulighed for at vide om man har alle væsentlige latter typer og deres eventuelle blandingsformer med i typologien.

Sprogforandringscentrets talesprogs korpus kan bidrage til løsningen af disse spørgsmål. Korpusset er som nævnt designet med henblik på studiet af indflydelsen på talesprogspraksis fra de traditionelle sociolingvistiske variable køn, alder, klasse, men især tid og sted. En refleksion og inddragelse af disse træk hører netop til ODT's konstituerende træk. Konkret: Bliver der leet mere af kvinder end af mænd, af unge end af ældre, af arbejdere end af middelklasse, i 70'erne end i 00'erne, i Vinderup end i København?

Det store problem ved en kvantitativ analyse af sprogets pragmatiske funktioner er at der ikke er et simpelt én-til-én-forhold mellem form og funktion. Man må derfor i hvert enkelt tilfælde finde og systematisere de forskellige typer af form-funktion-relationer der gør sig gældende for det konkrete, sproglige udtryk der skal undersøges (i dette tilfælde latter), og derefter beskrive hvilke funktioner udtrykket typisk udfylder. Her har vi adopteret konceptet *funktionsprofil* fra den engelske korpuspragmatiker Svenja Adolphs:

The development of functional profiles [...] tries to shift the focus from inference procedures to form/function composites. The relationship between form and function is here described in terms of recurrent associations in language-in-use. These associations become part of the functional profile [...]. (Adolphs 2008: 29-30)

Et vigtigt aspekt af ODT's fortsatte arbejde med latter, og med multifunktionelle talesprogsudtryk i det hele taget, bliver at udarbejde sådanne funktionsprofiler.

## Præsentations- og redigeringsværktøjer

Vi har Ole-Opfinderagtigt udviklet en række værktøjer mens vi venter på at der via eksterne midler kan knyttes leksikografisk it-kompetence til ordbogen. Nogle af værktøjerne vil vi integrere i portalen som præsentationsværktøjer, mens andre tjener som redigeringsværktøjer. Desuden har vi værktøjer af mere eksperimentel karakter, værktøjer som vi endnu ikke er sikre på hvordan vi vil bruge, hvis overhovedet. Vi befinder os som sagt i en eksplorativ fase.

På ordbogsportalens velkomstsider ovenfor så vi to menu punkter: *Konkordans* og *Statistik*. En reduceret konkordans med *ha*-forekomster ser således ud:

Fund	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
1.på	nogen	måder		nej		mm		ja	okay	ha		ja		ja		ja		okay		nej	
2	er	da	noget	forkalket	ham	den	ene	er	han	ikke	ha	hvad		nej	ja	vi	ville	sgu	ikke	med	
4																					
5	ja	nåh	jamen	det	kunne	også	være	meget	sjovt	ha	men	der	har	jeg	nemlig	sådan	en	her	øh	får	
6	nej	ha	vi	ville	sgu	ikke	med	ind	ja	ha	Lady	er	det	ikke	sådan	den	hedder	ha	ha		
7	nå	ej	hvør	irriterende	så	ja	ikke	med	ind	ja	ja	og	det	er	doktor	hjelpeles	ja	ha			
8	betændt	skal	vi	ikke	snakke	om	noget	andet	ha	ja	og	det	er	meget	fnst	det	er	meget			
9	ham	den	ene	er	han	ikke	ha	hvad	nej	ha	vi	ville	sgu	ikke	med	ind	ha				
10	ikke	det	er	lidt	tåbeligt	jo	ej	dog	ikke	ha	dog	ikke	nu	der	bryder	jeg					
11	hedder	ha	ha	ha	ha	ha	ha	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg			
12	sådan	den	hedder	ha	ha	ha	ha	ha	ha	ha	men	skriver	du	selv	sangtekst	da	for	eksempel	ja		
13	ha																				
14	Thor	Timler	han	kan	jo	en	masse		uh	ha	jamen	nu	har	han	jo	igennem	mange	mange	år		
15	ha	ha	ha	ha	ha	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg	var	på			
16	lektioner	som	vi	har	mm	mm	mm	mm	timer	ha	ha	mm	mm	mm	mm	mm	mm	sådan	det		
17	og	hvad	med	computer	bruger	du	det	meget	den	hedder	ha	ha	ha	ha	ha	ha	ha	ha	ha		
18	ha	Lady	er	det	ikke	sådan	det	er	hedder	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha		
19	på	en	eller	anden	dårlig	måde	og	ja	ha	der	har	også	de	ha	har	også	lavet	en			
20	det	ikke	sådan	den	hedder	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha		

Figur 2. Reduceret ha-konkordans

Som anført ovenfor er al latter i alle lydsekvenser der høres som latter, opmærket på samme måde, nemlig som *ha*. De tomme felter er pauser, men pauser skal i denne sammenhæng forstås som intervaller hvor nogen andre taler. De enkelte konkordanslinjer viser kun den enkelte talers tale – konkordansen består kun af tale, ikke af *samtale*. Et af de redskaber vi p.t. mangler i vores værktøjskasse, er således en let generérbar samtalekonkordans for at kunne klassificere *ha*-sekvenserne funktionelt. Her er der ikke bare tale om et internt redigeringsbehov men også om et brugerfladebehov: Brugeren skal kunne finde eksempler på de forskellige latterfunktioner og skal kunne vælge at høre lyden og/eller læse dialogen. Dels som belæg i forbindelse med den redigerede funktionsbestemmelse og funktionsprofilering men også som suveræn brugerkontrol af de redigerede fremstillinger i ordbogsartikler og bøger.

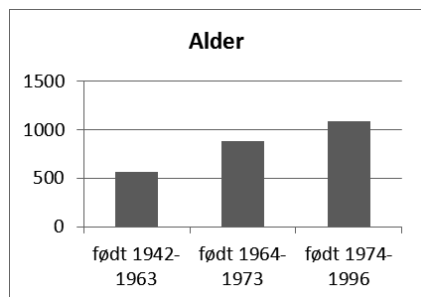
Konkordansen i Figur 2 er imidlertid også utilstrækkelig i en anden henseende. Hvis korpus virkelig skal ydes retfærdighed for sine sociolingvistiske variable skal det – ikke bare som redaktør men også som bruger – være muligt at generere en udvidet konkordans med de sociolingvistiske variable køn, alder, klasse, tid og sted samt andre opmærkninger:

	-2	-1	0	1	2	3	4	5	6	7	8	9	10	filename	project	recording	recording	participant	gender	birthyear	socialclass	xmin	
ja	okay	ha	ja	okay	ja	ja	ja	okay	nej	byso2ny:BySoct	ny	2006	XMM	MALE	0	1743	1886						
han	ikke	ha	hvad	nej	ha	vi	ville	sgu	ikke	med	vinderup2	Vinderup:gl	1978	AFH	FEMALE	1964	2024	84					
meget	sjovt	ha	det	er	okay	nogle	gange	fordi	man	får	amager3	Amager3	gl	2009	FWA	FEMALE	0	8,75					
ind	ha	ha	men	der	har	jeg	nemlig	sådan	en	her	øh	vinderup2	No name	ny	2007	XMF	FEMALE	0	3705	108			
ja	ha	ha	men	Lady	er	det	ikke	sådan	den	hedder	ha	vinderup2	Vinderup:gl	1978	AFH	FEMALE	1964	2030	031				
and	ha	ha	ja	og	det	er	doktor	hjelpeles	ja	ha	byso2ny:No name	ny	2006	XJM	MALE	0	2651	403					
nej	ha	ha	vi	ville	sgu	ikke	med	ind	ha	vinderup2	Vinderup:gl	1978	AFH	FEMALE	1964	2027	655						
dog	ikke	ha	men	der	har	jeg	nemlig	sådan	en	her	øh	vinderup2	No name	ny	2007	XMF	FEMALE	0	3705	108			
ja	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg	byso2gl:BySoct	gl	1987	MIP	MALE	1970	MIDDLE	C	3506	155		
ha	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg	vinderup2	Vinderup:gl	1978	AFH	FEMALE	1964	2073	075				
ja	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg	vinderup2	Vinderup:gl	1978	AFH	FEMALE	1964	2073	075				
ha	ha	ha	men	skriver	du	selv	sangtekst	da	for	eksempel	ja	byso2gl:No name	gl	2010	XMA	FEMALE	0	1493	07				
uh	ha	ha	jamen	nu	har	han	jo	igennem	mange	mange	år	byso2gl:BySoct	gl	1987	ATR	MALE	1932	WORKING	4469	42			
ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg	var	på	vinderup2	Vinderup:gl	1978	AFH	FEMALE	1964	2087	021				
timer	ha	ha	ja	ja	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm	mm
ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha
ja	ha	ha	der	har	også	de	har	også	lavet	en	odder3-05	Odder3	2006	SMP	FEMALE	1989	WORKING	2475	715				
ja	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha

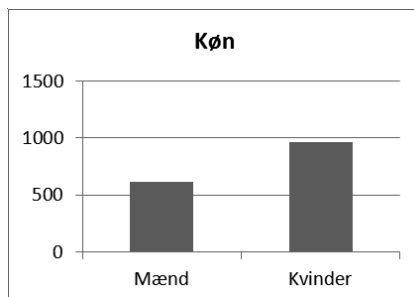
Figur 3. Udvidet ha-konkordans

De sociolingvistiske variable fortjener imidlertid at blive statistisk raffineret:

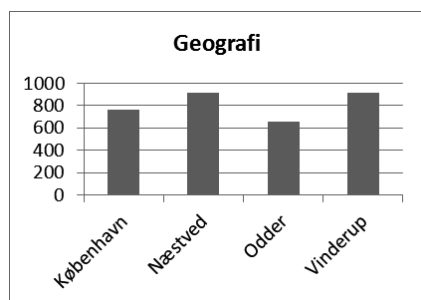
Figur 4–7: Relative frekvenser pr. 100.000 tokens for søgestrengen *ha*



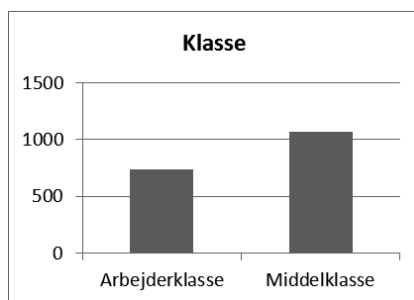
Figur 4



Figur 5



Figur 6



Figur 7

I Figur 4–7 ser vi i formidlet form de relative frekvenser for søgestrengen *ha* fordelt på nogle af de sociolingvistiske variable. Nu kan vi se hvem det er der ler: Unge ler mere end gamle, kvinder ler mere end mænd, middelklassen ler mere end arbejderklassen, og endelig bliver der tilsyneladende leet mere i Næstved og Vinderup end i København og Odder.

Nogle af de her viste forskelle skal nok forklares med at korpus er sammensat af geografiske og tidsmæssige punktnedslag og dermed i streng forstand ikke er repræsentativt for dansk talesprog. En forklaring på de geografiske forskelle kunne være at man har gennemført forskellige typer af sociolingvistiske undersøgelser på de forskellige lokaliteter. Men vi kan faktisk, i hvert fald i store træk, tegne en profil af hvilke sprogbrugere der ler mest, ligesom vi kan generere tilsvarende præsentationer for alle andre søgeord ved hjælp af samme skabelon. Den dag hvor vi får opmærket de forskellige latterfunktioner i korpus, kan vi tilsvarende generere individuelle præsentationer for hver latterfunktion.



## Afsluttende bemærkninger

Afslutningsvis vil vi pege på to udfordringer. Den første er af økonomisk art. Vi har her i projektets indledende eksplorative fase trukket på både centerinterne og centereksterne ressourcer. På længere sigt er det imidlertid en forudsætning at vi får tilført eksterne midler til et så ambitiøst projekt som ODT, både til det redaktionelle arbejde og til it-løsninger.

En anden udfordring i arbejdet med ODT består i en række teoretiske og metodiske afklaringer: Kan det forsvars udelukkende at udarbejde ordbogen på basis af en ortografisk transskription, eller skal lyd inddrages både i skrivning af artikler og i dokumentationen? Er det rimeligt at overtage det literale lemmabegreb, eller er det nødvendigt at etablere et oralt? Endelig er vi overbeviste om at ODT på længere sigt må inddrage multimodale dimensioner.

## LITTERATUR

- Adolphs, Svenja, 2008: *Corpus and Context. Investigating pragmatic functions in spoken discourse*. Amsterdam.
- Carroll, Noël, 2001: *Horror and Humor. I: Beyond aesthetics*. Cambridge. S. 235-254.
- Fiehler, Reinhard, 2005: *Die Gesprächspartikel. I: Duden Band 4, 7. Auflage*. S. 601-606.
- Fiehler, Reinhard (m.fl.), 2004: *Eigenschaften gesprochener Sprache*. Tübingen.
- Jefferson, Gail (m.fl.), 1987: *Notes on laughter in the pursuit of intimacy. I: G. Button and J.R.E. Lee (red.): Talk and Social Organization*. Clevedon. S. 152-205.
- Nørgård-Sørensen, Jens, 2001: *Plus at – en ny konjunktion i dansk*. *Danske Studier* 2001. S. 65-84.
- Schwitalla, Johannes, 2003: *Gesprochenes Deutsch. Eine Einführung*. Berlin.
- Therkelsen, Rita, 2002: *Hvad laver de der her'ere og der'ere egentlig? I: 9. Møde om Udforskningen af Dansk Sprog*.
- Thøgersen, Jakob & Nielsen, Søren Beck, 2009: *Ja men nej og nej men ja*. Når man hverken be- eller afkræfter ja/nej-spørgsmålets antagelser. Eller: Det fuldt udfoldede 'nja'. I: Hovmark, H., Stampe Sletten, I., & Gudiksen, A. (red.): *I mund og bog: 25 artikler om sprog tilegnet Inge Lise Pedersen på 70-årsdagen d. 5. juni 2009*. København. S. 337-350.
- Vøge, Monika & Wagner, Johannes, 2010: *Social Achievements and Sequential Organization of Laughter. I: Journal of Pragmatics* 42. S.1469-1473.

### Internet

Den Danske Ordbog: <http://ordnet.dk/ddo>

Sprogforandringscentret: <http://dgcss.hum.ku.dk>

*Carsten Hansen*

Sprogforandringscentret DGCSS, Københavns Universitet  
carhan@hum.ku.dk

*Martin H. Hansen*

Sprogforandringscentret DGCSS, Københavns Universitet  
martin\_h\_hansen@yahoo.dk

# Islandsk-tsjekkisk / Tsjekkisk-islandsk ordbok (IT/TI)

**En biskopal, bidireksjonal ordbok i stort lommeformat**

*Helgi Haraldsson*

The authors of the two-way Icelandic-Czech / Czech-Icelandic Dictionary are Helgi Haraldsson (Oslo) and Marta Bartošková (Prague). The estimated volume is about around 25,000 lemmas for each language.

The dictionary is bidirectional in the sense that it is intended to be equally useful to both Czech and Icelandic users. This entails concise phonetics and morphology for both languages. Lemmas and equivalents are given codes that refer to the morphology, thus providing adequate information on inflection and conjugation. Particular attention is paid to syntax, especially valency and the like. Various devices are applied to save space and avoid unnecessary repetitions; this includes the use of 'lexical sets' as described by Bo Svendsen in his *Practical Lexicography* (1993). Examples to elucidate the use of words in context are given insofar as space allows. (1003)

Existing Icelandic-Czech and Czech-Icelandic lexicographic works are typically monodirectional and do not meet the needs of Icelandic users.

*Nøkkelord:* Tospråklig leksikografi, tsjekkisk, islandsk, ordboksgrammatikk

## 1. Innledning

I 2005 henvendte dr. Marta Bartošková, Praha, seg til prof. Helgi Haraldsson ved Universitetet i Oslo med forslag om samarbeid ved utarbeidelse av en Islandsk-tsjekkisk/Tsjekkisk-islandsk ordbok. Dette førte omsider til at i november 2009 undertegnet disse to en avtale om utgivelse med forlaget „Leda“ i Praha. Manus skal leveres senest 31.12.2012.

## 2. Eksisterende ordbøker/prosjekter

- 1) Kupča, Vojtěch, 2008: Islandsko-český slovník. “Litera proxima. Průhonic”. 415 sider. Denne ordbok har en internetversjon:
- 2) Internetová verze: [www.isafold.cz/ordabok](http://www.isafold.cz/ordabok) som også har en tsjekkisk-islandsk del.
- 3) Videre fins det på internet en “skoleordbok”: A. Chejn: Islandsko-český studijní slovník.  
<http://www.hvalur.org/>. Denne nettsiden og verk nr. 2 er under fortløpende utvikling.

Felles for alle disse verk er at de er utpreget monodireksjonale, dvs. primært beregnet på tsjekkiskspråklige brukere. Morfologiske og syntaktiske opplysninger for islandskspråklige brukere er i praksis ikke-eksisterende, samt delbetydningsforklaringer i den islandske delen. Likeledes savner man stilmarkører.

Til gjengjeld har alle lemmaer i verk nr. 3 en automatisk link til Islands universitets nettside “Beygingar-lýsing íslensks nútímamáls” samt link med *viva voce* uttale (ved Jón Gíslason). Dette verket inneholder likedan vesentlig flere kombinatoriske opplysninger og brukseksempler enn de øvrige og er således et godt redskap for tsjekkiske brukere.

For islandske brukere er den til begrenset nytte.

Kupčas ordbok har presise oversettelser av teknisk og naturvitenskapelig ordforråd. Ellers inneholder den stor mengde helt marginale gloser mens en del aktuelle ord og uttrykk savnes. Feilaktige oversettelser forekommer.

## 3. Finansiering

Haraldsson er pensjonist og har frasagt seg slutthonorar fra forlaget (30.000 CZK, ca 1.226 Euro). Bartošková har tjent til livets opphold vha. oversettelser, tolking o.l. For å kunne avse tid til ordboken, måtte hun sikres minimal inntekt mens ordboksarbeidet pågår. Dette lyktes takket være rause islandske bevilgninger, bl.a. fra Det nasjonale jubileumsfond „Þjóðhátíðarsjóður“ (som ga 500.000 ISK, ca 3000 Euro) og Alting (1.000.000 ISK). Før dette (og før den sterke devalueringen av ISK i 2008!) bidro private personer med anselige beløp, f.eks. en høyskolerektor (500.000 ISK) og en pensjonert ingeniør (utdannet i Praha) med 600.000 ISK. Dermed er det grunn til å tro at finansieringen er sikret.

#### 4. Makro- og mikrostruktur

I en biskopal ordbok av beskjedent omfang er plassbesparing en sentral oppgave. Dette oppnås bl.a. ved nisjealfabetisk makrostruktur, der flere direkte på hverandre følgende sammensatte ord anføres i fortløpende alfabetisk rekkefølge, der tilda (-) representerer lemmadelen.

Artikkelstrukturen til et omfattende verb-lemma kan f.eks. se slik ut:

**verb**<sup>bøyningsindeks</sup> *reksjon (valens)* (dersom den gjelder alle delbetydninger, se pkt 7, vysypat) **1** (*delbetydningsforklaring*) ekvivalent(er)... **2** (*delbetydningsforklaring*) ... ◆ (romb) *fraseologi* (idiomer, ordtak, ordspråk, bevingede ord o.l.) ... ■ *partikkelverb* (kun i den islandsk-tsjekkiske delen) ... ▲ *henvisninger til andre artikler*.

Man regner med ca 25000 lemmaer hver vei, men bokens omfang gjør at plass for brukseksempler er begrenset. Dette stiller enda større krav til utformingen av slike eksempler i den grad det overhodet er rom for dem.

Jo flere brukseksempler, desto bedre ordbok. I større ordbøker kan man ofte tillate seg å bruke ett forklarende eksempel for å belyse ett viktig poeng, grammatisk, syntaktisk o.l, men i IT/TI er det uunngåelig å tilstrebe mest mulig polyvalente eksempler, dvs. "slå flest mulig fluer i ett smekk".

#### 5. Grammatiske opplysninger

For begge språk brukes bøyningsindekser (i superscript) som henviser til en kortfattet ordboksgrammatikk, mest i form av paradigmer. Indeksene består av tall og – i islandske indekser – evt. også bokstaver. For tsjekkiske substantiver og verb brukes i utgangspunktet samme morfologiske tabeller som i den biskopale ordboken *Praktický slovník norsk-český/česko-norský*. „Leda“ 2005, i noe modifisert form. Den islandske morfologiske oversikten bygger på ordboksgrammatikken i *Rússnesk-íslensk orðabók* (se også Haraldsson 2002), der den legger beslag på 48 sider. Det er derfor en komplisert oppgave å komprimere denne grammatikken til et omfang som er akseptabel for IT/TI, og arbeidet med dette pågår fortsatt. Likedan må avsnittet om islandsk fonetikk forkortes sterkt. I Valerij Berkovs islandsk-russiske ordbok (1962) er avsnittet om islandsk fonetikk på 23 sider.

#### 6. Eksempler på artikler

Fraseologiske enheter og andre faste uttrykk anføres kun én gang under det lem-maet som utgjør det (mest) sentrale ordet i ordforbindelsen. Fra artikler til

andre andre ord som inngår i frasen, henvises det etter et trekanttegn til den lemmaatikkelen som inneholder frasen. Dette for å unngå dobbeltoversettelse av same uttrykk/ordforbindelse. Eksempel:

**strach**  $m^5$  ótti  $m^1$ , hræðsla  $f^1$  [...] ▲ oko ...

viser til artikkelen **oko** som inneholder frasen „strach má velké oči“, ordrett „frykten har store øyne“, dvs. den fryktsomme overdriver alt hun/han er redd for:

**oko**  $n^1$  1 (*pl* oči) auga  $n^{1N}$  [...] ◆ [...] **strach má velké oči** deigum vex allt í augum ▲ kuří, písek, přimhouřit, volský, vraní

Samtidig ser vi at det henvises til flere artikler fra „oko“, f.eks. til kuří, uttrykket „kuří oko“ = liktorn:

**kuř|e**  $n^3$  kjúklingur  $m^{5ab}$  ◆ **neublížil by ani** ~ti hann gerir ekki flugu mein; ~í *adj*: ~í oko líkþorn  $n^4$ ; **náplast na** ~í oka líkþornaplástur  $m^4$ .

## 7. Ytterligere opplysninger

Reksjon og andre kombinatoriske data gis kursivert; fakultative elementer i parentes:

**žert**  $m^5$  ...; ~ovat<sup>3</sup> (*s T, o L*) spauga<sup>3</sup> (*við e-n; með e-ð*), gera<sup>2D</sup> að gamni sínu  
**vnutit**<sup>4P</sup> *D co neyða*<sup>2D</sup> {pröngva<sup>3</sup>} *e-u upp á A*  
**vysypat**<sup>1,5P</sup> 1 ... 2 (*A čim*) strá<sup>2D</sup> *e-u á A*

Store bokstaver representerer kasus (*A* = akkusativ, osv.). Dersom det er viktig å understreke at det dreier seg utelukkende om animata (primært personer) eller inanimata, brukes forkortelser med små bokstaver; det islandske **spauga** tar *með* + *akkusativ* hvis substantivet er inanimatum; *við* + *akkusativ* dersom det dreier seg om person; det tsjekkiske verbet **žertovat** styrer *o* + *lokativ* hvis man gjør seg lystig over noe/noen; *s* + *instrumentalis* hvis man spøker med noe/noen (žertovat s přítelem; žertovat s nemocí).

Hevet<sup>P</sup> i tsjekkiske verbindeksler betyr perfektiv aspekt, se *vnutit* og *vysypat*. Fravær av<sup>P</sup> betyr at verbet er imperfektivt (*žertovat*). I den islandsk-tsjekkiske delen gis gjerne en tsjekkisk verbekvivalent som aspektpar der det perfekte parverket kommer først og etter skråstrek følger det imperfektive:

**dangla**<sup>3</sup> *i e-ð po-/klepat*<sup>1,5</sup> *na co, št'ouchnout*<sup>2</sup>/*št'ouchat*<sup>5</sup> *do D, pohupovat*<sup>3</sup> *I*

Her er det overflødig å markere *št'ouchnout* som perfektivt; det perfekte parverket til *klepat* er *poklepat*.

„En for alle“ *prinsippet*, eller „lexical sets“, se Svensén 2009:394.

Visse grupper leksemer i samme betydningskategorier har parallelle kollokasjoner. Man velger ett av disse som representant for alle leksemer i hver gruppe og oppfører der alle konstruksjoner som er felles for denne gruppen.

I de øvrige artiklene dette gjelder peker en pil på denne „representanten“. Foruten ukedager o.l., månedsnavn, nasjonalitet o.l. gjelder dette bl.a. tall. La oss først se på „hovedrepresentanten“ for ordenstill, *pátý* ‘femte’:

**pát|ý** fimmti; **o** ~é (**hodině**) klukkan fimm; **po** ~é (**hodině**) eftir klukkan fimm, klukkan að ganga sex, á sjötta tímanum; (už to zkouší) **po** ~é (hann reynir þetta) í fimmta sinn; **jde na** ~ou ≈ klukkan er orðin fjögur; **do** ~é **odpoledne** til klukkan fimm síðdegis; **dnes je** {bylo to} ~ého května í dag er fimmti maí {það gerðist fimmta maí} ~é **přes deváté** í belg og biðu; **je** ~é **kolo u vozu** *e-m* er ofaukið ▲ kolona, mlít

Det er innlysende at det som står etter rombtegnet gjelder utelukkende lemmet *pátý* – ikke øvrige ordenstill.

**třetí** þriðji ○ *pátý*; **po** ~ í þriðja sinn ~ **svět** þriðji heimurinn; **do** ~ {třetice} **všeho dobrého** allt þegar þrennt er ▲ *pád*, říše, stuped

Andre uttrykk med *třetí* retter seg (mutatis mutandis) etter *pátý*: **o** *třetí* (**hodině**) klukkan þrjú; **dnes je** {bylo to} ~ho května í dag er þriðji maí {það gerðist þriðja maí}, osv.

## Uregelmessig bøyning

I de islandske bøyningssindeksene markerer en trekant alene ( ) at ordet tilhører heteroklitika (gjelder f.eks. sterke verb, substantivene *faðir*, *móðir*, *sonur*, *dóttir* o.fl.). Sammen med tall- og evt. bokstavsindeks signaliserer en slik trekant uregelmessigheter ved bøyning av et ord som ellers tilhører et vanlig paradigme. Disse avvikene forklares i ordboksgrammatikken. Det tsjekkiske indekssystemet tar ikke høyde for slike avvik. Det er derfor viktig at de belyses eksplisitt:

**aug|a** *n*<sup>IN</sup> 1 oko *n*<sup>1</sup>; **loka** ~**unum** {opna ~**un**} zavřít {otevřít} oči 2 (*gat*, *hola*) očko *n*<sup>1</sup>, otvor *m*<sup>5</sup>; ~ **á nál** {öxi} ouško *n*<sup>1</sup> jehly {oko sekery} 3 (*depill*) tečka *f*<sup>1</sup>; ~**a á teningi** tečka na hrací kostce ◆ **ganga** í ~**un** á *e-m* [...]. Av dette fremgår at flertallet av *oko* i delbetydning 1 er uregelmessig: *oči*. I de øvrige delbetydningene er flertallsformen regelmessig: *oka*.

I visse tilfelle er lemmatisering av uregelmessige former nødvendig:

oč.. *sjá* oko

De to halvfete punktene gjelder alle (uregelmessige) flertallsformer (oči, očí, očích, očima).

žij(..) *sjá* žít

Parentesen betyr, at **žij** alene eksisterer som en selvstendig form (entall imperativ). I tillegg anviser artikkelen presensformene **žiji/žiju – žiješ – žije** osv.

De tekniske grep som er belyst i pkt 6 og 7 stammer primært fra V. Berkovs leksikografiske virksomhet. En innovasjon består i følgende:

Som de fleste slaviske språk har tsjekkisk hverken bestemt eller ubestemt artikkel. Med tanke på tsjekkiske nybegynnere kursiveres den islandske bestemte artikkelen i den tsjekkisk-islandske delen med mindre den står i parentes, dvs. er fakultativ i det gitte tilfelle:

**želíz|ko** *n*<sup>1</sup> (lítið) járnstykki *n*<sup>6j</sup>, ≈ járn *n*<sup>4</sup> ♦ **mít více** *-ek v ohni hafa mörg járn eldinum*

**stúl** *m*<sup>5</sup> borð *n*<sup>4</sup>; **sedět za stolem** sitja<sup>Δ</sup> við borð(ið) ♦ **vyložít karty na** *-leggja*<sup>1</sup> *spilin á borðið* ▲ **prostřít, psací, sklídít** *aj*.

## 8. Prosjektets status

Den 15. november 2011 inviterte styret i ČNK (Český národní korpus) Haraldsson og Bartošková til å presentere prosjektet IT/TI på et seminar i Praha, der også forlaget LEDAs direktør var til stede.

Med hensyn til prosjektets omfang og ambisjonsnivå ble man enig om at manus skulle leveres først ved årsskiftet 2012-2013.

Våren 2012 er den tsjekkisk-islandske delen ferdigstilt og redigeres etter hvert av Bartošková. Hennes arbeid med den islandsk-tsjekkiske delen redigeres fortløpende av Haraldsson. Han har også utarbeidet de fleste ømtekster, f.eks. ordboksgrammatikk inkl. fonetikk for begge språk, brukerinstruks o.l.

## ANVENDT LITTERATUR

Berkov, Valerij, 1994: Russisk-norsk ordbok. Universitetsforlaget. Oslo.

(NB: En ny revidert utgave: Stor russisk-norsk ordbok kom i juni 2011 hos Kunnskapsforlaget. Oslo).

Berkov, Valerij et alii, 2003: Stor norsk-russisk ordbok. Kunnskapsforlaget, Oslo.

Haraldsson, Helgi, 2002: Rannsóknar- og fræðslurit 6. Beygingarkótar íslenskra orða.

Tilraun um orðabókarmálfræði með tilvísunarkerfi fyrir nafnorð, lýsingarorð og sagnir. "Orðabók Háskólans". Reykjavík.

Haraldsson, Helgi, 1996: Rússnesk-íslensk orðabók. Nesútgáfan. Reykjavík.

Kupča, Vojtěch, 2008: Islandsko-český slovník. Litera proxima. Praha.



Vrbová, Jarka et alii, 2005: Norština – Čeština. Praktický slovník norskó-český/česko-norský. Leda 2005. Praha.

Svensén, Bo, 2009: A handbook of lexicography : the theory and practice of dictionary-making. Cambridge University Press. Cambridge.

*Helgi Haraldsson*

Institutt for litteratur, områdestudier og europeiske språk, Universitetet i Oslo, Norge.  
helgi.haraldsson@ilos.uio.no

# Ensam fras söker kontext: Översättarens hantering av språkprov i en tvåspråkig ordbok

*Ylva Hellerud*

Working as a translator for the inter-Scandinavian dictionary project Islex, presents a number of interesting challenges due to the fact that the source language material is compiled and presented based on a monolingual semantic analysis. In this article I describe the process involved in choosing the most relevant translations of the examples provided for the lemmas *rekstrarerfiðleikar* and *veðrast*, focusing on the problem of lack of context. Finally I discuss the possibilities made available by digital publishing on the world wide web.

*Nyckelord:* översättning av språkprov, idiomatisk kontra ordagrann översättning, kultur-specifika språkprov

## 1. Bakgrund: Islex-projektets förutsättningar

Islexprojektet är ett interskandinaviskt ordboksprojekt, där ett källspråk – isländskan – utgör underlag för (från början) tre målspråk: danska, norska och svenska. Islex ska fungera som en allmänspråklig, tvåspråkig ordbok för respektive målspråk med syntagmatisk redovisning i form av kollokationer, idiom och språkprov, samt på en gång vara L1- > L 2- och L 2- > L 1-ordbok för språkparen isländska – svenska, isländska – norska och isländska – danska. I förlängningen ska källspråksmaterialet kunna användas för att skapa nya ordböcker på vilket målspråk som helst. Nyligen har också färöiska och franska tillkommit som målspråk. Med den utgångspunkten måste det kontrastiva perspektivet offras, och materialets organisering utgår därför från en isländsk betydelsestruktur. Språkproven i Islex är redaktionellt konstruerade, ibland med utgångspunkt i korpuset Íslenskt Textasafn.

## 2. Översättaren som lexikograf

Vad kan man som översättare bidra med i skapandet av en isländsk-svensk ordbok? Rent generellt kan man förstås säga att den erfarenhet man har av att hantera stora textmassor och överföra dem från det ena språket till det andra innebär att man aktivt har arbetat med att lösa ett stort antal problem som uppstår i mötet mellan två språk. Man vet att exakta ekvivalenter är sällsynta, men man vet också att med olika sorters omskrivningar kan man skapa en effekt som ligger mycket nära andemeningen i källspråket, det som brukar kallas pragmatisk ekvivalens. Man är dessutom medveten om att kontext är A och O. Och det handlar om kontext i alla dess betydelser: situationskontext, kulturkontext och så vidare. Ett utsnitt ur en roman kan vara omöjligt att översätta utan att ha läst hela verket, hela verket kan vara omöjligt att översätta utan att ha goda kunskaper i bland annat källspråkets litterära kulturarv.

Översättarens medvetenhet om kontextens betydelse kan förstås vara både en tillgång och en nackdel, i synnerhet när det gäller översättning av språkprov. Idiom och i viss mån kollokationer (såsom definierade i Svensén 2004) kan betraktas som lexikaliska enheter, en sorts utökade lemman, och utgör därför, förenklat sett, ett rent semantiskt problem; det är översättarens/lexikografens kunskap i källspråk och målspråk som avgör träffsäkerheten i valet av ekvivalent(er) eller, i förekommande fall, översättning. Språkproven, däremot, kan visa sig vara helt kontextberoende för att en adekvat översättning ska kunna åstadkommas.

## 3. Språkprov

I sin normala arbetssituation arbetar översättaren med en avgränsad, tematiskt och/eller konstnärligt sammanhållen skriven text (eller om det rör sig om film- och tv-översättning, talat språk som ingår i ett på samma sätt överordnat sammanhang). Det kan i sin tur finnas en eller flera överordnade kontexter (en artikel om avfallshantering kan t.ex. vara en replik på en tidigare publicerad artikel i samma ämne). Kunskap om de olika kontexterna kan vara helt avgörande för många formuleringar i en given text. Eftersom arbetet i Islex är organiserat så att den isländska redaktionen väljer allt isländskt material, kan det som översättare ibland vara problematiskt att sätta sig in i dels av vilken anledning språkprovet finns med (vilken funktion det har), dels i vilken kontext (tänkt eller verklig) språkprovet skulle kunna förekomma. Skillnaden mellan ledigt men vårdat tal-språk och formellt språk är dessutom mindre på isländska än på svenska, och utan kontexten som hjälp kan man som översättare hamna i valet mellan ett antal semantiskt korrekta men stilistiskt varierande översättningar.

### 3.1 Kontext och stil

Som exempel kan tjäna frasen

*rekstrarerfiðleikar leiddu til lokunar fyrirtækisins*

som är det språkprov som anges vid uppslagsordet *rekstrarerfiðleikar*. En någorlunda ordagrann översättning av frasen lyder

‘lönsamhetsproblem ledde till stängning av företaget’

men en sådan formulering skorrar i en översättares öra. I vilket sammanhang skulle en svensk uttrycka sig så? Det faktum att frasens subjekt står i artikellös form leder tankarna till någon sorts överskrift, t.ex. en tidningsrubrik (vilket dock inte anges eller antyds på något annat sätt i ordboksartikeln). En tänkbar översättning blir då:

‘lönsamhetsproblem bakom företagsnedläggning’

Här uppstår dock en konflikt mellan översättaren och lexikografen. Men hänsyn till att en ordbok också har en pedagogisk intention bör kanske för tydlighetens skull predikatet vara kvar?

‘lönsamhetsproblem ledde till företagsnedläggning’

Helt träffsäker är emellertid inte denna formulering heller ur en annan aspekt: en så generell utsaga skulle knappast förekomma som rubrik i en svensk tidning, utan man skulle troligen ange vilket företag det var fråga om. Dessutom finns det en pedagogisk poäng också i att illustrera svenska tidningsrubrikers långtgående elliptiska karaktär. Till sist illustrerar de två översättningarna det klassiska översättningsdilemmat mellan en friare – mer idiomatisk? övertolkande? – översättning å ena sidan och en mer ordagrann – stelare? okänsligare? – å den andra.

Utän att gå in detaljerat i fortsatta resonemang av den här typen kan ändå nämnas att ett antal andra översättningsmöjligheter också finns:

‘lönsamhetsproblem ledde till att företaget stängdes’

‘lönsamhetsproblem ledde till att företaget lades ned’

‘lönsamhetsproblem ledde till att företaget fick läggas ned’

‘problem med lönsamheten ledde till att företaget stängdes’

‘problem med lönsamheten ledde till att företaget lades ned’

‘problem med lönsamheten ledde till att företaget fick läggas ned’

osv.

Som översättare kan man i detta sammanhang reflektera över de möjligheter den digitala ordboken öppnar för. Att ange alla – eller åtminstone några av – de ovan nämnda översättningsmöjligheterna vore på sätt och vis önskvärt, men då måste också förklarande kommentarer infogas. Trots att det i en digital ordbok inte råder någon platsbrist, vore det ingen lyckad lösning att en sådan informationsmassa visades automatiskt vid en sökning på ett givet uppslagsord. Däremot kan man tänka sig att informationen ligger inbäddad och kan tas fram av den brukare som har intresse av den. Resultatet blir en hybrid, ett mellanting mellan ordbok och lärobok (och kanske också encyklopedi, vilket jag återkommer till) med möjlighet att också ange ett stort urval konstruktioner och fraser med utförliga kommentarer.

Ovanstående är exempel på en situation där språkprovet ur översättarperspektiv inte tillför något: lemmat *rekstrarerfiðleikar* kan översättas med 'lönsamhetsproblem' och språkprovet fungerar inte som någon semantisk fördjupning av uppslagsordet eller illustration av skillnad i konstruktion eller användning mellan isländska och svenska. De problem som belystes i mitt resonemang är av annan karaktär. Härnäst följer en redovisning av arbetet med uppslagsordet *veðrast*, där valet av språkprov bidrar till att en viktig betydelse av uppslagsordet blir uppmärksammas.

### 3.2 Kulturspecifik kontext

Verbet *veðrast* förekommer i Islex som eget lemma (och inte som deponensform av *veðra*). Någon ett-till-ett-ekvivalent existerar inte för ordet, men det är frågan om en lexikal lucka, inte en referentiell, och en definition kan lyda: "exponeras för och nedbrytas av väder och vind". Partiella ekvivalenter går att hitta, t.ex. 'vittra', 'förvittra', 'erodera', och hade det inte funnits något språkprov i denna artikel, hade lösningen kanske blivit att ange de nyss nämnda ekvivalenterna jämte den encyklopediska förklaringen. Det sätt på vilket arbetet med Islex är organiserat, nämligen att datainsamling och dataurval sker separat från översättning av lemmat och språkprov, ställer mycket höga krav på översättarens kompetens i källspråket. Den isländska redaktionens val av språkprov fungerar här som varningsklocka för översättaren, eftersom det blir uppenbart att ingen av de ovan nämnda ekvivalenterna fungerar när frasen

*þakið hefur veðrast og nú þarf að mála það*

ska översättas. En ordagrann översättning skulle i så fall bli

'taket har vittrat (alt. eroderat) och nu behöver det målas'

och det är inte en acceptabel formulering på svenska. En möjlig lösning med ledning av den encyklopediska förklaringen kunde vara

‘taket har exponerats för väder och vind och behöver nu målas’

men en sådan utsaga har för en infödd svensk något av “ingenjör Planertz” över sig, alltså en överdrivet formell utsaga om ett vardagligt fenomen. För att hitta fram till en godtagbar översättning krävs en kontext, eller rättare sagt två. Utsagan kan nämligen syfta på två olika förhållanden:

- 1: målarfärgen på taket har blivit sliten (av väder och vind) så det behöver målas om  
och
- 2: det nya förzinkade (galvaniserade) plåttaket har utsatts för väder och vind ett tag och därför är det nu dags att måla det för första gången.

Hade denna utsaga ingått i en text, hade det med största sannolikhet gått att förstå vilken av de två tolkningsmöjligheterna det gällde, men nu rör det sig alltså om en lösryckt fras som ska fungera som språkprov till ordet *veðrast* varför båda tolkningarna bör redovisas, i synnerhet som det visar sig att tolkningsmöjlighet 2 kräver en ekvivalent för *veðrast* som ligger långt ifrån de som föreslogs ovan. Ur svensk synvinkel kan det möjligen tyckas att tolkning 2 hänvisar till ett relativt snävt ämnesområde och därför skulle kunna utelämnas, men det finns goda skäl för att inkludera den – och här sätts översättarens kulturella kunskaper om källspråket på verkligt prov. Yttertak i förzinkad (eller galvaniserad) plåt är mycket vanligare på Island än i Sverige, och dessutom är förzinkad plåt också en allmänt förekommande fasadbeklädnad på Island. Mest framträdande är de privata villor från tiden kring förra sekelskiftet som ligger i Þingholt-området i centrala Reykjavík. Att använda omålad förzinkad korrugerad plåt vid renovering av dessa gamla hus rekommenderas av kulturhistoriska skäl – det finns annars färdigmålad plåt att köpa. Fabriksny förzinkad plåt bör utsättas för väder och vind ett par år innan den målas, vilket många islänningar känner till – även om de inte vet varför. Målning av förzinkad plåt är alltså ett långt vanligare fenomen i den isländska vardagen än i den svenska, varför det får anses befoget att inkludera tolkningsmöjlighet 2.

Som framgick ovan måste en omskrivning av någon sort tillgripas vid båda tolkningsmöjligheterna. Eftersom vi har att göra med en lexikal lucka kan man återigen använda sig av frågan: ”Hur skulle en svensk uttrycka sig i samma situation?” I det första fallet finns förutom förslaget ovan naturligtvis ytterligare några tänkbare översättningar, beroende bl.a. på vilken sorts tak det rör sig om, men det resonemanget utelämnas här för att ge plats åt det mer kulturspecifika.

Skälet till att förzinkad plåt som används för tak- och väggbeklädnad inte ska målas omedelbart är att man vill få till stånd en oxidation (som i sin tur har att göra med galvaniseringens rostskyddande egenskaper). En någorlunda ordgrann översättning av tolkningsmöjlighet 2 skulle därför bli

'taket har oxiderat och nu behöver det målas'

även om den tänkta kontexten fortfarande är lite oklar. Är det skriftspråk eller talspråk? Exakt när skriver/säger man något sådant? I en normal talsituation är troligen en eller flera faktorer redan kända vilket medför att en så fullständig sats inte uttalas. ("Varför målar de taket först nu? Det var ju klart för två år sedan." – "Det måste oxidera först.") En mer realistisk utsaga kunde vara en tänkt instruktion i en bok om byggnadsvård:

"När taket fått oxidera i cirka två år, är det dags att måla det."

En sådan översättning ligger dock alltför långt från originalet för att vara godtagbar. Däremot kan man genom andra kompensatoriska mindre tillägg, 'plåttaket *har fått* oxidera', dels förtydliga vilken sorts tak det gäller, dels antyda att oxidationen inte är en skada utan ett medvetet förfarande. En godtagbar lösning kan då bli:

'plåttaket har fått oxidera och måste nu målas'

Liksom i översättningen av det tidigare exemplet *rekstrarervfiðleikar leiddu til lokunar fyrirtækisins* hamnar översättaren i konflikt mellan en trovärdig svensk formulering och en pedagogisk översättning av mer ordagrant slag. Å andra sidan kan man fråga sig vilket förhållningssätt som är mest pedagogiskt (eller bättre för läsaren/mer respektfullt mot texten) – och i den debatten har ännu ingen vinnare utsetts.

#### 4. Befintliga hjälpmedel

Med detta sagt kan det vara intressant att resonera kring hur det hade gått till att få fram översättningen 'oxidera' till *veðrast* via de hjälpmedel som faktiskt finns tillgängliga idag. I den befintliga isländsk-svenska ordboken anges endast ekvivalenterna 'vittra', 'förvittra'. Det ger ju en vink om i vilken betydelsesfär vi befinner oss, men utan de ovan nämnda kunskaperna om förzinkad plåt går det inte att hitta fram till den svenska översättningen. En avancerad brukare med tillgång till den enspråkiga isländska ordboken, *Íslensk orðabók*, på papper eller digitalt, får en indikation i rätt riktning, för där framgår tydligt av det språkprov

man valt att det handlar om takplåt som måste exponeras *innan* den målas, vilket man också rätt snart sett om man som första åtgärd sökt uttrycket på nätet. Har man väl fått detta klart för sig, är det inte många musklick som ska till innan man hittar fram till kopplingen mellan takplåt, galvanisering/förzinking, målning och oxidering. Det visar sig också finnas gott om byggnadsvårdsintresserade privatpersoner i vårt land som använder nätet för informationsutbyte, och i diskussioner om förberedelserna för målning av plåttak använder de alla termen oxidera. Man har alltså som avancerad brukare tillgång till stora mängder text med kontext, på både källspråk och målspråk, vilket, om man hårdrar det hela, innebär att ordboken blir överflödig. Ett annat exempel kan tas från skönlitteraturens värld. I en översättning jag nyligen arbetat med utgjorde bank- och finansvärlden fonden för handlingen. Där förekom ett antal finanstermer, t.ex. ”derivathandel”, ”inflationindexerad” och ”kreditlina” som inte fanns i den isländsk-svenska ordboken. Någon isländsk-svensk specialordbok med bank- och finanstermer – eller ens allmänt ekonomiska termer – existerar inte. Det gick ändå att hitta fram till dem via jämförande textstudier på nätet, några isländsk-engelska ordlistor på isländska bankers hemsidor, jämte ett bekräftande telefonsamtal till en svensk språkontresserad bankman.

## 5. Internet – ett gigantiskt korpus. Har ordboken blivit överflödig?

Den avancerade brukaren har alltså idag möjligheter att använda sin egen dator som kombinerad ordbok, språkprovssamling och konsult i specialterminologi. Informationssökandet tar en del tid, men går betydligt fortare än att som förr leta upp specialordböcker, facklitteratur samt ta personlig kontakt med experter. I den mån specialordböcker inte finns att tillgå kan den professionelle översättaren i stor utsträckning klara sig med internet både för avkodning och produktion. För avancerade brukare kan mediet räcka till för avkodning, men kanske inte för produktion, och för nybörjare och vidarekomna är det förstas inget alternativ vare sig för avkodning eller produktion, eftersom arbetssättet kräver rätt omfattande kunskaper i såväl käll- som målspråk. Därför fyller en redaktionellt bearbetad, strukturerad och användarvänlig ordbok en viktig funktion. Den digitala publiceringen ger dessutom möjligheter till en långt utförligare pedagogisk funktion än som är genomförbart i ett verk tryckt på papper.



## 6. Faktotum för framtiden?

I ljuset av diskussionen ovan om de två språkproven i Islex, är det intressant att fundera kring de möjligheter som finns i digital publicering. I viss mån utnyttjas dessa i Islex, bland annat genom möjligheten till textsökning, där man enkelt kan få upp alla träffar på ett givet ord, alltså inte bara i lemmalistan, utan alla förekomster i kollokationer, idiom och språkprov på alla de språk som ingår. En bieffekt blir också att ordboken kan användas mellan språken på ett icke avsett vis, till exempel mellan norska och svenska. För övrigt är ordbokens struktur traditionell. Den digitala tekniken ger möjlighet att ha underliggande lager av information, till exempel konstruktion, fraseologi och stil, som i exemplet med *rekstrarerfiðleikar* och upplysningar av encyklopedisk karaktär som i exemplet med *veðrast*. Det ger brukaren möjlighet att välja hur mycket information hon vill ha vid ett givet tillfälle, och det skulle också underlätta för översättaren/lexikografen i det redaktionella arbetet, genom att gränserna blir tydligare mellan det rent lexikografiska arbetet – om gränsdragningen tillåts – och det pedagogiska.

## REFERENSER

- Íslensk orðabók (pappersutgåva), 1992. Red. Árni Böðvarsson. Reykjavík: Mál og menning
- Íslensk orðabók (digital utgåva). [www.snara.is](http://www.snara.is)
- Íslenskt textasafn. Stofnun Árna Magnússonar í íslenskum fræðum. [www.lexis.hi.is/corpus/](http://www.lexis.hi.is/corpus/)
- Norstedts isländsk-svenska ordbok. Sjunde oförändrade upplagan 2005. Red. Sven B.F. Jansson. Stockholm: Norstedts Akademiska Förlag.
- Svensén, Bo, 2004: Handbok i lexikografi. Stockholm: Norstedts Akademiska Förlag.

*Ylva Hellerud*

[ylva.hellerud@telia.com](mailto:ylva.hellerud@telia.com)

# Massive ordbokspalter i transparent presentasjon – NAOB på nett: samspillet mellom datastruktur, innhold og brukergrensesnitt

*Petter Henriksen*

The *Norwegian Academy's Dictionary* (NAOB) is planned to be published digitally in 2014. The transformation of the massive 20th-century dictionary *Norsk Riksmålsordbok* into a user-orientated publication on the Internet poses special challenges to data structuring and presentation techniques. The NAOB project aims to provide the Norwegian public and research community with an easily accessible and authoritative lexicographical documentation of the majority language variants Bokmål and Riksmål, and, also, a suitable platform for future updating and development. Internet publishing offers a new world of opportunities for user-friendly presentation compared to last century's book publishing format – if one makes the right choices regarding data structure. In this article, the author gives a systematic presentation of the NAOB project's insights and solutions, hopefully of relevance for other extensive national dictionary projects in the future.

*Det Norske Akademi's Store Ordbok* (NAOB) blir publisert i 2014, ikke på papir, men kun som en digital ordbok på Internett. Forrige gang dette innholdet ble publisert, var det kun på papir – det ærverdige gamle storverket Norsk Riksmålsordbok. Underveis har ordbokens innhold blitt radikalt transformert fra de massive bokspaltene som møtte brukerne i forrige århundre. Også vi selv i redaksjonen har blitt nokså transformert i løpet av prosessen, og forhåpentlig en del klokere. Vi vil gjerne dele våre erfaringer – kall det gjerne våre bedriftshemmeligheter – med de nordiske leksikografene i denne artikkelen.

Oppgaven er velkjent blant forlag som utgir ordbøker. Man begynner med (A) grafiske filer for bokutgivelse – eller enda mer primitivt som i vårt tilfelle – bare bokspalter (figur 1).

**slett**, adj., som adv. i bet. 2 b, b også (særll. fam.)  
**slettes (slets)** (bet. 3 lånt ell. påvirket fra ty. *schlecht*)  
 1) a) som (i sin overflate) er uten (større) forhøininger og fordypninger (og mere ell. mindre nærmer sig et plan); flat (1); jevn (1); glatt (1): *derute følger de slette marker* (Øverl., *Brød*, 134) / *slet som et gulv* (Hams., *Mark*, II, 168) / *(rideknekten) gikk ute i manegen og raket sugmuggen slet igjen* (Anker, *Søg.*, 36) / *(han) arbejdede sig opover det næsten slette fjæld* (B. B., *Forl.*, 229). b) i best. form uten foranstilt best. artikkel brukt fremhevende (med hbet. som *bare, nakne*): *legghugget inn i slette berget* (Brøgger, *Det norske f. i oldt.*, 76) / *slette bakken (var) mellem os* (Nans., *Poht.*, I, 130). b) (om hår) glatt (ikke krøllet ell. kruset) (Unds., *Kr.*, 76). 2) a) likefrem (2b og c); direkte (4); (som adv.): rent ut sagt; simpelt hen; uten videre; i forh. *reit og slett, slett og reit* (se *reit*, adj. 4). b) som adv. a) foreld., helt og holdent; ganske. b) brukt forsterk. foran nektende ord, med bet. 'aldeles, absolutt' (ofte utt. sammen med nektelsen som ett ord, med hovedtrykk på *slett*): *almeneheden behøver slet ingen nye tanker* (lbs., *Folk.*, 73) / *her er ikke så ilde her. Slet ikke så ilde* (lbs., *Vild.*, 76) / *det er slettes ikke noget*

Man skal ende med (B) en eller annen form for leksikografisk struktur som egner seg både for digital publisering og for fremtidig oppdatering og redigering. Hva er veien fra A til B og hvilke kritiske veivalg må man ta?

NAOB er en revisjon av den gamle Norsk Riksmålsordbok, enn så lenge Norges eneste komplette mangebinds ordboksverk. Norsk Riksmålsordbok ble utgitt fra 1937 til 1957, i fire bind med høyst varierende tykkelse, med det forvokste bind 4 som et ruvende monument over leksikografisk skaperfryd og forlagsadministrativ avmakt. I 1995 kom to nye bind med supplerende innhold, i et separat alfabet. På slutten av 1990-tallet startet prosjektet med å digitalisere verket med dét for øye å spleise de to alfabetene sammen i ett i en ny bokutgave til jubileumsåret 2014, samme år som Norge får sitt nasjonale ordboksverk for nynorsk – *Norsk Ordbok 2014*.

Underveis i prosjektet grydde erkjennelsen av at det ikke var et bokverk Norge ville ha, men en Internett-utgivelse. Og med det kom erkjennelsen av at dataenes struktur kom til kort, nå som man skulle publisere på den nådeløst avslørende arena som nettpublisering er. For ikke lenge siden fikk NAOB en fornyet datastruktur, og dataene gjennomgikk en storstilt konvertering. Resultatet er at vi nå har et redigeringsgrensesnitt og et grunnlag for et publiseringsgrensesnitt som fungerer godt.

Gjennom alt dette har vårt miljø fått finpusset sin metodikk for transformasjon av tunge ordbøker fra bokformat til Internett-format. Den presenteres her.

## Strategisk vurdering

Første steg er en strategisk vurdering. Man må bestemme om målet for prosjektet er *dynamisk*, dvs. et verk som skal leve videre gjennom stadig nye oppdateringer, eller *statisk*, dvs. et verk som egentlig bare skal gjenskapes slik det var på papir, med noen punktvisse oppdateringer. I fall målet er statisk, kan man

komme lettvint fra det, med en digital variant av faksimilen. Dette alternativet skal ikke drøftes her. Hvis målet er et dynamisk verk, må man åpne sinnet – og lommeboken – noe mer. Anbefalt metodikk er som følger, internt i prosjektet benevnt *drømmemodellen*:

- Man setter seg inn i *kildeverkets* innhold – hva som ligger der av artikkelinnhold å strukturere.
- Man legger så kildeverket bort og skaper sin idealversjon av det samme verket (“*drømmeverket*”), dvs. den ideelle oppbygning av verket og den enkelte artikkel.
- Man setter i gang med å omforme kildeverk til drømmeverk.

Med denne tilnæringsmåten tar man konsekvensen av at komplekse dataprojekter blir uoverstigelig vanskelige dersom man ikke frigjør seg fra forelegget. Drømmemodellen innebærer at man tar et oppgjør med klassiske motforestillinger i forlagsbransjen: Man ignorerer dem som sier at man ikke må miste informasjonen om det grafiske bildet i den gamle bokutgaven. Man ignorerer dem som sier at det blir mye å rette i de gamle dataene for å få dem til å konformere med idealet. Og hvis noen mener at man i utgangspunktet skal spare ressurser på ikke å implementere et system for lenkede henvisninger og automatisk betydningsnummerering – ignorerer man dem også.

## Tiltak<sup>1</sup>

Med den strategiske vurderingen tilbakelagt går man over til konkrete oppgaver. Vi skal fra *kildeverk* til *målverk*, fra historisk verk til drøm. Veien går gjennom:

1. Analyse av kildeartikkel
2. Design av målartikkel
3. Strukturering av målartikkel ved hjelp av DTD
4. Konvertering av kildedata, inkl. evt. optisk lesning
5. Spesifisering av stilark
6. Redigering

I det følgende gjennomløpes stegene med NAOB som eksempel.

---

<sup>1</sup> Viss information i nedanstående figurer har gått förlorad på grund av svart-vitt tryck istället för färg (red:s anm.).

## 1) Analyse av kildeartikkel

Analysen av kildeartiklene viser at man har velkjente leksikografiske kategorier som oppslagsord, ordklasse, varianter av oppslagsordet, etymologi, betydninger med definisjoner, henvisninger, betydningsskillere, sitater, fraser og mange flere. Kategoriene has i mente.

## 2) Design av målartikkel

Her skal man tenke resultatorientert, dvs. på vegne av de fremtidige brukerne man ønsker å hjelpe og glede. Man tegner ut i ikke-kodet form – gjerne i Word – hvordan man vil at artikkelen skal ta seg ut visuelt og funksjonelt når avdukingen finner sted og brukeren møter den for første gang på nettet. Siden skal man bevege seg bakover fra dette i kjede i en form for teleologisk metode. Her er eksempler fra NAOB på aspekter som ønskes i målartikkelens design:

A) Det ønskes at artikkelen skal se luftig ut, med tydelig markering av de leksikografiske elementene, med intuitivt forståelige ledetekster, med hierarkisk innrykking og nummerering av betydningsnivåene (figur 2).

**slett** adj.

BØYNING -

UTTALE [slet:]

ETYMOLOGI av norr. *slétt*, betydning 2 etter ty. *Schlichting*

### BETYDNING OG BRUK

#### 1 flat; jevn

##### SITATER

- *derute bølger de slette marker* (Øverl. *Brød* 134)
- *slet som et gulv* (Hams. *Mark. II* 168)

#### 1.1 I BEST. FORM UTEN FORANSTILT BEST. ARTIKKEL BRUKT FREMHEVENDE glatt og bar; naken

##### SITATER

- *tegn, hugget inn i slette berget* (Brøgger *Det norske f. i oldt.* 76)
- *slette bakken [var] mellom os* (Nans. *Polh. I* 130)

#### 1.2 OM HÅR glatt (ikke krøllete el. krusete)

#### 2 LTT. (svært) dårlig

##### SITATER

- *slette norske stiler* (Bull *Studier* 11)
- *[han] kunde gjøre et skjøn over bohavets beskaffenhed og [...] nedskrive i sin lommebog et af de tre ord: Godt, tåleligt, slet* (Sundt *Piperviken* 27)
- *stellet er slet* (Ibs. *Gynt* 237)

#### 2.1 umoralsk; fordervet] jf. forderve

##### SITATER

- *et slet menneske* (B. B. *SDv. I* 109)
- *en slet moral* (Ibs. *E. S. I* 452)

##### UTTRYKK

#### du slette tid (grunnbet. 'du dårlige, fordervede tid')

BRUKT SOM UTROP FOR Å UTTRYKKE OVERRASKELSE, FORSKREKKELSE EL. FORFERDELSE du store verden

- *du slette tid for et træf* (Scott *Kild.* 81)

B) Det ønskes at idiomene (de såkalte uttrykkene), skal plasseres under sine respektive betydninger og danne små subartikler, sublemmaer (figur 3).

2 UTT. (svært) dårlig  
SITATER

- *slette norske stiler* (Bull Studier 11)
- *[han] kunde gjøre et skjøn over bohavets beskaffenhed og [...] nedskrive i sin lommebog et af de tre ord: Godt, tåleligt, slet* (Sundt Piperviken 27)
- *stellet er slet* (lbs. Gynt 237)

2.1 umoralsk; fordervet| jf. forderve  
SITATER

- *et slet menneske* (B. B. SDv. I 109)
- *en slet moral* (lbs. E. S. I 452)

UTTRYKK

**du slette tid** (grunnbet. 'du dårlige, fordervede tid')  
BRUKT SOM UTROP FOR Å UTTRYKKE OVERRASKELSE, FORSKREKKELSE EL. FORFERDELSE du store verden

- *du slette tid for et træf* (Scott Kild. 81)

C) Det ønskes klikkbare henvisninger, som bringer brukeren til lenkemålet med markering av den delen av artikkelen som det vises til (figur 4).

**slett** adj.  
BØYNING -  
UTTALE [slet]  
ETYMOLOGI av norr. *sléttr*; betydning 2 etter ty. *Schlichting*

BETYDNING OG BRUK

1 **slet** jevn  
SITATER

- *derute følger de slette marker* (Øverl. Brød 134)

Klikk →

**jevn** adj.  
BØYNING -  
UTTALE [jevn]

BETYDNING OG BRUK

1

1.1 som har en overflate der er utan (stane) forhøyninger og/et. fordypninger, glatt (i bet. 1); plan  
EKSEMPEL

- *en jevn slette*  
SITATER
- *et jevnt snelag* (Sverdr. N. Land II 148)
- *min [Davids] fod står på jevn jord* (Sal 26:12)

1.2  
SITATER

- *lad mig [David] på den jevne st* (Sal 27:11; 1978/85; den rette) | den alminnelige bandede vei hvor det ikke er vanskeligheter el. farer
- *lad os gå med fred vor jevne vej* (lbs. K. K. 130)

1.3 [ ] STERK BENDNING MED S, A1 FORB. :  
UTTRYKK

**på det jevne**  
«på jordens», i det alminnelige, normale

2

2.1 (om lag o.l.) som har samme tykkelse, samme høyde over det hele  
EKSEMPEL

D) Det ønskes at artikkelledd som sitatseksjonen skal kunne kontraheres og ekspanderes (figur 5).

2 UTT. (svært) dårlig  
SITATER

↔

2 UTT. (svært) dårlig  
SITATER

- *slette norske stiler* (Bull Studier 11)
- *[han] kunde gjøre et skjøn over bohavets beskaffenhed og ... ord: Godt, tåleligt, slet* (Sundt Piperviken 27)
- *stellet er slet* (lbs. Gynt 237)

2.1 umoralsk; fordervet| jf. forderve  
SITATER

E) Det ønskes innholdsfortegnelser med lenkeklikk som fører til enkeltbetydninger og enkeltfraser (figur 6).

**slett** adj.

---

**INNHOLD**

Bøyning

Uttale

Etymologi

Betydning og bruk

1 flat; jevn

1.1 glatt og bar; naken

1.2 OM HÅR glatt (ikke krøllete el. krusete)

2 (svært) dårlig

2.1 umoralsk; fordervet

3 helt og holdent

3.1 aldeles; absolutt

3.2 BARNESPRÅK MED REFL. PRON.

---

**BØYNING -**

F) Det ønskes klikkbare lenker også til eksterne nettsteder, f.eks. for språkene omtalt i etymologiene (figur 7).

**slett** adj.  
**BØYNING -**  
**UTTALE** [slet:]  
**ETYMOLOGI** av norr. *slétt*; betydning etter tysk *schlecht*  
**BETYDNING OG BRUK**

1 flat; jevn



Med de ideelle ønskene ferdig designet i et tenkt brukergrensesnitt, utledes de forutgående stegene teleologisk av dette.

### 3) Strukturering av målartikkel ved hjelp av DTD

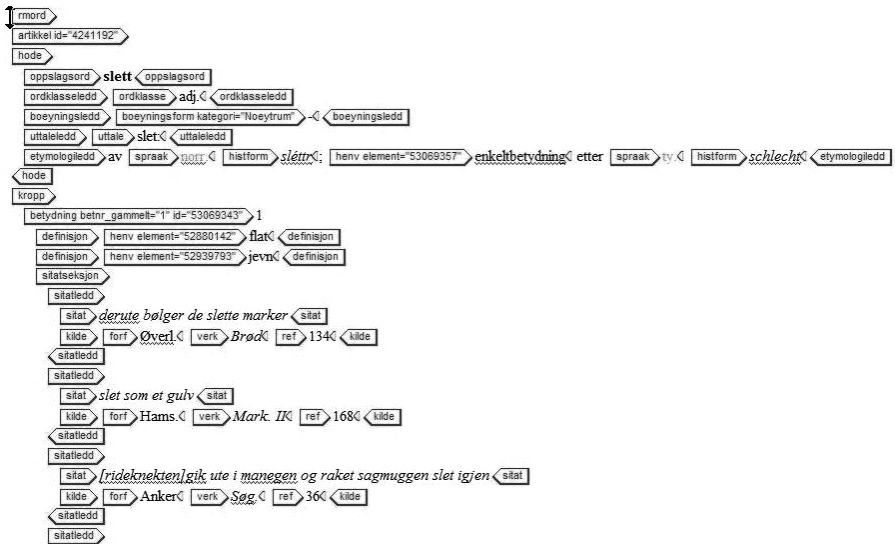
Neste steg er å bestemme hvordan dataene må være strukturert for å oppnå det ønskede visuelle og funksjonelle uttrykket. Verktøyet til dette er DTD-en, som er den formallogiske beskrivelsen av artikkelens innhold når innholdet har XML-format. DTD står for Document Type Definition, og brukes bak kulisene i redigeringsprogrammet til å regulere hvilke innholdstyper (leksikografiske kategorier) som er tillatt hvor i artikkelen. Figur 8 viser hele NAOBs struktur i pseudo-DTD-format.

**artikkel** = hode, kropp  
**hode** = oppslagsord, ordklasseledd?, boeyningsledd?, uttaleledd?, variantledd?, etymogiledd?  
**oppslagsord** = karakter-inlinetekst  
**ordklasseledd** = redaksjonsspråk-inlinetekst / ordklasse  
**ordklasse** = karakter-inlinetekst  
**boeyningsledd** = redaksjonsspråk-inlinetekst / genus / boeyningsform  
**boeyningsform** = karakter-inlinetekst  
**genus** = karakter-inlinetekst  
**grammatikk** = karakter-inlinetekst  
**uttaleledd** = redaksjonsspråk-inlinetekst / uttale  
**uttale** = karakter-inlinetekst  
**variantledd** = redaksjonsspråk-inlinetekst / variant  
**variant** = karakter-inlinetekst  
**etymogiledd** = redaksjonsspråk-inlinetekst / histform / grunnbetydning / appellativisering  
**spraak** = karakter-inlinetekst  
**histform** = karakter-inlinetekst  
**grunnbetydning** = karakter-inlinetekst  
**appellativisering** = karakter-inlinetekst  
**kropp** = betydning\*  
**betydning** = betydningsskiller\*, definisjon\*, utdypning\*, redeksseksjon?, sitatseksjon?, uttrykksseksjon?, betydning\*  
**betydningsskiller** = redaksjonsspråk-inlinetekst  
**definisjon** = redaksjonsspråk-inlinetekst  
**utdypning** = redaksjonsspråk-inlinetekst  
**redeksseksjon** = redeksledd\*  
**redeksledd** = innledning\*, redeks, utdypning\*  
**innledning** = redaksjonsspråk-inlinetekst  
**redeks** = redaksjonsspråk-inlinetekst  
**sitatseksjon** = sitatledd\*  
**sitatledd** = innledning\*, sitat?, kilde\*, utdypning\*  
**sitat** = typografi-inlinetekst  
**kilde** = redaksjonsspråk-inlinetekst  
**uttrykksseksjon** = uttrykksledd\*  
**uttrykksledd** = uttrykk?, variantledd?, etymogiledd?, uttrykksbetydning\*  
**uttrykk** = typografi-inlinetekst  
**uttrykksbetydning** = betydningsskiller\*, definisjon\*, utdypning\*, redeksledd\*, sitatledd\*

Her i NAOBs DTD finner man igjen karakteristiske strukturelementer som uttrykksseksjonen med idiomer (uttrykksledd med uttrykk) som subartikler, og det rekursive betydningshierarkiet (betydninger som kan inneholde betydninger).

Når artikkelen er strukturert som XML i henhold til DTD-en, omgis de ulike leksikografiske enhetene med tagger som forteller hva de er. Innenfor XML-terminologi kalles taggedede enheter for *elementer*. Figur 9 viser NAOB-data i XML-format.





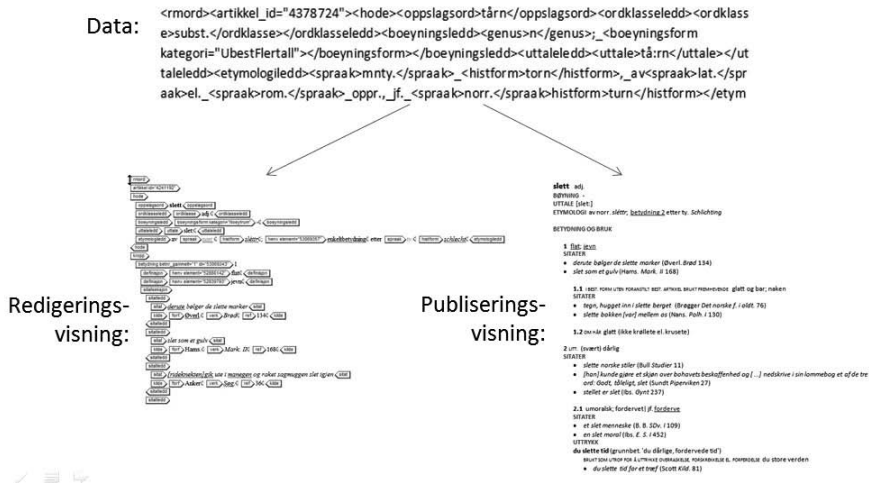
#### 4) Konvertering av kildedata

Neste steg er det brutale møtet med virkeligheten – transformasjonen av eksisterende data til drømmestruktur, eventuelt innledet av optisk lesning dersom utgangspunktet er bokspalter. Til dette lager man en så god oppskrift som man kan, og forsoner seg med at den i mange tilfeller vil misforstå og feiltolke, og at man siden må foreta en full gjennomredigering for å fange opp feiltolkingene. I kildedataene for NAOB er f.eks. sitater, kollokasjoner og idiomer alle kursiverte og opptrer på samme plass i betydningen. Uansett hvor fintfølende program man lager, kan man være sikker på at ikke det bare er idiomer som ender opp i uttrykksleksjonen. Konverteringsprogrammet treffer kanskje bare i 85 prosent av tilfellene. Det er en kompleks oppgave å programmere datatransformasjoner av denne typen. Ikke mindre kompleks er oppgaven som faller på fagleksikografen – å spesifisere de utallige stegene i transformasjonen så leksikografien ivaretas og på en slik måte at de lar seg omsette av en IT-konsulent til et dataprogram.

#### 5) Spesifisering av stilark

Den nest siste oppgaven er stilarkspesifikasjonene. Stilark er det regelsett som forteller hvordan datastrukturen skal vises frem, altså hvordan XML-koden skal

omsettes i grensesnittene hhv. i redigeringsprogrammet og i publiseringsvisningen for brukeren (figur 10).



## 6) Redigering

Med de fem første stegene tilbakelagt og med datakonverteringen gjennomført, er det bare redigeringen som står igjen. Den tar noen år.

## Samspillet mellom datastruktur og visning

I vårt arbeid med NAOB-dataene ble vi overrasket over hvor begrenset kjennskapen var i dataproseseringsmiljøet til grunnleggende sammenhenger mellom datastruktur og visning. Vi fant heller ikke noe om dette temaet i den leksikografiske litteraturen, så det er mulig at våre erkjennelser her ikke er allment tankegods. Derfor vies det plass i artikkelen til denne tekniske detaljen.

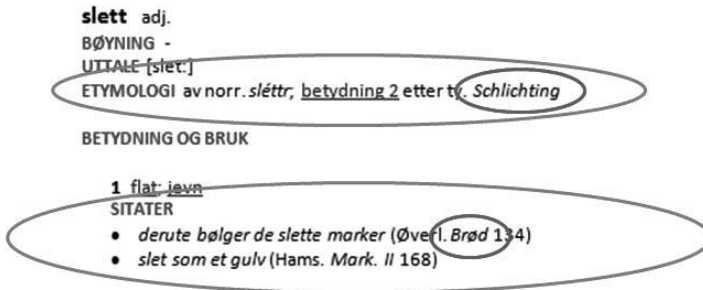
For å holde orden på en hvilken som helst ordbok, men særlig en kompleks, massiv en, må man håndheve strengt et dikotomisk skille mellom to elementtyper: *eskeelementer* og *innholdselementer*.

*Eskeelementer* er de strukturdannende elementene, metaforisk kalt esker fordi de liksom stables oppå hverandre og har innhold. Eksempler er *hode*, *kropp*, *etymologi*, *sitatseksjon*, *uttrykkseksjon*. I en luftig, digital publiseringsvisning kan eskeelementer gjerne vises med ny linje, kanskje også med en generert ledetekst så som ETYMOLOGI. Noen eskeelementer nær grunnplanet er slike som inneholder tekstlig innhold, slik tilfellet f.eks. er for *etymologi*, noen høyere

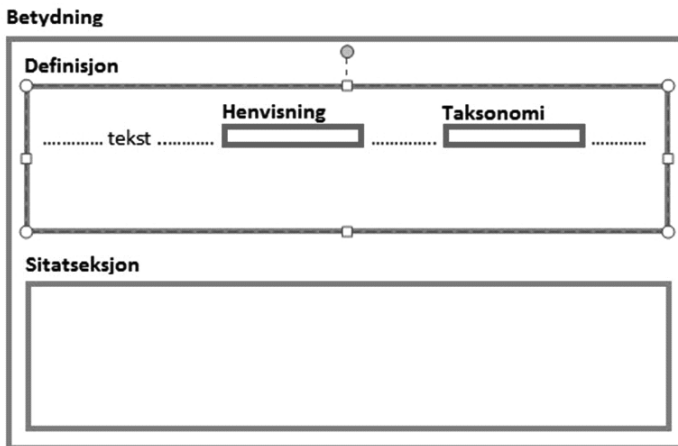
oppe i strukturhierarkiet har nye eskeelementer inni seg – “kinesiske esker” – slik tilfellet f.eks. er for *sitatseksjon*.

*Innholdselementer* opptrer på sin side inne i den løpende teksten som danner *innholdet* i eskene. Eksempler er *opphavsform*, *henvisning*, *utheving*, *stil* og *taksonomi*.

Figur 11 viser eksempler på eskeelementer ringet inn med grønt, innholdselementer med rødt.



I figur 12 er eske- og innholdsmetaforen tegnet skjematisk, igjen med eskeelementer i grønt og innholdselementer i rødt.



Håndhevingen av elementtypenes dikotomi medfører følgende regler:

- **DTD-plassering.** Eskelementer må ikke tillates i DTD-en i en innholds-kontekst, og innholdselementer må ikke tillates i DTD-en i en eskekon-tekst.

- **Skillere.** For eskeelementer skal omsluttende skillere (linjeskift, mellomrom, kommaer o.l.) ikke ligge i dataene, men genereres automatisk av stilarket. For innholdselementer skal omsluttende skillere ligge i dataene, dvs. de skal settes manuelt av redaktør (hardkodes).
- **Stilark for redigeringsgrensesnittet.** Eskeelementene og innholdselementene må vises ulikt i redigeringsstilarket for at distinksjonen skal være tydelig for redaktørene. Eskeelementene, som altså bygger opp artikkelen, vises på ny, innrykket linje og f.eks. med full slutt-tag. Innholdselementer, som altså inngår i den løpende teksten, vises omløpende inne i linjen (ikke ny linje) og f.eks. med kort slutt-tag.

## Konsekvenser for innholdet

Nettpublisering med et transparent, brukervennlig stilark er nådeløst avslørende. Nettpublisering behøver selvsagt ikke være transparent. Det finnes mange eksempler på nettpubliserede ordbøker der de gamle, massive tekstspaltene er gjengitt mekanisk på skjermen, tro mot den gamle boka. Publiseringer man ordboken sin på denne måten, blir man i det lange løp straffet, slik Internett-publisering nå utvikler seg. Med den ubegrensede plassen som Internett-publisering gir, forventer brukerne at grensesnittet skal være luftig, klart og lett å finne frem i. Uten dette uteblir trafikken, og det er det vanskelig å leve med i lengden.

Luftighet og klarhet er svakt strukturert leksikografis største fiende.

For NAOB gikk det som det måtte da tåken ble blåst bort – mye kronglete vegetasjon kom til syne. Vi måtte riktignok ta av oss hatten for den systematikken som de gamle leksikografer allikevel hadde klart å gjennomføre, med datidens verktøy og med 30 år mellom start og slutt. Men ja, det lot seg ikke lenger underslå at de gamle mestere begikk mange kunstneriske avvik. Vi kunne ikke lenger lukke øynene for fortidens dype og uforståelige betydningshierarkier, sitatløse sitater, blindhenvisninger og speilhenvisninger.

Transparent visning får frem struktur- og innholdsfeilene i avslørende relief, men den nyetablerte logikken gir også muligheten til å skape de verktøyene som redaktørene trenger for raskt å reparere feilene. For eksempel kan betydninger flyttes med “drag-and-drop” hvor som helst i hierarkiet og betydningsnummereringen oppdateres automatisk. Henvisningsmålene følger med når artikler skifter stavemåte eller når artikkeldeler flyttes mellom artikler. Man har makroer som omformer uttrykk til eksempler og eksempler til uttrykk. Samt mange andre hjelpemidler.

Om publikum blir bortskjemt av det digitale mediets transparens, blir også redaktørene det. Har man en gang sett innholdets strukturer i skarpt relieff, vil man nødig tilbake til tåkehavet.

*Petter Henriksen*

Kunnskapsforlaget, Oslo

petter.henriksen@kunnskapsforlaget.no

# Russin i kakan i Jesper Swedbergs *Swensk Ordabok*

*Lars Holm*

Among some men of linguistic distinction in Sweden, there is or rather used to be an ambition to revive words that have more or less fallen into disuse. Whenever such an attempt has been made the idea has been to make the language pithy and more vivid, more Swedish as it were. Among these men should be mentioned Stiernhielm (17th century), Spegel (18th century), Rydberg (19th century) and Collinder (20th century). However, none of them were particularly successful in their puristic efforts.

Reading and editing *Swensk Ordabok* tempted me to make a 21st-century endeavour in the genre. The words I have picked out belong in the semantic spheres of conversation; thought, knowledge, scholarship; people (ordinary people, professional people and social outcasts); food and eating. The essay hovers on the borderline between being serious and joking.

NYCKELORD: äldre nysvenska, Swedbergs *Swensk Ordabok*, svenska > latin, purism, nygamla ord, normativ/subjektiv språksyn, ironi, språkglädje

## Inledning

”Til hwad nytto?” var en fråga som Swedberg ibland ställde sig – och självsäkert gav flera goda svar på – när han skrev samman sitt storverk *Swensk Ordabok*. Sedan ordboken äntligen blivit tryckt (Swedberg 2009) efter nära 300 år, ställer jag ofta mig själv samma fråga. Denna uppsats är tänkt att ge *ett* svar. Ett annat ges i boken *Till bords med biskop Jesper Swedberg – mat och dryck i Swensk Ordabok (ca 1725)* (Lars Holm 2010). (Där introduceras forskningsgrenen *lexikologisk arkeologi*.)<sup>1</sup>

Att avsiktligt försöka i svenskan återinföra ord som förr hört hemma där men glömts bort och dött ut – det är ett vanskligt företag. Men flera språkmän

---

<sup>1</sup> Tack till Nils Tryding för idén, till Lillemor Santesson för förbättringar och till Östen Rudal för abstract.

och hängivna amatörer har ändå gett sig i kast med den uppgiften. Några av de mest aktiva och mest omtalade är – en per sekel – Georg Stiernhielm under 1600-talet, Haqvin Spegel under 1700-talet, Viktor Rydberg under 1800-talet och Björn Collinder under 1900-talet. Att viljan att återuppliva ord uppkom just under 1600-talet är ingen tillfällighet. Då blommade de ”götiska” tankegångarna upp för första gången. Fornsvenska och än mer fornisländska blev på modet i lärda kretsar, och man sökte efter lämpliga ord att blåsa nytt liv i. Det nationalistiska draget är framträdande; utomnordiska ord kom inte i fråga; inaktuella tyska och franska lån, som var legio, talade ingen sig varm för – det fanns ju många nog kvar i svenskan, till vissa bedömares förtret.

Fyra århundradens återupplivningsförsök av svenska ord manar alltså inte direkt till efterföljd. Trots sina starka och långvariga räddningsinsatser lyckades Viktor Rydberg – enligt sin utforskare Pelle Holm – bara med ett enda ord, *dryfta*, som ursprungligen betydde ’rensa säd’ (så i Swedbergs *Swensk Ordabok*) men som Rydberg gav betydelsen ’diskutera’ (Pelle Holm 1918: 111). Och frågan är om inte ordet *dryfta* nu är på väg att försvinna. 32 000 träffar på Google för *dryfta* mot nästan 8 miljoner för *diskutera* hade inte glatt Rydberg. (Bo Wendt ha sagt mig att *dödsbo*, i stället för sterbhus, är ett mera lyckat exempel.) Ännu sämre har det väl gått för den glödande puristen Björn Collinders många forninspirerade ordförslag, publicerade i flera ordböcker med manande titlar som *Svensk ordhjälp*. Han mötte nog mest löje och likgiltighet och torde nu vara tämligen bortglömd i rollen som svenskans räddare och förnyare. Att kalla *bacon* för *bräckefläsk* fick Collinder i alla fall varken Ica eller Konsum att göra (Collinder 1968: 27).

Men ändå! Under mångårig läsning i Jesper Swedbergs *Swensk Ordabok* har jag ibland hajat till och tänkt: Det ordet behövs ju idag också – så låt oss återuppliva det! Mina maktmedel är ringa, men ingen kan heller förbjuda mig att t.ex. kalla mig själv för **halfgubbe**. Av *Swensk Ordaboks* drygt 20 000 svenska ord har jag valt ut ett par dussin ”russin i kakan”. Jag skulle lätt och gärna välja ytterligare 50 eller 100. Urvalet är garanterat subjektivt. Men av varje ord kräver jag att det ska fylla minst ett, gärna flera, av kraven att vara *kort, användbart, laddat, roligt*. Jag har samlat orden i ett par löst sammansatta betydelsegrupper som tjänar som mellanrubriker nedan.

Jag har inskränkt urvalet av referensverk till det mest minimala. Men att skriva om svenska ord utan att stödja mig på *Ordbok över svenska språket utgiven av Svenska Akademien* (SAOB) eller *Svensk ordbok utgiven av Svenska Akademien* (SO) kan jag knappast tänka mig. Som *schibboleth* för om ett ord ”finns” i vår tids svenska eller inte litar jag på SO.

## Samtal i olika tonlägen

Innan elektriciteten, med läsljus och teve, blev varje familjs egendom, hade man nog mera tid för verbalt umgänge. När alla sysslor var utträttade, kunde man kura skymning och hänge sig åt **afftonsnack** eller **sqwallertime** (skvallertimme) (Uppslagsord i *Swensk Ordabok* skrivs här med **fetstil**.) Swedberg ger med sina latinska översättningsord (*confabulatoria vespertina* resp. *hora confabulationis*) inga negativa bibetydelser. Företeelsen finns nog fortfarande, men SO känns inte vid de två orden. *Pratstund* är en synonym, men det har en mycket allmänare betydelse. Jag finner Swedbergs båda ord väl värda ett återupplivningsförsök.

*Swensk Ordabok* har naturligtvis flera ord för *tal* och *samtal* och många ord för dem som samtalar och för hur man samtalar, och de allra flesta är lika levande nu som på Carl XII:s tid. Men det finns fynd att göra i *Swensk Ordabok*! Ett sådant är **almäle** (allmäle). SAOB definierar det 'allmänt rykte'; *mäle* för 'yttrande, det som sägs' är ju välkänt i genmäle, tillmäle etc. Swedbergs latinska definition (översatt till svenska) 'tal som förs runt genom allas mun' får en onekligen att tänka på en viss sida av den svenska kulturdebatten, "enfrågesidan". Ett annat kort och praktiskt ord är **missmäle**. SAOB definierar det som 'missägning' ('felsägning' låter bättre i mitt öra).

*Swensk Ordabok* innehåller en rad ord för personer på olika sätt invecklade i samtal. Några av dem som inte längre är i bruk är alldeles för uttrycksfulla för att inte dras fram ur glömskans mörker – och kanske brukas någon gång av den som känner sig manad. Vad en kalckstrykare är vet inte SAOB, men Swedberg översätter det på latin med *adulator*, dvs. 'smickrare'. Jag har inte lyckats spåra ordet någonstans, inte i Rietz' dialektlexikon, inte på Google, men nog ger det en maffig känsla av att någon brer på ... En synonym – det har samma latinska ekvivalent – är **trappedragare**. SAOB definierar det som 'lismare' och har ett sista belägg från 1840. Det finns nog fortfarande någon (några?) som löper i höga herrars och damers trappor – eller kanske hellre åker i deras hissar – för att tjata sig till förmåner. Så även om ordet är lite otympligt borde det återupplivas. En som däremot inte alls talar själv utan lyssnar till andra som talar kan vara en **lurant**. Denna lustiga kombination av en svensk stam och en främmande avledningsändelse förklaras av SAOB som 'spion'. Swedberg är som vanligt noggrann med sin latinska definition (översatt): 'den som lyssnar försåtligt'. Inte heller den sortens filur torde vara helt utdöd, så ordet kunde återinföras.

Swedberg hade en viss svaghet för det drastiska, det inte alldeles salongsfärdiga. Den som till äventyrs trodde att det karolinska Sverige var en fridfull idyll kan befrias från denna fördom genom läsning i *Swensk Ordabok*. Här finns ord



av flera ordklasser som kan användas av den som vill beskriva ett gräl. Verben **ordkastas** och **ordkeppas** (ordkämpas) säger ju mer än det slitna *debattera*. Man hör och nästan ser spydigheter och tillmälen hagla genom luften. De två orden utbjuds härmed till bruk av alla både tillfälliga och kroniska åsiktsvädrare. Som beteckning för tillhyggen erbjuds samtidigt de båda **bettord** (= bitande ord) och **pludder** ('snack') (ett motgift mot *bullshit*, kanske). Samtalstonen kunde vara fränare förr, mellan Jesper Swedberg och Urban Hierne så ärerörig att vissa av deras inbördes stridsskrifter aldrig fick tryckas. Sist de hithörande adjektiven. Flera av dem är ganska drastiska, vilket inte borde göra dem mindre användbara i vår tid heller. Den som är **leppelett** (läppalätt) har lätt för att slå käft, den som är **måldierf** (måldjäv) vågar säga vad som helst, och den som är *trutwig* är både läppalätt och måldjäv (allt enligt Swedbergs latinska ekvivalenter). Två adjektiv är bildade till ordet *ord*: **ordadryg**, med vilket Swedberg menar '(alltför) talför', och **ordfager**, som betyder 'vältalig', kanske med ironisk eftersläng. Vem av våra partiledare är mest ordfager, eller ska vi säga ordfagrast?

## Tanke, kunskap, lärdom

Ett ord i besserwissersammanhang är *efterklok*. Det saknas i *Swensk Ordabok* men här finns i stället det likabetydande **bakwis**. Det är kortare men har en annan viktigare fördel: det korresponderar perfekt med **framwis** som märkligt nog saknas i SAOB. Något "föreklok" eller "förevis" har jag inte stött på någonstans (annat än sporadiskt på Google). Den perfekta besserwissern är naturligtvis både bak- och framvis (att inte säga bakochframvis). Den lätt ironiska bismaken i dessa båda ord saknas helt i ett sedan länge utdött svenskt ord för 'förstånd', 'begåvning' i allmänhet: **nemme**; SAOB har varianterna *nimme* och *nämme*. Ordet kan utöva en särskild lockelse på personer med skånska rötter (som t.ex. mig). Det är besläktat med det dialektala adjektivet *nim* (kort i), 'skicklig', 'smidig' – kanske snarast i fingrarna men också i huvudet. Andra släktingar är *förninna* och tyskans *nehmen*, 'ta'. Det gäller alltså att ta för sig av kunskapen – och behålla den för framtida bruk.

Ett vanligt ord i danskan som inte utan vidare förstås av svenskar är *huske*, 'minnas'. En svensk synonym är *komma ihåg*, och den avslöjar vad som är gemensamt för svenskan och danskan: substantivet *hug* (numera *håg*) i betydelsen 'sinne', 'tanke(verksamhet)'. SAOB känner *hugsa* – konsonanterna i mitten har kastats om i danskan – med betydelsena 'överväga', 'minnas' etc. Swedberg, slutligen, levererar verbet **hugsar**, vars latinska ekvivalenter motsvarar betydelsena enligt SAOB. Varför inte återuppliva ordet, om inte annat så för danskanas skull? För att visa svenskans suveränitet kunde vi ju stava det *huxa*! Hol-

ländskt ursprung kan förmodas bakom verbet **bekaijar**. Swedbergs latinska översättningsord ger betydelsen 'fånga', mera precist 'snärja (som i nät)'. SO tar inte upp verbet men väl dess perfekt particip *bekajad* och ger betydelsen 'belastad'. Ordet är inte dödförklarat av SAOB (men bandet är tryckt för mer än hundra år sen). Jag kan bekänna att jag sedan länge kan fälla ett kritiskt omdöme som "Han är inte bekajad av några som helst hämningar" – och det fungerar alldeles utmärkt! Ytterligare två verb i *Swensk Ordabok*, vilkas betydelser ligger nära varandra och som har samma stamstavelse, är **misshoppas** och **wahn hoppas** (vanhoppas). Det första finns i SAOB, ges betydelsen 'hysa falskt hopp' men sägs vara dött sedan 1807; det andra har SAOB inte hunnit fram till än. Swedberg översätter båda till latin med *despero* – som vi känner igen från *desperat* – men om man på modern svenska vill tala om att hoppas förgäves, måste man tillgripa just den frasen eller andra omskrivningar. Så låt oss inte vanhoppas om språklig förnyelse!

Jag har valt ut fyra substantiv som jag vill räkna till denna grupp av återupplivningskandidater. I förra gruppen förekom ord för muntliga gräl, men det finns också ett utmärkt sådant för skriftliga konflikter: **pennekrig**, lätt moderniserat *pennkrig*. Swedberg tycks ha skapat ordet; första belägget finns i hans bok om svenska språket *Schibboleth* (1716), och sen har det använts sporadiskt. Det är värt ett bättre öde, för vem skulle hellre se och nyttja det sakligt mera adekvata "ordbehandlarkrig"? Men den psykiska kokpunkten kan uppnås utan strid mellan två eller flera – inom en enskild människa. Swedberg läste gärna poeten Samuel Columbus och citerar ett par gånger hans dikter i sin ordbok (inte helt bokstavstroget); så här skriver Columbus i sin dikt "Lustwin":

Hwirfwel-swäfwand Tancke-swärm/  
 Widt ock hårdt och swårt Bekymmer/  
 är den swarte Sky ock Skärm/  
 Som wår glädje Sool oss skymmer.

Det ord som fångat ordboks författarens uppmärksamhet är (med hans stavning) det högst uttrycksfulla **tanckeswerm**. SAOB tycks ha glömt eller gallrat ut ordet; jag har inte funnit det i någon enda ordbok. Men Google har en knapp handfull belägg från innevarande decennium! De handlar om själspövningar i vår tid – utan syftningar till Columbus – så jag tycks inte vara ensam om tanken att återinföra det i och för sig olustiga ordet *tanckeswärm*. Överhettade hjärnor lär inte vara ovanligare nu än på 1600-talet.

## Personer

*Swensk Ordabok* innehåller många personbetecknande ord, annars vore det ju ingen bra ordbok. De flesta av dessa ord är ännu fullt levande i svenskan, medan andra kanske syns till bara när äldre tider skildras. Jordbruk, många hantverk, politik och förvaltning, i närmiljön och på rikspanet, (osv.) har förändrats eller helt förvandlats. Ändå kan det finnas bortglömda personbeteckningar som skulle kunna göra tjänst efter inemot 300 års vintersömn. Jag har gallrat ganska strängt bland kandidaterna till återupplivning.

Först *vanligt folk*. Tre ord sätter människans ålder i fokus. Ett nygift – helst ungt – par kallas i *Swensk Ordabok* **vngfolck** (ungfolk). En nygift kvinna kallas **vngmor** (ungmor). Att ”vngfar” saknas i ordboken är nog bara en miss av Swedberg. Den förträffliga *Ordbok öfver svenska språket* av A. F. Dalin (tryckt 1850–55) tar upp alla dessa tre ord på ung-. De finns också i SAOB:s hittills sista häfte (U 305–309). Här dödförklaras ungfär (belagt 1675–1948) men märkligt nog inte ungmor i den aktuella betydelsen (belagt 1674–1871). Ungfolk finns så sent som i Göran Tunströms *Prästungen* (1976). Mitt tredje förslag från gruppen ”vanligt folk” är **någranne**, det vill säga ’granne som man (lätt) kan nå’. Om man ska kunna nå sin granne med rop från sin egen gård, eller till fots utan att behöva ta ytterkläder på, eller utan att behöva ta hästen ur stallet (bilen ur garaget), avslöjar dock inte ordet. Med vår tids urbaniserade boende kan betydelsen också vara mera psykologisk: ’granne som man står på talefot med’. Så jag tror att ordet behövs nu också. Men min främsta favorit bland ord som är värda att dammas av och åter tas i bruk – alla kategorier – är nog **halfgubbe** (halvgubbe); skälet kan möjligen vara strängt personligt. Enligt SAOB har Swedberg själv myntat ordet; han förser det med den något eufemistiska svenska förklaringen ’medelålders’ och den latinska ekvivalenten *semisenex*. Så sent som 1916 användes det av romanförfattaren Sigfrid Siwertz. Och på Google flödar beläggen. Ett genombrott kan väntas!

Så *yrkesfolk*. Vad kallas de, kvinnor och män, som bär fram maten och dryckerna till oss förväntansfulla kroggäster? Yrkesbeteckningarna *uppasserska* och *kypare* låter tämligen bedagade. *Servitris* och *servitör* låter möjligen något bättre men är omöjliga som tillrop. Vad säger man, när man vill påkalla deras uppmärksamhet? *Fröken* och *vaktmästarn* låter nästan som före bytet till högertrafik. Men *Swensk Ordabok* har lösningen: **kellarpiga** och **kellarswen** (källarsven)! Den ungdomliga touchen på de senare sammansättningsleden bör kunna tilltala även yrkesutövarna själva. Därmed över till en annan samhällssektor, både mera prestigeladdad och mera ifrågasatt: den konstnärliga. *Målare* (som ju också kan beteckna ett rent hantverksyrke, vanvördigt kallat ”plankstrykare”), ibland kompletterat med *landskaps-*, *porträtt-*, *akvarell-* och många andra förled,

har också en gammal synonym som finns i *Swensk Ordabok*: **conterfeijare** (konterfejare). Swedberg översätter till latin med *pictor*, så yrkesmannen är i gott sällskap med t.ex. en Albertus, med det epitetet, som prytt medeltidskyrkor i Uppland. Ingen lär idag reta sig på Albertus' och hans yrkesbröders alster, medan visst folkligt motstånd torde finnas mot de *objets d'art* som går under namnet *installationer*. Men vad ska man kalla dem som utför sådana? *Bildhuggare* låter helt malplacé. *Skulptörer* är föga bättre. Kanske skulle Lars Vilks hålla tillgodo med att kallas *installatör*? Men det är ju redan en samlande benämning för elektriker och rörmokare. *Swensk Ordabok* har lösningen: rondellhundens skapare är absolut en **belätesnidare**!

Sist *löst folk*. Swedberg var socialt högst medveten, han hade hjärta för de fattiga och elända både som biskop och som privatperson. Men samtidigt hade han en fäbles för ord för människor som hade halkat ur eller aldrig tagits upp i det sociala systemet, av nödvång eller fri (?) vilja. Söker man i *Swensk Ordabok* under latinets *homo nauci* (ungefär 'ynklig människa'), finner man tio ord från **askfis** till **wrak**. Men ofta låter lexikografen orden stå oöversatta och lämnar läsaren med en allmän känsla av att stå inför ett samhällets olycksbarn som en **ragaskerska** eller en (ett?) **wrongsnöpe**. Här har jag stannat för några personbeteckningar som kanske skulle kunna göra tjänst i den eviga diskursen om mänskligt elände. De är inte humana, inte moderna, men kanske i alla fall lite bistert roliga. De är ingenting för värstingkramare.

Allmänt nedsättande är **galgfogel** med synonymen **galgristare**. På Carl XII:s tid – och långt därefter – var både bödelsyxor och galgar i allmänt bruk, men de båda orden torde ha betecknat personer som ännu inte befann sig i omedelbar fara att avrättas. Galgfåglar finns fortfarande, och SO ger ordet en mycket snäll definition: "skojfrisk men ngt obehaglig person". Varför inte uppliva också det mera uttrycksfulla *galgristare*? Det är en man som skriver sina egna lagar och bildligt är fräck nog att gå fram och skaka den galge – som inte längre finns och som han inte riskerar att bli hängd i. I sällskap har galgristaren kanske en **luspung** eller en **rabald**. Den senare lever kvar i ordet *rabalder* – det oväsen och bråk som rabalden och hans gelikar ställer till med. En yngre orosstiftare kallades förr av äldre (vänner av ordning) för *slyngel*. SO varnar med skylten "starkt nedsätt[ande]", och ordet hörs sällan numera. *Swensk Ordabok* kan bidra med ett snarlikt ord: **snyffel**, som till latin översätts med *puer audaculus* (småfräck pojke). Men nu får jag snart Barnombudsmannen på mig ...

Lösdriveri har aldrig varit uppskattat av den överhet som vill ha full kontroll över oss (och de skattemedel vi helst ska bidra med). Men lagen var strängare förr; spinnhus var slutstationen för kvinnor, gruvor eller tvångsutskrivning till armén för män. Att söka i *Swensk Ordabok* på de latinska orden *ambulator*, *circulator* och *vagabundus* ger ett rikt utbyte i form av svenska ord för *lösdriverare*.

Frågan är om inte vissa av dem skulle kunna göra tjänst även i vår tid. Det första av tre utvalda är **gaturennare**, som knappast tarvar någon kommentar. Två ord som båda börjar på konsonantkombinationen sl- (liksom flera andra nedsättande ord) är **slens** och **slunt**, och de förklaras båda närmare i *Swensk Ordabok*. Det första kompletteras med frasen **junkar slens**, vilket antyder att lösdrivare åtminstone kunde utge sig för att vara ”fint folk”; tanken går osökt till poeten och bedragaren Lars Wivallius (som dog när Swedberg var 16 år) eller kanske ännu lättare till Thomas Manns romanfigur Felix Krull. Det andra ordet preciseras med ordet **hoffslunt** (hovslunt), och det aktualiserar det redan behandlade **trappedragare**. Det fanns alltså lösdrivare/lycksökare även på högsta sociala nivå. Undras hur det står till med det idag.

## Mat och ätande

Om den senast presenterade gruppen ord innehåller få – om ens något – ord som verkligen skulle fylla en lucka i 2010-talets svenska ordförråd, så vill jag nu redovisa en dryg handfull ord som jag verkligen tror skulle berika vårt språk. Först ett starkt kort inom betydelsegruppen *näringsställen*. (Just detta byråkratord skulle väl aldrig användas av någon hungrig uteätare: ”Nu går vi på ett näringsställe!” Helt omöjligt annat än som ploj.) Standardordet är väl *restaurang*, med *krog* tätt i hälarerna. Det senare ordet har under de senaste decennierna tappat det mesta av sin prägel av lågspråk. SO definierar det till och med som ”(finare) restaurang”! Men vad ska vi då använda för samlande term för just enkla matställen? *Swensk Ordabok* har svaret: **gårkök**. Bland de latinska ekvivalenterna finns *thermopolium*, vilket en av krogarna i *Fredmans epistlar* kallas. Förledet *går-* är besläktat med det tyska *gar*, ’färdigkokt, /-stekt’. SAOB ger ett yngsta belägg från Karlfeldt (1927) och definitionen ”sylta” (citationstecknen källans). Som *gårkök* kunde räknas kvarterskrogar, vägre Restauranger, gatukök, ”hål i väggen” etc., alla med den gemensamma nämnaren *enkel* (och rimligen också *billig*). Har man tur kan man kanske även i ett gårkök bli serverad **lecker-mat**. Observera att detta är *ett* ord med huvudtryck på första stavelsen, inte ett exempel på felaktig sammanskrivning. Ordet kan få det att vattnas i munnen, åtminstone på mig, och det kan ersätta det längre och lite slitna *delikatesser*.

Men mat är inte bara näring. Det finns och fanns ord för de yttre formerna kring dess intagande. Studenter och gymnasister åt förr ofta i *matlag*, som höll till i en professionell ”matmamas” privata kök eller matsal. I stället för detta ord innehåller *Swensk Ordabok* ordet **matskap**. Enligt de latinska ekvivalenterna har detta ord ett vidare betydelseomfång, ’personer som äter tillsammans’. Det är tillämpligt inte bara på studenter i matlag utan också på t.ex. åtta pensione-

rade f.d. kolleger som förljuvar sin ålderdom med att gemensamt äta sin onsdagslunch på ett noga utvalt *gårkök*. (Man har sett exempel på detta i 2010-talets Lund.)

Så sitter man då där och äter och dricker. Det finns beteenden som är ”fina” och ”mindre fina”. Några enhetliga matbordsnormer lär inte finnas i vår tid – och ännu mindre några som skulle gälla både 1700-talet och 2000-talet. Inte desto mindre innehåller *Swensk Ordabok* ett par ord om bordskick som har kommit ur bruk – men som vore väl värda att tas i bruk igen. Det första är **magahof** – SAOB har formen *maghov*, vilket förklaras som ’måttlighet med mat’. Ordet blir genast mera begripligt, när man tänker på ord som *housam* och *behov*, bildade till ett gammalt *hof*, ’mätta’. Swedbergs latinska definitioner är så eleganta att de förtjänar att citeras: *modus in edendo et bibendo, mensura stomachi, satietatis modus*. (Det räcker med att kunna engelska för att förstå det latinet.) Ett ord inom samma betydelsefält är adjektivet **fagermått**. Det tillhör de ganska talrika uppslagsord som Swedberg inte översätter eller definierar (på latin eller svenska); han hann väl inte. Men SAOB, som belägger ordet mellan 1604 och 1850, står till tjänst med en definition: ”som (av önskan att visa fina eller goda seder) äter endast litet eller låtsar sig vara mått ...”. Ordets tillämpning torde vara både köns-, ålders- och klassbunden; det har ringa chans till någon renässans bland tonåriga kampsportskillar eller frossande gourmander på herrmiddag. Ett annat av Swedberg oförklarad ord är **machtbete**. Ordet torde vara en variant av vad SAOB kallar *maktbit*, den sista biten bröd (eller kött ...) på matbordet; den bör helst sparas, för den innehåller en extra, samlad kraft. (På tyska talar man om *Anstandsbrocken*, den sista bit som ingen får röra – eftersom det vore en insinuation mot värdfolket för att ha bjudit snålt.) Sist i ordraden adjektivet **matfrågen**, som ju helt enkelt betyder ’hungrig’ men som säger så mycket mer om hur gärna man vill sätta sig till bords.

Återstår två ord inom matsfären som inte alls har med fint sätt vid matbordet att göra men som skulle berika svenskans ordförråd, om de (äter)infördes. Ända sedan jag var barn har jag retat mig på namnet på ett av ”de fyra sädesslagen”: *korn*. Men det finns ju *vetekorn*, *rågekorn* och *havrekorn* också! (Knappast ”kornkorn”; det låter för tokigt.) Eftersom det sädesslag vi nu kallar korn förr var det mest odlade, var det sädesslaget *an sich*. Men som namn på ett visst sädesslag konkurrerade det länge med **biugg** (bjugg), ett namn som efterhand glömdes bort i riksspråket. En återupplivning av *bjugg* som ’sädesslag odlad främst för ölframställning’ och en precisering av *korn* som ’frö av sädesslag’ skulle skapa ordning och reda på fältet. – Sist ett uttryck i *Swensk Ordabok* som jag finner helt outhärligt: **kriska äplen**. SAOB upptar adjektivet och definierar det som ’syrlig om frukt’. Men *syrlig* får mig inte att höra hur tänderna klyver

äpplet med ett litet brak och saliven forsar fram i munnen så som *krisk* gör. Ljudsymbolik?

Det finns mycket mer att hämta i Swedbergs *Swensk Ordabok* – tag och läs!

## KÄLLOR OCH LITTERATUR

- Collinder, Björn, 1968 (etc.): Svensk ordhjälp för skolorna och den skrivande allmänheten. Stockholm.
- Dalin, A. F., 1850–55: Ordbok öfver svenska språket, 1–2. Stockholm.
- Holm, Lars, 2010: Till bords med biskop Jesper Swedberg – mat och dryck i Svensk Ordabok (ca 1725). Skara.
- Holm, Pelle, 1918: Viktor Rydberg som språkrensare. Lund.
- Rietz, Johan Ernst, 1867: Svenskt dialekt-lexikon eller Ordbok öfver svenska allmogespråket. Lund.
- SAOB = Ordbok över svenska språket, utg. av Svenska Akademien, 1893 ff. Lund.
- SO = Svensk ordbok, utg. av Svenska Akademien, 2009. Stockholm.
- Swedberg, Jesper, 2009: Svensk Ordabok, utg. efter Uppsala-handskriften, med tillägg och rättelser ur övriga handskrifter, av Lars Holm. Skara.

Google (diverse ord; sommaren 2010.)

*Lars Holm*

[lars.holm17@gmail.com](mailto:lars.holm17@gmail.com)

# Norsk dialektatlas i digital utgåve og språkdata på kart

*Jon Holmen & Oddrun Grønvik*

This paper discusses the mapping of linguistic data as a support for research in spoken language. Maps link linguistic phenomena to place from the specific level (dialect forms), to the general (demarcation of language families). Lexicography integrating speech data uses geographical mapping of linguistic data to supplement place-name lists. The technology behind digital mapping is in rapid development, and visualization of linguistic data in maps is a professional growth area.

The University of Oslo owns a collection of 596 dialect maps, based on data from a systematic registration 1950–1970.

Digitizing the maps started in 2008, aiming at making the map data searchable by linguistic criteria. In digitized maps, layers add isoglosses and symbols to place. Maps can be searched with multiple arguments and present new results.

The model for map visualization is generalized. The maps link dialect forms, informants and places. The model simplifies virtual storage of linguistic data linked to place and informant, and generates map information (areas and isoglosses) from the detailed information.

The chosen tool chosen is ArcMap. The map information is stored in Oracle Spatial. Tool choice is relevant to the planned product and is commented on below.

*Nøkkelord:* datasamling, dialektatlas, digitalisering, talemål, isogloss, kart, kartverkty, målføre

Språkvitenskapleg framsyning av data på kart har pågått frå slutten av 1800-talet i Europa (Hoff 1954: 1 ff.). Hovudføremålet har vori å heimfesta talemålsdata. Dette arbeidet har resultert i fleire store europeiske utgjevingar, særleg på kontinentet, og i minst eitt stort pågåande prosjekt (Atlas Linguarum Europae – ALE).



## Kvifor språkkart?

Kart syner geografisk utbreiing og førekomst på ein meir effektiv og lettfatteleg måte enn tekst gjer, og eit bilete av utbreiinga av eit fenomen er likeins lettare å minnast. Dette gjeld alt som har geografisk utbreiing, også språk. Mange leksikografiske verk bruker derfor kart i større eller mindre grad som supplement til tekst i ordartiklar.

Den viktigaste hindringa for bruk av kart over har tradisjonelt vori at dei er dyre og krevjande å laga på papir. Kvart fenomen krev sitt kart, og karta måtte teiknast for hand. Dersom ein ønskjer kart som kombinerer ulike fenomen, måtte dei teiknast på ny – det finst ingen enkel måte å smelta saman to papirkart på. Det finst prenta samlingar av kart og plansjar på gjennomsliktig plastfolie som kan leggjast oppå einannan, men karta er førehandsdefinerte. I fall dei er innbundne, er rekkjefølgja fastsett av utgjevar.

## Det ideelle digitale kartverktøyet

Det kan vera nyttig å tenkja seg korleis det ideelle kartverktøyet burde vera. Der burde den som treng kartet, sjølv kunna velja kva fenomen som skal kartleggjast, pluss framvisingsmåte og kartkode, og så produsera kartet sitt ved hjelp av innmata datasett og nokre enkle menyar. Frå norsk synsvinkel kunne ein byrja med eit digitalt matrikkelkart over heile landet (irekna grenseområde mot granneland) med alle historiske administrative grenser, med tidfesting. Ein bør kunna velja kva tidspunkt ein ønskjer å ta grensene frå. Vidare bør ein kunna velja kartutsnitt, kva fenomen ein vil syna, og om fenomenet skal visast som flate, linje eller punkt på kartet. For publikasjonsføremål må ein kunna velja fargar, stiler for linjer, fonter og punktsymboler fritt. Ein må også ha tilgang til kartdata fra andre fagfelt for relevant annan informasjon som ein ønskjer å visa dei utvalde dataa i samanhang med.

Det er eit stykke fram før dette kartverktøyet ligg i hendene på norske – og utanlandske – lingvistar og leksikografar, men digitaliseringa av Norsk dialektatlas (ND) er ei byrjing, og ideen om det generelle og multifunksjonelle kartverktøyet ligg i botnen for arbeidet.

## Data attom digitale språkkart

Kva slags data ligg attom vurderingar av geografisk utbreiing av eit språkleg fenomen? Basisdata er alltid ord, anten ord i kontekst eller einskildord. Alle framstillingar av overgripande språklege fenomen, t.d. infinitivsending eller

tjukk l i norske målføre, må vera dregne ut av dokumenterte samlingar av ord og tekst. Overgripande språklege fenomen kan i sin tur ha ulike merkelappar etter om framstillinga har eit lokalt eller landsomfattande perspektiv, eller er synkront eller historisk innretta. Såleis har mange av karta i ND-samlinga titlar av typen "Gamalnorsk –*ir* i fleirtal". Denne tittelen fortel kva føremål kartet har tent, men dømetilfanget er ordformer samla inn mellom 1950 og 1970.

Dersom ein skal kunna samordna ulike datasamlingar og bruka dei til å kartleggja språklege fenomen, må derfor alle vera indekserte utfrå dømeord, og det må vera mogeleg å sjå kva for datasamling dømeorda kjem frå. I tillegg kjem sjølvstøtt tittelen på kartet, som syner kva fenomen kartet skal kasta lys over, og andre peikarar til det aktuelle fenomenet.

## Sjølve kartsamlinga

Kartsamlinga omtala som ND omfattar 596 kart og kartskisser på papirark i storleik A3 eller mindre. Dei er teikna om lag 1950-1975, nokre få etter dette. Dei fleste karta tek utgangspunkt i eit kommunekart over Noreg som syner kommunegrensene i 1940, men somme er teikna frå botnen. Mange av karta har eigne, til dels omfattande, maskinskrivne kartnotat, i tillegg til overskrift og kartnøkkel som er skriven på papirkarta. Kartsamlinga med følgjenotat har aldri vori klargjord til publisering i bok. Einskildkart er nytta og publiserte i andre samanhengar, og ein del av kartsamlinga (den delen som teiknarane rekna seg ferdige med) er blitt kopiert og utlevert til somme andre språkvitskaplege institusjonar. Betre tilgjenge til kartsamlinga har lenge vore ein prioritet for norsk målføregransking (Dalen 2001).

## Norsk dialektatlas – kort historikk

Det første initiativet til å få eit norsk dialektatlas vart teki i 1951 (Hoff 1959: 595). Prosjektet vart lagt til Norsk målførearkiv (NMA) ved Universitetet i Oslo (1936-2002), som tilsette kartteiknar i 1952 (Wiggen: 108). Det storstilla føremålet med ND var todelt. For det første ville ein byggja opp ei kartsamling – eit atlas – som heimfeste målføreinndelinga i Noreg (Hoff 1959: 598 f.). For det andre ønskte nok NMA å stø opp om bruken av heimfesting på kart som ordinær vitskapleg metode for synkron talemålsforskning. Hindringa for at dette kunne gjerast i større omfang, var nettopp kostnaden ved å utarbeida papirkart.

Etter den opphavlege planen skulle ND ha tre band, dei to første med hovudgrensene for fonologi og morfologi i norske og tilgrensande målføre, det tredje med ord- og omgrepsgeografi (Hoff 1959:598). Referanseverk for både

ND og Målføresynopsisen<sup>1</sup> var Storms kortare ordliste og Sigurd Kolsruds *Nynorsken i sine målføre*. Tema, dømeord og prioritering av dei er såleis i stor grad henta frå Storm-lista, medan Kolsrud vart hovudreferanse for syntetisering av målføredata uttrykt i områdeinndeling.

Det var ikkje midlar til å samla inn eigne data til ND. Målførearkivet nytta difor synopsistilfanget, pluss tilfang som kom inn i samband med særgranskningar. Tilfanget var for det meste var samla inn i perioden 1945-1959, og var så godt verifisert som dei tilsette greidde å få det (Hoff 1959).

Fleire kartteiknarar har signert ND-kart, men den mest produktive var utan tvil Johan Anton Schulze (1924-), som vart tilsett på prosjektet i 1952. Han har teikna 302 av dei 596 papirkarta som no ligg i arkivet. Ein annan viktig kartteiknar var Ingeborg Hoff (1911-1993), fyrstearkivar ved NMA (1948-1972).

Etter at NMA vart nedlagt som eiga universitetseining ved Universitetet i Oslo (Borg 1996: 108) vart arkiva i 2002 overførte til dei språksamlingane ved UiO som nynorskleksikografane forvaltar. Formelt er det Institutt for lingvistik og nordistikk (ILN) som har ansvaret for arkiva til NMA. Kart og tilhørande notat er no lagra i stålskap i kjellararkivet til Norsk Ordbok 2014.

Frå 1976 og fram til NMA var nedlagt, var ND tenkt som grunnlag for eit norsk bidrag til Atlas Linguarum Europae (ALE). Dette tiltaket, kombinert med utgjevingsplanen for ND, fekk som følge at kartsamlinga vart omstrukturrert (ND-nummerering) i tillegg til den serielle kartnummereringa som var nytta til då. ND-nummereringa følgjer i hovudsak den opphavlege utgjevingsplanen, med fonologi og morfologi først. Lenger ut i nummerrekkja ser ein at karta truleg er teikna i samband med ymse særgranskningar, og tilfanget er ikkje så strengt ordna.

---

<sup>1</sup> Målføresynopsisen består av 43 handskrivne protokollar som er resultatet av ei stor undersøking av uttalen av om lag 1500 utvalde ord frå alle herada i Noreg, utført om lag 1950 til 1970. Det er frå ein til tre protokollar per fylke. Målføresynopsisen er ei unik kjelde til innsyn i norske målføre på 1900-talet. Både i høve til eldre og yngre forskning i norsk talemål utgjer Målføresynopsisen eit målepunkt som ein ikkje kan sjå bort frå. Særleg verdfullt er det at Målføresynopsisen ligg føre som ein samla dokumentasjon, med alt tilfanget ordna etter same skjema med tanke på jamføring og systematisering. Protokollane er lagra ved Intitutt for lingvistik og nordistikk ved Universitetet i Oslo, som del av dei nynorsk språksamlingane. Dei er fotoskanna og lagde i database i 2006, og er fritt tilgjengelege på nettet.

## Norsk dialektatlas som kartsamling

ND er blitt til ved éin institusjon, med få teiknarar og eit etter måten einheitleg datatilfang som utgangspunkt. Det har tidleg vori tankar om ein utgjevingsplan, bygt på to sentrale referanseverk. Gjevi dette må ein, når ein ser på kartkatalogen, undra seg over at karttilfanget framstår som såpass ueinsarta. Karta er såleis ikkje i innhald knytte opp mot paragrafar i Storm-lista, slik Målføresynopsis er, og korkje første eller andre nummerering følgjer Storm-lista – eller utgjevingsplanen – heilt ut.

Sjølve kartteikninga er også svakt standardisert. Det er eit minus at det ikkje ligg føre ein skriven, sams kartteiknarinstruks som slår fast konvensjonane for teiknbruk (linjer, fargar eller symbol), og teiknordning. Kartsymbola (teikn og linjer) er viste i kartnøkklane, og har uttyding, men dei har ikkje namn. Dette har ikkje noko å seia for bruken av einskildkart, men når ein skal samordna karta, og prøva å syntetisera kartteiknar-praksisen, står ein andsynes ei utfordring. Det kan sjå ut som om karta er blitt til eitt for eitt, og at tanken på heile verket ikkje har vori heilt til stades i arbeidet med detaljane.

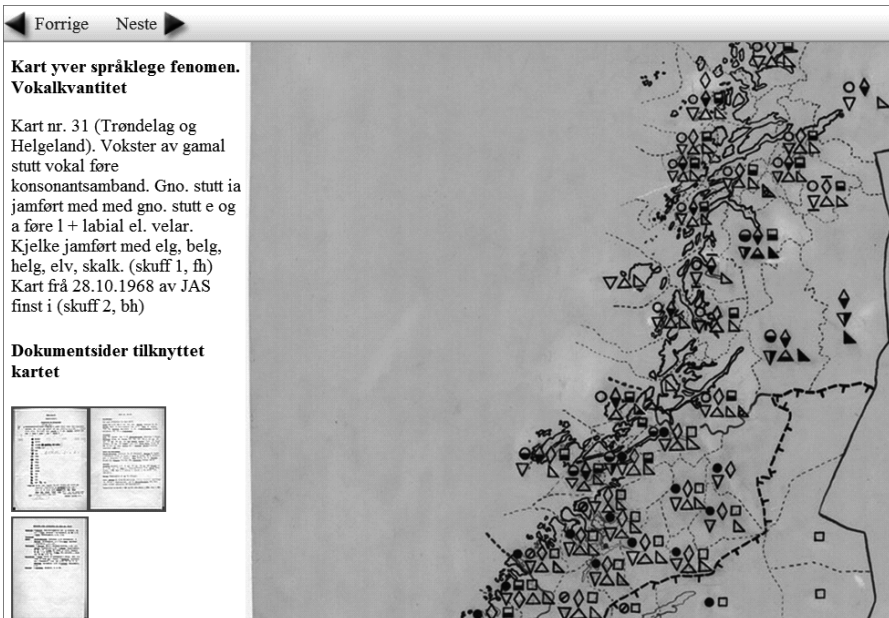
## Første digitalisering av Norsk dialektatlas

Papirutgåva av ND vart aldri fullført etter planen, og det tilfanget som ligg føre, er ikkje gjort klar til prenting. I nokre år etter at NMA i praksis var nedlagt, låg samlinga daud i kjellaren på UiO. Arbeidet med å registrera og digitalisera karta tok til i 2006 på initiativ frå Kristin Bakken, dåverande direktør for Prosjekt NO2014, og i samarbeid med Eining for digital dokumentasjon (EDD) ved ILN (Tangen 2006: 1). Karta vart fotoskanna 2006-2007, og lagra av EDD, likeins følgjenotata. Det vart utarbeidd eitt grundig kartregister, som har opplysning om ND-nummer, område, emne og emnehierarki. Systematikken for emnehierarkiet er teki frå Storm (1884) og Kolsrud (1951).

Det vart også laga ein database som gjer det mogleg å søkja i kartregisteret, og så sjå karta som bilete, i lag med eventuelle følgjenotat, som også er viste som bilete. Databasen med nokre kart vart tilgjengeleg på internett i 2007 (sjå URL i litteraturlista). Våren 2011 vart alle 596 kart lagde ut. Ein kan søkja i kartarkiv, kartnamn, karttype og kartkommentar og ND-nummer, alt dette på grunnlag av kartregisteret (Tangen 2006). Dermed er råtilfanget til det som skulle blitt ND publisert og fritt tilgjengeleg for ålmenta, og det er eit viktig framsteg, gjevi at denne kartsamlinga rpresenterer mange tital forskarårsverk. Illustrasjon 1 nedanfor syner søkjevindauga med søkjekategoriar, illustrasjon 2 syner eit tilslag i dagens nettutgåve.



Illustrasjon 1. Søkjevindauga kartdatabasen i dag.



Illustrasjon 2. Døme på tilslag

## Digitalisering av innhaldet i ND-karta

Sidan 2008 er det arbeidd med å få til ein meir effektiv presentasjon av kartinformasjonen i ND. For dei to hundre første karta (etter ND-nummereringa) er teksten på karta – titlar og kartnøklar – skriven inn, likeins teksten i følgjenotata. Innskrivarane har – så godt det er mogleg – skildra symbol og symbolbruk. Isoglossar er sette inn med koordinat slik at område utgjer delpolygon av noregskartet.



Illustrasjon 3. Døme på problematisk kart

Transkripsjonen er til dels vanskeleg av di karta er teikna i liten skala og dermed unøyaktige. Karta er heller ikkje alltid er så tydeleg som ein kunne ønskja, sjå illustrasjon 3 ovanfor, der kart nd004 (Målføreinnndeling) er synt. Dette er eit av dei eldste karta, frå 1951.

## Programvare til framvising av karta

Som nemnt skal karta visast fram som ulike kartlag oppå et grunnlagskart frå Statens Kartverk. Kartdataa blir lagra i Oracle Spatial med uthenting via Oracles søkemotor. Føremonen med dette er trygg lagring og kraftig søkjemotor.

Karta skal visast på internett ved bruk av OpenLayers. Dette er ei Open Source programvare som er nytta mellom anna av Google Map og Statens Kartverk. Fordelen med denne programvara er at ho er gratis, har mange brukarar (eit stort arbeidsmiljø) og tilnærma uavgrensa tilpassings- og variasjonsmåtar.

## Føremål og metode for vidare digitalisering

Per dato er det slik at kartinformasjonen berre finst i karta. Føremålet med den vidare digitaliseringa er å gjera det mogleg å gjera såkalla romlege søk, det vil seia gjera sjølv kartinformasjonen søkbar. Vi ønskjer å laga ein kartdatabase der ein kan (1) søkja i *sjølv karta* etter (a) språkleg fenomen (ned til lågaste oppgjevne nivå, som ordform), (b) stad (ned til 1940-kommune), (c) isogloss og (d) kartsymbol; (2) kombinera søk i kart og kartregister; (3) leggja kart oppå kvarandre og få opp nye kombinasjonar av opplysningar; (4) splitta opp aggregert kartinformasjon i einskildkart (til dømes visa eitt og eitt dømeord, ein og ein isogloss). Denne tilarbeidinga av kartinformasjonen vil føra oss eit langt steg nærare det ideelle kartverket som er omtala innleiingsvis.

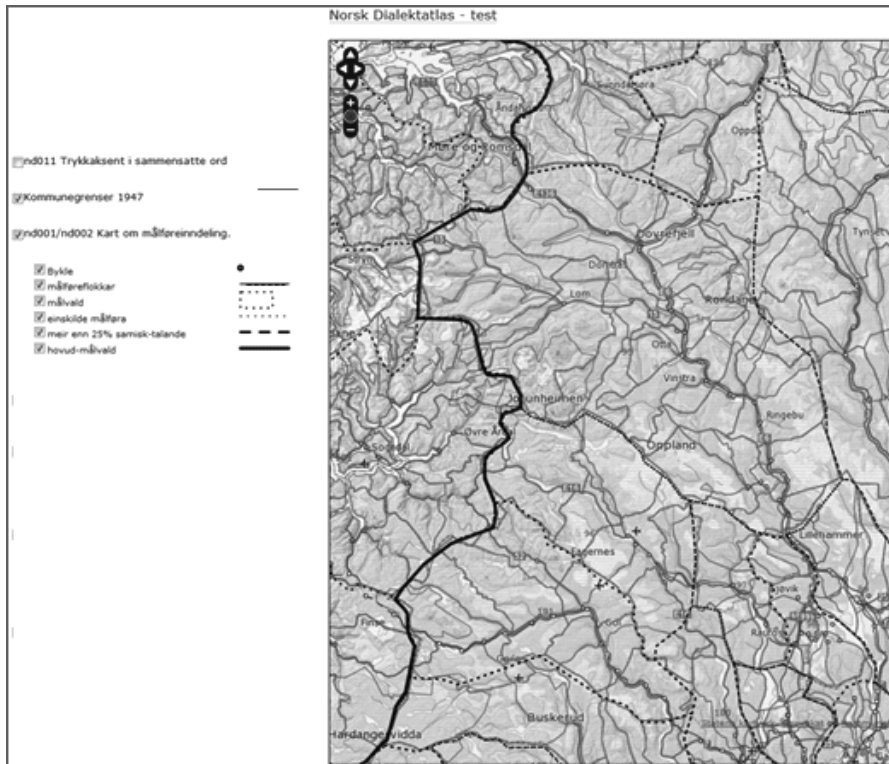
For å få til det, må vi standardisera og digitalisera grensemerking og symbolbruk på karta. Prosessen kvart kart må gjennom, er denne: (1) georeferering av manuskarta – det vil seia at ein må setja koordinat for hjørnepunkt og kartprosjeksjon. Dette er no gjort for heile ND. (2) Eigenskapane på kvart manuskart må vektoriserast til geometriske objekt (punkt, linjer og polygon). Dette arbeidet blir utført i ArcGIS, og dataa blir lagra i Oracle Spatial. (3) Neste steg er å transkribera kartnøklane på karta og følgjedokumentasjonen, og kopla dette saman med dei geometriske objekta (som nemnt ovanfor er ein god del av transkripsjonen alt gjort).

Attende til isoglossane: Innsetjinga av koordinat på isoglossane må bli unøyaktig i og med at skalaen er såpass liten. At dei er det, kjem fram når ein forstørrar karta eller legg digitaliserte isoglossar frå fleire kart oppå kvarandre. Då ser vi at isoglossar som skulle liggja oppå kvarandre, ikkje gjer det. Det blir skapt inntrykk av gråsone der ein veit at det ikkje er gråsone, og kartdokumentasjonen syner at kartteiknarane ikkje har meint å laga det. Kva gjer vi så med det?

Frå Statens kartverk har vi fått kommunekartet av Noreg frå 1940 i digital utgåve. Dette kartet syner grensene for den tids 750 kommunar (i dag om lag

400). Tidlegare målføregransking i Noreg har synt at målføregrensene fell etter måten godt saman med administrative grenser (Kolsrud: 63). Dersom digitaliseringa av isoglossane på ND-karta stadfester dette, er kommunekartet frå 1940 fintmerkande nok til det meste av dei dataa som skal registrerast. Ein mogleg framgangsmåte, som ville forenkla innteikninga av isoglossar, er derfor å leggja til grunn at isoglossar fell saman med administrative grenser. Då kan ein merka opp målføreområde ved å knyta kommunar til ulike språklege fenomen. Det må framleis vera høve til å markera gråsoner og trekkja grenser på tvers av administrative grenser, men det vil ikkje ofte vera bruk for dette i høve til det eldre datatilfanget.

Dømet nedanfor syner korleis framvisinga kan bli når ein legg eit målførekart oppå kommunekartet frå 1940.



Illustrasjon 4. Hovudinndeling mellom austnorsk og vestnorsk (knytt til utviklinga av ord med kort rotstaving i norrønt). Utsnittet syner i aust området frå Numedal til Sør-Trøndelag.

Kartet viser isoglossen frå det teikna kartet med hovudinndelinga mellom austnorsk og vestnorsk, lagt oppå det digitale kommunekartet. Her er også dei

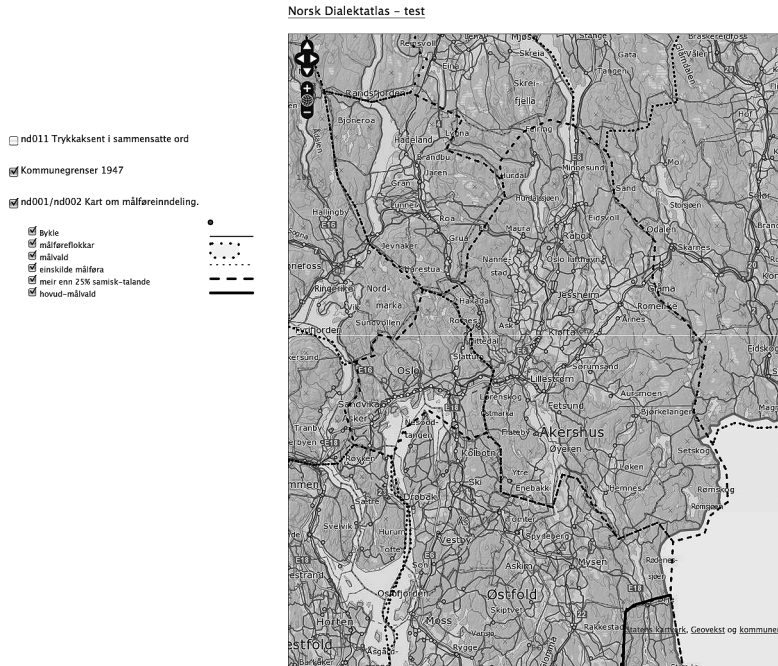


teikna inn grenser mellom ulike målføreområde (prikka linjer). Dei teikna grensene fell godt saman med kommunegrensene. Dette resultatet stadfester at det kan vera forsvarleg å bruka kommunekartet som utgangspunkt for lenking av målføreinformasjon, slik at den språklege opplysninga blir knytt til eit ferdig definert polygon.

Det er ikkje alltid samanfall. Rett til høgre, eller aust for, jamvektsgrensa ligg Gudbrandsdalen, austanfor Gudbrandsdalen ligg Østerdalen, og denne grensa er innteikna. Ho skjer gjennom kommunar i to tilfelle. Det eine stemmer med målføreopplysningar annan stads frå (Folldalen), det andre tilfellet er meir tvilsamt (Sollia). Dette kan vera ein digitaliseringsfeil, eller ein feil i det opphavlege kartet. Er det det, kan det hanga saman med skalaen i det opphavlege kartet. Det vil uansett vera bruk for å kunna dra isoglossar på tvers av kommunegrenser.

## Målføreområde og endring over tid

Kartutsnittet nedanfor syner Sørøstlandet rundt Oslofjorden. Her er også kommunekartet frå 1940 lagt i botnen med hovudkartet for inndeling av målføreområde oppå. Målføreområdet for oslomål er her teikna med sørgrense i Oslofjorden, på nordsida av Nesodlandet. Dei sørlegaste bygdene i Akershus, Follo, har i følgje dette kartet ei dialekt som høyrer nærare saman med dialektane i Østfold enn med oslomålet. Denne vurderinga av fottomålet stemmer med det Sigurd Kolsrud seier i boka Nynorsken i sine målføre frå 1951, og både kart og bok byggjer på innsamlingar i dei føregåande tiåra.



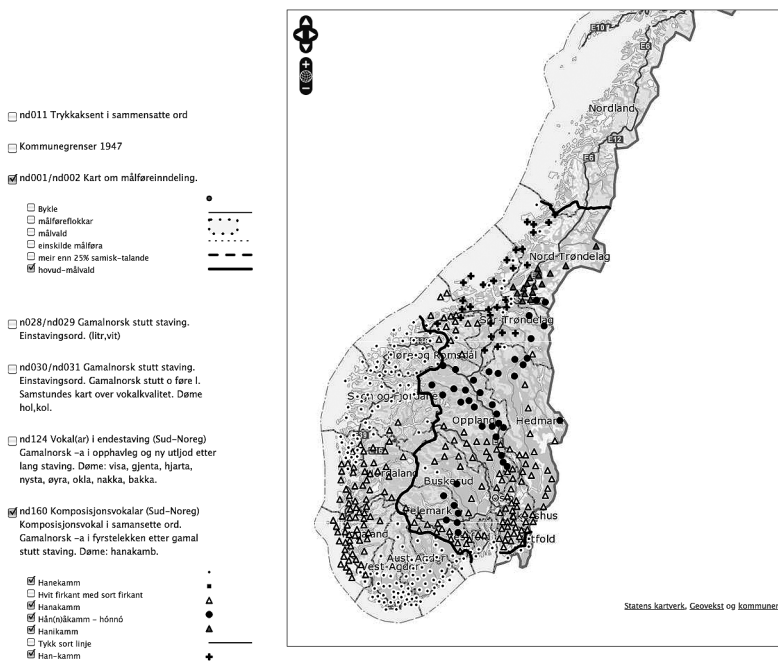
Illustrasjon 5. Sør-austlandet. Hovedinndeling lagt oppå kommunegrenser av 1947

Follobygdene har etter 1950 blitt del av omlandet for hovedstaden, med sterk innflytting og mykje arbeidspendling. I dag blir det opplevd slik at grensa mellom oslomål og østfoldmål går om lag på fylkesgrensa mellom follobygdene og Østfold, altså rett nord for Moss.

Dermed kan ND-karta også nyttast som tidsbilete. Dei gjev eit målepunkt for stoda i dag mot stoda i 1950. Dette kan gjelda også for andre språkgrenser. Det vil bli mogleg å leggja kart for ulike tidspunkt oppå kvarandre og dermed syna korleis isoglossar flyttar seg.

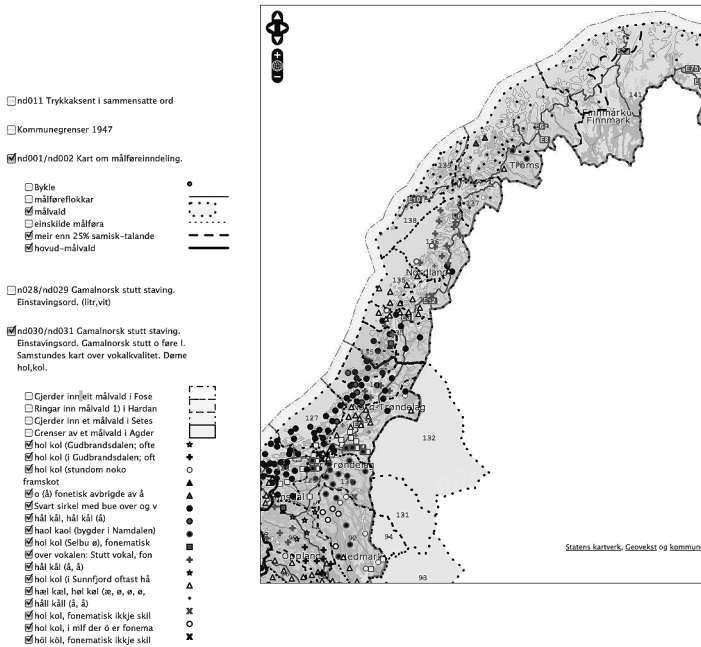
## Eit par døme på kombinerte målførekart i digital versjon

Nedanfor er det synt døme på samankopling av to målførekart. Det eine, nd001, syner hovudskiljet mellom austnorsk og vestnorsk, altså grensa for jamvekt (jf. illustrasjon 1 ovanfor). Det andre, nd160, syner fugevokalen i samansetningar med førsteledd som er gammalt tostava maskulinum med kort rotstaving. Dømeordet er ”hanekam”.



Illustrasjon 6. Utvalte dialektfenomen mot hovedinndeling og målvald. Hanekamm realisert som *hanekamm*, *hanakamm*, *hånnåkamm*, *han-kamm*.

Kartet syner at forma *hanakamm* – market med trekant med spiss oppover – er godt representert på båe sider at hovudskiljet mellom det sentrale austlandet i aust, og a-målsområdet i vest. Men hovudskiljet gjer at forklaringa på *a*-fugen må vera ulik. I vest er det *a*-mål, og bindevokalar av norrøn genitiv på *-a* er bevart. I aust er det jamvekta, realiserte som *a-a*, som gjer at bindevokalen *a* har haldi seg i svake maskulinum. Dette kartet er dermed eit døme på korleis ein med føremon kan kombinera fleire kart for å få fram eit poeng.



Illustrasjon 7. Utvalde dialektfenomen mot hovedinndeling og målvald. Tjukk l og gammel kort o i Trøndelag og Nordland. Dømeord *hol, kol*

I denne illustrasjonen er hovudinndelingskartet nd001 lagt under nd030 og nd031. Kartet syner eit utsnitt frå Nord-Trøndelag og over i Nordland. Igjen er hovudgrensene lagde under eit kart som syner formdistribusjon for eitt fenomen, nemleg norrøn stutt o framfor l, i dømeorda *hol* og *kol*. Kartet syner dermed også vokalkvalitet.

Svart prikk, som tyder 'stutt o framfor tjukk l', går eit stykke opp i det norsk-norske området – forbi grensa for jamvekt. Formene *håll, kåll* er einerdande i Troms og Finnmark bortsett frå Bardu. Dette kartet gjev opplysning om svært mange ulike former av dømeorda, og er så rikt på kartsymbol at det ikkje lett å lesa. Den digitale utgåva bruker raudt og svart, og er dermed lettare å lesa enn papiroriginalen, men det er framleis for mykje informasjon her. Men i digitaliserte kart vil det vera mogleg å slå av og på dei ulike symbola. Dersom ein berre er interessert i ei eller to ordformer, kan ein syna berre dei. Mange stader er òg dømeorda *hol* og *kol* realiserte litt ulikt. Med litt eksta arbeid er det mogleg å lenka ordspesifikke symbol til ordform. Da kan ein få eitt kart for *hol* og eitt for *kol*.

## Framtidsmusikk

Når kartsamlinga attom ND er ferdig digitalisert som skildra ovanfor, kan ein gå vidare til neste steg, som er automatisk kartgenerering. Om ein slik funksjon vart knytt til nettutgåva av Norsk Ordbok 2014 ville ein enkelt kunna syna utbreiing for ordformer og tydingar – ei utbreiing som i dag kjem fram i form av (til dels svært lange) lister med namn på kommunar og fylke.

Vidare kan ein sjå for seg at kartgenerering blir eit standardverktøy for å syna utbreiing av språklege fenomen. Det som trengst, er at kartfunksjonen blir kombinert med eit indekseringsverktøy og med høve til å byggja omgrepshierarki.

## LITTERATUR

- Borg Arve 1996: Quo Vadis Målførearkivet? I: Talatrusten 1996 s. 107-109.  
 Dalen, Arnold, 2001: Rapport om innhaldet i Målførearkivet (upublisert)  
 Hoff, Ingeborg, 1954: Um dei europeiske målatlasa – teknikk og fyremål (stensil)  
 Hoff, Ingeborg, 1959: Norsk dialektatlas. I: NTS 19. S. 595–622.  
 Kolsrud, Sigurd, 1951: Nynorsken i sine målføre. Oslo. Dybwad.  
 Storm, Johan, 1884: Kortere Ordliste med Forklaring af Lydskriften. Kristiania.  
 Tangen, Håvard, 2006: Rapport um fyrebuingi til innlysingsarbeidet med norsk dialektatlas (upublisert)  
 Venås, Kjell, 2005: Sigurd Kolsrud. Novus forlag. Oslo.  
 Wiggen, Geirr, 2006: Dei dialektologiske arbeida til Johan Anton Schulze. I: Telemark Historie. Tidsskrift for Telemark historielag nr 27. S. 106-119.

## ARKIVTILFANG

- Digitalisering av karta til Norsk dialektatlas. Referat frå synfaring til HW C 021 31.8.2006  
 Norsk Dialektatlas – biletdatabase: <http://www.edd.uio.no/perl/search/search.cgi?appid=210&tabid=2338> om Atlas Linguarum Europae sjå <http://ale.lingv.ro/>  
 Målføresynopsisen sjå <http://www.edd.uio.no/synops/work/hovedside.html>

*Jon Holmen*

EDD/ILN, Universitetet i Oslo.  
 jon.holmen@iln.uio.no

*Oddrun Grønvik*

NO2014, Universitetet i Oslo.  
 oddrun.gronvik@iln.uio.no

# SAOLHist – alla upplagor av SAOL i en och samma databas

*Louise Holmer*

The Swedish Academy Glossary (SAOL), has so far been published in 13 editions since 1874. The goal of the SAOLHist project is to put all editions of SAOL, some of which exist only on paper, into a single database.

The number of lemmas of the Glossary has varied substantially over the years. The first edition comprised about 35,000 lemmas, while the latest has 125,000. In every new edition, new words have been added and obsolete ones have been excluded.

With SAOLHist, it will be possible to search for a word's timeline. Since all editions are searched simultaneously, information about the year of the first appearance of a word in the Glossary, or when it was subsequently removed, will be readily available.

*Nyckelord:* SAOL, SAOLHist, historisk lexikografi, svenskans ordförråd

## 1. Inledning

Sedan ett par år tillbaka pågår projektet SAOLHist vid redaktionen för *Svenska Akademiens ordlista* (SAOL) i Göteborg. Projektet går ut på att digitalisera alla de gamla upplagorna av ordlistan och sammanföra dem med de senare upplagorna i en databas som ska göras allmänt tillgänglig på nätet. De gamla upplagorna har skannats in och bearbetats digitalt vid redaktionen. Ansvarig för projektet är Sven-Göran Malmgren. Övriga redaktionsmedarbetare i SAOLHist är Daniel Berg, Monica von Martens och Louise Holmer. Skanning och bearbetning har utförts av Karin Malmgren, Joel Rogström och Oscar Sönnergren.

I den här artikeln presenteras projektet SAOLHist. I början beskrivs och exemplifieras det stora och diversifierade material som ligger till grund för databasen, och senare beskrivs digitaliseringsprocessen med den skanning och korrekturläsning som arbetet har inneburit. Avslutningsvis kommenteras projektet i sin helhet och tänkbara forskningsområden med anknytning till projektet behandlas.

## 2. Bakgrund till projektet SAOLHist

SAOL publicerades första gången 1874 och har sedan dess sammanlagt kommit ut i tretton upplagor, den senaste publicerad 2006. SAOL ger uppgift om uppslagsordens stavning och böjning och kommenterar i förekommande fall bruklighet och eventuellt det fackområde som uppslagsordet hör till. Däremot är det ingen definitionsordbok även om ungefär en femtedel av orden är försedda med definition eller någon annan sorts kommentar. Lemmaantalet är betydligt större än i andra enspråkiga svenska ordböcker, vilket framför allt beror på ett fylligt urval av sammansättningar (se vidare inledningskapitlet i SAOL13). För en fördjupning i bakgrunden till SAOL och ordlistans kännetecken hänvisas till exempelvis Gellerstam (red., 2009) och Johannisson & Mattsson (1974).

De olika upplagornas lemmaantal har varierat relativt mycket under åren, vilket åskådliggörs i tabell 1 nedan. Upplaga 2–5 utgör i stort sett bara nytryckningar av upplaga 1, medan upplaga 6 är den första som skiljer sig från föregångarna, framför allt genom sitt högre lemmaantal. Därför har vi låtit skanna in upplaga 1, 6, 7, 8, 9 och 10, medan upplaga 11–13 redan finns tillgängliga för redaktionen i digitalt format.

Upplaga	År	Huvudredaktör	Lemmaantal
1–5	1874–1883	F.A. Dahlgren	35 000
6	1889	F.A. Dahlgren	40 000
7	1900	Otto Hoppe	71 000
8	1923	Ebbe Tuneld	85 000
9	1950	Pelle Holm	155 000
10	1973	Gösta Mattsson	135 000
11	1986	Martin Gellerstam	115 000
12	1998	Martin Gellerstam	120 000
13	2006	Martin Gellerstam	125 000

Tabell 1. SAOL:s upplagor med årtal, huvudredaktör och lemmaantal (efter Gellerstam 2009: 56).

De olika upplagorna och deras respektive omfång och utformning har styrts såväl av redaktörerna och deras principer som av Svenska Akademiens önskemål. Antalet lemman är något som kommenteras i förordet till varje upplaga, och de flesta redaktörerna beskriver hur de har mönstrat ut föråldrade ord och tagit in nya. SAOL har hela tiden haft som uppgift att spegla det moderna ordförrådet. Det gör att en stor del av arbetet inför varje ny upplaga har bestått i att

lägga in nya ord såväl som att mönstra ut föråldrade ord (se vidare Gellerstam 2009; inledningen till SAOL13 2006; Berg, Holmer & Sköldberg 2010). Ordböcker i allmänhet tenderar att öka i omfång för varje ny upplaga, åtminstone tills något slags tak är nått beträffande sidantal, ordbokens vikt och liknande (Atkins & Rundell 2008: 22 f.). Detta stämmer till stor del för SAOL – ordlistans fysiska tjocklek har ökat med åren och lemmaantalet har ökat sedan den första upplagan. Den första upplagan var den lemmanumerärt minsta med ca 35 000 lemman, medan den nionde upplagan från 1950 är den hitintills mest omfattande med 155 000 lemman. I och med den elfte upplagan hamnade ordlistan i Göteborg med Martin Gellerstam som huvudredaktör, och där produceras den fortfarande. Arbetet med den fjortonde upplagan pågår nu under ledning av Sven-Göran Malmgren (huvudredaktör) och Sture Berg (biträdande redaktör).

### 3. Tidigare digitalisering av SAOL

SAOL har digitaliserats på olika sätt tidigare. Genom projekt Runebergs försorg finns upplaga 6 (1889) och 8 (1923) tillgängliga på internet, inskannade och delvis korrekturlästa (Projekt Runeberg 2007). Upplaga 11 (1986) finns i den allmänt tillgängliga Språkbankens konkordanser (Språkbanken 2011). Upplaga 12 (1998) gavs ut som CD-version och fanns också på internet under tidigt 2000-tal.

Upplaga 13 finns på CD och på nätet. CD-versionen heter SAOL Plus och kom 2007. Den har betydligt mer information och fler sökmöjligheter än pappersversionen. Tillägget ”Plus” betecknar framför allt att den innehåller samtliga *teoretiskt* tänkbara böjningsformer för varje uppslagsord, i motsats till den tryckta ordlistan som bara ger ett fåtal böjningsformer för de flesta ord och ofta inga alls vid sammansatta ord. Användaren kan i SAOL Plus välja att söka bland uppslagsord, i artikeltexten eller bland böjningsformer (se vidare Berg, Holmer & Hult 2008; Martola 2008).

Nätversionen av den trettonde upplagan publicerades i mars 2009 i en faksimilversion (Svenska Akademien 2011). Eftersom ordlistan bara ligger som bilder i pdf-format och inte i en databas går det inte att göra några förfinade sökningar såsom att söka på delar av ett ord eller i artikeltext eller liknande.

Under 2011 påbörjades också, på initiativ av Svenska Akademien, arbetet med att ta fram en app-version av SAOL13 för smarta mobiltelefoner. Appen är i skrivande stund inte färdig, men dess lansering är beräknad till november 2011.



#### 4. Lemmaantalet och omsättningen av lemman mellan upplagorna

Överlappningen av lemman mellan de olika upplagorna är stor, men uppskattningsvis har ca 200 000 lemman vid ett eller flera tillfällen funnits med i ordlistan. Med projektet SAOLHist blir det möjligt att se i vilka upplagor ett visst ord har funnits med och därigenom när det kom in i ordlistan och när det (eventuellt) ströks ur ordlistan. Det blir också möjligt att se vilka ord som bara funnits med i en enda upplaga. Vidare kan man också få en uppfattning om ett ”basordförråd”, alltså lemman som funnits med i alla upplagor, från den första till den sista. Det handlar alltså om sådana ord som hittills inte har mönstrats ut av bl.a. åldersskäl, utan som har behållit sin aktualitet i språket och därmed i ordlistan.

För att få en fingervisning om hur lemmauppsättningen har ändrats mellan upplagorna gjordes en mycket liten undersökning vid redaktionen hösten 2009. Lemman på *Fa*-undersöktes i den första, nionde och trettonde upplagan. Totalt handlade det om 440 lemman, där 80 av dem fanns med i upplaga 1, 320 i upplaga 9 och 305 i upplaga 13. Av dessa var 50 lemman gemensamma för de tre upplagorna, t.ex. *fabel*, *fabrik*, *falsk* och *fagott*. Så många som 150 lemman av de 440 fanns inte längre kvar i SAOL13. Exempel på sådana utmönstrade ord är *fadersinstinkt*, *fajanskopp* och *fallvirke*.

Här bör tilläggas att varje enkelt (osammansatt) ord för med sig sammansättningar och avledningar. Ett exempel är lemmat *fabrikation* vilket saknas i första upplagan. I nionde upplagan finns *fabrikation* med och har ett flertal sammansättningar, t.ex. *fabrikationsfel* och *fabrikationshemlighet*. Varje nytt lemma för alltså också ofta med sig flera andra lemman.

Även om undersökningen som beskrivs ovan är minimal säger den ändå något om den stora omsättningen av lemman som har ägt rum mellan de olika upplagorna. Se tabell 2 för en enkel uppställning över lemmaförekomsten.

#### 5. Utvecklingen av ordlistans artikelstruktur

Till och med 11:e upplagan placerades de flesta sammansättningar in under sina respektive huvudord, vilket gjorde att ord som etymologiskt hörde ihop också återfanns på samma ställe i ordlistan. Med 12:e upplagan infördes i stället en sorteringsordning som innebär att alla lemman placeras i bokstavsordning oavsett etymologi (Gellerstam 2009: 71–73, SAOL12: VI f.). Dessutom infördes den s.k. lemma-lexemmodellen, vilken innebär att tidigare olika lemman med samma böjning sorteras in under samma huvudord och blir olika lexem. Tidi-

gare hade homografer dessutom ofta sorterats in under samma lemma även om böjningen skilde dem åt (t.ex. *ljus* a. och *ljus* s., se SAOL12: X). I och med 12:e upplagan gjordes också en noggrannare genomgång av varje lemmas ordklass-tillhörighet (SAOL12: XIX).

Lemma	SAOL1	SAOL9	SAOL13
falsifierbar			X
falsifiering			X
falsifikat		X	X
falsifikation			X
falsk	X	X	X
falskant		X	
falskdeklarant		X	X
falskdeklaration		X	X
falskeligen	X	X	X
falskhet	X	X	X
falsklarm			X
falsklegg			X
falskmyntare		X	X
falskmyntarlīga		X	

Tabell 2. Exempel på lemmaförekomst vid *Fa-* i tre olika upplagor av SAOL.

Många av upplagorna uppvisar stora skillnader i sättet att presentera artiklarna. I de tidiga upplagorna anges exempelvis genus för substantiven. Ett ord som *björk* ser ut enligt följande i första respektive sjunde upplagan:

SAOL1 Björk (pl. *-ar*) s. f.  
SAOL7 björk (*-en, -ar*) rf.

I första upplagan ges alltså först pluraländelsen och därefter ordklassmarkeringen ”s. f.” vilket står för substantiv, femininum. I sjunde upplagan ges förutom pluraländelsen även ändelsen i bestämd form. Ordklassangivelsen är också modifierad till ”rf.”, vilket står för reale, femininum (se vidare Malmgren 2002: 10–12). Inledningen till sjunde upplagan bjuder på följande förklaring:

Utanför de levande varelsernas område, t. e. i fråga om *mur, vägg*, nyttjar man i de bildades språk numera icke ofta *han* och *hon*; i stället säges *den*. Med andra ord: maskulinum och femininum hafva här blifvit ersatta med ett nytt genus, som på sista tiden fått namnet *realgenus*, och som nu i Ordlistan betecknas med *r*. (SAOL7: VIII)

Hur verben presenteras skiljer sig åt mellan de olika upplagorna. Faktum är att verb är den ordklass vars presentationsstruktur uppvisar störst och flest skillna-

der. För att åskådliggöra några av olikheterna mellan upplagorna har jag valt ett deponentiellt verb från första konjugationen, *hoppas*. Överst i uppställningen visas den senaste upplagan och sedan följer de andra i omvänd kronologisk ordning.

- 13, 12 **hoppas** v. *hoppades* till <sup>1</sup>*hopp1*
- 11 **hoppas** -ades v. dep.
- 10 **hopp|as** -ades v. dep.
- 9 **hopp|as** -ades itr. dep.
- 8 **hopp|as** (-ades) i. o. t. dep.
- 7 **hopp|as** (-ades) i. dep.
- 6 **Hoppas** (-ades) dep.
- 1 **Hoppas** (-as, -ades, -ats), v. dep.

Ett antal detaljer syns tydligt i uppställningen ovan. För det första är det lemmats utseende och presentation: i upplaga 1–6 användes versal initialt, något som togs bort i 7:an. I upplaga 7 infördes däremot lodstreck efter ordstammen, en detalj som fick vara kvar i flera upplagor. Från upplaga 11 och framåt finns inget lodstreck i den här typen av enkelt (icke sammansatt) uppslagsord.

I den tolfte och trettonde upplagan har varje verb, även om det är sammansatt, ordklassbeteckningen "v." Först ges ordklassuppgift i förkortad form och därefter preteritumformen *hoppades* fullt utskrivet. I alla de tidigare upplagorna ges bara böjningsändelsen i preteritum, eller som i första upplagans fall, böjningsändelserna i presens, preteritum och supinum (-as, -ades, -ats).

Den första och de sista fyra upplagorna har ordklassbeteckningen "v." för verb medan upplagorna däremellan har beteckningen "dep." för deponens och dessutom en transitivetsbeteckning: "i." för intransitivt och "t." för transitivt verb. Där förutsätts alltså läsaren veta vad förkortningarna står för och dessutom vad de innebär. I dagens ordlista är det oklart om man skulle nå förståelse hos användarna om man skrev "i. o. t. dep." i stället för "v.", men äldre tiders användare hade troligtvis lite bättre kunskap om grammatisk terminologi.

Definition saknas i alla upplagor. Upplaga 12 och 13 har däremot en hänvisning till substantivet *hopp* och dess första lexem ("1. förhoppning"), vilket indirekt antyder betydelsen av verbet *hoppas*.

## 6. Digitaliseringsarbetet

Skanningen och bearbetningen av det inskannade materialet har gjorts vid ordlisteredaktionen. Lite förenklat kan processen sägas ha gått till enligt följande: originalen har skannats in, tolkats av programmet Omnipage och därefter kontrollerats manuellt. Sedan har filerna exporterats till xml-format. Därefter går det

att se materialet i en databas eller visa dem på en webbsida. I ett inledande skede har endast lemmat och ordklassuppgifterna extraherats (i den mån ordklass finns med, vilket inte alltid är fallet i de äldre upplagorna).

Originalsidorna bjuder på en typografisk mångfald. Typsnitten varierar mellan upplagorna, bruket av kursiv och rak stil växlar liksom teckengraden inne i artikeltexten. Detta har medfört vissa svårigheter i efterarbetet. Tolkningen av texten i Omnipage har ibland haft svårigheter med att identifiera just det som det är programmerat att identifiera, nämligen ord med fetstil (lemmat och dess sammansättningar och avledningar) och ordklassbeteckning (som, vilket visats tidigare, varierar en hel del mellan upplagorna). Om då programmet ”missat” att hämta upp ett fetstilt ord som dessutom har många sammansättningar blir resultatet att efterleden kopplas ihop med helt fel förled, alternativt saknas helt. Se figur 1 för exempel på en sida ur upplaga 7 som skannats in, tolkats av programmet och korrekturlästs.

F ¶	
<p><b>f</b> (<i>fet</i>; pl. =, best. <i>f-en</i>) n. <i>F.dur. f-dursskala. f-ljud. ¶</i></p> <p><b>fabel</b> (-n; <i>fabler</i>) r. - <b>aktig</b> (~t; ~are) a. -</p> <p><b>bok. - diktare. - hjälte. - samling. -</b></p> <p><b>fabricer</b>   a (-ade) t. - <b>ing</b> rf. <b>fabrik</b> (-en; -er) m. <b>fabriks</b>   <b>alster. - anläggning. -</b></p> <p><b>arbetare. - arbete. - arbetska. - byggnad. - distrikt. - drift. - flicka. - idkare. - inspek-tion. - inspektör. - märke. - mässig</b> (-t; ~are) a. - <b>ort. - pris. - rörelse. - skorsten. - stad. - vara. -</b></p> <p><b>fabrik</b>   <b>ant</b> (-en; -er) m. - <b>at</b> (-et; pl. = 1. ~er) n. - <b>ation</b> (-en; ~er) r. -</p> <p><b>ör</b> (-en; ~er) m. ¶</p> <p><b>fabulös</b> (-t; ~are) a. ¶</p> <p><b>facit</b> n. - <b>bok. - tabell. ¶</b></p> <p><b>fack</b> (-et; pl. = n. - <b>bildning. - förening. - in-sikt(er). - kunskap. - lärare. - man. - mäs-sig</b> (~t; ~are) a. - <b>skola. - skrift. - studium. - term. - tidskrift. - verk.</b> <i>Korsvirke. ¶</i></p> <p><b>fackl</b>   a (-an; -or) rf. <b>fackell</b> <b>belysning.</b></p>	<p><b>fadder</b> (-n; <i>faddrar</i>) mf. - <b>gåfva. - kyss. - fadderskap</b> (-et; pl. =) n. ¶</p> <p><b>faddhet</b> (-en; -er) rf. ¶</p> <p><b>fader, smdr. far</b> (<i>fadem, äfv. ffordren; fäder, best. fäderna, äfv. poet. fädren</i>) m. - <b>lös. - mord. - mördare. - mörderska.</b></p> <p><b>faders</b>   <b>glädje. - hem</b> <i>relig.</i> - <b>hjärta. - hus</b> <i>relig.</i> - <b>kärlek. - namn. - sinne. - väld.</b> - <b>far</b>   <b>bro</b>   <b>de</b>   r. - <b>fa</b>   <b>de</b>   r. - <b>föräldrar. - mo</b>   <b>de</b>   r. <b>farsgubbe</b> (n) <i>skämts. o. hvard. -</i> <b>fader</b>   <b>lig</b> (-t; ~are) a. - <b>ligen</b> (o. - <b>ligt</b>) adv. - <b>skap</b> (-et) n. -</p> <p><b>fadervår</b> (pl. =) n. ¶</p> <p><b>fager</b> (-t; <i>fagra; fagrare</i>) a. - <b>hyllt</b> (n. ¶ =) p. a. - <b>kindad</b> (-t) p. a. <i>poet.</i></p> <p><b>faggorna</b> best. pl. <i>Hafva-ngt i f.</i> ¶</p> <p><b>fagott</b> (-en; -er) r. - <b>blåsare</b> (-n; pl. =) m. - <b>stämma. fagottist</b> (-en; -er) m. -</p> <p><b>fajans</b> (-en; -er) r. - <b>fabrik. - fat. - kärl. - fakir</b> (-en; -er) m. ¶</p> <p><b>faksimile</b> (-t; pl. = 1. -n) n. - <b>maner.</b></p>

Figur 1: Sida ur SAOL7 efter skanning, tolkning och kontrolläsning.

Markeringarna i figuren visar olika ställen där programmet har haft svårt att tolka tecknen, exempelvis vid lodstrecken i **fabriks**|**alster** och **faders**|**glädje** och vid tecknet ~ på ett flertal ställen, t.ex. efter **fabelaktig** och **fabriksmässig**.

I den färdiga databasen kommer man att kunna söka med modern stavning och med originalstavning. Det innebär att om man söker på lemmat *avhandla* ska man få resultatet att lemmat funnits med i alla upplagor, trots den tidigare stavningen *afhandla*. Vid lemmen med variantstavningar som t.ex. *schyst* och *juste* ska man kunna söka på båda stavningarna. I resultatlistan visas sedan båda stavningsvarianterna med respektive upplaga (*juste* kom in i den 8:e upplagan och stavningen *schyst* i den 11:e).

## 7. Vad betyder ordlistans variation för SAOLHist?

I artikeln har jag gett ett flertal exempel på hur varierad ordlistan har varit i sitt sätt att presentera artiklarna i de olika upplagorna. Vad får det för inverkan på projektet SAOLHist? För det första är det ett stort material att arbeta med för alla inblandade. Det gör att det praktiska arbetet såsom inskanning och efterarbete har varit relativt omfattande. Att sedan läsa slutkorrektur blir också ett mycket stort arbete även om delar av det kan automatiseras (t.ex. genom att de olika upplagorna körs mot varandra). De olika sätten att markera ordklass påverkar också. Många lemmen i de tidigare upplagorna saknar helt ordklassangivelse och många av dem som har ordklassmarkering har olika sätt att markera det på. Det leder till en sorts normalisering beträffande ett flertal av lemmanas ordklassuppgifter. Lemman som tidigare har haft ordklassmarkeringen ”i. o. t. dep.” för intransitivt och transitivt deponensverb får i stället bara ett ”v.” för verb.

## 8. Sammanfattning och framtida användningsområden för SAOLHist

I den här artikeln har jag beskrivit de olika upplagorna av SAOL och hur dessa har insortrats i projektet SAOLHist. I och med SAOLHist kommer i princip samtliga upplagor att finnas sökbara i en och samma databas. Även om SAOL har givits ut i elektronisk form tidigare blir det här det första projekt som samlar alla upplagor på ett ställe. Med SAOLHist kommer man att kunna söka fram ord som funnits med i alla upplagor såväl som ord som bara funnits med i en upplaga innan de har blivit utmönstrade. Totalt handlar det om ca 200 000 olika lemmen, däribland ett basordförråd som utgörs av ord som funnits med i alla upplagor.

Förhoppningen är att databasen ska kunna ligga till grund för både språk-historisk, lexikografisk och lexikologisk forskning och att även allmänheten kommer att kunna ha glädje av den. Kanske blir också en del ordlistanvändare

mindre upprörda när ord stryks ur den aktuella upplagan av ordlistan, eftersom SAOLHist kommer att utgöra en permanent viloplats för utmönstrade ord.

## KÄLLOR OCH LITTERATUR

- Atkins, B.T. Sue & Michael Rundell, 2008: *The Oxford Guide to Practical Lexicography*. Oxford.
- Berg, Sture, Louise Holmer & Ann-Kristin Hult, 2008: SAOL Plus – a New Swedish Electronic Dictionary. I: Bernal, E. & J. DeCesaris (eds.), *Proceedings of the XIII Euralex International Congress (Barcelona 15–19 July 2008)*, S. 291–296. (Cd-rom.)
- Berg, Sture, Louise Holmer & Emma Sköldberg, 2010: Time to say goodbye? On the exclusion of solid compounds from the Swedish Academy Glossary. I: Dykstra, Anne & Tanneke Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*, S. 567–576
- Gellerstam, Martin, 2009: SAOL i många upplagor. I: Gellerstam, Martin (red.), S. 53–83.
- Gellerstam, Martin (red.), 2009: SAOL och tidens flykt. Några nedslag i ordlistans historia. Stockholm.
- Johannisson, Ture & Gösta Mattsson, 1974: Svenska Akademiens ordlista under 100 år. Skrifter utgivna av Svenska språknämnden 55. Stockholm.
- Malmgren, Sven-Göran, 2002: Normering i Svenska Akademiens ordlista 1874–1950: principer och resultat. I: *LexicoNordica* 9. S. 5–20.
- Martola, Nina, 2008: SAOL Plus – SAOL på cd-rom. I: *LexicoNordica* 15. S. 261–278.
- Projekt Runeberg 2007: Svenska Akademiens Ordlista. <http://runeberg.org/saol/> (september 2011)
- SAOL1–13 = Svenska Akademiens ordlista 1–13 uppl., 1874–2006.
- SAOL Plus = Svenska Akademiens ordlista över svenska språket. 13 uppl. Cd-rom. 2007: Stockholm.
- Språkbanken 2011: Konkordanser – SAOL 11. <http://spraakbanken.gu.se/konk/> (september 2011)
- Svenska Akademiens 2011: Svenska Akademiens ordlista. <http://www.svenskaakademien.se/>
- [svenska\\_sprakets/svenska\\_akademiens\\_ordlista/saol\\_pa\\_natet/ordlista](http://svenska_sprakets/svenska_akademiens_ordlista/saol_pa_natet/ordlista) (september 2011)

*Louise Holmer*

Lexikaliska institutet, Inst. för svenska språket, Göteborgs universitet  
 louise.holmer@svenska.gu.se

# Data og repræsentativitet i ordbogsarbejdet

*Henrik Hovmark*

Nowadays, lexicography is closely associated with information technology and language technology tools. However, until fairly recently lexicography was based on traditional printing and collections of paper slips. Taking a description of the working processes in lexicographic work based on collections of paper slips, I discuss the relations between empirical data, work flow and final product in lexicography in general. I focus on the question of representativeness: When is a collection of paper slips representative? How and to what extent is it possible to make generalizations based on data which is highly qualitative? I point to the fact that any kind of data, text corpora as well as collections of paper slips, is the outcome of a selection and should be scrutinized critically by lexicographers.

Jeg arbejder til daglig på Ømålsordbogen (ØMO), en ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer i perioden 1750-1945. Indsamlingen til ØMO startede i 1909 (jf. Gudiksen & Hovmark 2009), og kildegrundlaget er en seddelsamling på ca. 3 mio. sedler. Vi benytter også et mindre korpus af udskrevne båndoptagelser på ca. 1,3 mio. løbende ord som et yderst vigtigt supplement (jf. Gudiksen & Hovmark 2008), men seddelsamlingen er det primære udgangspunkt.<sup>1</sup>

Når jeg og mine kolleger i dag fortæller om hvordan vi henter sedler frem fra seddelkasserne, forsøger at tyde karakteristiske håndskrifter, og sorterer sedlerne i bunker på skrivebordet, er reaktionen ofte at det må være et spændende og fascinerende arbejde. Men undertiden kan man også notere en vis undren: Hvorfor arbejder man ikke med et tekstkorpus? Og er håndskrevne sedler, i visse tilfælde kun en håndfuld eller færre, virkelig tilstrækkelige til at udarbejde en hel ordbogsartikel? Reaktionen er interessant og vil danne udgangspunkt for den følgende artikel. Først og fremmest gør reaktionen opmærksom på et vigtigt aspekt ved ordbogsarbejdet, nemlig omgangen med kilder og det empiriske

---

<sup>1</sup> Jeg har tidligere arbejdet som redaktør ved Den Danske Ordbog (DDO), der jo bygger på et tekstkorpus på ca. 40 mio. ord, jf. DDO, bind 1: 54.

grundlag for de ordbøger der produceres. Men den vidner også om hvor gennemgribende en forandring hele ordbogsfeltet har gennemgået siden it-teknologien for alvor holdt sit indtog i 1980'erne og begyndte at ændre datagrundlag og arbejdsmetoder væsentligt. Til trods for at seddelsamlinger – og den dataindsamlingsmetode der ligger bag, nemlig excerpering – indtil for få år siden var enerådende inden for leksikografien og stadig danner grundlag for en række ordbogsprojekter, er der mange leksikografer der i dag ikke har kendskab til seddelsamlingernes virkelighed: deres tilblivelse og indhold, og arbejdet med dem. Måske kender man dem kun fra Simon Winchester bog om tilblivelsen af Oxford English Dictionary: *The Meaning of Everything* (Winchester 2003).

Jeg vil i denne artikel åbne døren ind til seddelsamlingerne og støve nogle af sedlerne af. Formålet er imidlertid ikke primært at fortælle en historie om gamle dage. Det er derimod min hensigt at bruge eksemplerne fra seddelsamlingerne til at kaste et blik på et mere generelt leksikografisk spørgsmål, nemlig spørgsmålet om data og repræsentativitet i ordbogsarbejdet, et spørgsmål som ordbogsredaktører må forholde sig til uanset om de data der arbejdes med foreligger i form af en seddelsamling eller et tekstkorpus – eller evt. begge dele, sådan som det er tilfældet for fx ØMO. Jeg vil hente mine eksempler fra ØMO's seddelsamling, og på den måde kan artiklen forhåbentlig samtidig tjene til at give et indtryk af seddelsamlinger og arbejdet med dem.<sup>2</sup>

## 1. Seddelsamlinger og korpus som data

Der er ikke nogen tvivl om at seddelsamlinger i en vis forstand er forældede, og at fremkomsten af tekstkorpora er et uhyre stort fremskridt i ordbogsarbejdet. Man fornemmer en næsten beruset begejstring for de nye muligheder i it-teknologien i den første udgave af Svenséns *Handbok i lexikografi* fra 1987:

... datorn [har] revolutionerat själva det lexikografiska arbetet: den låter oss göra saker med materialet som vi tidigare inte ens kunnat föreställa oss som möjliga, och den tvingar oss att på ett mycket välgörande sätt ifrågasätta och revidera mycket av det vi ansett som en gång för alla givet. (Svensén 1987: 243)

Begejstringen bliver ikke mindre i den reviderede og udvidede udgave fra 2004, hvor Svensén skriver følgende om tekstkorpora og brugen af dem:

<sup>2</sup> Denne artikel bygger på mit foredrag ved leksikografikonferencen i Lund, men af pladsmæssige hensyn vil fokus ligge på den del af foredraget der handlede om data og repræsentativitet. Diskussionen af termen korpus, herunder om en seddelsamling kan kaldes et korpus, berøres kun indirekte, og eksempelmaterialet er reduceret.



Det stora utbudet av texter i elektronisk form och möjligheten att lagra och analysera stora textmängder på ett ekonomiskt sätt har revolutionerat inte bara lingvistikens utan också lexikografin. Stora elektroniska korpusar har givit lexikografer möjlighet att fatta beslut baserade på språket i verkligt bruk snarare än på intuition och användning av sekundärkällor. (Svensén 2004: 58)

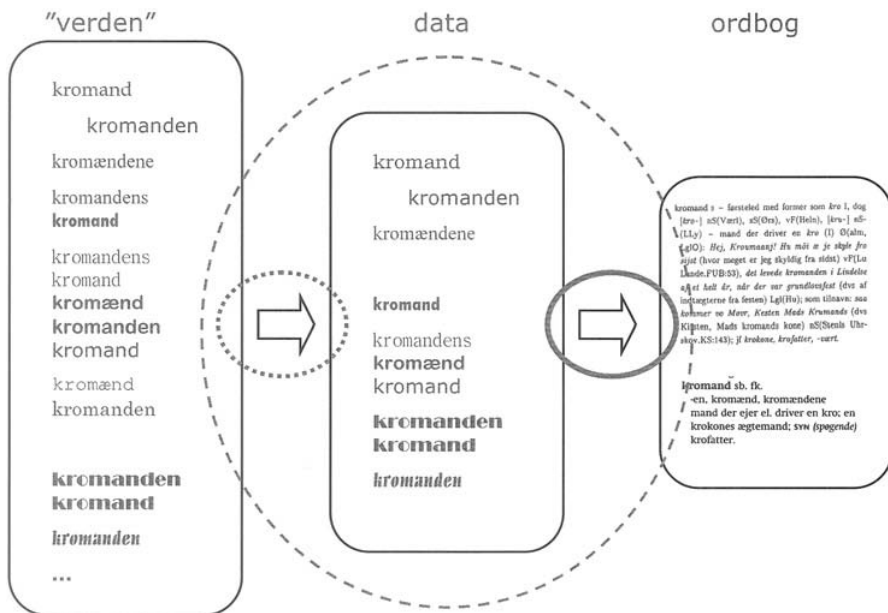
For ubefæstede sjæle kan Svenséns beskrivelse forstås på den måde at seddelsamlinger er baseret på intuition eller sekundærkilder. Det er næppe meningen – Svensén er selv kritisk over for en unuanceret tilgang til (brugen af) tekstkorpora i både positiv og negativ retning, jf. afsnit 3.1.3.8 (Korpusseptiker og korpusfundamentalister) i Svenséns bog, og følgende opsummering:

Att någonting *inte* existerar i språket kan man aldrig få bekräftelse på, inte ens om man baserar sin undersökning på de mest omfattande och mångsidigt sammansatta korpusar och har de mest sofistikerade analysverktyg till sitt förfogande. Och det är fortfarande människor som skriver ordböcker, inte korpusarna. (Svensén 2004: 73; Svenséns fremhævelse)

Men der er ikke nogen tvivl om at det er fristende og nærliggende at anse seddelsamlinger for at være et ringe(re) kildegrundlag end tekstkorpora, bl.a. fordi korpora rummer mulighed for kvantitative undersøgelser og giver et bedre indblik i fx syntaktiske forhold og mere upåfaldende ord og betydninger der ikke altid er blevet indstreget når man har excerperet til en seddelsamling (jf. Svensén 2004: 57, Gudiksen & Hovmark 2008: 176). Jeg vil imidlertid gerne argumentere for det frugtbare i ikke at fokusere for snævert på datagrundlagets form (seddelsamling eller tekstkorpus). Det er i alle tilfælde vigtigt at have en kritisk tilgang til det specifikke datagrundlag man arbejder med, dets tilblivelse og sammensætning, og brugen af det i det daglige arbejde, set i forhold til den generelle problematik: data og repræsentativitet.

Man kan som et banalt eksempel tage ordet *kromand* der optræder i både ØMO og DDO (jf. Figur 1). ØMO's beskrivelse bygger på en seddelsamling, og DDO's på et tekstkorpus. Men der er i begge tilfælde tale om at produktet, dvs. ordbogsartiklerne, bygger på et udsnit af virkeligheden, ikke den fulde virkelighed. Ordet *kromand* findes i en mængde former og tilfælde i virkeligheden eller "verden", og må antages at følge bestemte, genkendelige mønstre. Det som ordbogsredaktørerne har bygget på, er imidlertid "kun" en delmængde af en vis beskaffenhed af denne virkelighed – en delmængde som jeg vil kalde data. Data kan være en seddelsamling eller et tekstkorpus – eller en enkelt ordbogsforfatters intuitive viden om og fornemmelse for sproget. Samlingen af data kan være mere eller mindre statisk eller dynamisk, lukket eller åben. I løbet af redigeringsprocessen bliver data omsat til et bestemt produkt i en karakteristisk, ofte

stærkt konventionaliseret form, nemlig en ordbogsartikel (fx *kromand*). I såvel udvælgelsen af data som i behandlingen af data i forbindelse med udarbejdelsen af ordbogsartiklen, skal der foretages vurderinger og træffes valg. Redaktøren må være opmærksom på 1) om data er repræsentative for den store virkelighed, og 2) opbygge nogle kritiske arbejdsrutiner i omgangen med data og omsætningen af det til en ordbogsartikel. Det vil aldrig være hele den sproglige verden man indfanger i et korpus, og der træffes nogle beslutninger, menneskelige beslutninger som Svensén siger (2004: 73; jf. citatet ovenfor), både i udarbejdelsen af korpus og undervejs fra korpus til ordbogsartikel. Det er disse to processer jeg vil se nærmere på i det følgende, primært med udgangspunkt i det helt konkrete arbejde med en seddelsamling, dels i forbindelse med indsamling af data (excerpering), dels i forbindelse med vurderingen af dataenes repræsentativitet.



Figur 1: Ordet *kromand*, fra "verden" over data til ordbogsartikel (i ØMO og DDO).

## 2. Udvalgelse af data i en seddelsamling: excerpering

Hvad er det der står på sedlerne i en seddelsamling? Der står naturligvis bestemte uddrag af virkeligheden eller "verden" (jf. Figur 1), men hvilke? Og hvordan er de blevet indsamlet eller udvalgt? Den særlige udvælgelsesproces som ligger til grund for seddelsamlinger, kaldes traditionelt for excerpering. Ter-

merne excerpt og excerpering dækker over lidt forskellige ting afhængigt af hvilken historisk kontekst og tekst der er tale om, men i ordbogssammenhæng er følgende definition et godt udgangspunkt: "A citation is a short extract from a text which provides evidence for a word, phrase, usage, or meaning in authentic use" (Atkins & Rundell 2008: 48). Definitionen peger på at et excerpt til en ordbog skal give et autentisk belæg, ikke kun på et ord, men på en eller anden dimension af sproget. Definitionen åbner op for en lidt bredere opfattelse af hvad et excerpt rent praktisk er, som vil vise sig nyttig når vi dykker ned i ØMO's seddelkasser. Der står nemlig mange forskellige typer oplysninger på sedlerne i ØMO's seddelsamling (og tilsvarende samlinger).

Figur 2 giver et eksempel på et klassisk excerpt. Et antal ord eller ordforbindelser i en trykt tekst er indstreget, og der er efterfølgende udskrevet en seddel på hvert belæg. ØMO er en talesprogsordbog, så det er naturligt nok replikkerne fra den pågældende tekst (en roman af Morten Korch) der har haft redaktørens interesse.

82

»Stabbe er jo ette saaden aa bryde i, Søren, og Jeppe magter jo ikke en hel Del nu; men sid nu hen til Bordet, saa kommer jeg med Kaffe,« sagde Hanne.

»Ja, Tak, Hanne, det var nu ellers ette Meningen, vi skulde ha Kaffe.«

»Jov, vist saa, vi skal da ha Kaffe ilejn, Jeppe og jeg. Sid nu bare ner begge to.«

»Ja skal det være, saa lad os sætte vos, David.«

»Det maa vi vel heller, Hanne vil jo ha sin Villed frem.«

»Hi hi, ja det passer. Du kender Mor, David.«

»Det gør jeg, og jeg kender min egen, det er jo for saavidt let selleje saame alle Steder.«

»Ja, I har let ved at ralliere over Fjuntimmerne; men I ka jo itte undvære vos ilejn.«

»Poj! — Nej, det er et sandt Ord, Hanne.«

»Ja, nu kommer jeg leje med Kaen. Væsko spis nu noget til.«

Hanne skænkede Kaffe og hød Kage.

»Tak! Tak!«

»Hm! — Poj! — Det var da ellers en dejlig Regn, vi fik først i Ugen. — Havri derude ved Skoven, jeg skal løve for, han har rettet sig, og nu fjoger det jo i Dav igen,« begyndte Sognefogden.

Fjog  
af nu fjoger det jo i  
Dav igen  
Gottlieb Afskr.af  
Korch: G-d.t. 82

Figur 2: Excerpering af ordet *fjoge* 'småregne' fra en roman af Morten Korch.

Men eksemplerne fra ØMO viser også at sedlerne i denne samling ikke kun rummer ukommenterede sprogbrugseksempler (svarende til en slags konkordanslinjer ved søgninger i et tekstkorpus), men også kan rumme input af mere kommentaragtigt tilsnit til andre dele af en ordbogsartikel, fx definitionen eller de særlige encyklopædiske, kulturhistoriske afsnit som er et særkende for ØMO (jf. "evidence .. for use, or meaning", Atkins & Rundell 2008: 48). Det gælder fx et belæg på ordet *guldgaldon*, udskrevet fra en bog fra 1924 om fester og højti-

der. Den omgivende tekst er også medtaget: ”de ældres Huer var med Guldgalon og sorte Baand”, men ikke som eksempel på en sproglig kontekst for ordet *gulddgalon*. Den omgivende tekst er taget med fordi den bidrager til betydningsbeskrivelsen af ordet. På sedlen er der en henvisning til kilden, hvor redaktøren kan læse mere om især den kontekstuelle brug af gulddgaloner.

Der kan altså optræde mange typer sproglige data på en ordbogsseddel, nøjagtig lige så mange som man har ønsket at beskrive i ordbogen, og som man derfor har haft brug for at nedfælde oplysninger om på sedlerne. I talesprogsordbogen ØMO’s samlinger indgår der således mange sedler hvor det alene er udtalen af et ord der anføres. Én enkelt seddel kan naturligvis også rumme oplysninger om flere sproglige elementer på samme tid (fx udtale, syntaks og betydning). Det gælder fx følgende seddelbelæg på ordet *fjoge* (Figur 3), hvor ordet er lydskrevet i citaterne, og hvor citaterne desuden er medtaget i den redigerede artikel.

Kert. 357  
 R.G. K.H. 1766  
 del begynde a fjog. hi med for  
 småregne  
 KOMPL. LISTE 466  
 fjoge, o. Boraringe 908  
 Kasmas T PA 87  
 (Sp. : Typ = Støvregne)  
 Lan: 'fjas. ræjn, del fjog. a ilid  
 KOMPL. LISTE 385

fjoge III v – [fjaːw; sjæld fjaw] øF(alm, jf FØF:-273), sF(VSk), m acc 1 T(Bjb), [fjōːw] vF(Br, Søs, Ejlbj), [fjāːw] vF(Lu). Bøjn: ede-bøjn. – småregne, støvregne, fjog(e)regne (jf fnuge 2) øF(alm), vF(Lu, Br), sF, T(alm), Æ(Bregn): det begyndte [a fjaw] lige ret før øF(Kert), [de fjāːw ilid] øF(Revn), [de æ någ bliw räjːnvæːs / jaː de fjāːw ən bidə] vF(Lu); jf fjog II || hertil fjog(e)-regn, -regne, -vejr.

Figur 3: Talesprogseksempler på ordet *fjoge* 'småregne', medtaget som citater i ØMO.

Undertiden kan de sprogrøver man møder på sedlerne have et kunstigt præg eller være resultatet af en forholdsvis stram redaktionel styring (jf. Hovmark 2004), men begge citater er i dette tilfælde ganske spontane og autentiske – glimt af ordet *fjoge* i typisk løbende tale.

Når man har et tekstkorpus, har man hentet større, sammenhængende stykker af den sproglige virkelighed ind i ordbogsværkstedet, hvor alle deres dele konstant kan hentes frem og studeres i deres kontekst. Når man har en seddel-samling, har man ikke på samme måde de sammenhængende stykker tekst og

konteksten til stede – man har udvalgt det man har fundet interessant og relevant. Dette forhold bliver særligt påtrængende når der er tale om en talesprogsordbog hvor data i høj grad er indsamlet før båndoptagerens tid, for her står den samlede tekst ikke længere til rådighed når først den er blevet ytret. Det betyder til gengæld at man er nødt til at gå direkte til primærkilden, ”verden”, hvis der er noget man vil slå op. Man har ikke ”fastfrosset” større dele af konteksten, men man har mulighed for at gå tilbage til de informanter man har kontakt til, og som på en måde udgør ordbogens levende korpus. Det er netop hvad man ofte gjorde på ØMO (og stadig gør i begrænset omfang): Som arbejdet med ordbogen og samlingerne skred frem, stødte man uundgåeligt på huller i materialet, ubelyste spørgsmål osv., hvorefter man henvendte sig til relevante informanter for at komplettere og supplere datagrundlaget for ordbogen (jf. Bévort 1965, Gudiksen & Hovmark 2009). De to seddelbelæg på ordet *ffoge* ovenfor (Figur 3) stammer netop fra sådanne kompletteringsrunder (jf. markeringerne ”KOMPL. KLADDE 4686” og ”KOMPL. LISTE D 385”), udført af redaktørerne ”K.M.” (Kristen Møller) i 1966 og ”PA” (Poul Andersen, den mangeårige leder af ØMO) i 1937. I ØMO’s tilfælde har redaktørerne været vant til at skulle bevæge sig ud over det forhåndenværende seddelkorpus når man ønskede at undersøge noget, og man er som redaktør på ordbogen stadig nødt til at være opmærksom på sammenhængen mellem ”verden” og det aktuelle datagrundlag.

### 3. Arbejdet med data i en seddelsamling

Når man udarbejder artikler på basis af en seddelsamling, er der ofte kun ganske få belæg, undertiden endda kun et enkelt, til rådighed for redaktøren. Det er dog ikke pr. definition ensbetydende med at der altid er flere belæg på et ord i et tekstkorpus, det kommer helt naturligt an på samlingen og korpusset. Fx findes der 39 sedler på *kromand* i ØMO’s samlinger, men faktisk kun 24 belæg på samme ord i Korpus90 der rummer størsteparten af det oprindelige korpus til DDO (jf. ordnet.dk/korpusdk/). Man må også være opmærksom på at flere belæg kan stamme fra samme kilde – det gælder både seddelsamling og tekstkorpus (der er fx kun 13 kilder til ordet *kromand* i Korpus90). Men det er rigtigt at datagrundlaget for artikler og betydninger i ØMO og tilsvarende ordbøger kan forekomme temmelig spinkelt. Men netop derfor kan seddelsamlinger og arbejdet med dem være et udmærket udgangspunkt for en diskussion af data og repræsentativitet, herunder de konkrete kildekritiske overvejelser det kan være relevant at gøre sig som ordbogsredaktør når man arbejder med sine data.

Lad mig som eksempel tage skældsordet *krapillik*. Dette belæg er blevet underkastet den samme kildekritiske vurdering som alle andre belæg i ØMO.

Ender vurderingen med et ja, vil belægget komme med i ordbogen – det gælder også hvis sedlen står alene og altså er det eneste belæg på et ord. Det er med andre ord ikke antallet af belæg, men kvaliteten af et belæg der er afgørende. Og kvaliteten vurderes på basis af velkendte kildekritiske faktorer: Hvem er meddeleren? Hvordan har indsamlingssituationen været? Og hvem er optegneren?

En vurdering af meddeleren stod i centrum helt fra projektets start, dvs. under indsamlingsarbejdet. Knappe resurser gjorde det nødvendigt at finde de bedste meddelere, og man førte kartotek over meddelerne og deres kvaliteter – eller mangel på samme (jf. Gudiksen & Hovmark 2009, Køster 2009). Meddelerne blev vurderet på to skalaer: om de var gode (1. eller 2. classes) sprogligt henholdsvis sagligt (dvs. hvad angår den kulturhistoriske viden om den ældre bondekultur). En vurdering af meddeleren er også i dag en naturlig del af arbejdet. Et belæg fra en dårlig meddeler kommer ikke med. Hvad angår meddeleren til ordet *krapillik*, Lars Hansen, beskrives han som en ”meget interesseret Meddeler” og ”I-Klasses” hvad angår det saglige. Der nævnes ikke noget om det sproglige, men der står at meddeleren har været bofast i det pågældende sogn hele livet bortset fra de senere år, hvilket er et plus i karakterbogen i ØMO-sammenhæng, for det kan tyde på et stabilt og forholdsvis upåvirket dialektalt sprog. Karakteristikken er ikke overvældende, men der er heller ikke noget der taler imod at acceptere belægget på *krapillik*. Derudover tilhører *krapillik* en bestemt orddannelsestype, dannet med suffikset *-ik(e)* der bruges i afledninger med emotiv betydning, i dette tilfælde en række af synonymer brugt som nedsettende betegnelse for en gammel udslidt hest. Ordet *krapillik* står derfor ikke helt alene alligevel.

Ud over meddeleren vurderes også indsamlingssituationen og optegneren. Indsamlingssituationen kan have præget de sproglige data i mere eller mindre autentisk retning (jf. Atkins & Rundell 2008: 48), dvs. gjort data, fx sprogeksempler, mindre autentiske eller ”naturlige” (jf. det såkaldte ”observer’s paradox” inden for fx talesprogsforskningen og etnografiske videnskaber, dvs. at forskerens tilstedeværelse påvirker informanten og gør dennes optræden og sprogbrug mindre spontan og naturlig; jf. fx Hovmark 2004). Og optegnavner og -holdninger kan også spille en rolle, fx i forbindelse med en udtaleangivelse (udtaleangivelserne i ØMO er normaliserede efter strukturalistiske principper, og materialet kan tolkes lidt forskelligt; jf. fx Andersen 1958: 13-16, Gudiksen 2001: 43f.).

Den kildekritiske vurdering henholder sig til en række objektive faktorer, men i sidste instans kommer man ikke uden om et vist skøn: Skal man tro på hvad der står på en seddel, eller skal man ikke? Svaret vil til en vis grad afhænge af den enkelte redaktør og dennes erfaringer og vurderinger. Det kan fx være en vurdering af om et ord er centralt eller perifert i forhold til begreber som folke-

ligt sprog vs. rigsmålspræget sprog. Under alle omstændigheder kræves der et indgående og detaljeret kendskab til stoffet og kilderne når man arbejder på ØMO, dvs. ikke kun sproglige forhold, men også mere interne forhold omkring indsamlingen, meddelelserne og optegnerne. Og det er vigtigt at have løbende, fælles diskussioner af typiske problemstillinger på redaktionen for at etablere faste rammer for omfanget af skøn.

Spørgsmålet om data og repræsentativitet trænger sig også på i forbindelse med angivelsen af et ords eller en betydnings udbredelse. ØMO dækker et stort geografisk område med mange forskellige dialekter. I alt er der afgrænset 14 hovedområder (fx Sjælland, Falster og Fyn) og 15 underområder (fx Sydvestsjælland, Nordfalster og Østfyn) (jf. ØMO Tillægsbind: 37). Det er en obligatorisk del af arbejdet at stedfæste alle de data der præsenteres i ordbogen. De fleste optegnelser er blevet stedfæstet til et bestemt sogn, og ofte vil man opremse de pågældende sogne i ordbogsartiklen. Men hvis man vurderer at materialet rækker til det, vil man foretage en generalisering af udbredelsen til et eller flere hovedområder – eller evt. til hele ømålsområdet. Det sidste er fx tilfældet ved ordet *kromand*. Her er der anført: ”Ø(alm)”. Det betyder at ordet vurderes til at være almindeligt i hele ømålsområdet. Mere forsigtige markører er ”vist alm”, der står for ’vist almindeligt’, og ”spor opt”, der ofte bruges og betyder ’sporadisk optegnet og muligvis almindeligt’. Endelig kan man bruge markøren ”spor” ’sporadisk og ikke almindeligt’, men det kræver at redaktøren har belæg for at ordet er mindre almindeligt end et andet (eller flere andre) ord, der da efter reglerne skal nævnes eksplicit i artiklen (jf. ØMO Tillægsbind: 22).

Hvad skal der fx til for at et ord som *kromand* kan opnå stemplet Ø(alm)? Det er klart at antallet af belæg, dvs. et kvantitativt element, spiller en vis rolle her. Ved *kromand* var der 39 belæg, et ganske hæderligt antal i ØMO-sammenhæng. Ganske vist er der kun anført udtaleoplysninger på 22 af de 39 sedler, men da der kun er én betydning, er dette ikke noget problem. Men naturligvis spiller fordelingen på de forskellige områder også en rolle. Også her består *kromand* testen: materialet er pænt fordelt på hovedområderne: 14 belæg fra Sjælland m.v., 5 belæg fra Møn og Lolland-Falster, og 20 belæg fra Fyn, Tåsinge, Langeland og Ærø m.v. (man forsøgte hurtigt at tilrettelægge indsamlingsarbejdet så der blev indsamlet data fra alle hovedområder, jf. fx Gudiksen & Hovmark 2009). Endelig spiller meddelelserne også en rolle, men her er der flere kendte, solide meddelere repræsenteret. I vurderingen indgår imidlertid også et lidt større skøn. Ordet *kromand* er almindeligt i rigsmålet, og der foreligger ikke andre væsentlige alternativer til ordet, bortset fra *krofatter* der bruges uformelt på Fyn og *krovært* der er langt svagere belagt og kun opnår betegnelsen ”spor opt”. At drage den konklusion at *kromand* er almindeligt udbredt i alle ømaal er derfor ikke voldsomt dristigt, heller ikke selvom der fx mangler belæg fra et par

af de mindre hovedområder. Ud fra kendskabet til seddelsamlingens tilblivelse og karakteristika er der ikke nogen tvivl om at data er tilstrækkelig repræsentativt, og at *kromand* er almindeligt i ømålene.

Som nævnt udelades belæg fra dårlige meddelere, og der bliver ikke oprettet en artikel hvis et enkeltstående belæg fra en sådan meddeler står alene. Et eksempel på en dårlig meddeler er den såkaldte Valentiner, der på kartoteks-kortet kort og kontant karakteriseres som ”uden synderlig interesse”. Men hvorfra ved vi at Valentiner er dårlig? Det slutter vi os bl.a. til ud fra et basalt kendskab til hans baggrund: han var født på egnen syd for København, mellem Køge og Roskilde, men han havde i sit voksne liv sejlet som kaptajn og noteres i kartoteket som bosiddende i det indre København. Men vi ved det også ud fra erfaringer med de oplysninger han har bidraget med. At en person flytter til et andet sted, er ikke pr. definition ensbetydende med at den pågældende mister følingen med barndommens sprog. Men de oplysninger som Valentiner kom med i 1940’erne forekom allerede dengang at være spredte og tilfældige, og der blev sået tvivl om deres kvalitet. Og når man konfererede med andre meddelere fra det levende korpus, viste tvivlen sig ofte at være velbegrundet. Valentiner havde fx leveret et enkeltstående belæg på ordet *langlars* ’doven person’. Tre andre belæg fra samme område er imidlertid såkaldte ”minus-sedler”. På spørgsmålet om hvorvidt en doven person kan ”kaldes for en Langlars?”, svares der henholdsvis: ”ukendt”, ”Det har jeg aldrig hørt” og ”en doven Person kaldes for en Dovenlars”. Resultatet er at vi i dag normalt ser bort fra belæg fra Valentiners hånd, med mindre de er underbygget af andre kilder.

#### 4. Afslutning og perspektivering

Jeg har i denne artikel diskuteret spørgsmålet om data og repræsentativitet på basis af eksempler fra ØMO’s seddelsamling. Seddelsamlinger er forældede som datatype, men det er problematikken omkring data og repræsentativitet ikke: Uanset om data foreligger i form af en seddelsamling, et tekstkorpus, en kombination eller noget helt fjerde, vil der være tale om et udsnit af virkeligheden som skal vurderes kritisk som kildegrundlag. I 1987 nævner Svensén at den nye it-teknologi kan give et frisk pust til vante forestillinger og traditionelle arbejdsrutiner inden for ordbogsarbejdet (jf. afsnit 1 ovenfor, Svensén 1987: 243). Eksemplerne fra ØMO’s seddelsamling kan måske bruges som et muligt afsæt for generelle overvejelser omkring data og repræsentativitet, også i den it-styrede ordbogsvirkelighed baseret på tekstkorpora.

Eksemplerne fra ØMO har fx peget på at en kritisk omgang med data omfatter to relationer eller overgange: relationen mellem virkeligheden og data,



og mellem data og ordbogsprodukt (artikel). Netop fordi seddelsamlinger rummer relativt små og selektivt udvalgte udsnit af virkeligheden, vil man ofte efterspørge en større kontekst. I ØMO's tilfælde er dette særlig påtrængende fordi data her sjældent foreligger i trykt form, men eksisterer som en flygtig virkelighed der kun kan gribes i farten – og i de første mange år indfangede man ikke engang denne virkelighed med båndoptagere, men kun med pen og papir. At foretage en kildekritisk vurdering af de glimt af virkeligheden som har været til rådighed i seddelsamlingen har derfor været og er stadig en indarbejdet rutine, ligesom det har været en nødvendig del af indsamlingsarbejdet og arbejdet med seddelsamlingen jævnlige at gå tilbage til virkeligheden og spørge den til råds. Denne opmærksomhed kan utvivlsomt stadig være nyttig i vore dage hvor man med de store tekstkorpora i højere grad kan have oplevelsen af at have om ikke den fulde, så dog den tilstrækkelige og nødvendige virkelighed liggende på computeren (jf. Svensén 2004: 72 med henvisninger). I hvor høj grad er tekstkorpus repræsentativt? Hvilke sprogtyper og genrer er repræsenteret, og i hvilket forhold?

Overvejelser af denne art er naturligvis ikke ukendte når man arbejder med et tekstkorpus (jf. fx Asmussen (2006) om KorpusDK). Et andet forhold der kommer til syne når man ser på arbejdet med seddelsamlinger er måske mere interessant at overveje. Det gælder det meget indgående arbejde med den enkelte kilde, først og fremmest dens ophav, men også forholdene omkring indsamlingen og indsamleren. Denne fornemmelse for den enkelte kilde opnås ikke på samme måde ved arbejdet med tekstkorpora. Den store mængde data med mulighed for undersøgelse af kontekstuelle forhold, frekvens osv., ofte på basis af statistiske værktøjer og komplekse søgninger (jf. Atkins & Rundell 2008: 103ff., DDO, bind 1: 54-59), er den store og revolutionerende styrke ved tekstkorpora – men kan det undertiden også være en svaghed? Det resultat man præsenteres for er anonymt, og man har ikke altid behov for at komme tættere på den enkelte kilde. Det er i visse sammenhænge en fordel fordi det sætter fokus på generelle og typiske træk i sproget, og mindsker risikoen for at særtræk og måske ligefrem idiosynkratisk sprogbrug kommer til at fylde for meget i ordbogen. Men det kan også skubbe den enkelte kilde og fornemmelsen for den i baggrunden i den daglige arbejdsrutine.

En af de store udfordringer i arbejdet med ØMO's seddelsamling er vurderingen af de meget kvalitativt udvalgte datas repræsentativitet i redigeringen af materialet til en artikel, eksemplificeret ved muligheden for at generalisere udbredelsesangivelser. Her har man en stor fordel når man har et tekstkorpus og statistiske værktøjer til rådighed der kan give mere velfunderede svar på den slags spørgsmål. Når blot man husker at data i en seddelsamling ikke er fuldstændig tilfældige – i ØMO's tilfælde blev data fx indsamlet systematisk ud fra

kriterier om repræsentativitet (geografisk fordeling, meddelerkarakteristik), hvilket giver mulighed for at nå frem til et svar der også er kvalificeret. Og når man husker at de resultater som maskinen spytter ud ikke uden videre kan eller bør tages som udtryk for en endegyldig eller eneste sandhed (man kan jo fx have formuleret sin forespørgsel på en uhensigtsmæssig måde, ligesom de forskellige korpusværktøjer kan anvende forskellige statistiske metoder; jf. Hovmark 2009). De er yderligere input til at give en så fyldestgørende og præcis fremstilling af den sproglige virkelighed som muligt, og de kan med fordel gøres til genstand for kildekritisk opmærksomhed – i lighed med oplysningerne i en seddel-samling.

## KILDER OG LITTERATUR

- Andersen, Poul, 1958: Fonemsystemet i Østfynsk. København.
- Atkins, B.T. Sue & Michael Rundell, 2008. *The Oxford Guide to Practical Lexicography*. Oxford.
- Bévort, Inger, 1965: Indsamlingen af materiale til Ømålsordbogen. I: *Danske Folkemål* 19. S. 239-250.
- DDO = Den Danske Ordbog 1-6, 2003-2005. København.
- Gudiksen, Asgerd, 2001: -vorn's lydlige form. I: *Danske Talesprog* 2. S. 27-76.
- Gudiksen, Asgerd & Henrik Hovmark, 2008: Båndoptagelser som kilde til Ømålsordbogen. I: Svavarsdóttir, Ásta m.fl. (red.): *Nordiske Studier i Leksikografi* 9 (Skrifter udgivet af Nordisk Forening for Leksikografi 10). S. 171-180.
- Gudiksen, Asgerd & Henrik Hovmark, 2009: Måske husker De noget alle andre har glemt. I: Gudiksen, Asgerd m.fl. (red.): *Dialektforskning i 100 år*. København. S. 13-64.
- Hovmark, Henrik, 2004: Jagten på det autentiske citat. I: *Danske Talesprog* 5. S. 3-21.
- Hovmark, Henrik, 2009: Hund og menneske imellem. I: Farø, Ken (red.): *Sprogvidenskab i glimt*. Odense. S. 397-402.
- Køster, Finn, 2009: Fra fortid og nutid. I: Hovmark, Henrik m.fl. (red.): *I mund og bog*. København. S. 193-204.
- Svensén, Bo, 1987: *Handbok i lexikografi. Principer och metoder i ordboksarbetet*. Stockholm.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm.
- Winchester, Simon, 2004: *The Meaning of Everything. The Story of the Oxford English Dictionary*. Oxford.
- ØMO = Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer, 1-, 1992 ff. København.
- ØMO, Tillægsbind, 1992. *Indledning, nøgler, forkortelseslister, kort*. København.

*Henrik Hovmark*

Ømålsordbogen, Nordisk Forskningsinstitut, Afdeling for Dialektforskning, Københavns Universitet  
hovmark@hum.ku.dk

# De første trykte islandske ordbøger

*Gunnlaugur Ingólfsson*

During the mid-seventeenth century two dictionaries of the Icelandic language by Icelandic scholars were compiled. One of them, *Specimen Lexici Runici*, was originally written by Magnús Ólafsson (†1636), priest at Laufás in Northern Iceland, the other, *Lexicon Islandicum*, was written by Guðmundur Andrússon (†1654), who lived in Copenhagen from 1649 until his death. Both of these works were undertaken on the initiative of the Danish professor Ole Worm (†1654), who was an ardent researcher of Danish and other Nordic antiquities. After preliminary editing, revision and additions to the original glossary, *Specimen* was published in Copenhagen by Worm in 1650. This work concentrates on classical Old Icelandic literature, the Eddas, sagas and skaldic poetry, with a few words and idioms, proverbs and poetry from the contemporary language of the author.

The other work, *Lexicon Islandicum*, although completed in 1654, was not published until 1683 by PHI. Resen, professor at the University of Copenhagen. This work is considerably more comprehensive than *Specimen*. The basic vocabulary actually stems from Old Icelandic literature, but a host of words and phrases are taken from the author's contemporary language, both from poetry and prose. So many expressions and idioms appear here and are explained for the first time that this work is very important for the history of Icelandic vocabulary.

In the paper these works are described more closely and compared to each other.

*Nyckelord:* historisk ordbog, periodeordbog, den islandske leksikografis historie, islandsk sproghistorie, islandsk ordforråd

I det syttende århundrede blev der forfattet og udgivet to ordbøger over islandsk sprog. De omfattede især det gamle norrøne eller islandske sprog, det poetiske sprog såvel som det prosaiske, det vil sige over eddadigtning og skjaldedigtning, sagatekster og lovtækster. Alligevel omfatter de også en del ord fra almuesproget, ikke blot enkelte ord, men også idiomer og ordsprog, samt forskellige ord fra forfatternes samtidige digtersprog. Disse ordbogsværker vedrører derfor ikke kun ældre lærdom, men har også betydning for islandsk sproghistorie efter den senmiddelalderlige periode.

Disse to ordbøger er på den ene side *Specimen Lexici Runici* (*Specimen*) af pastor og provst, Magnús Ólafsson (1573–1636) i Laufás i Nordisland, og på den anden side er der tale om *Lexicon Islandicum* (*Lexicon*) af Guðmundur Andrússon (†1654), der var beskæftiget med forskellige opgaver indenfor nordisk filologi i København omkring og efter midten af det syttende århundrede i hvilke år han tillige forfattede ordbogen.

Begge disse værker blev forfattet på initiativ af den danske læge og oldtidsforsker Ole Worm (1588–1654), som var professor i medicin ved Københavns Universitet. Worm selv udgav Magnús Ólafssons *Specimen* i året 1650, men Guðmundur Andrússons *Lexicon* blev ikke udgivet før året 1683 af den retslærde historiker, P.H. Resen (1625–1688), som også var professor ved Københavns Universitet.

På Magnús Ólafsson og Guðmundur Andrússons tid havde humanismens lærde i Norden stiftet omend kun et overfladisk så dog et kendskab til de nordiske rigers historie, som var bevaret i de islandske middelalderhåndskrifter. Men de lærdes viden om håndskrifternes originalsprog var meget begrænset, og de var derfor afhængige af de kundskaber, som samtidige islændinge havde i det gamle sprog.

I det følgende vil jeg gøre rede for indsamlingen af materiale samt bearbejdningen og præsentationen i de to ordbogsværker og sidst gøre et forsøg på at sammenligne dem. Jeg begynder med en redegørelse for *Specimen*, da det værk er ældst og udkom først.

## Om Specimen Lexici Runici (1650)

Allerede i året 1635 anmodede Ole Worm i et brev til Magnús Ólafsson om, at han skulle forfatte en ordsamling over ord fra digtersproget og andre svært forståelige ord i de norrøne tekster (Jakob Benediktsson 1948a: 235). Men Magnús Ólafsson døde godt et år senere og efterlod en ordsamling, som hans søn og efterfølger, pastor Jón Magnússon, renskrev og sendte til Ole Worm i årene 1638 og 1639.

Magnús Ólafssons originale udkast såvel som Jón Magnússons kopi er nu tabt, men der er bevaret en afskrift af Jón Magnússons kopi i Uppsala i Sverige under katalognummeret De la Gardie 55 (DG 55). Worm har tilføjet adskilligt stof, blandt andet fra biskop Brynjólfur Sveinsson. Ved at sammenligne dette håndskrift med den trykte udgave, kan man se, at der er tilføjet en del materiale inden udgivelsen i 1650, bl.a. et antal citater fra eddadigte og Snorres Edda samt fra lovtekster og nogle islandske sagaer, fx Njals saga, og endvidere nogle ordsprog, som ikke findes i DG 55. Man antager at *Lexicon Islandicum*s forfat-

ter, Guðmundur Andrússon, som på den tid var begyndt sin filologiske virksomhed i København, har udført dette arbejde, eftersom nært beslægtet stof fremkommer senere i hans egen ordbog; men stoffet kan i øvrigt ikke stamme fra *Specimen*, blandt andet fordi Guðmundur Andrússons citater i *Lexicon* tit er udførligere og henvisningen til dokumenter mere præcis. (Faulkes 2010: xii–xiii).

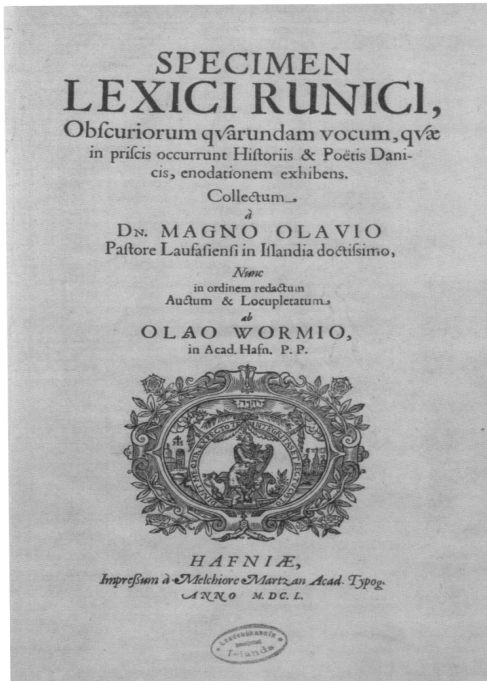


Fig. 1. *Specimen Lexici Runici*, København 1650, titelblad

Worm har i den trykte udgave omskrevet opslagsordene til runeskrift, der trykkes først, men de efterfølges af ordet med latinske bogstaver. Desuden har Worm ladet forklaringer og citater oversætte til latin, men udeladt forklaringer på dansk, som der ellers er en hel del af i håndskriftet. I ordbogen er der flere hundrede citater fra sagaer, lovbøger, digte o.s.v. De islandske sagaer er de oftest citerede, men derudover citeres kongesagaer, Orkneyingasaga (kaldt Jarla saga i *Specimen*), Snorres Edda og eddadigte. Senere tiders værker citeres også, såsom bibeloversættelsen af 1584 og islandske digte og rímur fra det sekstende og syttende århundrede.

Der findes også almensproglige udtryk i ordbogen, hvis kilde ikke anføres, ligesom der angives mundheld og ordsprog, som ikke stammer fra ældre skrifter, men har levet på folkets læber i forfatterens samtid.

### Megastruktur

Værkets hovedindhold, ordbogsartiklerne, lemmalisten, er på 144 hele sider i folioformat. Foran hovedindholdet kommer, efter titelbladet, en indledning af Worm, kildehenvisninger samt litteraturliste, *Syllabus Autorum*, og sidst en tilegnelse og en lovprisning af udgiveren, Worm. Men hoveddelen indledes med overskriften *Lexici Runici Specimen* o.s.v. Ordbogen begynder med bogstavet A og slutter med Þ.

### Mikrostruktur

Opslagsordene er klart adskilt fra andre oplysninger: først kommer opslagsordet med runeskrift efterfulgt af samme med latinske bogstaver i kursiv. Lemmaen adskilles fra resten af ordbogsartikkelen med et tegn, som t.eks. komma, kolon eller semikolon. Dernæst følger grammatiske oplysninger som ordklasseinddeling, bøjning m.m. Siden følger ordets forklaring på latin og derefter et ordret citat fra en kilde (lovbog, saga, digt eller et direkte citat fra en navngiven digter). Til sidst oversættes citatet til latin, men ikke altid ordret, nogle gange endda meget frit.

Alle disse oplysninger gives ganske vist ikke konsekvent i hver enkel ordbogsartikel. Der er ikke fuldstændig overensstemmelse i angivelsen af oplysninger, undertiden fordi selve opslagsordet ikke giver anledning til, at oplysningerne gives: Opslagsordet kan være en kollokation, en talemåde eller et ordsprog, som bevirker at ordforklaringen er vigtigere end de grammatiske oplysninger.

## Lexicon Islandicum (1683)

Jeg går nu over til ordbogen *Lexicon Islandicum*, som blev udgivet i 1683. Forfatteren Guðmundur Andrésson kom til København i efteråret 1649 og blev immatrikuleret ved Københavns Universitet tidligt på sommeren i 1650. Han udførte samtidig forskellige forskningsopgaver for Ole Worm indtil sin død i 1654. Blandt hans opgaver fra disse år er ordbogen, *Lexicon Islandicum [...] vel Lingvæ Septentrionalis Dictionarium*, den vigtigste.

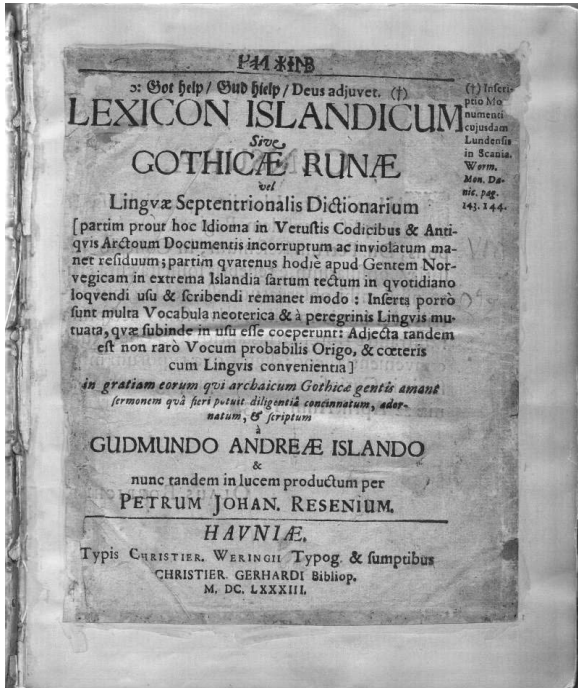


Fig. 2. *Lexicon Islandicum*, København 1683, titelblad

Ordbogen udkom knap 30 år efter forfatterens død. Førnævnte P.H. Resen forestod udgivelsen og skrev indledningen. Udgaven lider under nogle skavanker, den vrimler med trykfejl og andre fejl, og ifølge Jón Helgason (1926: 97) har værket været „fordærvet under trykningen med den største rigdom af trykfejl, som nogen islandsk bog kan byde på“. Der er gjort forsøg på at råde bod på dette med en rettellesliste med rettelser efter ordbogens hoveddel; men listen er ikke omfattende nok til at råde bod på manglerne.

## Megastruktur

*Lexicon Islandicum* er i kvartformat, og hoveddelen, lemmalisten, er på 269 sider i to spalter på hver side. Foran lemmalisten er der, foruden Resens indledning om Guðmundur Andréssons liv og virke, adskilligt andet indhold, bl.a. censur, digtning tilegnet Resen, to portrætter (den ene af biskop Guðbrandur Þorláksson, den anden af provst Arngrímur Jónsson den lærde) samt tekst og digtning – det hele naturligvis på latin. Efter disse indledende kapitler følger siden ordbogens hoveddel, der følger en særskilt sidenummerering.



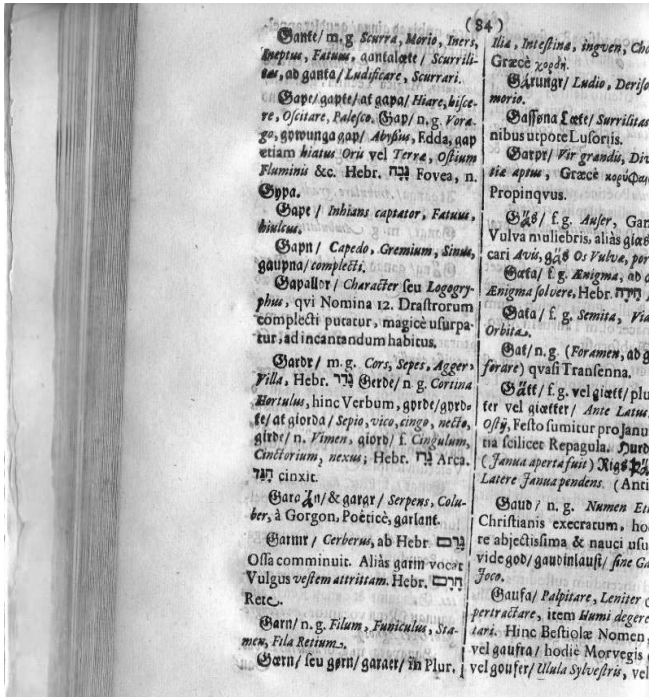


Fig. 3. En side fra *Lexicon Islandicum af Guðmundur Andrússon*.

Efter selve hoveddelen følger fire upaginerede sider med forskellige småting. På en ny side begynder så førnævnte liste med rettelser, *Errata Typographica*, som strækker sig over to og en halv side.

## Makrostruktur

Nu går vi over til ordbogens hoveddel og ser først på den alfabetiske orden: den er A, B, C o.s.v. og slutter med P. I og J underordnes samme bogstav, ligesom V og U (Ú), da vokalerne, med og uden accenttegn, klassificeres som samme bogstav, *a, á; i, í* o.s.v. I denne alfabetiske inddeling behandles bogstavet *æ* som *ae*, f.eks. står ordet *klæði* foran ordet *klafi*. Ord, der begynder med bogstavet *ö*, findes tit under *o* (*ó*), men står ellers med *a*, især hvis ordet med begyndelsesbogstavet *ö* er knyttet til ord med *a* i ordstammen på grund af omlyd: *öflugur* s.v. *afli*; *öðuskel* s.v. *aða* o.s.v.

Den alfabetiske orden inden for hvert bogstav er tit så mangelfuld at det kan tage lidt tid at finde frem til det man søger. Men i hovedtræk er inddelingen den såkaldte ABC-rækkefølge, det vil sige, at den er inddelt efter det første, andet og

tredje bogstav o.s.v. Men det må dog understreges, at den alfabetiske orden tit er usikker, for eksempel på grund af tvivlsom etymologi.

I begyndelsen af hvert bogstav er der sædvanligvis en kort indledning om bogstavet, dets udtale og hvilke bogstaver det svarer til i henholdsvis latin, runeskrift, græsk og hebraisk. Det er forskelligt hvor udførlig denne behandling af bogstaverne er, bogstavet A optager størst plads, mens N og S savner derimod en sådan behandling.

### Mikrostruktur

Som allerede nævnt, inddeles ordbogsartiklerne på hver side i to spalter. Opslagsordene skrives med fede typer og gotisk skrift. Målsproget, latinen, er trykt med latinske bogstaver, græske ord med græske typer og hebraisk med hebraisk skrift. Den egentlige oversættelse til latin, hvad enten det drejer sig om enkelte ord, kollokationer eller sætninger, trykkes med kursiv. Foruden de fede typer er opslagsordet endvidere afgrænset fra andet stof med en skråstreg, „/“. Skråstregen adskiller desuden verbernes bøjningsformer og adjektivernes køn. Græske og hebraiske ord anføres tit, eftersom forfatteren forsøger at give en etymologisk forklaring med henvisning til græske og hebraiske ordstammer, som forklares på latin.

I ordbogen er ordklasserne opdelt på traditionel måde, selv om forkortelserne ikke altid er de samme. Pronomener, for så vidt de forekommer i bogen, oversættes kun med tilsvarende latinske ord. Men de personlige pronomener, *eg*, *þú*, *hann/hún*, betegnes med ordet *pronomén* foran oversættelsen. Når verbernes bøjningsformer angives, udelades ordklassebetegnelsen tit. Det samme gælder adjektiver, der gerne anføres i alle tre køn.

### Retskrivning og opsætning

Som sagt er både redigering og korrekturlæsningen mangelfuld, jf. citatet ovenfor af professor Jón Helgason. Selve trykningen af bogen, håndværket, er endvidere mangelfuld, idet trykningen af bogstaverne ikke er tydelig nok, de er somme tider delvis brækket af eller faldet helt bort. Man har måttet udforme nogle gotiske bogstaver specielt til formålet, og de passer ikke godt til den gotiske type, da de rager ud over af linjerne, både for oven og for neden. For øvrigt er selve trykket ikke så godt som ved de almindelige bogstaver. Endelig er der mange steder fedtfingre og klatter.

Nogle fejl minder faktisk om OCR-fejl, f.eks. trykkes (læses) *n* tit som *u*, *m* > *in/ni*, *n* > *ri* o.fl. Af alt dette er det klart, at både sætter og korrekturlæsere, hvis der har været nogen, har haft ringe eller ingen kendskab til islandsk sprog.

## MS Junius 120

I Bodleian Library i Oxford er der bevaret et håndskrift af Guðmundur Andrés-sons ordbog under katalognummeret MS Junius 120. Dette er tydeligvis en kopi af et håndskrift, der muligvis har været Guðmundur Andréssons autograf, men det drejer sig næppe om et trykfærdigt manuskript.

Der er ikke anledning til at behandle dette håndskrift nærmere her, men det må bemærkes, at MS Junius 120, som vrimler med fejl og misforståelser, ikke var til den nytte ved nyudgaven af Guðmundurs *Lexicon* som man håbede, da dets eksistens først blev almindelig kendt i 1950'erne (Gunnlaugur Ingólfsson 1999: xiv–xv).

## Materiale – Ordforråd – Kilder

Guðmundur Andrés-son har hentet materialet til sin bog nogle steder, som jeg nu vil gøre rede for i store træk.

Ordforrådet er dels hentet fra gamle skrifter, lovtekster, sagaer og eddaer – og kilderne nævnes tit. På den anden side er det klart, at en del af ordforrådet er hentet såvel fra senere tiders digtning som samtidens talesprog.

Guðmundur citerer tit digtning med verslinjer og brudstykker af vers, som stammer fra rímur og danseviser. Navnene på disse kilder anføres sjældent, men siges mere alment, at de stammer fra et gammelt vers, et ældre kvad eller digt.

Han bruger gerne ordsprog og idiomer, som muligvis bygger på hans boglige viden, fordi det meste af materialet er kendt fra andre kilder, mens andet kun kendes fra hans ordbog. Disse citater betegnes med latinske ord som *vulgo* 'almuesprog', *aiunt* 'siger folk' o.s.v. I ældre og yngre skrifter finder man ganske vist noget af dette stof, men her har vi dog et mere håndgribeligt bevis for, at dette i hvert fald har været udbredt i Guðmundur Andrés-sons tid.

## Specimen og Lexicon Islandicum

Hvis man vil sammenligne disse to værker, så kan det nævnes, at *Specimen Lexici Runici* er det første forsøg på at forfatte en ordbog over det gamle norrøne sprog, oldislandsk. Ordbogen er ganske vist begrænset, ordbogsartiklerne er kun godt ét tusinde, og ved arbejdet anvendte man håndskrifter og mange ste-

der citeres ord fra dem i kontekst, selv om det ikke altid er lige så nøjagtigt, som man nu ville kræve af citater. Magnús Ólafsson ser ud til at have anvendt pergamenthåndskrifter, som siden er gået tabt og disse citater kan have en selvstændig tekstkritisk værdi i arbejdet med de pågældende værker og være til nytte ved tekstanalyse og kaste et lys over håndskrifternes historie. Selv om værkets fokus er det gamle klassiske sprog, finder man i *Specimen* forskellige ord, idiomer og ordsprog fra forfatterens samtid.

*Lexicon Islandicum* af Guðmundur Andrússon er et langt mere omfattende værk, selv om det ikke anses for at være stort efter nutidens målestok. Ordbogsartiklernes antal er lidt over fire tusinde. Guðmundur Andrússon anvender, som nævnt, de ældre skrifter, forklarer ordene og citerer nogle steder ordret. Men ordforrådet er meget større end i *Specimen*, og man har især „anset bogen for at være en vigtig kilde til islandsk sproghistorie, fordi mange af ordene er trykt her for første gang foruden, at mange sjældne ord bliver forklaret“, som filologen Jakob Benediktsson (1948b: XX) formulerer det i indledningen til sin udgave af Guðmundur Andrússons polemiske skrift *Deilurit (Discursus Oppositivus)*.

## LITTERATUR

- Benediktsson, Jakob. 1948a. *Ole Worms Correspondence with Icelanders*. Bibliotheca Arnamagnæana VII. Hafniæ.
- Benediktsson, Jakob. 1948b. *Guðmundur Andrússon. Deilurit*. Íslenzk rit síðari alda. 2. bindi. Kaupmannahöfn: Hið íslenska fræðafélag.
- Faulkes, Anthony. 2010. *Magnús Ólafsson of Laufás. Specimen Lexici Runici and Glossarium Priscæ Linguae Danicæ*. Edited by Anthony Faulkes and Gunnlaugur Ingólfsson. Orðfræðirit fyrri alda V. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum. London: Viking Society For Northern Research.
- Helgason, Jón. 1926. *Jón Ólafsson frá Grunnavík*. Safn Fræðafjelagsins um Ísland og Íslendinga V. bindi. Kaupmannahöfn: Hið íslenska fræðafélag.
- Ingólfsson, Gunnlaugur. 1999. *Lexicon Islandicum. Orðabók Guðmundar Andrússonar*. Ný útgáfa. Gunnlaugur Ingólfsson og Jakob Benediktsson önnuðust útgáfuna. Reykjavík: Orðabók Háskólans.

*Gunnlaugur Ingólfsson*

Stofnun Árna Magnússonar í íslenskum fræðum/Háskóli Íslands  
gi@hi.is

# Etymologien til ordet ”jordmor”

*Ernst Håkon Jahr*

*Dette arbeidet er dedisert til minnet om mor mi, sjukepleier og jordmor Else Jahr f. Aale (1917-2002), overjordmor ved Sarpsborg sykehus 1967-73.*

All standard Norwegian dictionaries today give the same etymology of the word *jordmor* (‘midwife’): “From *jord* (‘earth’), because in earlier times women gave birth on dirt floors.” This contribution contests this etymology and discusses four different etymologies suggested by various scholars. Special focus is given to a suggestion that *jordmor* originates from Old Norse *jóð* (‘foetus, small child’) via *\*jóðamóðir/\*jóðsmóðir* > *jord(e)mor* (with an unetymological *r*). Another Scandinavian word for midwife, *ljosmor/ljysmor*, can also be accounted for by this origin: *\*jóðsmóðir* > *josmor* (with the loss of *ð*), which was then reinterpreted as *ljosmor* (‘light mother’).

NØKKELOD: jordmor, ljosmor, gjordmor, jordgolv-etymologien, jordånd/Moder Jord-etymologien, jordmor, madderakka

## Innleiing

Ordet ’jordmor’ blir i *Norsk ordbok* (bd. 5) forklart slik: ”av *jord* av di fødslar tidlegare gjekk føre seg på jordgolv”. Denne etymologien finner vi i dag i de fleste oppslagsverk. I dette bidraget vil jeg diskutere holdbarheten av denne og andre etymologiske forklaringer som er foreslått.

Jordmoryrket er som kjent den eldste kvinneprofesjonen vi vet om. Jordmødre er omtalt allerede i Det Gamle Testamentet (2. Mosebok 1, 15-21), og fra så tidlig som det 2. århundre stammer den første læreboka i jordmorkunst, skrevet av legen Soranos fra Efesos i Lilleasia, ei bok som var i bruk i mer enn 1400 år.

Jordmoras oppgave er å hjelpe den fødende å forløse barnet. Filosofen Sokrates sin mor var jordmor, og Sokrates (ca. 470-390 f. Kr.) så på sin egen samtalegjerning som en jordmorpraksis: Slik ei jordmor hjelper kvinner å forløse sine barn, hjelper filosofen sine samtalepartnere å ”forløse” – dvs. sette ord på, uttrykke sine egne tanker og ideer.

Betegnelse for kvinner som hjelper til ved fødselen, er ulike i forskjellige språk, og de tar sitt utgangspunkt i flere av de funksjoner og/eller handlinger som fødselshjelperen har eller utfører:

Gresk (klassisk):  $\mu\alpha\tilde{\iota}\alpha$ . Uklar etymologi, ordet kan også brukes om andre kvinner, helst eldre og nærstående, og som en vil vise respekt. Har gitt oss lånordet *majeutikk* (jordmorkunst). I gresk mytologi er Maia navnet på guden Hermes sin mor. Et forsøk på etymologisk forklaring gir Hjalmar Frisk i *Griechisches etymologisches Wörterbuch*. Han mener at  $\mu\alpha\tilde{\iota}\alpha$  er danna fra et slags barneord,  $\mu\tilde{\alpha}$ , for 'mamma', med tillegg av nominalsuffikset *-ja*, som er et av de vanlige suffiksene i feminine substantiv i gresk. Denne forklaringa virker som en rein teoretisk konstruksjon på grunnlag av 'μᾶ' i begynnelsen av ordet. 'μᾶ' er nemlig ytterst sparsomt belagt i de bevarte tekstene fra klassisk tid. Men om det likevel er rett, har også det greske ordet elementet *mor* i seg, slik vi finner i flere andre språk, og i *jordmor*.<sup>1</sup>

Latin: *obstetrix* (av verbet *obstare*, stå foran, stå imot; betegner vel da den kvinne som står foran den fødende under fødselen).

Norrønt: *nærkona*, *náverukona* (betegner den kvinne som er nærværende ved fødselen), *yfirsetukona* (jf. *sitja yfir konu*, betegner den kvinne som overvåker eller hjelper til ved fødselen), *bjargrygr* (< *bjarga*, hjelpe).

Italiensk: *ostetrica* (jf. latin *ovafor*), også *levatrice* (av latin *levare*, løfte, heve opp; betegner da den kvinne som løfter opp barnet etter fødselen, jf. også at i det gamle Rom het fødselsgudinna *Levana*).

Spansk: *comadre/comadrona* (fra latin *\*conmater* > *commater*, 'med-mor').

Fransk: *sage femme*, 'klok kvinne'.

Tysk: *Hebamme* (gammelhøytysk *bevianna*; fra verbet *heben*; middelnedertysk *hevemôder*; betegner den kvinne som løfter opp barnet etter fødselen, jf. italiensk *ovafor*).

Engelsk: *midwife* ('med-kvinne', jfr. spansk *ovafor*).

Nederlandsk: *vroedvrouw* ('klok kvinne', < *vroed*, klok, og *vrouw*, kvinne; reknes som oversettelselån fra fransk; dialektvarianter: *vroedwif*, 'klok kone', *vroedmoeder*, 'klok mor').

Frisisk: *froedfrou* ('klok kvinne', < *froed*, klok, jf. norr. *fróðr*, og *frou*, kvinne; reknes som oversettelselån fra fransk).

Svensk: *barnmorska* (folkeetymologisk, muligens opprinnelig fra nedertysk *bademôdersche*, som der betegner den som bader det nyfødte barnet). Også *jord(e)gumma* og *jord(e)moder* fins i svenske dialekter, og er også belagt i trykt skrift fra 1500-tallet og framover.

Islandsk og færøysk: *ljósmóðir*, *ljósa* (jf. nedafor).

Dansk: *jordemoder* (jf. nedafor).

Norsk: *jordmor* (i dialektene (jf. Norsk Ordboks ordsedler) finner vi også belegg på *ljøsmor*, *nærkone*, *stråmor*, *linnmor*, *krafsekjerring*, jf. nedafor. Aasen 1873 har *jordmoder*, *ljøsmoder*, *nærkona*, *straamoder*).

Som det framgår av denne lista, har flere forhold ved fødselshjelperen vært med på å bidra til de betegnelsene som blir brukt om henne i de ulike språka. Det er enten snakk om egenskaper ved henne eller om ei fysisk handling hun utfører. Det er ei klok kvinne [fransk, nederlandsk, frisisk] som er med [spansk, engelsk] eller nærværende [norrønt] ved fødselen, som står foran den fødende [latin] og hjelper til [norrønt], som løfter barnet opp etter at det er født [italiensk, tysk], som bader barnet etter fødselen [nedertysk]. Alt dette er lett å forstå som grunnlag for betegnelser for fødselshjelperen. Men verken *jordmor* eller *ljøsmor* faller naturlig inn i dette mønstret. Verken jord eller lys har noe direkte med det å føde å gjøre, og henviser heller ikke til noen hjelpende funksjon ved fødselen. (*Krafsekjerring* ser ut til mest å ha blitt brukt om hjelpekvinner til jordmødrene, koner som blei tilkalt hvis ei jordmor ikke var å få tak i. Aasen (1873) gir denne definisjonen på verbet *krafsa*: ”gramse, gribe efter noget, bevæge Hænderne meget; ogs. grave, skrabe.”)

Det vi må prøve å komme fram til, er ei forklaring på ordet *jordmor* som både er innholdmessig og språklig-historisk plausibel.

## Fire etymologier

Sjøl om de norske standardverka i dag er i full overensstemmelse når det gjelder forklaringa på opphavet til ordet *jordmor*, betyr likevel ikke det at forklaringa med jordgolv har vært den eneste foreslåtte etymologien. I det følgende skal vi se på i alt fire ulike etymologiske forklaringer som sia 1800-tallet er framsatt for ordet *jordmor*.

### 1. *jordmor* = *gjordmor*

Det er naturlig å begynne med en etymologi som alle synes å avvise i dag, nemlig at opphavet egentlig er *gjordmor*, fra å ”gjorde”, dvs. omgjorde. Dette skulle da bety at fødselshjelperen ”om-gjorder” den fødende, enten ved å binde noe rundt midjen hennes, eller ved å holde armene rundt henne for å støtte og eventuelt hjelpe til med å trykke. En kunne også tenke seg at det er snakk om å omgjorde det nyfødte barnet (som gjerne omtales som å ”linne” barnet, jf. *linnmor*, som oppgis bl.a. fra Jæren for *jordmor*). I romanen *Konerne ved vandposten*

(1920) bruker Knut Hamsun skrivemåten med 'gj': "Saa i nat hører han barneskrik i huset, Mattis op og rænder til gjordmor, rænder til doktor" (Hamsun 1974 [1920]: 140).

I sedlene til *Norsk Ordbok* finner vi også ordet skrevet med *gj-*, og av ingen ringere enn dialektologen Hallfrid Christiansen. Fra Gimsøy i Lofoten har hun notert ordet på denne måten: "jordmor f. [jo:r`mo:r] Hallfrid Christiansen. Gimsøy." (Men så "jordmor f." fra Eidsfjord i Hadsel, også oppskrevet av Hallfrid Christiansen.)

Mor mi fortalte at da hun gikk på jordmorskolen i Bergen, det var det siste krigsåret 1944-45, fikk de dosert at den riktige historiske skrivemåten var med 'gj', altså *gjordmor*, og da nettopp med "omgjorde" som begrunnelse for slik skrivemåte.

Da Jordmorskolen i Oslo feira sitt 175-årsjubileum 1993 med et eget festskrift (*Jordmorutdanning gjennom 175 år*; red.: Knut Bjørø et al.), satte de dette øverst på omslaget: Gjordemor – Jordemor – Jordmor. Med det mente de muligens at det har vært ei slik språklig utvikling, men dette er verken omtalt eller diskutert i boka ellers. Det er da heller ikke riktig.

Dette er en etymologi som blir fullstendig avvist (jf. Falk 1900: 62). Grunnen til det er at skrivemåten med bare 'j' er belagt før 'gj' og 'j' falt sammen i uttalen. Hvis *gjordmor* var det rette, skulle vi vente å finne belegg på skrivemåten med 'gj' fra før dette lyd-sammenfallet var skjedd. Isteden finner vi bare skrivemåte med 'j' belagt. Dermed kan ikke *gjordmor* med 'gj' være etymologisk rett, og følgelig har ordet ikke noe med "omgjording" å gjøre.

## 2. Jordgolv-etymologien

Den andre etymologien vi skal se på, er den som nå altså er å finne i alle standardverka (*Norsk Ordbok*, *Bokmålsordboka*, *Nynorskordboka* o.fl.), nemlig at fødselshjelperen blir kalt *jordmor* fordi fødselen var på jordgolv i eldre tid. Eli Ellingsve, som redigerte *jordmor*-artikkelen i bind 5 av *Norsk Ordbok*, har opplyst til meg (muntlig) at hun verken hadde tid eller mulighet til å foreta egne undersøkelser, men måtte i all hovedsak bygge på andre ordbøker. At ulike ordbøker i stor grad bygger på hverandre, er ikke uproblematisk reint faglig, for det betyr naturligvis også at gale opplysninger og etymologier står i fare for å bli kopiert fra ordboksverk til ordboksverk. Når så en leser finner samme etymologi oppgitt i alle tilgjengelige verk, framstår opplysningene som langt sikrere og underbygde enn de kanskje er. Jeg mener nok at det er tilfellet med ordet *jordmor*, som dermed godt illustrerer det generelle problemet: kopiering av usikre opplysninger fra ordbok til ordbok, fordi forfatterne ikke har tid til å gjøre egne undersøkelser, men isteden i stor grad må stole på tidligere verks opplysninger om de enkelte orda.



Hva er det så som støtter jordgolvetymologien til ordet *jordmor*? Ordet settes ofte i betydningsmessig sammenheng med *Hebamme* på tysk. *Hebamme* blir forklart som ”løftekvinne”, altså den som løfter opp barnet etter at det er født. På samme vis tenker en seg at jordmora kalles så fordi hun løfter barnet fra et sted, nemlig jordgolvet (jf. norrønt *hóf hann af jorðu*).

Det er også slik at ordet ”golv” inngår i flere uttrykk i forbindelse med fødsel i eldre tid. Norrønt har uttrykket *liggja á gólfi* for det å føde, og i flere dialekter er uttrykket blitt brukt opp til moderne tid: *liggjelgå i golve* (Hordaland og Jæren). Flere kilder beskriver førkristen skikk med at når et barn blei født, skulle det legges på jorda inntil det var bestemt om det skulle ”settes ut” eller ikke. Jf. Keyser (1867: 4): ”Det nyfødte Barn henlagdes af den Kvinde, der havde hjulpet Moderen ved Fødselen, paa Jorden, og blev der liggende uden at røres af nogen af Forældrene, indtil Bestemmelsen, om det skulde opføres eller udsættes, var tagen.” Og videre (s. 8): ”Det var naturligviis det sjældnere Tilfælde, at Forældre udsatte sit Barn; oftest blev det opfødt. Barnet blev da taget op fra Jorden og baaret hen til Faderen.”

Medisineren A. L. Faye (1885: 699) gir en tilsvarende beskrivelse, etter først å ha gjort greie for ulike fødselstradisjoner rundt om i verden: ”Til en knælende eller ganske liggende Stilling af den Fødende paa Gulvet svarer godt den senere Benævnelse *Jordemoder*, – Jordgumma, – om Fødselshjelpersken. Denne Betegnelse har nemlig rimeligvis fra Begyndelsen af været brugt om den Kvinde, som først bistod den Fødende og derefter løftede det nyfødte Barn op fra Jorden (Gulvet) og rakte det hen til Faderen, at han skulde afgjøre, om det skulde opføres eller ikke.” Faye henviser til Keyser (1867) som belegg på dette: ”Paa denne, som det forekommer mig, mest utvungne Maade forklarer R. Keyser det omtvistede Ord, *Jordemoder*, idet han til støtte for denne Opfatning ligeledes henviser til det tydske *Hebamme*, der peger i den samme retning” (Faye 1885: 700).

I Nordland har *stråmor* vært brukt for jordmor. Hvis grunnen til dette navnet er at fødselen skjedde på et strå-underlag, vil det naturligvis kunne refereres til som støtte også for jordgolvetymologien til *jordmor*. Slik blir da også *stråmor* brukt, f.eks. av Falk (1900:64): ”Dette ord henviser da selvfølgelig til et straaieie som fødested.” Falk (1900: 63) konkluderer: ”Herefter betegner jordmor den kvinde hvis funktion det er at hæve det nyfødte barn op fra jorden.”

Det er likevel litt merkelig at de fleste forfatterne bygger så mye på henvisninga til tysk *Hebamme*. Det er nemlig utvilsomt stor forskjell på å betegne fødselshjelperen som den som *løfter* barnet opp etter fødselen, og det at hun betegnes ved *utgangspunktet for handlinga*. Naturligvis må det finnes et slikt utgangspunkt, et sted barnet er, og som fødselshjelperen løfter det fra, men mens flere språk, som vi har sett, bruker oppløftinga av barnet til å betegne jordmora, er

det ikke noe annet språk som lar det stedet barnet løftes fra, få slik betydning at fødselshjelperen betegnes ved det.

Det første danske belegg vi har på ordet *jordmor*, er fra 1535 (Hans Tavsens, luthersk reformator) og i Bibelen av 1550 (jf. Falk 1900: 63): *jordemoder*. Men 25 år før er ordet *iordagumma* belagt i svensk (1510, i Jomfru Mariæ Yrtagardher, jf. Falk, sm.st.). Det kan argumenteres for at dette er ganske seint. Når en vil forklare ordet slik Keyser og Faye gjør (jf. ovafor), med tilslutning altså av samtlige av dagens ordbøker, er rimeligvis det lange tidsskillet mellom den egentlig førkristne skikken som fortelles, og det at ordet først dukker opp på 1500-tallet, et problem. Riktignok blir det hevda at skikken med å ligge på gulvet og føde, evt. at barnet løftes opp (til faren) etter fødselen, har holdt seg langt opp mot vår egen tid. Men likevel kan det tidsmessige være et problem her. Mange fødsler på 1400- og 1500-tallet foregikk slett ikke på jordgolv. At jordmora skulle betegnes slik på grunn av tidligere tiders mer kummelige forhold for hele befolkninga, kan vanskelig være riktig.

Det kan iallfall ikke være den fulle og hele begrunnelse. Alf Sommerfelt, som var ansvarlig for artikkelen som beskrev sammensetninger med *jord-* i *Norsk riksmålsordbok* (bd.1, 2; 1937), har tydeligvis ikke vært helt bekvem med den "reine" jordgolvs-etymologien. Han fører oss således direkte over til etymologi nr. 3, når han skriver om *jordmor*: "så kalt fordi fødselen tidligere foregikk på jorden (gulvet) forat barnet kunde styrkes av jordkraften".

### 3. *Jordånd/Moder Jord-etymologien*

Det Alf Sommerfelt bringer inn i artikkelen i *Norsk riksmålsordbok*, er en helt annen og viktigere grunn – hvis den er rett – til hvorfor en fødsel skulle foregå på et jordgolv. Det var ikke slik at det skjedde fordi en ikke hadde annet golv, men fordi sjøve jorda blei forbundet med viktige krefter, som det var avgjørende at det nyfødte barnet fikk ta del i. Etymologiforklaring nr. 3 dreier seg derfor om akkurat det: er *jord-* i *jordmor* å forstå som noe annet enn "mold"?

Vi bruker jord-symbolikk den dag i dag, naturligvis: "Av jord er du kommet, til jord skal du bli", i kirkas begravelsesrituale. Møller (1924) legger an en argumentasjon som legger mye vekt på jord-symbolikken i en diskusjon om opprinnelsen til ordet *jordmor*, og han kan sies å dele oppfatning med det vi mer antydningvis finner i *Norsk riksmålsordbok*.

Folkelivsgranskeren Nils Lid (1946b, cf. også 1946a, 1947) tar imidlertid dette momentet helt ut, og mener at ordet *jordmor* er "a transfer from a female earth genius dwelling in the houseground" (1947: 356), altså en slags Moder Jord-forklaring. (Ideen var imidlertid diskutert tidligere, se Setälä 1912: 209.) Lid legger særlig vekt på ordet for *jordmor* i samisk, *madderakka*, som, ifølge han, samtidig er navnet på ei jordgudinne som lever under samehytta: "Madda-

rakka is originally also a variant of the widely known Great Mother. The earth-mother, *jord-moder*, for midwife, must be explained as a transference from a spirit named like that" (Lid 1946b20). I et foredrag i Det Norske Videnskaps-Akademi har han denne konklusjonen: "Ein må, etter det framførde, tolka hjelpekona *jordmor* som ei overføring frå eit vette med same namnet, som er tenkt å vera hjelpeånd for barnet" (Lid 1946a:19).

Den forklaringa vi har sett på her, legger altså et helt annet innhold i førsteleddet *jord-* i *jordmor* enn det vi så i forklaring 2. Her er det snakk om "Moder Jord" og jord-ånder, som skal beskytte barnet, og som jordmora er den legemlige representant for og formidler av. Med tanke på jordmoras viktige funksjon og sosiale status i mange samfunn er dette ei forklaring som betydningsmessig må sies å være mer troverdig enn jordgolvetymologien alene.

#### 4. *jordmor* = *jodmor*

Den siste etymologiske forklaringa vi skal se på, blei først foreslått av Cleasby og Vigfusson i deres islandsk-engelske ordbok (1874: 326). Der finner vi først ordet for nyfødt foster, lite barn, norrønt *jóð* n.<sup>2</sup>, og deretter dette oppslagsordet:

*jóð*-moðir, f. [Dan. corrupt *jorde-moder*], a midwife.

Vi finner det samme både i 1800-tallsutgava og i 1900-tallsutgava ("Uggleupplagen") av den svenske *Nordisk familjebok* (bd. 7, 1884, sp. 1336; bd. 13, 1910, sp. 141):

Jordgumma: da. *jordemoder* (förvrängning af isl. *jóðmóðir*, af *jóð*, afkomma, barn), barnmorska.

Det som er spesielt interessant med ei slik forklaring, er at med den kan både *jordmor* og *ljosmor*, det siste en mye brukt betegnelse i norske dialekter for fødselshjelperen, forklares med samme opphav. Vi må i begge tilfelle via folkeety-mologiske utviklinger. Som følge av at ordet *jóð* ikke lenger blei forstått, blei betegnelsen for fødselshjelperen utvikla i to retninger: én gav *jordmor* som resultat (med tilknytning til ordet *jord*, som folk kjente og forbandt noe med), den andre gav *ljosmor* (uttalt [ju:smu:r], med tilknytning til ordet *lys/ljos*, som folk også kjente). Hjalmar Falk (1894: 37f) henviser til *jóð*, og nevner flere sammensetninger med dette ordet: "jóðsjúk bruges om kvinden som er i barnsnød, jóðsótt om fødselssmerter. Denne udledning bekræftes, for ikke at sie bevises, af det nyislandske *jóðmódir*, som bruges ved siden af *ljósmódir* og *ljósa*. Herefter skulde da jordemor egentlig hede jodemor, i gammel form: *jóða-módir*, og *ljosmor* skulde stamme fra et ældre *jóðs-moðir*."

Den utviklinga Hjalmar Falk ser for seg her, er altså slik: Utgangspunktet er et ikke belagt norrønt ord: *\*jóðamóðir*/*\*jóðsmóðir*, men danna av et godt belagt ord, *jóð*, foster/nyfødt barn. Utviklinga til *jordmor* og *ljosmor* blir da slik:

- *jóðamóðir* > *jordemor* > *jordmor* (med et innskudd av /r/ for å gjøre ordet forståelig, jf. to andre ord med innskutt uhistorisk /r/: norr. *dáligr* > *dårlig*, *tysdag* > *tirsdag*).
- *jóðsmóðir* > *jósmor*, med bortfall av /ð/, og så, folkeetymologisk, forstått som *ljosmor*/*lysmor*.

Det tiltalende ved denne forklaringa er at den er både språklig og saklig plausibel. Vi slipper å argumentere med fødsler på jordgolv, heller ikke behøver vi å appellere til Moder Jord-mystikk og -symbolikk. At ei jordmor har med et foster/nyfødt barn (*jóð*) å gjøre, er som det skal være, og språklig er det ikke usannsynlig med ei utvikling som den antyda ovafor. Folkeetymologiske omlaginger når betydninga av et ord(element) – her *jóð* – ikke lenger blir forstått, fins det mange eksempler på. At ordet *jordmor* først dukker opp i denne forma så seint som på 1500-tallet, passer også godt inn i denne forklaringa. Når det gjelder *ljosmor*, er det slik at jordmora/ljosmora fra gammelt av ikke har hatt spesielle funksjoner som direkte kan knyttes til noe med lys. Iallfall kjenner vi ikke noen slik tradisjon fra gammelnorsk tid. Nils Lid (1946a, 1946b, 1947) argumenter riktignok sterkt for en lystradisjon i forbindelse med fødsler, men han antar at den må være yngre enn den jordånd-tradisjonen (og Moder Jord-tradisjonen) han anser for å være den opprinnelige (jf. ovafor). Det kan derfor tenkes at det som eventuelt måtte finnes omtalt av seremonier med lys i forbindelse med fødsler, og Nord-Norge blir nevnt spesielt av Lid (1946b: 7ff), er blitt innført etter at ordet *ljosmor* kom i bruk, og folk mente det hadde noe med lys å gjøre. Når fødselshjelperen blei kalt *ljosmor*, var det naturlig å utvikle noe med lys omkring hennes aktivitet. Alf Torp (1919) avviser imidlertid at *ljosmor* har noe med *lys* å gjøre. Han henfører isteden forleddet *ljós-* til *laus* og *løysa*, og viser til et norrønt belegg: *leysa kind frá konum*.

I sitatet ovafor fra Falk (1894: 37f) sier han altså at det nysislandske *jóðmóðir* beviser at dette er den riktige etymologien. Falk bygger rimeligvis på ordboka til Cleasby og Vigfusson, som altså oppgir *jóðmóðir* for islandsk, men Falk gir i 1894 ingen henvisning. Når han så seks år seinere igjen skriver om ordet *jordmor* (Falk 1900: 62-64), nevner han, overraskende nok, ikke *jóð* eller *jóðmóðir* i det hele tatt. Nå er det isteden jordgolvetytologien han forfekter, som også sitert ovafor: ”Herefter betegner jordmor den kvinde hvis funktion det er at hæve det nyfødte barn op fra jorden” (Falk 1900: 63).

Hvorfor Falk så plutselig skifta standpunkt, vet vi ikke. I Falk og Torps etymologiske ordbok (1903-06 på norsk og 1910-11 på tysk) er det også bare jordgolvetymologien som gjelder, og det har sikkert bidratt avgjørende til at den er enerådende i ordbøkene i dag.

## Konklusjon

Verken etymologien som tilsier skrivemåten *jordmor*, eller jordgolvetymologien står etter dette særlig sterkt. Av de fire forklaringene på ordet *jordmor* jeg har drøfta her, står nr. 3 og nr. 4 sterkere. Jeg vil tro at Den Norske Jordmorforening vil ha, og mor mi ville hatt, stor sans for nr. 3, og – som nevnt – samsvarer den egentlig godt med den sterke stilling og solide sosiale status jordmødre har hatt ned gjennom historia. Som språkforsker er jeg likevel tiltalt av den fjerde og siste forklaringa. Den gjør det forståelig at ordet først dukker opp på tidlig 1500-tall, og forener både språk og sak på en tilforlatelig måte, slik vi gjerne vil at etymologiske forklaringer skal gjøre.

Dagen etter at jeg hadde holdt foredraget mitt på konferansen, kommenterte en av de kvinnelige tilhørerne at hun alltid hadde tenkt på jordmora som den som hjelper barnet til Jorda, til Verden, og at dét var opphavet til sjøve ordet *jordmor*. Ettersom folkeetymologi spiller ei helt avgjørende rolle i forklaring 4, er det vel som vi skulle vente at stadig nye og kreative forslag dukker opp, spesielt når det gjelder et så sentralt område som her, fødsler. Jeg tror nok likevel det er slik at antall mulige forklaringer på ordet *jordmor* er uttømt med de fire jeg har diskutert her.<sup>3</sup>

## Noter

1. Takk til professor Jerker Blomqvist, Lund, for hjelp med klassisk gresk i dette avsnittet.
2. Cleasby og Vigfusson (s. 326) skriver dette som en kommentar til *jōð*: "this interesting word is prob. akin to óðal, auðr, eðli, referring to an old strong verb, jóða, auð, throwing light upon the sense of these words". Dette avvises av Torp (1896: 172), som isteden antyder at *jōð* kan komme fra indoeuropeisk *\*jeuto-m*, partisipp av rota *\*jeu*, sanskrit *yáuti* "forbinde", og dermed bety "det som er forbundet med mora", altså i mors liv.
3. Takk til Marit Hovdenak, som alt i 1980 sendte meg kopier av Norsk Ordboks sedler med belegg på *jordmor*, *ljosmor*, *nærkone*, *stråmor*. Det skulle altså gå lang tid før jeg fikk brukt dem!

## LITTERATUR

Bjæro, Knut et al. (red.), 1993: *Jordmorutdanning gjennom 175 år*. [Oslo.]

- Cleasby, Richard og Gudbrand Vigfusson, 1874: Icelandic-English dictionary. Oxford: Clarendon.
- Falk, Hjalmar, 1894: Vanskabninger i det norske Sprog. Populære Foredrag. Kristiania: Cappelen.
- Falk, Hjalmar, 1900: Kulturminde i Ord. Kristiania: Aschehoug.
- Falk, Hjalmar og Alf Torp, 1903-06: Etymologisk ordbog over det norske og det danske sprog 1-2. Kristiania: Aschehoug. Tysk utg. 1910-11: Norwegisch-dänisches etymologisches Wörterbuch 1-2. Heidelberg: Winter. Nytt opplag av den tyske utg. ved Universitetsforlaget, Oslo 1960.
- Faye, A. L., 1885: Oplysninger om Forhold og Skikke vedrørende Svangerskab og Fødsel hos de gamle Nordboere. I: Norsk Magazin for Lægevidenskaben. Tredie Række, bd. 15.
- Frisk, Hjalmar, 1954-72: Griechisches etymologisches Wörterbuch 1-3. Heidelberg: Winter.
- Hamsun, Knut, 1974 [1920]: Konerne ved vandposten. Oslo: Gyldendal.
- Keyser, Rudolf, 1867: Nordmændenes private Liv i Oldtiden (= Æfterladte Skrifter af R. Keyser. Andet Bind anden Afdeling). Christiania: Mallings.
- Lid, Nils, 1946a. Ljosmor og jordmor. [Sammendrag av foredrag holdt i Det Norske Videnskaps-Akademi.] I: Det Norske Videnskaps-Akademi, Årbok 1945. Oslo: Jacob Dywad i kommisjon, s. 18-19.
- Lid, Nils, 1946b: Light-mother and Earth-mother (= Studia Norvegica. Ethnologica & folkloristica, vol. I, No. 4). Oslo: Aschehoug.
- Lid, Nils, 1947. [Sammendrag av Nils Lids artikkel Light-mother and Earth-mother = Lid 1946b]. I: Norsk Tidsskrift for Sprogvidenskap 14, 1947, s. 356.
- Møller, J. S., 1924: Oprindelsen till Ordet 'Jordemoder'. I: Tidsskrift for Jordemødre 34, 168-172.
- Nordisk familjebok. Konversationslexikon och realencyklopedi. Första utg. 1876-1899, bd. 7 (Hufvudskål – Kaffraria), Stockholm 1884: Gernandts boktryckeri-aktiebolag, og Andra utg. 1904-1926, bd. 13 (Johan – Kikare), Stockholm 1910: Nordisk familjeboks förlag.
- Norsk Ordbok 1966-, bd. 5 (harm – jåttut). Oslo 2005: Det Norske Samlaget.
- Norsk Riksmålsordbok 1937-1957, vol. 1, 2. halvbind (For- – låvevegg). Utarb. av Trygve Knudsen og Alf Sommerfelt, utgitt av Riksmålsvernet. Oslo: Aschehoug.
- Setälä, Emil, 1912: Aus dem Gebiet der Lehnbeziehungen. I: Finnisch-ugrische Forschungen 12, 161-289.
- Torp, Alf, 1896: Bidrag til germansk, fornemlig nordisk Ordforklaring. I: Sproglig-historiske Studier tilegnede Professor C[arl] R[ichard] Unger. Kristiania: Aschehoug.
- Torp, Alf, 1919: Nynorsk etymologisk ordbok. Kristiania: Aschehoug.
- Aasen, Ivar, 1873: Norsk Ordbog. Christiania: Mallings.

*Ernst Håkon Jahr*

Universitetet i Agder

ernsthakon.jahr@uia.no

# Parallellkorpusen som resurs i lexikografiskt arbete

*Håkan Jansson*

This paper discusses the use of parallel corpora as a tool in bilingual lexicography. After a brief background sketch with reference to literature and (lack of) reported ongoing projects, the text briefly considers the notion of a parallel corpus, before it reaches the main subject: a hands-on study of the opportunities connected to the use of parallel corpora in lexicography.

The corpus used for the study is a relatively small, about 115,000-token, ad-hoc corpus compiled of texts from the Nordic Council. The languages investigated are Icelandic on the one hand, and Danish, Norwegian and Swedish on the other. One advantage of parallel corpora seems to be the possibility to compare concordances of n-grams between the studied languages. The two Icelandic verbs probed as examples were shown to be part of hitherto unnoticed equivalent pairs, where one or both sides turned out to be MWEs. Co-occurring patterns between active or passive voice and certain translation equivalents have also been noticed.

*Nyckelord:* korpusbaserad lexikografi, parallellkorpus, tvåspråkig lexikografi, n-gram, flerordsuttryck

## Inledning

Den lexikografiska litteraturen har sedan ett antal år noterat möjligheterna att använda parallellkorpusar i lexikografiskt arbete. Två exempel på tvåspråkiga lexikonprojekt där parallellkorpusar i någon mån varit aktuella är dels *The Oxford-Hachette French dictionary* (Corréard et al. 2007), dels den kommande *New English-Irish Dictionary* senast beskriven i Atkins et al. (2010). Än så länge har det dock inte publicerats några beskrivningar som presenterar bruket och nyttan av parallellkorpusar i dessa projekt.

Det här bidraget avser att beskriva arbetet med en förhållandevis liten korpus sammansatt av texter på isländska, danska, norska och svenska från Nordiska Rådet. Studien kommer att utgå från översättningar till isländska, där två

isländska verb i alla dess böjningsformer väljs ut för närmare undersökning med utgångspunkt i de parallella texterna. Särskilt intresse riktas mot återkommande ordsamband. Korpusens parallella texter används för att hitta kandidater till ekvivalentpar, som sedan i någon mån jämförs med de förslag till ekvivalenter och konstruktioner som påträffas i lexikon.

## Bakgrund

Allt sedan Svensk ordbok (SOB) (1986) och Collins-COBUILD English language dictionary (COBUILD) (1987) kom ut, har arbetet med korpusar som bas för enspråkig lexikografi blivit en alltmer självklar förutsättning. I tvåspråkig lexikografi har utvecklingen mot korpusbruk inte varit lika tidig och tydlig, men Svensén (2004: 59) rapporterar att *Oxford-Hachette French dictionary* (Corréard & Grundy 1994) för första gången 1994 använde två enspråkiga korpusar, en engelsk och en fransk, till stöd för arbetet med källspråket och målspråket var för sig. För att fullt ut kunna jämföra två korpusar vid själva valet av översättningsekvivalenter krävs dock någon typ av parallellstruktur mellan korpusdata i källspråket och i målspråket. Atkins & Rundell (2008: 473 ff.) beskriver tre olika nivåer av korpusstöd vid arbetet med tvåspråkiga lexikon:

1. *Parallella konkordanser*. Man undersöker konkordansen för det lemma som ska analyseras i ett fönster, samtidigt som man undersöker konkordansen för den tilltänkta översättningsekvivalenten i ett annat fönster. Det finns inte någon matchning mellan texterna i korpusarna.
2. *Parallella korpusar* med matchande texter. Ingen av de ingående texterna är en översättning av den andra, men de är matchade med hänsyn till genre och/eller domän eller ytterligare relevanta faktorer.
3. *Parallellkorpus* med länkning på meningsnivå. Här är en av texterna en översättning av den andra, och meningarna har länkats till varandra. De översättningsekvivalenter som dyker upp här, har valts ut av en översättare med tanke på att de ska passa i den aktuella kontexten.

Vid arbete med den första och andra typen av korpusar måste ordboksredaktören själv ha en idé om vilken översättningsekvivalent som ska användas, medan den tredje typen genom konkordansen kan ge förslag på översättningsekvivalenter.

Benämningarna på dessa tre typer av korpusar är inte helt entydiga i den internationella litteraturen. Typ 2, *parallella korpusar* kallas ibland för *parallellkorpus*, medan typ 3 *parallellkorpus* ibland kallas för *översättningskorpus* (*translation corpus*). Den undersökning som här presenteras kommer uteslutande att



gälla korpusar av typ 3, och de kommer genomgående att benämnas *parallellkorpusar*.

Komparativa fördelar respektive nackdelar med arbete med de tre ovanstående korpus typerna har diskuterats av bl.a. McEnery et al. (2006) och Atkins & Rundell (2008). En viktig fördel med den tredje typen är att den kan ge förslag på översättningsekvivalenter och konstruktionsmönster. En nackdel kan vara att översatta texter ofta kan innehålla uttryck och konstruktioner som är mer typiska för originaltexten än för det språk den ska översättas till. Detta har bl.a. påtalats av Gellerstam (1996) samt av Borin & Prütz (2001). Man måste dessutom vara medveten om att den enskilde översättaren inte sitter inne med någon gudomlig sanning i frågan om vilken av flera tänkbara ekvivalenter som är mest lämplig i olika fall. Men i och med att översättare alltid arbetar med hela texter som är relaterade till en viss kontext, kan studier i deras val av översättning ge lexikografen uppslag till kontextuella perspektiv på valet av lexikonekvivalenter.

Det mesta som har skrivits om parallellkorpusar har skrivits inom språkteknologi, lingvistik eller inom vissa specifika språk. I Norden är särskilt kontrastiva studier inom anglistik relevanta. Bland beskrivningar av metoder och möjligheter med parallellkorpusar kan särskilt Johansson & Hofland (2007) och Hasselgård (2010) vara av intresse här. Såväl Johansson & Hoflands som Hasselgårds texter kan dessutom ge en grund för att förstå metoden bakom arbetet med flera stora parallellkorpusar där ett av de ingående språken är nordiskt. Bland de korpusar som nu finns tillgängliga med ett ingående nordiskt språk är *Europarl* (da-en, fi-en samt sv-en; se vidare Koehn 2005 för beskrivning av Europarl). Vidare kan *English-Norwegian Parallel Corpus*, beskriven av Johansson, Ebeling & Oksefjell (1999/2002), nämnas. Ytterligare en korpus, beskriven av Altenberg, Aijmer & Svensson (2001), är *The English-Swedish Parallel Corpus*.

Det finns dock ännu inte någon större korpus där nordiska språk jämförs med varandra. Arbete med att ta fram en sådan korpus pågår dock vid Språkbanken vid Institutionen för svenska språket vid Göteborgs universitet. De principer som styr arbetet med denna korpus är desamma som de som ligger till grund för SALT-projektet – som beskrivs av bl.a. Rawoens (2003). De ingående språken i den internordiska korpusen är danska, färöiska, isländska, norska och svenska.

## Material och metod

Materialet i den här studien kommer från en mindre *ad hoc-korpus* bestående av texter från Nordiska rådet. Beslutet att välja texter från Nordiska rådet är i huvudsak grundat på att texterna finns lätt tillgängliga på Internet och att de är fria ur copyrightsynpunkt. Korpusen består av nyhetsbrev från Nordiska Rådet från perioden 2006–2009 samt en kortare rapport, *Friskare nordiska hav* (2010). Gemensamt för alla texter är att de är översatta *till* isländska<sup>1</sup>. Språket i källtexterna varierar mellan danska, norska och svenska i nyhetsbreven, medan språkparet i rapporten är svenska-isländska. Storleken på korpusen är ca 115 000 löpord fördelat på ca 15 000 ordformer. Domänen för korpusen bedöms vara samhällsdebatt medan genretillhörigheten är facktext/populärvetenskap.

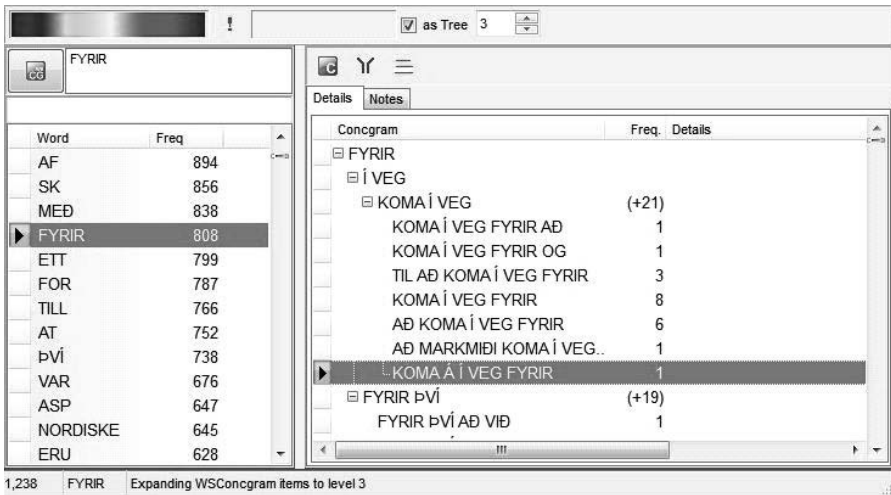
Arbetet med att föra samman texterna till en korpus har gjorts med hjälp av programvaran WordSmith 5.0, hädanefter WS5, vilken närmare beskrivs av Scott (2008). Eftersom vare sig nyhetsbrev eller rapport hade en helt identisk layout i de olika språkversionerna var det nödvändigt att för-redigera texterna, så att deras respektive textstycken hjälpligt motsvarade varandra, innan de parallelliserades på meningsnivå i WS5.

Ett område där utnyttjandet av parallellkorpusar kan vara av avgörande betydelse är när man vill hitta ekvivalentpar där en av, eller båda, de ingående ekvivalenterna är ett flerordsuttryck. För att särskilt kunna studera sådana ekvivalentpar användes den funktion i WS5 som kallas *Conc-gram*. WS5:s *Conc-gram*-funktion är en tillämpning av det som i litteraturen i dag normalt benämns *n-gram*, men som även varit känt under namn som *bi-gram* och *skip-gram*; se t.ex. Cheng et al. (2006) för en översikt över utvecklingen av funktionen i olika applikationer.

Trots att isländska i det här fallet är målspråk, kom alla sökningar att göras med utgångspunkt i isländska. Det är annars vanligt att man gör sökningar som dessa med utgångspunkt i källspråket, men eftersom källspråken förutom svenska även inkluderade norska och danska var isländska här det enda gemensamma språket.

Eftersom ekvivalentparen kan bli betydligt fler och betydligt mer komplicerade vid flerordsuttryck koncentrerades undersökningen på isländska flerordsuttryck med verb *koma fyrir* och *standa fyrir*. Figur 1 visar ett exempel på hur en sökning på *fyrir* i olika kombinationer i *Conc-gram*-funktionen i WS5 kan se ut. Lägg särskilt märke till det markerade resultatet *koma á í veg fyrir* som visar att den sökta ordkombinationen inte måste vara kontinuerlig med de ingående komponenterna i någon viss ordning.

<sup>1</sup> Enligt e-postkommunikation med Michael Funch, huvudredaktör för nyhetsbrevet.



Figur 1. Översikt över uppfångade flerordsuttryck med fyrir i experimentkorpusten

När man har hittat ordkombinationer av ett slag man vill undersöka kan man klicka sig vidare till en KWIC-konkordans (se figur 2). Vi ser här att den ordkombination som var markerad i figur 1 återkommer i kontext i figur 2.

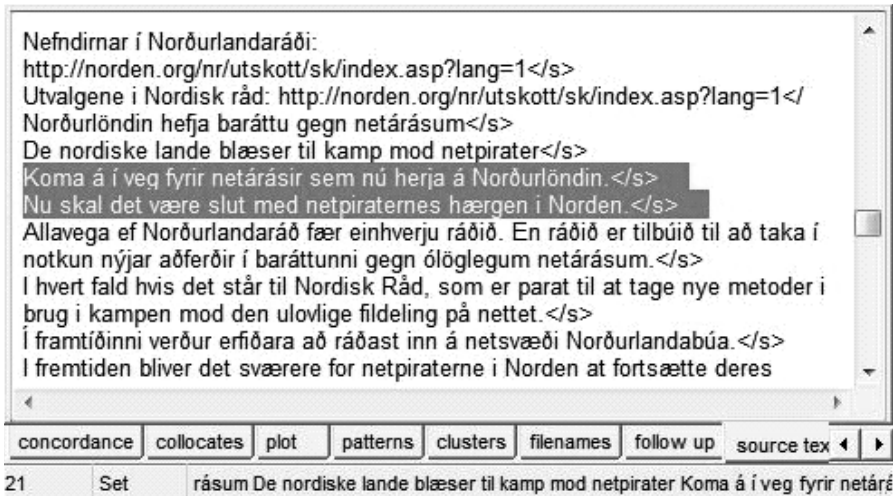
N	Concordance	File	%
1	samstarf til að koma í veg fyrir mansal, fíkniefnasmygl, netglæpi,	2009-09-01.txt	96%
2	samstarf til að koma í veg fyrir mansal, fíkniefnasmygl, netglæpi,	2009-09-01.txt	96%
3	samstarf til að koma í veg fyrir mansal, fíkniefnasmygl, netglæpi,	2009-09-01.txt	96%
4	til að draga úr og koma í veg fyrir olíuleka og afleiðingar þeirra á	Renare_hav.txt	95%
5	en er ekki jafnmikilvægt að koma í veg fyrir minni háttar losun í tengslum við	Renare_hav.txt	92%
6	en er ekki jafnmikilvægt að koma í veg fyrir minni háttar losun í tengslum við	Renare_hav.txt	92%
7	munu löndin einnig koma í veg fyrir stjórnsýsluhindranir í viðskiptum	2009-02-16.txt	90%
8	borde utökas. Vilji menn koma í veg fyrir frekari mengun og eitrun	Renare_hav.txt	77%
9	í sveitarfélögum en að koma í veg fyrir leka frá ökrum? Á det då bättre att	Renare_hav.txt	64%
10	til kamp mod netpirater Koma á í veg fyrir netárásir sem nú herja á	2009-02-02.txt	56%
11	öryggisráðstafanir til að koma í veg fyrir að lagning gasleiðslu frá Rússlandi	2006-09-11.txt	55%
12	öryggisráðstafanir til að koma í veg fyrir að lagning gasleiðslu frá Rússlandi	2006-09-11.txt	55%
13	öryggisráðstafanir til að koma í veg fyrir að lagning gasleiðslu frá Rússlandi	2006-09-11.txt	55%
14	tilraun ESB til þessa til að koma í veg fyrir hnignun vatns og vatnakerfa.	Renare_hav.txt	45%
15	tilraun ESB til þessa til að koma í veg fyrir hnignun vatns og vatnakerfa.	Renare_hav.txt	45%
16	skattaparadísum til að koma í veg fyrir undanskot í hinu alþjóðlega	2009-05-25.txt	42%
17	skattaparadísum til að koma í veg fyrir undanskot í hinu alþjóðlega	2009-05-25.txt	42%
18	skattaparadísum til að koma í veg fyrir undanskot í hinu alþjóðlega	2009-05-25.txt	42%
19	kvinnehandel Hvernig má koma í veg fyrir verslun með konur? Hvordan kan	2006-10-02.txt	34%
20	með það að markmiði að koma í veg fyrir grófa glæpastarfsemi eins og	2009-09-07.txt	31%
21	með það að markmiði að koma í veg fyrir grófa glæpastarfsemi eins og	2009-09-07.txt	31%

concordance collocates plot patterns clusters filenames follow up source text notes

21 Set rásum De nordiske lande blæser til kamp mod netpirater Koma á í veg fyrir netárásir sem nú herja á Norðurlöndin. Nu skal det

Figur 2. Kwic-konkordans med koma í veg fyrir som sökbegrepp

Från konkordansen i figur 2 kan man i sedan i sin tur klicka sig till de länkade parallelltexterna (se figur 3).



Figur 3. Länkade parallelltexter

I dessa länkade parallelltexter kan man så undersöka hur verksamma översättare träffar sina val om översättningsekvivalenter. Dessa val kan sedan jämföras med resultaten av de olika översättningarna som görs av ordboksredaktionerna.

## Resultat: undersökning av isländska flerordsuttryck

Den här studien bygger på en närmare undersökning av isländska flerordsuttryck som bildas med två vanliga isländska verb som bas: *koma* [sv. komma] och *standa* [sv. stå]. Experimentkorpuser innehåller 148 belägg för verbet *koma* i alla dess ordformer, varav 15 fall av flerordsuttrycket *koma í veg fyrir* [sv. förhindra]. *Koma í veg fyrir* är den ekvivalent som Svensk-isländsk ordbok (Holm & Davíðsson 1982), hädanefter SvIO, ger till uppslagsordet *förhindra*; jämför även ISLEX (2011) under uppslagsordet *koma*. Följande språkprov ur parallellkorpuser ger en tydligare bild av grundbetydelsen:

- (1) Är det då bättre att satsa på kommunala reningsverk eller på att *förhindra* urlakning från åkrarna?  
Borgar sig þá frekar að koma upp hreinsunarstöðvum í sveitarfélögunum en að *koma í veg fyrir* leka frá ökrum?

De femton fall av *koma í veg fyrir* som finns i korpusen har också *förhindra* som enskilt vanligaste källspråksord. Andra ord som översätts med *koma í veg fyrir* är *stoppa* exempel (2) och *forebygge* exempel (3):

## (2) stoppa (sv)

*/.../ minska och stoppa oljeutsläppen och deras effekter.*

*/.../ draga úr og koma í veg fyrir olfuleka og afleiðingar þeirra /.../*

## (3) forebygge (no)

*/.../ forebygge menneskehandel, ulovlig handel med narkotika, internetkriminalitet, miljøkatastrofer og pandemier /.../*

*/.../ koma í veg fyrir mansal, fíkniefnasmygl, netglæpi, umhverfisslys og farsóttir /.../*

Bland de övriga källspråksformuleringar som översätts med *koma í veg fyrir* hittade vi uttryck som *bekjempe* exempel (4) och *komme /.../ til livs* exempel (5), vilket visar att alla parallella uttryck i korpusen inte är direkta kandidater till att bli översättningsekvivalenter i ett lexikon, utan snarare får ses som ett försök att skapa dynamisk ekvivalens från översättarens sida:

## (4) bekjempe (no)

*/.../ bekjempe den grove kriminaliteten som menneskehandel er,*

*/.../ koma í veg fyrir grófa glæpastarfsemi eins og mansal,*

## (5) komme /.../ til livs (no)

Hvordan kan man *komme* kvinnehandel *til livs*?

Hvernig má *koma í veg fyrir* verslun með konur?

Det finns även friare formuleringar än dem vi ser i (4) och (5) som leder till *koma í veg fyrir*. Ett intressant sådant exempel är (6), där *oppná fordeler* [rörande] *förhindringer* omtolkas till *koma í veg fyrir /.../ hindranir* med omvänd polaritet:

(6) Ved å samles om ett ernæringsmerke vil alle landene også *oppná fordeler når det gjelder grenseforhindringer* ved handel.

Með því að sameinast um eitt næringarmerki munu löndin einnig *koma í veg fyrir stjórnsýsluhindranir* í viðskiptum milli landanna.

Den vanligaste ekvivalenten till *koma í veg fyrir* är alltså *förhindra*, och om man undersöker detta lemma i samma korpus ser man att det förekommer tio gånger, och översätts fyra gånger med *koma í veg fyrir*, fyra gånger med *stöðva* [sv. stoppa] (7) respektive en gång vardera med *hindra* (8) [sv. hindra] och *sporna við* (9) [sv. spjärna emot].

## (7) stöðva (isl)

Nordiskt projekt för att *förhindra* skatteflykt fortsätter  
Samstarf um leiðir til að *stöðva* skattaundanskot heldur áfram

## (8) hindra (isl)

som huvudsakligen sprids via /.../ båtbottnfärger som *förhindrar* algpåväxt.  
sem dreifist aðallega með /.../ botnbátamálningum sem *hindrar* þörunga-  
vöxt.

## (9) sporna við (isl)

/.../ och hur vi ska *begränsa* och *förhindra* framtida katastrofer.  
/.../ og hvernig hægt er að *sporna við* stórslysum í framtíðinni.

Korpusexemplen har visat att *koma í veg fyrir* är en högfrekvent flerordsenhet, som, om den av redigeringsprinciper inte ges lemmastatus, åtminstone borde ges en framträdande plats i mikrostrukturen. Ekvivalentparet *förhindra* ↔ *koma í veg fyrir* finns med i såväl SvIO som ISLEX, men analys av korpusexemplen visar att det kan vara motiverat att även ge *stoppa* och *förebygga* som ekvivalenter till *koma í veg fyrir*. Det är viktigt att understryka att existensen av parallella uttryck i källspråk och målspråk i korpustexterna inte i sig utgör ett bevis för att visst ord ska få en bestämd översättningsekvivalent. Översättarnas val måste analyseras. Exempel (2) och (3) ovan – med *stoppa* och *förebygga* – visar på översättningar som är tydliga kandidater till att tas med i lexikon, medan (4) och (5) är resultat av arbete med dynamisk ekvivalens och därmed inte relevanta för en ordboksartikel.

Ett annat isländskt högfrekvent flerordsuttryck är *standa fyrir*. Av 88 belägg för verbet *standa* i alla dess former i korpusen gäller 24 olika former av *standa fyrir*. Den svenska grundbetydelsen för *standa fyrir* är 'arrangera', 'stå för' (Jfr ISLEX, 2011 *standa*). SvIO saknar dock *standa fyrir* som ekvivalent till *arrangera*, och *stå för* finns inte med i SvIO, vare sig som eget lemma eller som kollokation under *stå*. Följande korpusexempel visar hur *arrangera/standa fyrir* typiskt används:

## (10) arrangera (sv)

Nordiska ministerrådet och NordGen, /.../, *arrangerar* den 13 maj ett evenemang i FN-byggnaden /.../

Norræna ráðherranefndin og NordGen, /.../, *standa fyrir* viðburði í byggingu Sameinuðu þjóðanna þann 13. maí /.../

En genomgång i parallellkorpusen ger också vid handen att *arrangera* är det källspråksord som oftast ger resultatet *standa fyrir*. Tabell 1 (nedan) visar vilka danska, norska och svenska verbkonstruktioner som har översatts med *standa fyrir*.

Danska källspråksord	Norska källspråksord	Svenska källspråksord
organisere (2 ex)	arrangere (5 ex)	arrangera (6 ex)
arrangere	være arrangør	stå bakom
koodinere (i passiv med <i>blive</i> )	være ansvarlig	genomföra
lancere	byde ... på	ha (2 ex)
stå som vært		

Tabell 1. Verbkonstruktioner som översätts med *standa fyrir*

Tabellen visar det för norska och svenska tydliga mönstret att det är *arrangera/arrangere* som leder fram till ekvivalenten *standa fyrir*. Bland andra ekvivalentpar märks bl.a. följande:

(11) byde ... på (no)

På kvelden *byr* informasjonskontoret *på* konsert /.../

Um kvöldið *stendur* upplýsingaskrifstofan *fyrir* tónleikum /.../

(12) stå som vært (da)

/.../ Nordisk Ministerråds kontorer *står som værter*.

/.../ Norræna ráðherranefndin *stendur fyrir* námstefnunum.

(13) ha (sv)

Nordiska ministerrådet *har* också en rad olika program och aktiviteter som sträcker sig långt utanför de nordiska länderna.

Norræna ráðherranefndin *stendur einnig fyrir* fjölmörgum áætlunum og aðgerðum sem ná langt út fyrir Norðurlönd.

I exempel (11) ser vi att den norska och den isländska satskonstruktionen är identisk, även om det nog kan anses att det är en liten semantisk skillnad mellan *stå för* [isl. *standa fyrir*] och *bjuda på*. Exempel (12) visar hur skillnaden i föredragen aktantuppsättning mellan *stå som vært* och *standa fyrir* leder till att den isländska översättaren ser sig nödd till att explicit införa det implicita led, *seminarierne* [isl. *námstefnunum*], som tidigare förekommit i den danska kontexten. Beträffande exempel (13) är det inte självklart att översättaren menar att *standa fyrir* egentligen ska ses som ekvivalent till *ha*, utan snarare tolkar 'ha' som 'stå bakom', 'stå för'.

När vi nu har vänt på resonemanget och ser på hur *arrangera/arrangere* översätts till isländska framträder ett intressant mönster. Vi kan nämligen se det som Sinclair (2004:142) kallar *colligations*. De isländska ekvivalenterna är här: *standa fyrir* (12 ex.), *halda* (16 ex.), *efna till*, *hafa umsjón*, *skipuleggja*. Mönstret gäller valet mellan *standa fyrir* och *halda*. Det visar sig nämligen att i fallen med

*standa fyrir*, som i (14), så har källspråket i överväldigande antal ett verb i aktiv diates (10 aktiva verb, 2 passiva verb):

(14) *arrangera / standa fyrir*

De nordiske landene *arrangerer* også en rekke nasjonale seminarer under FN's kvinnekommisjons sesjon.

Norrænu ríkin *standa* einnig hvert fyrir sig *fyrir* nokkrum námstefnum á kvennaráðstefnunni.

I frågan om *halda* (15) gäller det omvända (4 aktiva verb, 12 passiva verb):

(15) *arrangera / halda*

Konferansen som heter 'Energidagan', *ble* i år *arrangert* for tredje gang.

Ráðstefnan, sem bar nafnið 'Orkudagurinn', *var* nú *haldin* í þriðja skipti.

Sinclair brukar framhålla att det endast är genom korpusstudier man kan studera *collogations*, d.v.s. samvariation mellan olika grammatiska former hos sökorden och den kontext de förkommer i. Parallellkorpuser blir därmed ovärderliga för att hitta mönster som detta i valet mellan *standa fyrir* och *halda* som ekvivalent till *arrangera*.

Sammanfattningsvis kan vi säga följande om *standa fyrir*: Korpusexemplen visar att *standa fyrir* är en högfrekvent flerordsenhet som, om den inte kan få lemmastatus i kommande ordböcker, åtminstone bör få en framträdande plats i mikrostrukturen. Vi har också sett att *arrangera/arrangere* oftare översätts med isländska *standa fyrir* i aktiv diates, medan översättningen i passiv diates blir *halda*. I ett lexikon från ett skandinaviskt språk till isländska borde nog brukarna upplysas om det.

## Slutord

Huvudsyftet med det här konferensbidraget har varit att visa på möjligheten med (och kanske i någon mån behovet av) att använda parallellkorpuser i tvåspråkig lexikografi. Den korpus som här har använts är naturligtvis för begränsad för fullskalig lexikografisk tillämpning – den borde vara minst 100 gånger så stor och innehålla fler typer av text. Icke desto mindre har exempelstudierna kunnat visa på hittills förbisedda perspektiv på behandlingen av de isländska orden *koma* och *standa* i relation till deras ekvivalenter i de skandinaviska språken.

Principellt har vi sett att ekvivalenspar där flerordsuttryck är inblandade är svåra att undersöka utan en parallellkorpus med möjlighet att upprätta konkordanser av s.k. n-gram. Parallellkorpuser gör det vidare möjligt att analysera de



val som kompetenta översättare har träffat vid arbetet med att finna ekvivalenter i kontext. Vi har hittat ett fall av samvariation mellan verbets diates i källspråket och den ekvivalent som valdes.

Tillgången till parallellkorpusar varierar naturligtvis beroende på vilka språkpar som ska undersökas, men mellan engelska och finska/norska/svenska finns det idag korpusar som används inom språkvetenskap i allmänhet. Det vore en angelägen uppgift att nu ta dem i bruk för lexikografin. Som nämnts ovan pågår arbete med att ta fram en korpus med texter från de fem nordiska språken. När det gäller andra europeiska språkpar har t.ex. Koehn (2005) beskrivit det arbete som pågår med att upprätta en parallellkorpus för elva europeiska språk, inklusive danska, finska och svenska, med material från Europaparlamentet. Det är således rimligt att anta det kan finnas såväl korpusar som parallelltext att tillgå redan nu, och i ännu högre grad om några år. Det som återstår, är att göra dessa resurser tillgängliga, dels genom överenskommelser som gör dem tillgängliga, dels genom att utveckla korpusverktyg som gör det lätt att söka i dem.

## LITTERATUR

- Altenberg, Bengt, Karin Aijmer & Mikael Svensson, 2001: The English-Swedish Parallel Corpus (ESPC) : Manual of enlarged version. [tillgänglig elektroniskt: [http://sprak.gu.se/digitalAssets/1333/1333431\\_manual\\_espc.pdf](http://sprak.gu.se/digitalAssets/1333/1333431_manual_espc.pdf)]. (sept 2011)
- Atkins, B. T. Sue, Adam Kilgarriff & Michael Rundell, 2010: Database of ANalysed Texts of English (DANTE): the NEID database project. I: Dykstra, Anne (ed.), *Proceedings of the Fourteenth EURALEX International Congress*. Leeuwarden. S. 549–556.
- Atkins, B. T. Sue & Michael Rundell, 2008: *The Oxford guide to practical lexicography*, Oxford.
- Borin, Lars & Klas Prütz, 2001: Through a glass darkly: Part of speech distribution in original and translated text. I: Daelemans, Walter, Khalil Sima'an, Jorn Veenstra & Jakub Zavrel (ed.), *Computational linguistics in the Netherlands 2000*. Amsterdam.
- Cheng, Winnie, Chris Greaves & Martin Warren, 2006: From n-gram to skipgram to conngam. I: *International journal of corpus linguistics*, 11. S. 411–433.
- COBUILD = Sinclair, John, 1987: *Collins COBUILD English language dictionary*, London.
- Corréard, Marie-Hélène & Valerie Grundy, 1994: *Oxford-Hachette French dictionary : French-English, English-French*, Oxford; Paris.
- Corréard, Marie-Hélène, Valerie Grundy, Jean-Benoit Ormal-Grenon & Nicholas Rolin, 2007: *Le grand dictionnaire Hachette-Oxford: français-anglais, anglais-français = The Oxford-Hachette French dictionary: French-English, English-French*, Oxford; Paris.

- Friskare nordiska hav, 2010. Nordisk Ministerråd. København. [tillgänglig elektroniskt: <http://www.norden.org/sv/publikationer/publikationer/2010-732>]. (sept 2011)
- Gellerstam, Martin, 1996: Translations as a source for cross-linguistic studies. I: Aijmer, K, B. Altenberg & M. Johansson (ed.), *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*. Lund.
- Hasselgård, Hilde, 2010: Parallel corpora and contrastive studies. I: Xiao, Richard (ed.), *Proceedings of UCCTS 2010; Using Corpora in Contrastive and Translation Studies*. Edge Hill University, Ormskirk.
- Holm, Gösta & Aðalsteinn Davíðsson 1982: *Svensk-isländsk ordbok: Sænsk-íslensk orðabók*, Lund; Reykjavík.
- ISLEX, 2011: ISLEX : en tvärnordisk, flerspråkig onlineordbok med isländska som källspråk och svenska, norskt bokmål, nynorska och danska som målspråk. Reykjavík, Bergen, Göteborg, København. [tillgänglig elektroniskt: <[http://islex.lexis.hi.is/islex/](http://islex.lexis.hi.is/islex/;)>; <<http://islex.se/>>]. (sept 2011)
- Johansson, Stig, Jarle Ebeling & Signe Oksefjell, 1999/2002: *English-Norwegian Parallel Corpus: Manual*. [tillgänglig elektroniskt: [www.hf.uio.no/ilos/english/services/omc/enpc/ENPCmanual.pdf](http://www.hf.uio.no/ilos/english/services/omc/enpc/ENPCmanual.pdf)]. (sept 2011)
- Johansson, Stig & Knut Hofland, 2007: Towards an English-Norwegian parallell corpus. I: Teubert, Wolfgang & Ramesh Krishnamurthy (ed.), *Corpus linguistics: Critical concepts in linguistics*. London ; New York.
- Koehn, Philipp, 2005: *Europarl: a parallel corpus for statistical machine translation*. I: *Proceedings of the tenth Machine Translation Summit*. Phuket, Thailand. [tillgänglig elektroniskt: <http://www.mt-archive.info/MTS-2005-Koehn.pdf>]. (sept 2011)
- McEnery, Tony, Richard Xiao & Yukio Tono, 2006: *Corpus-based language studies : an advanced resource book*, London.
- Rawoens, Gudrun, 2003: *Korpuslingvistik och kontrastiv språkbeskrivning. Ett svensk – nederländskt projekt*. *Tijdschrift voor Skandinavistiek*, 24.
- Scott, Mike, 2008: *Developing WordSmith*. *International Journal of English Studies [Special issue]: Monograph: Software-aided Analysis of Language*, edited by M. Scott, P. Pérez-Paredes & P. Sánchez-Hernández, 8, s.95-106. [tillgänglig elektroniskt: <http://dialnet.unirioja.es/servlet/articulo?codigo=2729155>]. (sept 2011)
- Sinclair, John McHardy, 2004: *Trust the text : language, corpus and discourse*, London.
- SOB = *Språkdata*, Göteborgs universitet, 1986: *Svensk ordbok : [ordens betydelse och användning] : [mer än 100.000 ord och fraser]*, Solna.
- Språkbanken: <http://spraakbanken.gu.se/> (sept 2011)
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*, 2 uppl. Stockholm.

*Håkan Jansson*

Institutionen för svenska språket, Göteborgs universitet.

[hakan.jansson@svenska.gu.se](mailto:hakan.jansson@svenska.gu.se)

# Bruklighetsuppgifter i ordböcker: ett retoriskt perspektiv

*Lars-Erik Johansson*

This study examines the background of the homogenous structure of usage labels in dictionaries. A typological analysis of the rhetorical virtues (*elocutionis virtutes*) and the levels of style (*genera dicendi*) in classical rhetoric theory encompasses prescriptive as well as descriptive elements and shows that this rhetorical system, emanating from Roman rhetoric, is the base for the still used taxonomy of usage labels.

NYCKELORD: ordböcker, bruklighetsuppgifter, retorisk dygdelära, retorisk stillära

Ordböckernas bruklighetsbeteckningar, dvs. uppgifter om användningsområde eller kontextuella uppgifter, exempelvis om stilnivå, fackområde, frekvens, geografiska eller tidsmässiga inskränkningar, bildlig eller överförd betydelse, har gamla anor. De stammar från de århundraden då retoriken i hög grad har styrt produktion av text. Klassisk retorisk teori och praxis har redan under antiken utvecklat klara förhållningssätt till hur ord kan och bör användas, synsätt som relativt oförändrade traderas i ordböckernas bruklighetsmarkeringar. Att det finns en förbindelse mellan klassisk retorik och ordboksutformning är emellertid inte något som har tilldragit sig lexikografers intresse. Således finns i sakordsregistret till det mäktiga internationella trebandsverket *Wörterbücher* (Hausmann 1989-91) termen *Rhetorik* över huvud taget inte förtecknad, trots att lexikografins relationer till flera andra discipliner utförligt berörs. Ändå har retorikstudiet, i huvudsak inom ramen för en omfattande latinläsning, utgjort kärnan i gymnasial och akademisk utbildning allt från renässansen fram till 1800-talet. Latinet har varit *commune artium vehiculum*, lärdomens universella språk, och retoriken den givna disciplinen för ett skickligt handhavande av detta språk.

Denna studie belyser i huvudsak typologiskt – alltså inte i form av en diakronisk framställning – hur bruklighetsbeskrivningen, en viktig del av ordboksartikeln, i detalj återgår på romersk retorisk teori om ordval i språklig produk-

tion, sådan denna har utformats inom ramen för en av de centrala retoriska bearbetningsfaserna (*elocutio*). Studiens centrala parti utgörs av en analys av den romerska teoribildningen. Därefter visas hur ordböckernas system för bruklighetsmarkeringar mycket nära överensstämmer med teorins huvuddrag.

## Retorikens dygdelära

Den romerska retorikens lära om *elocutio* är en teori om språklig-stilistisk formulering. Teorin består av två innehållskomplex; ett behandlar allmänna språkbruksfrågor (innefattande dels en s.k. dygdelära, dels en stilartslära), ett annat specifikt strukturerade formuleringsmodeller. I denna studie berörs det första delkomplexet (framställningar om retorikens teori för *elocutio* är exempelvis Volkmann 1885, Leeman 1963, Lausberg 1990). Hänvisningar till antika källor görs främst till Quintilianus' normerande och syntetiserande handbok för lärare, *Institutio oratoria* (förkortad Q.); i några fall ges hänvisningar till Ciceros *De oratore* och till den anonyma handboken *Rhetorica ad Herennium* (förkortad Her.). Det retoriska systemet är inte enhetligt utan har förändrats över tid och har i viss mån olika utformning hos skilda teoretiker.

Produktionen av språkliga alster innehåller enligt klassisk retorisk teori fem bearbetningsfaser (s.k. *rhetorices partes*): *inventio*, *dispositio*, *elocutio*, *memoria* och *pronuntiatio*. I fasen *elocutio* får det innehåll som har utvecklats inom *inventio* och som har erhållit en viss ordning i *dispositio* sin språkliga gestaltning. Eftersom *elocutio* berör den språkliga formuleringen, har denna del av retoriken nära samband med grammatikämnet. Ett textinnehåll betecknas med termen *res* 'sak, tankar; eg. begrepp' och dettas språkliga utformning med termen *verba* 'ord'; inom *elocutio* kläds alltså ett begreppsligt innehåll i ord. En i huvudsak "konformistisk" grundinställning bildar utgångspunkt. En framställning skall vara kortfattad, klar, korrekt och engagerande för att inte trötta och irritera mottagare. För att uppnå detta kan avsändare byta ut vardagliga och slitna språkliga uttryck mot ovanliga och överraskande. De skall dessutom välja en stil som passar ämne, mottagare och situation.

Målet för *elocutio* är generellt att uppnå träffande formuleringar och idiomatisk korrekthet. Utgångspunkt är läran om talets dygder och fel, dvs. *elocutionis virtutes et vitia*. Den retoriska dygdeläran utgör del av ett större antikt tankekomplex, emanerande från Aristoteles, där dygdebegreppet har såväl filosofiska och etiska som retoriska dimensioner; den retoriska dygdeläran ger en generell etisk orienteringsram för talarens handlande (denne skall vara "vir bonus dicendi peritus", Q. 12,1,1). Dygden är alltså en egenskap hos avsändaren, som vidarebefordrar denna till sitt verk; därmed blir *virtus* också en textegenskap.

Det finns flera olika retoriska dygder, såsom *prudentia* 'vishet', *iustitia* 'rättvisa', *temperantia* 'besinning'.

Den del av det retoriska värderingskomplexet som är knuten till *elocutio* innehåller fyra dygder för språk- och stilkvaliteter: *latinitas* 'renhet', *perspicuitas* 'klarhet, tydlighet, språkriktighet', *ornatus* 'smyckning' och *aptum/decorum* 'anpassning' (jfr Q. 1,5,1). Ibland förekommer även andra termer och indelningar (som *elegantia*, *compositio* och *dignitas* i *Her.* 4,12,17). Detta traditionella system kan utökas med ytterligare dygder, exempelvis *brevitas* 'korthet'. Brott mot dygdeläran kan innebära att oriktiga/olämpliga ord eller ordfogningar används eller att en språklig-stilistisk företeelse har för låg eller för hög frekvens. Sambandet mellan dygder och fel är dock inte logiskt utan är i viss mån asymmetriskt utformat.

Den retoriska dygdeläran har ett allmänt produktionsetetiskt syfte och utgör grunden för den elementära retoriska utbildningen. Teorin berör de generella kommunikationsnormer som reglerar en avsändares språkbruk och har en pragmatisk-normativ inriktning. Den utgör en sammanställning av ett antal preskriptiva, generella och värderande språk- och stilkvaliteter, ofta ordnade i begreppsliga motsatspar. Dygdeläran är således en blandning av språkliga (kodrelaterade), stilistiska/estetiska, kommunikativa och samhällliga (sociokulturella) normer. Denna retoriska lära innebär mindre en teoretisk beskrivning av dygderna (en sådan ges exempelvis av *latinitas* inom grammatikstudiets ram) utan en mera övergripande klassificering av och benämningar på den mångfald av språkliga och stilistiska uttrycksmöjligheter som *elocutio* erbjuder. Inte heller kriterierna bakom de olika normerna diskuteras explicit (annat än i en huvudsakligen grammatiskt inriktad diskussion av vad som betingar *latinitas*). Vad retoriken erbjuder är ett system av från didaktisk och samhälllig synpunkt direkt tillämpbara normer, ett normsystem som huvudsakligen återgår på kanoniska mönster, dvs. författarskap som värderas högt. Dygdeläran hänger nära samman med andra normsystem inom retoriken, framför allt trestilsläran och läran om *copia* 'mångfald'.

Bedömningen av vad som är korrekt och felaktigt språkbruk är inte självklar. En orsak till detta är att de olika dygderna inte sällan kan komma i konflikt med varandra; exempelvis kan en träffande retorisk figur inom *ornatus* innebära ett tillåtet brott mot kravet på språklig korrekthet inom *latinitas*. Ofta ses *aptum* som ett slags "överdygd" vid en sammanvägning av konfliktskapande normideal. Avgörande är i sista hand avsändarens *iudicium* 'omdöme' och anspråk på *ingenium* 'originalitet'. Eftersom retoriken i högre grad ger generella råd än formulerar absolut tvingande regler, blir gränsen mellan *virtus* och *vitium* flytande. Således tillåts skickliga avsändare att göra avsteg från de generellt formulerade dygderna. Sådana avsteg kallas *licentiae* 'friheter' och kan accepteras exempelvis

inom litteratur, särskilt poesi, och i starkt känsloladdade texter. Elevens *vitium* kan vara mästarens *virtus* eller annorlunda uttryckt: det som hos flertalet är ett fel är hos mästaren endast skenbart ett *vitium* utan i stället en *virtus*. Det gäller emellertid att noga känna gränserna inom normsystemet.

De olika dygderna uppfattas och tillämpas under olika tider på olika sätt; ett exempel är den under renässansen laddade striden om ciceronianismen, dvs. om man endast fick anknyta till Ciceros språkbruk eller om man dels eklektiskt kunde välja från flera antika förebilder, dels anpassa sitt språkbruk, bland annat i fråga om terminologi, till sin samtid (med Bembo som en företrädare för den förra uppfattningen och Erasmus för den senare).

## Latinitas

Normen *latinitas* (jfr grek. *hellenismós*), som i detalj behandlas i den tredje delen av romerska grammatikor, styr ett idiomatiskt korrekt uttryckssätt, dvs. språkriktighet i generell mening. I Quintilianus' normerande framställning utgör följande fyra kriterier norm för *latinitas*: *ratio* 'omdöme', *vetustas* 'ålder, tradition', *auctoritas* 'auktoritet' och *consuetudo* '(sed)vana, språkbruk' (Q. 1,6,1). Det viktigaste och slutligt avgörande kriteriet för språkriktigheten är *consuetudo*. Väsentligt är också avsändarens omdöme (*iudicium*) som möjliggör en medelväg mellan logiken i *ratio*, förebildernas *auctoritas* och samtidens språkliga *consuetudo*.

Med kriteriet *ratio* görs en bedömning av språkriktighet utifrån logisk konsekvens. Utgångspunkten för kriteriets tillämpning är dels analogi, dels etymologi (Q. 1,6,1–38). Analogitänkandet å ena sidan utgår från att språk har en logisk uppbyggnad och att man utifrån generella mönster kan dra slutsatser i enskilda fall; för den som är osäker på genustillhörigheten hos ett ord som *funis* 'rep', kan ordet *panis* 'bröd', som är maskulint, vara ett mönster (Q. 1,6,5). Quintilianus påpekar dock – i anslutning till striden mellan analogister och anomalister – att ett analogiskt betraktelsesätt inte får genomföras konsekvent (Q. 1,6,12); den givna kompromissen uppnås genom tillämpning av *consuetudo* (Q. 1,6,16). Etymologin å andra sidan söker ett ursprungligt samband mellan ordform och ordbetydelse och kan exempelvis vara ett hjälpmedel för att skilja barbarismer från korrekta former; Quintilianus är relativt skeptisk till ett etymologiserande synsätt och hänvisar i stället åter till *consuetudo* (Q. 1,6,29 f.).

Även *vetustas* (ibland *antiquitas*) kan användas som kriterium för språkriktighet, något som gör det möjligt att ta ställning till det som brukar kallas "gesunkenes Kulturgut". Avgränsningen mot det tredje kriteriet, *auctoritas*, är dock inte alltid klar. Quintilianus menar att ord som hämtas från gammal tid

(*verba vetera*) kan ge framställningen (stilen) en särskild storhet, inte utan visst behag ("a vestustate repitita .... adferunt orationi maiestatem aliquam non sine delectione"). Men försiktighet är en dygd: sådana ord får inte bli för många eller för iögonenfallande (ty inget är tröttsammare än affektion) och de får absolut inte hämtas från avlägsna och nu glömda tider ("iam oblitteratis repitita temporibus"). De bästa gamla orden är således de nyaste (Q. 1,6,39–41). I det poetiska språket accepteras emellertid arkaismer på ett annat sätt än i prosa (om synen på arkaismer under romersk tid se Lebek 1970).

Det tredje kriteriet, *auctoritas*, utgår från antikens positiva uppskattning av sociala egenskaper som ålder, visdom, levnadserfarenhet och ställning (personlig auktoritet); sådana egenskaper ger normeringskraft (jfr Heinze 1927). Inom *elo-cutio* avses språkbruket hos erkända författare, auktoriteter. Emellertid bör man enligt Quintilianus begränsa sig till författare, vilkas språk är förankrat i samtidens allmänna språkbruk; även detta kriterium modifieras alltså av det samtida bruket. Efter guldåldern börjar dock samtidens språkbruk att uppfattas som lågt stående och som felaktigt och korrupt; då blir en mera rättrogen anslutning till *auctoritas*, sådan denna manifesteras i guldålderns språk och litteratur, en självklar norm.

*Consuetudo* (*usus* eller *usus cotidianus*) är det fjärde och överordnade kriteriet. Det avser samtidens språkbruk. Härmed menas dock inte massornas tal; majoriteten har sällan rätt. Man får aldrig acceptera som eftersträvansvärda språknormer dåliga vanor som tillämpas av många människor. *Consuetudo* är med andra ord elitens språkbruk ("consensum eruditorum", Q. 1,6,43–45). En avvikelse från detta bildade språkbruk är alltid en *licentia*. Cicero hävdar i sin behandling av *consuetudo* det primat som stadens språk (*urbanitas*) har över arkaismer och rusticismer (Cicero i *De oratore* 3,44); rusticismer är snarast sidordnade med barbarismer och solecismer, dvs. *vitia* till *urbanitas* och *latinitas*. Uttalad måttstock för språkbruket skall vara språkbruket hos de bildade i storstaden (Rom); allt som uttrycks väl sägs på ett "urbant" manér ("quidquis bene dicetur et urbane dicti", Q. 6,3,106 f.). Ett urbant språkbruk innehåller inget missljudande, lantligt, opolerat eller främmande i betydelse, ordval eller uttal. Det uttrycks alltså inom romersk retorik en klar uppfattning om den standardiseringsfaktor som språkbruket i en storstad (huvudstad) bör utgöra.

Ett brott som sker mot *latinitas* kallas för enskilda ord *barbarismus* eller *barbarolexis* och för ordfogningar *solecismus*; en genom *licentia* godkänd barbarism benämns *metaplasmus* och en accepterad solecism betecknas *schema (figura)*. Med *barbarolexis* avses bruk av olatinska ord och ordbetydelser. Hit hör bland annat (oassimilerade) lånord (*verba peregrina* i motsättning till *verba nostra*); besläktade med dessa är provinsialismer (*verba gentilia*). De senare fördras i viss mån av Quintilianus, i synnerhet om de kommer från Italien, men mot de förra

är han negativ, förutom mot grekiska ord som ibland måste accepteras i avsaknad av lämpliga inhemska uttryck (en typisk schablon i antika latinska texter är uttrycket "ut vocant Graeci") (Q. 1,5,55–58). Med *barbarismus* i inskränkt mening avses ord som bryter mot latinets fonologi, men termen kan även avse samma begrepp som *barbarolexis* betecknar.

## Perspicuitas

*Perspicuitas* (jfr grek. *saphéneia*) är den andra dygden inom *elocutio* och har som sin förutsättning *latinitas*. Målet för denna dygd är den intellektuella förståelsen; det som formuleras skall kunna förstås av en mottagare, även om denne inte är helt koncentrerad på textförståelsen: det skall i princip vara omöjligt att inte förstå (Q. 8,2,23). Begreppet *perspicuitas* är ett samlingsbegrepp för textelement som gynnar bättre förståelse. Det förenar ett antal olika kategorier: språklogiska (entydighet i ordval och ordfogning), sakorienterade (detaljnoggrannhet) och verkansinriktade (åskådlighet, estetisk effekt, förståelighet).

För att uppnå klarhet skall en avsändare i kraft av sitt *iudicium* bruka korrekta ord (*verba propria*). Det medvetna valet (*voluntas*) av ett träffande ord (*electio verborum*) kräver att avsändaren till sitt förfogande har ett stort ordförråd (*copia verborum*), ur vilket han kan välja det rätta ordet ("quo nihil inveneri possit significantius", Q. 8,2,9); till varje åsyftat innehåll (*res*) hör ett passande *verbum*. Egenskapen hos ett ord att till en bestämd sak vara *verbum proprium* benämns *proprietas* (motsatsen är *improprium*). Som *verba impropria* betraktas ord som inte passar, ord som inte längre är i bruk, nyskapade ord (*verba ficta*), regionala ord ("verba ... regionibus quibusdam magis familiarie", Q. 8,2,13, jfr Q. 11,3,10 om lantligt ordförråd och uttal) och i viss mån fackord (dvs. ord som hänger samman med speciella yrken, "*artium propria*", Q. 8,2,13). Inställningen till fackord är i viss mån ambivalent; samtidigt som teorin betonar vikten av ett exakt ordval, något som i princip skulle kunna gynna facktermer, är den normalt negativ mot sådana ord, inte minst av det skälet att fackorden genom sin sociolektiska karaktär (många fackord är knutna till de agrara näringarna) dels lätt blir svårförståeliga, dels därtill ofta kan ha en olämplig lantlig eller dialektal prägel. Varningar ges även för anstötliga uttryckssätt; då är omskrivningar att föredra (se exempelvis *Her.* 4,45). En konsekvens av målet att eftersträva det träffande ordet (och undvika det felaktiga) blir att det är viktigt att ha god kunskap om synonymer. Brott mot *perspicuitas* benämns *obscuritas* 'dunkelhet'; sådan dunkelhet kan i vissa fall emellertid betraktas som *licentia*.



## Ornatus

*Ornatus* (grek. *kósmos*, *kataskeué*) anses normalt vara den mest eftersträvade, glansfulla och effektiva dygden och avser god utformning, god ordning och konstfull gestaltning av texter i fråga om ordval, satsbyggnad, klang och rytm. Målet är att hos mottagarna förmedla en positiv känsla, ett behag (*dilectatio*), som gör att syftet med framställningen stärks och publiken inte känner någon leda (*taedium*). Ett smyckat tal lyssnar mottagare gärna på ("libenter audiunt") med stor uppmärksamhet ("magis attendunt"), de förstår innehållet lättare ("facilius credunt") och låter sig ryckas med (Q. 8,3,5). Den språkliga prydnaden har ytterligare kvaliteter; Quintilianus nämner egenskaper som skarpsinnighet, strålkraft, rikedom, munterhet, ljuvhet och noggrannhet ("Nam primum acuto, secundum nitido, tertium copioso, deinceps hilari iucundo accurato diversum est", Q. 8,3,49).

Först och främst måste talaren välja det riktiga ordet bland ett stort antal synonyma uttryck. Flera ord kan ha samma betydelse men kan till karaktär (stilvärde) ändå vara olika; somliga är i högre grad högtidliga, upphöjda, glänsande, behagliga eller välljudande än andra ("honestiora, sublimiora, nitidiora, iucundiora, vocaliora", Q. 8,3,16). Man kan ibland genom att tillämpa ornatusdygden överskrida de stränga gränser för språklig korrekthet och förståelse som *proprietas* drar upp utan att förfalla till *vitium* eller *improprietas*; denna dygd är nämligen i viss mån relativ och kan tolkas på ett vidare sätt än som knuten till enbart ord (Q. 8,2,6). Med stöd i ett sådant synsätt kan skickliga avsändare använda sig av arkaismer, neologismer och troper (*verba translata*). De senare är ett sätt att säga något genom att avsändaren använder ett annat slags utsaga än den direkta; man använder alltså ett annat ord eller uttryck än det normalt brukliga. Tropen ger således ord en ny och för kontexten speciell innebörd. Troper kan vara alltifrån unika nyskapelser till stelnade uttryck i umgängesspråket. Till troper räknas *metaphora*, *metonymia*, *synechdoche*, *emphasis*, *hyperbole*, *antonomasia*, *ironia*, *litotes* och *periphrasis*.

## Aptum eller decorum

Dygden *aptum* (Quintilianus' term) eller *decorum* (Ciceros term) (grek. *prépon*), ett under antiken relativt oskarpt begrepp med både etiska, estetiska och retoriska dimensioner, speglar generellt förmågan att få olika delar att förena sig till ett harmoniskt helt. I princip hör alltså *aptum* inte endast till *elocutio* utan till samtliga produktionsfaser. En konsekvens blir att dygden relaterar till såväl inre aspekter hos texten och textframställningen (s.k. inre *aptum*) som sociala och samhälleliga förhållanden (s.k. yttre *aptum*); härigenom introduceras i retori-

kens produktionsetestetiska system den för all normering centrala dualismen mellan lingvistisk beskrivning och sociala värderingar (jfr Q. 11,1,1–93). Ett typiskt tillämpningsområde är retorikens stillära (*genera dicendi*), där valet av stil betingas av såväl språkliga som samhälleliga aspekter (jfr nedan).

Tillämpning av *aptum* inom *elocutio* berör självfallet samma områden som behandlas under dygderna *latinitas*, *perspicuitas* och *ornatus*. Generellt kan man säga att läran om *aptum* ger en bedömningsgrund för hur en avsändare skall ställa sig till sitt val av exempelvis sociolekter och dialekter, skrift- och litteraturspråk och olika diafasiska varieteter. Grundläggande är att avsändaren har ett överflöd av ord att välja mellan och att han väl känner skillnaden mellan olika synonymer, så att han alltid kan välja det exakt riktiga ordet. Avsändaren har därvid i viss mån möjlighet att nybilda ord (neologismer) och att medvetet använda gamla ord (arkaismer), troper, lånord, regionalt färgade ord och facktermer. Dessa möjligheter måste dock brukas med yttersta försiktighet; den självklara utgångspunkten är att bruket av sådana ordkategorier bör betraktas som *vitia*. Språkligt-stilistiska brott mot *aptum*, beroende på olämpliga förbindelser dels mellan *res* och *verba*, dels mellan *verba* och *verba*, betecknas med termer som *ineptum* eller *indecorum*. Generellt gäller att man inom poesi tillämpar en mildare bedömning av vad som skall anses passande än inom prosa.

## Stilartsläran

Den arsenal av språkliga verktyg som teorierna i *elocutio* lyfter fram får emellertid inte användas urskillningslöst. Enligt läran om *aptum* måste varje framställning utformas i en viss stilart. Regler för detta ges inom den s.k. trestilsläran (*dicendi genera*, *elocutionis genera*, gt. *charaktéres* resp. *plásmata tés léxeos*). Trestilslärebegreppet har i den teoretiska litteraturen knappast erhållit någon pregnant bestämning utan präglas av fyra sammanblandade kriterieområden i anknytning till företeelser som *aptum*, *ornatus*, genre och en hierarkiskt utformad samhällsstruktur. Stilartsläran, som i mer eller mindre utvecklad form kommer leva kvar långt in på 1900-talet, styr ytterst valet av språklig ”hyperstruktur”, dvs. val av stilnivå.

*Genus grande*, den upphöjda stilen, kännetecknas av upphöjda och vackra ord i en jämnt flytande ström; den är svårbehärskad men verksam. *Genus medium*, mellanstilen, består av ord av en lägre, dock inte lägsta eller allra mest vardagliga dignitet. *Genus subtile*, den enkla stilen, utmärks av att den sänker sig till den alldagligaste varianten av naturligt tal; den är fattig på utsmyckningar, saknar i huvudsak figurer och har en enkel syntax (jfr *Her.* 4,8,11).

Trestilsläran utgör inget tvingande schema som detaljreglerar en avsändares stilistiska val. I stället visar läran i grova drag på ett antal stilistiska möjligheter. Inom varje stilart finns en betydande bredd och olika stilarter kan blandas med varandra (Q. 12,10,67).

## Dygde- och stilartsläran som mönsterbildande system

Dygde- och stilartsläran är alltså central i retorikens teori för textproduktion och textanalys. Karakteristisk är dess blandning av deskriptiva och framför allt preskriptiva inslag. Det har varit viktigt att känna till och acceptera grunddragen i regelsystemet, eftersom detta från Quintilianus' tid och framåt i stor utsträckning har haft en normerande innebörd. Den romerska retoriken har tidigt skapat ett avancerat system för såväl språklig analys som textplanering och språkvård. Systemet har innefattat medvetenhet om dels ett slags språklig elitistisk standard, företrädesvis omnämnd med varietetsbeskrivande termer som *sermo latinus*, *sermo urbanus* och *consuetudo*, existerande som normideal liksom som iakttagbar bruksvarietet, med en viss bredd och i viss mån präglad av sin användning i rum, tid och register, dels ett större antal varieteter som av olika skäl inte har kunnat jämföras med den prestigefyllda varieteteten (betecknade med termer som *sermo rusticus* och *sermo vulgaris*).

Denna teoribildning har, inom ramen för en samlad latinsk varietetsarkitektur, innehållit ett urskiljande av och en värdering av diatopiska, diakroniska och diafasiska varieteter med klart iakttagbara ”nivågrupperingar”. Bland annat har teorin omfattat uppfattningen om att tal och skrift är skilda från varandra och produceras under olika betingelser, att ordförrådet kan delas in i å ena sidan en ”allmän” del och å andra sidan främmande ord och lånord, fackord och socialt markerade ord, att det litterära (poetiska) språket ger språkbrukare större friheter, att språkbrukare genom att använda troper, dvs. bruka ett annat slags ord än vad de vet är det normalt brukliga, kan markera sin attityd till det som behandlas (genom exempelvis ironi, skämt, överdrift och metaforiskt eller metonymiskt ordbruk) och att frekvenser i språkanvändningen har betydelse för värdering av gott språkbruk. Systemet har därtill bidragit med ett stort antal praktiska termer för att beteckna oriktigt eller oönskat språkbruk (*vitia*).

I första hand har detta system tagits i bruk för att det har utgjort ett genomtänkt ramverk för en taxonomisk gruppering av språkbruk, för klassifikation av ord och texter, med tydliga begreppskategorier och i viss mån även en adekvat terminologi; det är denna grundläggande retoriska taxonomi som ligger till grund för ordböckernas bruklighetsmarkeringar (liksom för terminologin i äldre grammatikor). Systemet har under lång tid, i princip in på 1900-talet,

fungerat som en i det närmaste obligatorisk taxonomi, oberoende av den språksyn som systemanvändarna har haft. Exempelvis urskiljs inom systemet kategorier som arkaismer och neologismer med hjälp av termer som kan upplevas som användbara och neutrala, oberoende av inställning till de begrepp som termerna speglar.

Systemet har emellertid även erbjudit en befintlig och beprövad om än elitistiskt grundad teori för språknormering och språkvård, en både realistisk och nyanserad grund för bedömning av vad som är kommunikativt, språkligt, stilistiskt och estetiskt korrekt och av vad som är socialt lämpligt. Det har således kommit att ligga till grund för renässansens återupplivande av guldålderslatinets med Cicero som huvudmönster och med Quintilianus' handbok som främsta rättesnöre. Det har dessutom varit bekvämt att tillämpa i de folkspråkliga skriftdiskurser som successivt har vuxit fram från tidigmodern tid; kort sagt, språkligt intresserade individer har kunnat känna igen sig och har på sitt eget språk och genom etablering av en inhemska auktoritativ kanon kunnat tillämpa en elaborerad och välbeprövad taxonomi, i all synnerhet som målet för en utveckling av det egna folkspråket normalt har varit att om möjligt stå på de gamla mästarnas skuldror och från denna position efterbilda guldålderns skimrande och lockande latin: *scire latine* har varit en förutsättning för en önskvärd uppodling av folkspråket.

Under århundradena fram till nutiden har detta retoriska analys- och produktionssystem givetvis kommit att tolkas på olika sätt, ibland troget retorikens regler, ibland i medveten opposition mot dessa. Senare tiders tolkningar av systemet har dessutom medfört olika resultat (jfr barockens och klassicismens helt olikartade texter trots återopande av samma regelsystem).

## Retorikens dygde- och stilartslära och ordböckernas bruklighetsuppgifter

Följande kortfattade genomgång visar hur retorikens begreppsbildning (och i viss mån terminologi) har sin motsvarighet i ordböckernas bruklighetsmarkeringar. För överskådlighetens skull relateras den klassiska retorikens begreppsbyggande till ett modernt begreppssystem, utgående från olika typvarieteter (diatopiska, diastratiska, diafasiska etc.), i form av termer som innehåller *dia-* som prefix. Exempel på sådana taxonomier finns hos Hausmann 1977: 112–143 om diasystematiska bruklighetsbeteckningar, hos Hartmann & James 1998: 151 om "usage labels" och hos Svensén 2004: 379 och framför allt 390–397 om diasystematisk markering (jfr Ludwig 1991 om en kommunikativ "Prädispositionsmodell" och Wiegand 1979 om pragmatisk information). Sådana taxono-

mier präglas givetvis av att gränserna mellan olika grupper måste bli oskarpa; en bruklighetsangivelse kan klassificeras på olika sätt och i sig förena två eller flera olika perspektiv (exempelvis 'äldre dialektbruk' som innefattar såväl ett diatopiskt som diakroniskt perspektiv). I det följande anges några typiska bruklighetsbeteckningar (de först angivna exemplen kommer från äldre svenska ordböcker).

Romersk retorik skiljer mellan gamla ord och nya, dvs. ord som skall undvikas eller användas med måtta; ord som kommit helt ur bruk skall aldrig användas. Det är således viktigt att känna till vilka ord som betraktas som arkaismer (paleologismer) och neologismer (motsv). Typiska *diakroniska* bruklighetsbeteckningar är *obsoletum*, *antiquum*, *ålderdomligt*, *arkaiserande*, *obrukligt*; *nyord*, *nutida språkbruk*, *modernt*, *dagligt*.

Retoriken gör vidare en distinktion mellan stadens och landsbygdens språk. Storstadens urbana språkbruk framställs i retorisk teori som mönstergillt, medan lantliga ("rustika") uttryckssätt bör undvikas. Karakteristiska *diatopiska* ordbokstermer är *dialektalt*, *regionalt*, *provinsiellt*, *i vissa trakter*, *bygdemålsfärgat*, eventuellt förknippat med en närmare geografisk bestämning till region (*sydsvenskt*, *finlandssvenskt*) eller till landskap. En tredje distinktion är den mellan inhemska och främmande ord; den retoriska grundregeln är att främmande ord bör undvikas, om de inte är nödvändiga, exempelvis då det egna språket saknar adekvat terminologi. Brukliga *diaintegrativa* bruklighetsbeteckningar är *vox peregrina*, *lånord* eller – vilket är vanligast – helt enkelt en uppgift om långivande språk (av eng., av ty. etc.).

En fjärde ordboksdistinktion, benämnd *diamedial*, görs mellan talspråk och skriftspråk; i retoriken understryks att det inte får låta som om en talare läser ur ett manuskript. Nära samband med denna distinktion har den *diafasiska* dimensionen, som anger hur olika typer av kommunikationssituationer kan anses befinna sig på en skala mellan formellt och informellt. Ofta blandas *diamediala*, *diafasiska* och i viss mån även *diaevaluativa* bruklighetsuppgifter. Vanliga beteckningar är *i talspråk* (inte sällan försett med inskränkningar som *dialektalt*, *regionalt*, *okonstlat*, *lägre*, *bildat*), *talspråklig stil*; *i skriftspråk* (ofta med frekvensangivelser som *vanligt*, *mindre brukligt*), *skriftspråklig konstruktion*, *prägel* eller *stil*, *informell*.

Ytterligare en retorisk distinktion görs mellan olika sociala skiktningar i språkbruket. Typiska *diastatiska* beteckningar är *vulgo.*, *vox plebeja*, *bildat*, *familjärt*, *gement*, *grövre*, *populärt*, *vulgärt*, *finare*, *lägre umgängesspråk*, *det lägre språket*, *kan väcka anstöt*. Retorikens bedömning att det litterära språket ger stor frihet återspeglas i ordböckernas *diatextuella* beteckningar, som hänför sig till texttyper och genrer och som kan exemplifieras med (*skön*)*litterärt*, *poetiskt*, *vit-ter stil*, *litterärt färgat skriftspråk*, *bibliskt*.

Retorikens ambivalenta inställning till fackord återfinns i de *diatekniska* markeringarna. Hit hör en beteckning som *fackspråk(lig)* men även flera beteckningar som relateras till vetenskapsområden och yrken: *tekn., vetenskapl., bot., glasm, gulds*. Retorikens synpunkter på att ord kan användas alltför sällan eller ofta speglas i de *diafrekventa* beteckningarna, som täcker en skala från vanligt till ovanligt: *mycket ofta, ganska ofta, ofta, numera ganska ofta, sällan, sällsynt, brukas mindre*. Någon gång anges ett ”neutralt mellanläge” *allmänt*.

Retorikens bruk av troper för att överföra bland annat olika typer av känslor, sinnesstämningar och attityder motsvaras av de *diaevaluativa* markeringarna, som innehåller en stor grupp termer, exempelvis *eufemistisk, pejorativ, ironisk, skämtsam, skeptisk, föraktlig, gäckande, försmädlig, föraktfull, nedsettande, förtrolig, förklenlig*. Vissa typer av troper, framför allt metaforer och metonymier, återfinns dock inte bland bruklighetsmarkeringarna utan utgör i stället del av ordböckernas betydelseanalys.

*Dianormativa* bruklighetsangivelser slutligen, som på ett övergripande sätt ansluter till dygdeläran, ger normativa bedömningar av språkbruk. Typiska termer är *ej korrekt, orätt för, i mindre övertänkt språkbruk, felaktig användning, stridande mot gott svenskt språkbruk*.

## Slutreflexion

Det seglivade systemet med bruklighetsbeteckningar, emanerande från en i grunden preskriptiv romersk retorisk tradition, används alltså fortfarande efter 2 000 år, trots att synen på varieteter och stilmönster i hög grad har förändrats, exempelvis genom att i dagens standardvarietet diatopiska och diasociala varieteter alltmer blir stilistiska markörer i en bred registeranvändning, gränsen mellan tal- och skriftspråk blir oskarp, stilblandning blir allt vanligare och det poetiska språket blir allt mindre litterärt särpräglat. Kanske är tiden nu inne för lexicografen att utveckla en ny taxonomi för bruklighetsbeteckningar?

## LITTERATUR

Cicero. De oratore.

Her. = Rhetorica ad Herennium.

Q. = Quintilianus. Institutio oratoria.

Hartmann, Reinhard R. K. & James, Gregory, 1998. Dictionary of lexicography. London.

Hausmann, Franz Josef, 1977. Einführung in die Benutzung der neufranzösischen Wörterbücher. Tübingen.

- Hausmann, Franz Josef (red.), 1989-1991. Wörterbücher. Berlin.
- Heinze, R., 1925. Auctoritas. I: Hermes 60. S. 348-366.
- Lausberg, Heinrich, 1990. Handbuch der literarischen Rhetorik. 3. Aufl. Stuttgart.
- Lebek, Wolfgang Dieter, 1970. Verba prisca. Göttingen.
- Leeman, A. D., 1963. Orationis ratio. Amsterdam.
- Ludwig, Klaus-Dieter, 1991. Markierungen im allgemeinen einsprachigen Wörterbuch des Deutschen: ein Beitrag zur Metalexikographie. Tübingen.
- Svensén, Bo, 2004. Handbok i lexikografi: ordböcker och ordboksarbete i teori och praktik. 2. uppl. Stockholm.
- Volkman, Richard, 1885. Die Rhetorik der Griechen und Römer. 2. Aufl. Leipzig.
- Wiegand, Herbert Ernst, 1979. Pragmatische Informationen in neuhochdeutschen Wörterbüchern. I: Germanistische Linguistik. S. 139-271.

*Lars-Erik Johansson*

Kommunikation och information (nordiska språk), Högskolan i Skövde  
lars-erik.johansson@his.se

# ISLEX – en flersproget nordisk ordbog

*Halldóra Jónsdóttir & Þórdís Úlfarsdóttir*

The Icelandic-Scandinavian dictionary project ISLEX has been going on since 2006 and is now in its final stages. The outcome of the project is twofold. Firstly, a long-awaited Icelandic-Scandinavian dictionary has been compiled, specially designed for the web and its features (e.g. sound, images, links). Secondly, the project has provided valuable experience in building up such work jointly in four countries, as the Icelandic source language is processed in Reykjavík and the target languages in Copenhagen, Bergen and Göteborg, respectively.

The ISLEX dictionary describes modern Icelandic. It contains a range of collocations, idioms and examples of use, all of which are translated into the target languages: Danish, Norwegian (Bokmål and Nynorsk) and Swedish. ISLEX is thus a multilingual dictionary within which are contained four bilingual dictionaries. Moreover, the Faroese joined the editorial team at a later stage, so Faroese will become the fifth target language.

The user interface offers search in both the source and target languages, as well as direct comparison between the target languages' equivalents. The project has possibilities for research that could benefit Nordic linguistic collaboration, and its material is also a good base for inter-Nordic projects in education and language technology.

*Nøgleord:* Flersproget, Elektronisk ordbog, Islandsk, Skandinavisk

## 1. Indledning

Islex er en fællesnordisk, elektronisk ordbog udarbejdet af *Árni Magnússon-instituttet for islandske studier* (SÁ) i samarbejde med fire andre universitets- og forskningsinstitutioner i Norden, i Danmark, Norge, Sverige og på Færøerne. Samarbejdspartnerne er *Det Danske Sprog- og Litteraturselskab* (DSL) i København, *Institutt for lingvistiske, litterære og estetiske studier* (LLE) ved universitetet i Bergen, *Institutionen för svenska språket* ved universitetet i Göteborg og *Fróðskaparsetur Føroya* i Tórshavn (fra 2011).



Islex har islandsk som kildesprog. Den indeholder ca. 50.000 islandske opslagsord med oversættelser på dansk, norsk (både bokmål og nynorsk), svensk og færøsk. Islex er den eneste nordiske ordbog hidtil der forener så mange sprog i ét værk. Ordbogen omfatter først og fremmest moderne islandsk, og opslagsord, sprogbrugseksempler og ordforbindelsere tager udgangspunkt i samtidens sprog. Værket er tænkt som et svar ved det store behov for handy ordbøger mellem islandsk og de øvrige nordiske sprog, da de fleste af de eksisterende ordbøger er blevet forældede og ikke længere lever op til tidens krav. Samtidig er formålet med projektet at udarbejde et mangfoldigt, moderne og gedigent ordbogsværk, der kan være til gavn for flest mulige brugere, islændinge såvel som udlændinge, studerende, lærere, oversættere og alle andre som behøver sproglige oplysninger. Tidsrammen for projektet er perioden 2006-2011.

## 2. Fremstilling og arbejdsmetoder

Islex-ordbogen var fra begyndelsen tænkt som en webordbog da ved projektets start var der nye tider indenfor islandsk ordbogsproduktion. Udviklingen indenfor informationsteknologi var hastig, ikke mindst webbens opståen og dens utallige muligheder. På kun få år var den almene befolkning blevet vant til umiddelbart at kunne finde frem til oplysninger på nettet. Det var ikke længere nok at ordbøgerne var 'elektroniske', de skulle også være tilgængelige på webben og helst billige eller helt gratis.

Islex har af den grund måttet svare tidens krav. Dette er i sig selv ikke helt enkelt, da ordbogsarbejdet i sin natur er en langsommelig process, hvad enten fremstillingen er elektronisk eller ej, foruden at der er fare for at ordbogens tekniske design hurtigt bliver forældet og utidssvarende. Ved arbejdet med Islex har man forsøgt at imødekomme disse behov ved stadig at forny og forbedre databasen og brugergrænsefladen, alt afhængigt af omstændighederne og projektets økonomi, da under de bedst mulige forhold må elektroniske ordbøger altid være under teknisk revision (bortset fra opdatering af deres indhold).

Islex-projektet opererer i en specialdesignet webbaseret database, hvis system opereres af SÁ.

Fire til fem redaktioner i lige så mange lande arbejder på Islex-projektet. Hver gruppe er en selvstændig arbejdsenhed der varetager og er ansvarlig for sit sprog i ordbogen. Ifølge sagens natur er den islandske redaktions arbejde vidt forskelligt fra de øvrige skandinaviske redaktioners, hvor deres hovedopgave først og fremmest er at oversætte det islandske materiale til målsprogene. Den islandske redaktion har defineret det islandske ordforråd og redigerer det inden

målsprogsredaktørerne tager fat. Arbejdet har altid foregået samtidig i alle landene, med den undtagelse at Færøerne først kom med i 2011.

Hver institution har copyright til det stof som dens medarbejdere indsætter i databasen. Således har SÁ retten til hele det islandske indhold, DSL har retten til det danske materiale o.s.v. Projektet repræsenteres af et særligt samarbejdsudvalg fra alle medlemslandene, som varetager den formelle kontakt mellem institutionerne.

## 2.1 Arbejdsmetoder i Island

Bearbejdningen af den islandske del af Islex-webordbogen indebærer forskellige komponenter. I de tilfælde som dette er relevant inddeles ordene efter betydning, der udarbejdes en ordforklaring på islandsk for hver betydning, der indføres sprogbrugseksempler efter behov, faste udtryk bliver standardiseret og de får en forklaring. Der tilføjes faglige etiketter og henvisninger hvis dette anses for nødvendigt, og nogle steder tilføjes der en illustration.

I redigeringsarbejdet har man ikke fulgt alfabetet, men i de fleste tilfælde arbejdet efter de såkaldte semantiske felter. De tre største ordklasser, substantiver, adjektiver og verber, er næsten udelukkende blevet bearbejdet på den måde og desuden en stor del af adverbierne. På den måde bliver f.eks. alle ord tilhørende de semantiske felter *dyr* eller *jura* bearbejdet som en helhed. Med denne arbejdsmetode opnås der en bedre overensstemmelse mellem ordbogsartiklerne. Mindre ordklasser, pronominer, præpositioner, konjunktioner og udråbsord er dog for det meste udarbejdet som en helhed.

## 2.2 Arbejdsmetoder i Skandinavien

Der er fire redaktioner der arbejder på målsprogene i Islex. Redaktørernes opgave består hovedsagelig i at oversætte det islandske materiale (opslagsord, sprogbrugseksempler og ordforbindelser) til deres eget sprog, og tilføje yderligere oplysninger eller forklaringer efter behov. I Bergen står den samme redaktør for oversættelsen til de to norske sprog. Ligesom deres kolleger i den islandske redaktion har de nordiske redaktører arbejdet efter den semantiske inddeling af ordforrådet ved deres oversættelser. Med denne arbejdsmetode får oversætteren et bedre overblik over udvalget af de synonymmer som han/hun kan vælge imellem.

Oversætterne har ligeledes taget en aktiv del i opbygningen af Islex. Der er tale om en bestemt form for kvalitetskontrol, der uden tvivl har medført et

bedre ordbogsværk. Oversætterne har bl.a. tit foreslået nye opslagsord, som i de fleste tilfælde er blevet tilføjet ordbogen; de har efterlyst sprogbrugseksempler til bestemte ord og foreslået ændringer på eksempler som de har anset for at være ufuldstændige. Desuden har de undervejs påpeget indholdsmæssige mangler og systematiske fejl, både hos hinanden såvel som i den islandske del.

### 2.3 Kommunikation

Som før nævnt arbejdes der på Islex-ordbogen i fem lande samtidig. Det er derfor meget vigtigt for et projekt af denne type at man kan kommunikere på en hurtig og bekvem måde. Foruden daglig elektronisk kommunikation er der årligt blevet afholdt to arbejds møder. Møderne har været gavnlige for formidling af oplysninger, diskussioner og planlægning og deltagerne er enige om at kommunikation gennem computere ikke helt kan erstatte personlige samtaler.

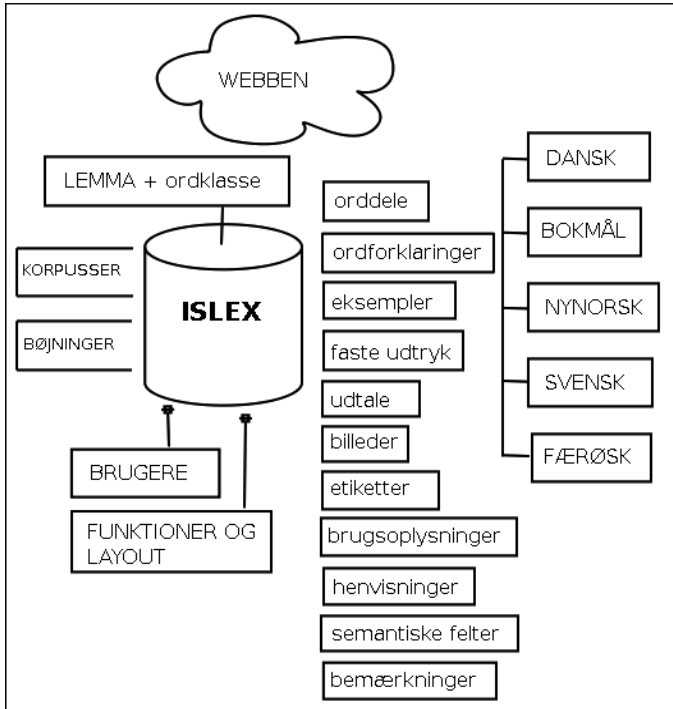
## 3. Materialet i Islex

### 3.1 Antal og omfang af dataenheder

I Islex-databasen er der forskellige data som alle er nøjagtig defineret. Billede 1 viser de vigtigste komponenter som indgår i projektet.

Islex er en mellemstor ordbog hvis man ser på antallet af opslagsord, med ca. 48.000 lemmer. Ordbogsartiklerne indeholder store mængder oplysninger, d.v.s. det materiale som fremstår under hvert enkelt lemma. Materialet bygger på systematisk inddelte oplysninger, i den islandske del såvel som på målsprogene, sprogbrugseksempler er f.eks. en særskilt kategori og henvisninger en anden. Enhederne har hver sin betegnelse alt afhængigt af den rolle som de indtager i ordbogen. I alt er der defineret 80 typer data i ordbogen, deraf er omkring halvdelen oplysninger om islandsk foruden systemoplysninger, men den anden halvdel omfatter forskellige data der angår de fem målsprog. Her er oversættelserne selvfølgelig de vigtigste.

Følgende tabel viser antallet dataenheder i ordbogens islandske del samt systemoplysninger. Det drejer sig om runde tal og listen er ikke tømmende, f.eks. nævnes der ikke redaktionelle bemærkninger og flere af den slags som ikke kan falde under det vigtigste indhold.



Billede 1. Islex-ordbogens opbygning.

*Databasen er betegnet med den store tønde. Over den er der en kasse med opslagsord og ordklasse, som er databasens kernepunkt. Til venstre er datakilder: korpusser og bøjningsoplysninger. Til højre ser man kasser med de vigtigste dataenheder i Islex, og længst til højre er målsprogene, dansk, bokmål, nynorsk, svensk og færøsk. Under ISLEX-tønden er der tabeller der gemmer oplysninger om brugere, samt indstillinger for dataenhedernes funktion og udseende.*

kategori	antal	forklaring
BETYDNING	13.860	betydningsnummer
BILLEDE	3.260	billede
BRUG	1.730	brugsbestemmelse
BØJNING	42.230	link titl BÍN
EKSEMPEL	31.050	sprogbrugseksempel
ETIKETTE	2.450	markering af fagområde
FAST UDTRYK	4.370	ordforbindelse
FORKLARING	58.310	islandsk ordforklaring

HENVISNING	4.820	intern henvisning
KASUSSTYRNING	3.080	kasusstyrning i islandsk
ORDDELE	34.790	orddele i sammensætninger
ORDFORBINDELSE	5.310	kollokation/ordforbindelse
SEMANTIK	50.840	semantisk felt
STADIUM	48.120	ordbogsartiklens stadium
SUBJEKT	270	subjektets kasus
VERBOVERSKRIFT	850	verbets undertitel
VERBTYPE	4.409	verbets syntaktiske klasse
VERBUDTRYK	4.280	kollokation med verbum

Tabel 1. *Datatyper for islandsk samt systemoplysninger.*

### 3.2 Antal ord efter ordklasser

Antal opslagsord i Islex er som før sagt ca. 48.000. Som i traditionelle ordbøger er substantiverne flest, ca. 68 % af ordforrådet, adjektiverne indtager ca. 17 %, verberne tæller omkring 9 %, adverbier ca. 4 % og ord af andre ordklasser er meget færre. Ordforrådet fordeles på ordklasserne på følgende måde (tabel 2):

Ordklasse	Antal ord	% af ordforrådet
Substantiver	32.826	68%
Adjektiver	8.353	17%
Verber	4.230	9%
Adverbier	1.814	4%
Talord	71	< 1%
Pronominer	125	< 1%
Præpositioner	233	< 1%
Konjunktioner	93	< 1%
Præfikser	136	< 1%
Udråbsord	62	< 1%
Forkortelser	162	< 1%
<b>I alt</b>	<b>48.100</b>	<b>100%</b>

Tabel 2. *Inndeling af ordforrådet i Islex efter ordklasser*

Antal ord i hver ordklasse fortæller dog ikke hele historien om omfanget af ordklassens data. Selv om navneordene er langt flere end verberne så er antallet af sprogbrugseksempler med verberne større (med verberne følger der 10.000 eksempler imod 9.000 med navneordene). Grunden til dette er at beskrivelsen af verber kalder på eksempler i højere grad end i navneordene.

### 3.3 Ordforbindelser

Ordforbindelser er en vigtig del af Islex-ordbogen og der lægges større vægt på at beskrive deres betydning og brug end man tidligere har kendt til i islandske tosprogordsbøger. Grunden dertil er især at ordforbindelser er en stor del af sproget og mange faste udtryk kan være ugenomsigtige og svære at forstå for udenlandske brugere. I Islex bliver alle ordforbindelser oversat til målsprogene, i nogle tilfælde med en tilsvarende ordforbindelse på et andet sprog.

Begrebet *ordforbindelse* har en bred definition, d.v.s. en fast forbindelse mellem to eller flere ord. I Islex gøres der forskel på tre typer ordforbindelser som har arbejdstitlerne *osamb*, *ostæða* og *sostæða*. *Osamb* defineres som idiom eller en metaforisk forbindelse, hvor de enkelte ord ofte har lidt eller intet at gøre med ordforbindelsens betydning.

*Ostæða* står for *orðastæða* 'kollokation'. Her forekommer opslagsordet sammen med faste følgeord, et eller flere (f.eks. *pommes frites, børn og unge*).

*Sostæða* er arbejdstitlen på *sagnorðastæða* og svarer til andre kollokationer på nær at her er verbet i hovedrollen (f.eks. *begå et mord, fjerne ukrudt*).

Antallet af idiomatiske ordforbindelser i Islex er ca. 4.400, der er ca. 5.300 kollokationer og verbale kollokationer er ca. 4.300. En stor del af disse tre typer ordforbindelser i Islex er hentet fra *Stóra orðabókin um íslenska málnotkuns* database (Jón Hilmar Jónsson 2005) med forfatterens tilladelse, og det er ham selv som står for udvælgelsen af navneordsforbindelserne. Nogle ordforbindelser er dog kommet ind i ordbogen ad andre veje.

### 3.4 Sprogbrugseksempler

Der er mange sprogbrugseksempler i Islex-ordbogen. Deres funktion er at vise opslagsordet i en typisk sammenhæng og at skærpe distinktionen mellem ordets forskellige betydninger, hvor der er flere end en betydning. Her er dog ikke tale om autentiske sprogbrugseksempler med en kildehenvisning som man kender fra nogle ordbøger, men de er lavet specielt til dette formål. Redaktørerne har bestræbt sig på at have sprogbrugseksemplerne så korte og rammende som

muligt. Islandske korpuser har været til stort gavn ved udformningen af sprogbrugseksemplerne eftersom deres indhold er en udtømmelig kilde om sproglig sammenhæng. Sprogbrugseksemplerne er på moderne islandsk og oversættes alle til målsprogene ligesom ordforbindelserne.

### 3.5 Bøjninger

Islex-ordbogen anfører ikke bøjningsendelser eller bøjningsklasser efter hvert opslagsord som man tit ser i andre ordbøger, men derimod anføres der ved hvert ord et link til webben *Beyingarhlýsing íslensks nútímamáls* (<http://bin.arnastofnun.is>) som er en database med bøjninger af det islandske ordforråd. Ved klikket åbnes det valgte ords bøjningsparadigme.

### 3.6 Semantiske felter

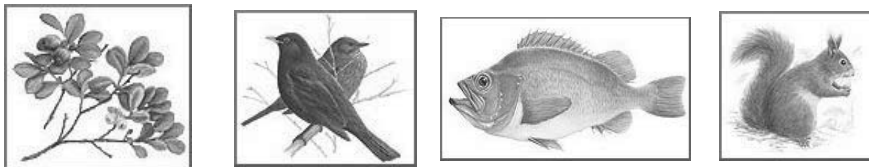
Hele ordforrådet i Islex er inddelt efter semantiske felter. I kapitler 2.1 og 2.2 fremkom det at arbejdet med de islandske ordbogsartikler såvel som oversættelserne er sket efter dette princip. De semantiske felter er grupper af betydningsbeslægtede ord, som varierer meget i størrelse, det kan dreje sig om nogle få ord op til flere hundrede, men et typisk semantisk felt omfatter nogle dusin ord. Eksempler på semantiske felter er *dyr, tøj, boliginventar, biler* (mest konkrete substantiver); *lyd, uvejr, nationalitet, slægtsforhold* (mindre konkret betydning); *fysik, geologi, medicin, grammatik, matematik* (specielle fagområder). Omfattende semantiske felter blandt adjektiverne er *alder, tilfredshed, tilstand, uduelighed, økonomi, sundhed, farve, form og udseende*. Verbernes semantiske felter er f.eks. *ændring, død, ophold, bevægelse, berøring, søvn, nedbør og vind*. De semantiske felter overlapper tit hinanden, f.eks. tilhører ordet *gulrót* ('gulerod') det semantiske felt *botanik* såvel som *mad*; og *kennari* ('lærer') tilhører både felterne *stillingsbetegnelse* og *skole*.

Ved at fokusere på én semantisk gruppe ad gangen opnår redaktøren et helhedsoverblik over et bestemt område inden for ordforrådet hvor et skiftende antal semantisk beslægtede ord forekommer fra hele alfabetet. Denne ordning sikrer en bedre sammenhæng i den endelige redigering af ordbogsartiklerne.

Til trods for at redaktørerne ikke bearbejder ordforrådet i en alfabetisk rækkefølge kan de altid danne sig et overblik over hvor langt redigeringen af hvert ord er kommet inden for hvert sprog da det er markeret i et særligt felt når redigeringen af ordbogsartiklen er færdig. Hvis et ord f.eks. er blev oversat til fuldstændighed til dansk så markerer redaktøren dette for at den danske redaktion ikke behøver at tænke mere på ordet i denne omgang.

### 3.7 Billeder

Der er et stort antal billeder i Islex. Deres formål er at være til gavn og glæde for brugeren, men de indtager også en mere omfattende rolle. For bestemte ord giver et billede den bedste forklaring af ordets betydning (*anker, æble, briller, egern, saks, tandhjul, trædrulle, tøjklæmme*), og somme tider, når det drejer sig om flere end én betydning, er det praktisk at bruge billeder for at adskille de forskellige betydninger. Desuden gør billederne ordbogen mere attraktiv i unge brugeres øjne. I Islex er der omkring 3200 fotografier og forskellige slags illustrationer. Deraf er der ca. 500 tegninger af planter, fugle, fisk og andre dyr af kunstneren Jón Baldur Hlíðberg.



Billede 2. Nogle illustrationer fra dyre- og planteverdenen af Jón Baldur Hlíðberg.

### 3.8 Målsprogene

Målsprogene i Islex-webordbogen er som før sagt dansk, norsk bokmål, nynorsk, svensk og færøsk. Her bliver målsprogene ikke behandlet på tømrende vis, på nær nogle punkter (se nærmere Aldís Sigurðardóttir *et al.* 2008: 784-786 og Aldís Sigurðardóttir *et al.* 2012).

#### 3.8.1 Oversættelserne: en krævende opgave

Islex-webordbogen stiller store krav til målsprogenes redaktører, d.v.s. oversætterne, som skal berberske islandsk såvel som målsproget. Hun/han må være opmærksom på begge sprogs nuancer og ferm til at finde frem til de rigtige ækvivalenter. Oversætterne har en forskellig baggrund, nogle af dem har arbejdet som udsendte lektorer ved Islands Universitet og lært islandsk ved den lejlighed. Andre derimod er islændinge der har boet størstedelen af deres liv i et af de nordiske lande.

Der gælder ingen særlig regel for i hvilken rækkefølge målsprogene bliver oversat i databasen, nogle gange kan de svenske oversættelser komme først, andre gange de norske o.s.v. Målsprogene støtter hinanden i mange tilfælde under oversættelsesarbejdet, men de kan også have forstyrrende effekt på hinan-



den. Da kan det være til hjælp for oversætteren at kunne se ordbogens brugergrænseflade, hvor man kan se ét sprog ad gangen uforstyrret af de øvrige sprogs tilstedeværelse. I databasens arbejdsgrænseflade ser redaktørerne alle sprogene på en gang. Oversætterne synes tit det er godt at åbne et rent og uberørt lemma, men i andre tilfælde kan de drage nytte af en anden oversætters forskningsarbejde hvis hun/han har oversat ordet før, f.eks. hvis det drejer sig om et fagudtryk. Mange fagtermer og andre internationale ord ligner hinanden på de nordiske sprog på nær islandsk.

Nogle ting er mere vanskelige end andre ved oversættelserne af ordforrådet i Islex. Det islandske sprogs fagterminologi, d.v.s. ord der falder ind under fagområder, kan være meget svær at behandle for oversætterne og nogle af redaktionerne har hentet faglig bistand udefra for at arbejde på denne del af ordforrådet. I nogle af landene har man konsulteret fagfolk ved oversættelsesarbejdet. Blandt krævende fagområder er *metrik*, *geologi* og *jura*, især da der ikke findes mange islandske håndbøger med udenlandske indekser indenfor disse specialområder.

### 3.8.2 Yderligere oplysninger om ækvivalenterne

Når der vælges flere end én ækvivalent som oversættelse af et bestemt islandsk ord tilføjes der tit oplysninger om hver ækvivalents anvendelse. Dette er til stor hjælp for dem der skriver en tekst på et skandinavisk sprog, som ikke er deres modersmål. Følgende ord på svensk forklarer dette:

**rymja**, v bröla (om tjur), rāma (om ko), skria (om åsna)  
**fimmkall**, n m femma (om myntet); fem spänn (om beloppet)

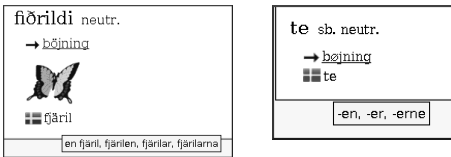
På samme måde kan man mange steder finde oplysninger om det islandske opslagsord som giver en nærmere forklaring på ordets betydning eller brug for udenlandske sprogbrugere. Følgende danske oversættelser beskriver dette:

**kvæntur**, adj gift (om mand)  
**skólasytir**, n f skolekammerat (om pige)

Alle disse ekstraoplysninger bliver defineret under et særskilt navn i værket database. Det danske ord *skolekammerat* er en bestemt type data og *om pige* er en anden type. På denne måde holdes alle typer oplysninger så 'rene' som muligt inden for databasen.

### 3.8.3 Målsprogenes morfologi

Når en islandsk bruger anvender Islex-webordbogen kan det være gavnligt for hende/ham at vide f.eks. om det danske målsprogsord er fælleskøn eller intetkøn eller hvordan et svensk verbum er i datid. Disse oplysninger finder man i ordbogens database. De er ikke synlige på overfladen, men fremstår når man bevæger musen over en ækvivalent hvorefter der springer en boks frem med ordets bøjningsendelser. Billede 3a viser svenske bøjningsendelser for ordet *fjäril* 'sommerfugl' og billede 3b viser bøjningsendelser for det danske ord *te*.



Billede 3a. *Opslagsordet fjörildi med en svensk oversættelse. Boksen med bøjningsendelser springer frem når man bevæger musen over det svenske ord. Så kan man se ordet med artikel, ordets flertal og flertal med artikel. Under det islandske opslagsord kan man se et link over i ordets böjning.*  
 Billede 3b. *Opslagsordet te med en dansk oversættelse. Her har man bevæget musen over det danske ord således at ordets böjningsendelser fremstår.*

Der arbejdes på at føje norske bøjningsendelser til Islex-databasen, men det er usandsynligt at færøsk morfologi vises på den måde. Det færøske sprogs morfologi ligner islandsk og af den grund er det ikke bekvemt at vise bøjningerne i så knap form. Alle disse ting er dog under udvikling når dette skrives og den endelige udformning er ikke afgjort endnu.

## 4. Brugeradgang til Islex

Islex-ordbogen fremstår på webben i en form der naturligvis byder på en anden og på mange måder en mere livlig fremstilling end det er muligt i en traditionel, trykt bog. Der kan f.eks. vises billeder og animationer, gives lydseksempler såvel som åbnes for interne links mellem lemmaerne foruden at der er mange slags søgefunktioner. Billede 4 viser ordbogens udseende på webben.

Billede 4. Den danske brugergrænseflade i Islex. Søgefunktionerne er øverst i spalten til højre. Der vises opslagsordet *eldgos* og dets ordbogsartikel, hvor man ser et billed og den danske oversættelse af ordet. Spalten til højre viser ordet i alfabetisk orden sammen med naboordene for oven og for neden.

Når en bruger slår op i ordbogen indstiller hun/han selv hvilken ordbog han ønsker at bruge (d.v.s. på hvilket sprog oversættelserne fremstår), og det er muligt at se flere end ét sprog samtidig. Billede 5 viser ordet *ljósapera* hvor alle målsprogene fremstår på en gang.

Billede 5. Ordet *ljósapera* med oversættelser på dansk, norsk (bokmål og nynorsk) og svensk.

Islex-ordbogen har fem forskellige 'profiler', d.v.s. at hjemmesidens tekster findes på fem sprog: islandsk, dansk, norsk bokmål, nynorsk og svensk (bruger-

grænsefladen på færøsk er under forberedelse). Man kan komme ind på Islex-webben på fire forskellige måder: *www.islex.hi.is*, *www.islex.dk*, *www.islex.no* og *www.islex.se*, d.v.s. at projektet har sin egen webadresse i hver land. Hvis f.eks. brugeren kommer ind på webadressen *www.islex.se* fremstår ordbogens svenske profil automatisk. Brugeren kan dog altid selv indstille det sprog hun/han ønsker at websiden fremstår på.

## 5. Projektets status og videre udvikling

Mulighederne for at optage flere sprog i Islex-databasen er langt fra brugt op og hvis der fremkommer passende samarbejdspartnere er det oplagt at indlemme flere sprog. Sprogene engelsk, fransk og tysk er blevet nævnt i denne sammenhæng som spændende muligheder.

Da samarbejdet med svenske og norske universitetsinstitutioner omkring Islex blev indledt var planen at udgive trykte ordbøger mellem henholdsvis islandsk og svensk, og islandsk og norsk. Disse planer kan godt realiseres på et senere tidspunkt, men da må man tage i betragtning at både layout og trykning af ordbøger er dyr.

Islex er ikke kun en ordbog, men under arbejdet er der også opstået en stor database over de nordiske sprog. Hvert islandsk opslagsord er knyttet til ækvivalenter på dansk, norsk bokmål og nynorsk, svensk og færøsk igennem databasen. Hele dette materiale med nøjagtig definerede oplysninger bevirker at Islex-databasen kan danne grundlaget for al slags sprogteknologiske projekter inden for de nordiske sprog. På grund af det store ordforråd inden for målsprogene kan Islex-materialet nemt bruges til ordbøger i den modsatte retning, f.eks. en svensk-islandsk ordbog o.s.v. Tiden vil så vise hvilke sproglige projekter der vil udspringe af Islex.

## LITTERATURLISTE

- Aldís Sigurðardóttir, Anna Hannesdóttir, Halldóra Jónsdóttir, Hákan Jansson, Lars Trap-Jensen, Þórdís Úlfarsdóttir, 2008: ISLEX – An Icelandic-Scandinavian Multilingual Online Dictionary. Proceedings of the XIII Euralex International Congress: 779-789. (på CD)
- Aldís Sigurðardóttir, Anna Hannesdóttir, Margunn Rauset, 2012: Ein-, to- eller fleirspråkleg ordbok? I: Nordiska studier i Lexikografi 11. (Skrifter utgivna av Nordiska föreningen för lexikografi 12.) Lund.
- Jón Hilmar Jónsson, 2005: Stóra orðabókin um íslenska málnotkun. JPV, Reykjavík.

*På webben*

Beygingarlýsing íslensks nútísmáls. Árni Magnússon-instituttet for islandske studier:  
<http://bin.arnastofnun.is>.

Islex-ordbogen. Árni Magnússon-instituttet for islandske studier: [www.islex.hi.is](http://www.islex.hi.is),  
[www.islex.dk](http://www.islex.dk), [www.islex.no](http://www.islex.no) og [www.islex.se](http://www.islex.se).

*Halldóra Jónsdóttir*

Árni Magnússon-instituttet for islandske studier  
[halldo@hi.is](mailto:halldo@hi.is)

*Þórdís Úlfarsdóttir*

Árni Magnússon-instituttet for islandske studier  
[disa@hi.is](mailto:disa@hi.is)

# Adverb og adverbialer: En forsømt ordklasse i ordbøkene

*Jón Hilmar Jónsson*

In most traditional dictionaries the status of adverbs and adverbials is weak compared to other word classes. This applies to the selection of lemmas where adverbs seldom have high priority, partly because of the dominance of adjective lemmas of the same word stem. It also applies to the status of many multi-word adverbials which can be difficult to detect within the entries of nouns, verbs and adjectives, especially when the adverbial has an independent meaning. Syntactically, adverbs and adverbials are most closely related to verbs but their occurrence and description in verb entries is *inevitably limited* because of the condensation of the dictionary text.

In electronic dictionaries adverbs and adverbials can assume a stronger position. Multi-word adverbials, like other multi-word units, can be directly searchable and even have lemma status on the same basis as single-word lemmas. Grammatical tagging of such lemmas also allows search and classification independent of their lexical representation.

Grammatical sorting and classification also facilitates the semantic analysis, with regard to synonym and antonym relations and to the grouping of related lemmas in larger semantic fields.

The article discusses this theme with reference to Icelandic dictionary material and an ongoing lexicographical project.

*Nøkkelord:* ordklassefordeling, adverbialer, lemmatisering, flerordslemmaer, ordnett, grammatisk tagging, semantisk klassifisering

## 1. Innledning

De fleste ordbøker inneholder en hel mengde forskjellige leksikografiske data som ligger på ulike plan i ordbokens struktur. Sammensetningen av disse dataene og deres posisjon og plassering i ordboksteksten er ikke alltid klar, i hvert fall ikke for brukerne, og det kan forekomme at heller ikke leksikografen har full oversikt.

Brukeren får gjerne, i hvert fall ved første øyekast, et ufullkomment bilde av disse forholdene. I reklamen står det ganske ofte ikke mye annet enn at ordboken inneholder en mengde oppslagsord og ordforbindelser, eller fraser. I omteksten til ordboken, innledningen og brukerveiledningen, kan man forvente å finne mer utførlig informasjon men heller ikke den gir alltid en dekkende skildring av de enheter ordboksteksten er sammensatt av, og av hvordan de er plassert i forhold til hverandre.

Lemmaene og lemmalisten utgjør ryggraden i enhver ordbok. I de fleste ordbøker består den av enkeltord fra ulike ordklasser. Samtlige lemmaer får en ordklasseangivelse som delvis tjener det formål å skille mellom homonymer. Makrostrukturen er som regel alfabetisk, og lemmaene blir fremstilt som uttrykksrelaterte og dermed gjerne flertydige enheter. I allmennspråklige ordbøker og i de fleste større ordboksverk kan man regne med at alle ordklasser er representert selv om det kan være av svært ulikt omfang. Substantivene utgjør som regel det overveldende flertall, adjektivene kommer på andreplass, etterfulgt av verbene som den tredje største ordklassen. Nederst på denne lista står de lukkede ordklassene, pronomen, tallord, preposisjoner og konjunksjoner. Og midt imellom står det en ordklasse som riktignok er åpen, men som vil ha vesentlig mindre omfang og vekt enn de større ordklassene, dvs. adverbene.

Jeg vil i denne artikkelen rette oppmerksomheten mot denne sistnevnte ordklassen og prøve å belyse hva som bidrar til at den gjerne står litt avsides innenfor lemmalista og i ordboksteksten, og hvordan den kan få en mer aktiv posisjon i en ordboksmodell der flerordsenheter kan få lemmastatus, og der betydningsbærende adverbialer også er inne i bildet som lemmakandidater (jf. også Jónsson 2009).

## 2. Plassering og relasjonsforhold

I en tradisjonell allmennspråklig ordbok blir betydningsbærende flerordsenheter underordnet ettordslemmaene. Det betyr at i mange tilfeller vil deres plassering være uklar og avhengig av ulike forhold i de enkelte artikler. Denne uklarheten bidrar også til at de får en uregelmessig og varierende form og fremstilling i ordboksteksten. Deres underordnete posisjon i ordboksstrukturen understrekes ytterligere ved at de ikke får noen grammatisk markering på linje med enkeltordenes ordklasseangivelser.

Morfologisk er adverbene sterkt knyttet til adjektivene, og det betyr at de gjerne opptrer som sublemmaer underordnet adjektiver av samme ordstamme, selv om en slik lemmatisering sjelden er regelmessig og konsekvent. Syntagmatisk derimot er adverbene sterkest knyttet til verbene, og dermed vil verbartik-

lene best illustrere deres egenskaper og karakteristiske bruk. Paradigmatisk og semantisk er adverbene relatert til adverbiale flerordsenheter av ulike slag, som i den grad de overhodet blir angitt i ordboksteksten, primært opptrer som elementer i substantivartikler. Leksikalsk og funksjonelt står mange adverb dessuten i klar relasjon til preposisjoner av samme ordstamme.

Alt dette bidrar til at adverb og betydningsbærende adverbialer inntar en beskjeden plass i de fleste ordbøker. Behovet for å angi og beskrive adverb og adverbialer kan også begrenses av det forhold at de i språkbruken tilhører remadelen av de kommunikative enhetene, der tekstsammenhengen er med på å avdekke den aktuelle betydningen.

### 3. Ordklassenes styrkefordeling, en sammenlikning av noen sentrale nordiske ordbøker

Så lenge den trykte ordboken var enerådende, var det ikke realistisk og nesten ikke aktuelt å skaffe seg et helhetlig bilde av de enkelte ordklassers profil og karakter, for hvert enkelt oppslag og hver enkel observasjon er begrenset til en bestemt artikkel, og den alfabetiske sorteringen i de fleste tilfeller det eneste synlige bindeleddet mellom lemmaene. For leksikografen kunne det også være problematisk å skaffe seg en tilstrekkelig innsikt i dette forholdet, og å formulere og følge klare retningslinjer for utformingen av ordboksteksten til lemmaer av en gitt ordklasse, siden etappene i redaksjonsarbeidet sjelden er konsentrert om en bestemt ordklasse uavhengig av alfabetet.

Ordklassenes størrelsesforhold har heller ikke fått noen særlig oppmerksomhet, og de blir sjelden fremhevet i redegjørelsen for innholdet i en ordbok. Når det gjelder nyere ordbøker, kan denne informasjon hentes frem fra databasen og gi et interessant grunnlag for vurdering og sammenlikning av ulike ordbøker.

Jeg vil i første omgang sammenlikne noen nordiske ordbøker som dels kan defineres som allmennspråklige ordbøker, og som i alle tilfeller dekker det allmenne ordforrådet.<sup>1</sup> Som det fremgår av tabell 1 er kategoriseringen i ordklasser ikke helt parallell, og visse klasser og underkategorier forekommer bare i én av ordbøkene. Men det spiller en mindre rolle for det som her er aktuelt, nemlig sammenlikne de åpne ordklassene og se hvordan adverbene er plassert. For å få det til, har jeg måttet slå sammen kategorier som tilhører samme ordklasse og i enkelte tilfeller forenkle klassifiseringen.

<sup>1</sup> Jeg takker mine kolleger i Danmark, Norge og Sverige, Henrik Lorentzen i København, Oddrun Grønvik i Oslo og Sven-Göran Malmgren i Göteborg, for å ha formidlet informasjon om de skandinaviske ordbøkene.



	Substantiver	Adjektiver	Verb	Adverb	Antall lemmaer
Íslensk orðabók	62.100 (74%)	13.300 (16%)	7.300 (8.6%)	1.300 (1.4%)	84.000
Den trykte DDO	42.600 (72%)	8.800 (15%)	7.000 (11%)	990 (2.0%)	59.390
DDO på nettet	67.400 (77%)	12.200 (14%)	7.050 (8.0%)	990 (1.0%)	87.640
Retskrivningsord- bogen	43.100 (69%)	10.700 (17%)	7.500 (12%)	800 (2.0%)	62.100
Norsk ordbok 1-9	202.400 (77%)	38.000 (14%)	20.400 (7.7%)	3.500 (1.3%)	264.300
Metaordboka	438.000 (79%)	79.300 (14%)	30.000 (5.5%)	6.200 (1.5%)	553.500
Svenska Akademiens ordlista	97.230 (77%)	18.800 (15%)	9.000 (7.0%)	1.560 (1.0%)	126.590
Svensk ordbok	43.200 (69%)	9.550 (15%)	8.700 (14%)	1.170 (2.0%)	62.620

Tabell 1. Ordklassefordeling i noen nordiske ordbøker

Adverbene utgjør i alle tilfeller en ganske liten del av lemmabestanden, og som forventet er prosenttallet lavest for de ordbøkene som inneholder de fleste lemmaer. Hvis vi ser på de andre ordklassene så er prosenttallet for substantivene nokså konstant, fra 69 til 79 % og det øker jo flere lemmaer ordboken har. Antall adjektiver er enda mer konstant og utgjør fra 14 til 17 % av lemmabestanden. Den største variasjonen finner vi derimot i verbene, fra 5,5 % i Metaordboka til 14 % i Svensk ordbok. Det ville være interessant å se nærmere på dette forholdet, ikke minst på det som gjør at Svensk ordbok gir verbene en forholdsvis stor vekt, mens substantivene står relativt svakt.

En slik sammenlikning gir naturligvis et ganske grovt og i og for seg ufullstendig bilde av det som karakteriserer en ordbok med hensyn til de leksikalske enheter den inneholder og beskriver, og deres innbyrdes forhold og relasjoner. Her burde mange av de enheter som tradisjonelt tilhører ordbokens mikrostruktur, også være inne i bildet og da først og fremst de betydningsbærende flerordsenheter som fungerer som adresser for semantiske angivelser, forklaringer og/eller illustrerende brukseksempler. Disse enhetene blir gjerne plassert under substantivlemmaer ut fra det prinsipp at substantivet utgjør det sterkeste elementet i forbindelsen selv om forbindelsen i sin helhet står som verbal eller adverbial. Denne lemmaplasseringen betyr at slike flerordsenheter ofte står ganske isolert i ordboksteksten, og at deres semantiske og syntagmatiske egenskaper ikke får noen aktualitet for helheten.

#### 4. Forholdet i konstruksjonsordbøker, ordboken Orðastaður

Når det gjelder adverb og adverbialer så ligger de i større grad enn andre ordklasser skjult i ordboksteksten slik at de bare til en viss grad er direkte tilgjengelige som lemmer, og forekommer i større grad som elementer i beskrivelsen av andre ordklasser. Det gjelder især verbene. Det betyr at en helhetlig oversikt over denne utvidete ordklasse og dens karakteristiske egenskaper oftest er utelukket, og at brukeren får nøye seg med mer eller mindre tilfeldige glimt.

De ordbøker vi hittil har sett på, kan grovt sett karakteriseres som allmennspråklige ordbøker. I andre typer ordbøker, spesialordbøker av ulike slag, vil ordklassedelingen se annerledes ut. Med hensyn til adverb og adverbialer er konstruksjonsordbøker spesielt aktuelle i denne sammenheng. I den islandske ordboken Orðastaður (1994, 2. utg. 2001), som kan defineres som en kollokasjons- og konstruksjonsordbok, får vi et helt annet bilde av ordklassene, noe som fremgår av nedenstående tabell. Her er substantivene fortsatt den største ordklassen, men den er adskillig mindre enn i de allmennspråklige ordbøkene og overskrider så vidt 50 %. Verbene står derimot tilsvarende sterkere og ligger omtrent på samme nivå som adjektivene. En annen stor forskjell, og det som kan overraske mest, er at adverbene ikke lenger blir nevnt; det fins altså ingen adverbiale lemmer i ordboken. Men hvis vi ser på ordboksteksten, så er det klart at adverb og adverbialer inntar en betydelig plass i beskrivelsen av de enkelte lemmer, til og med slik at det aktuelle lemmaet primært fungerer som adgangsnøkkel til informasjon om varierende og semantisk sammenhengende adverbiale enheter, enkeltord og flerordsforbindelser. Det forutsettes altså at brukeren finner den logiske veien til denne informasjon gjennom kjente basisenheter som adverbene og adverbialene står sammen med. Med hensyn til det ville lemmatisering av de adverbiale enhetene være ulogisk og overflødig innenfor en alfabetisert lemmaliste med enkeltord.

Vi legger også merke til hvor få lemmer ordboken inneholder, drøyt 11.000 i alt. Men også her dekker lemmerne enheter som ligger tett inntil overflaten som en slags latente lemmer som ville utvide lemmalisten kolossalt, dvs. semantisk klassifiserte sammensetninger av de enkelte lemmerne i en særskilt komponent, i alt bortimot 100.000 ord. Men den sentrale kjernen i ordboken er ordforbindelsene den inneholder, omtrent 45.000 kollokasjoner og konstruksjoner.

Antall lemmaer	11.200	
Substantiver	6.200	56%
Adjektiver	2.200	19.5%
Verb	2.400	21%
Forledd	400	3.5%

Tabell 2. Orðastaður: Sammensetningen av lemmabestanden

Det kan sies å være noe paradoksalt i det å holde adverbene utenfor lemmalisten samtidig som de får en spesiell fokusering i ordboksteksten. Og selv om det i og for seg er logisk i en ordbok som Orðastaður, understreker det hvor vanskelig det er å komme fram til en tilfredsstillende lemmatisering og presentasjonsmåte innenfor rammen av en trykt semasiologisk ordbok.

De hindringer og problemer det her dreier seg om, har naturligvis med struktureringen av ordboksteksten å gjøre, ikke minst det avgjørende skillet mellom ordbokens makro- og mikrostruktur. Hvis adverbene skal kunne komme sterkere til syne og adverbialene trekkes frem som selvstendige og aktive enheter, må de opptre på samme plan som de tradisjonelle ettordslemmaene. Og det er fullt ut mulig hvis vi tar skrittet fra den trykte til den elektroniske ordboken.

## 5. Adverb og adverbialer i Íslenskt orðanet

I de siste årene er det på Island blitt bygget opp en leksikografisk database og ut fra den en ordboksmodell som jeg og min medarbeider Þórdís Úlfarsdóttir har gitt navnet Íslenskt orðanet (se Jónsson og Úlfarsdóttir 2011). Her utgjør data-samlingen til Orðastaður grunnlaget sammen med data fra andre kilder, bl.a. en fraseologisk begrepsordbok ved navnet Orðaheimur som utkom som en slags fortsettelse av Orðastaður, der de faste ordforbindelser og idiomatiske uttrykk står i fokus, og der hovedlemmalisten består av begrepsoverskrifter som forener semantisk beslektede fraser. Ordnettet utgjør nok et skritt i omformingen av lemmastrukturen fra det semasiologiske aspektet der enkeltordene er enerådende, til det onomasiologiske aspektet der oppmerksomheten gjelder lemmaenes innbyrdes relasjoner (leksikalske, syntagmatiske, paradigmatisk, grammatisk og semantisk), og der flerordsenheter er inkludert som likestilte enheter. Her er det ikke plass til å gi en helhetlig beskrivelse av ordnettet, men jeg vil fremheve noen karakteristiske trekk.

Nedenstående tabell gir en oversikt over ordklassene og deres størrelsesforhold i Íslenskt orðanet. Foruten det høye antall lemmaer (ca. 260.000) legger vi merke til hvor stor plass verbene opptar i lemmalisten. Adverb og adverbialer (som en sammensatt ordklasse) skårer også adskillig høyere enn adverbene i de nordiske ordbøkene som ble omtalt tidligere.

Antall lemmaer	255.000	
Substantiver	127.500	50%
Adjektiver	37.000	14.5%
Verb	79.000	31%
Adverb og adverbialer	11.500	4.5%

Tabell 3. Íslenskt orðanet: Sammensetningen av lemmabestanden

I ordnettet er flerordslemmaer aktuelle innenfor alle ordklasser, men de er spesielt karakteristiske for verb og adverb(ialer). Den leksikografiske analysen gjelder lemmaenes innbyrdes relasjoner, især de semantiske, der det forutsettes en omfattende entydiggjøring av såvel ettords- som flerordslemmaer. Denne analysen er også verdifull for en målbevisst betydningsbeskrivelse av lemmaene.

### 5.1 Adverbiale flerordslemmaer, leksikalske og grammatiske mønstre

Innenfor mikrostrukturen, og sammen med andre elementer i ordboksteksten er de ulike adverbialene ganske uhandgripelige enheter. Det avgjørende ved overføringen til makrostrukturen og lemmaplanet er å forsyne hvert enkelt lemma med en grammatisk kode som gjengir hvilke enheter det er sammensatt av. Det gjelder ellers for alle flerordsenheter i ordnettet og ble først utført for verbene som dermed fikk et nytt klassifiseringsgrunnlag. For at den grammatiske markeringen skal fungere ved sortering og klassifisering av lemmaene, må de få en standardisert fremstilling. De grammatiske taggene gjelder i hovedsak ordklasse, kasus og bestemthet, men dessuten får ledd med varierende innhold en avgrensende markering i form av hakeparentes og en passende proform eller kontekstangivelse med typiske ord og kombinasjoner.

Dermed får de adverbiale lemmaene en grammatisk angivelse på linje med ettordslemmaenes ordklasseangivelse, noe som i grunnen gir et enda bedre sorteringsgrunnlag.

En sortering av de adverbiale flerordslemmaene, alfabetisk så vel som etter grammatiske tagger, avdekker grammatiske og leksikalske mønstre samtidig som den kan gjenspeile semantiske relasjoner. Figur 1 inneholder to alfabeti-

serte lister med adverbialer. Den ene innledes med preposisjonen *frá* ‘fra’ i temporal betydning (*frá barnsku* ‘fra barndommen’, *frá fæðingu* ‘fra fødselen’ osv.). Den andre innledes med konjunksjonen *þegar* ‘når’. I en elektronisk ordbok kan slike kombinasjoner i stor utstrekning lemmatiseres og kobles sammen på ulike måter, enten det gjelder relasjoner av leksikalsk, grammatisk eller semantisk karakter. I ordnettet er lemmalisten i prinsippet åpen og lemmatiseringen motiveres av at vedkommende lemmakandidat står i en bestemt semantisk eller syntagmatisk relasjon til andre lemnaer.

14 frá barnsbeini ao ffs no-d	78 þegar tímar líða ao ffs no-n so
15 frá barnsku ao ffs no-d	79 þegar tímar líða fram ao ffs no-n so ao
16 frá barnsku ao ffs no-d	80 þegar um hægist ao ffs so
17 frá birtingu til sólarlags ao ffs no-d fs no-g	81 þegar upp er staðið ao ffs so so
18 frá blautu barnsbeini ao ffs lo-d no-d	82 þegar út á líður ao ffs so
19 frá blábernsku ao ffs no-d	83 þegar úthallar degi ao ffs so no-d
20 frá byrjun ao ffs no-d	84 þegar úthallar slætti ao ffs so no-d
21 frá eilífi til eilíðar ao ffs no-d fs no-g	85 þegar úthallar sumri ao ffs so no-d
22 frá forðum tíð ao ffs ao no-n	86 þegar úthallar vetri ao ffs so no-d
23 frá fornheionum tíma ao ffs lo-d no-d	87 þegar verst gegnir ao ffs so
24 frá formu fari ao ffs lo-d no-d	88 þegar þar að kemur ao ffs so
25 frá formöld ao ffs no-d	89 þegar þörf beiðir ao ffs no-n so
26 frá forsögulegum tíma ao ffs lo-d no-d	90 þegar þörf krefur ao ffs no-n so
27 frá fyrsta fari ao ffs lo-d no-d	91 þegar öll kurl komu til grafar ao ffs fn-óákv-n no-n so fs no-g
28 frá fyrstu qerð ao ffs lo-d no-d	92 þegar öllu er á botninn hvolft ao ffs fn-óákv-d so fs no-g so
29 frá fyrstu hendi ao ffs lo-d no-d	93 þegar öllu er til skila haldið ao ffs fn-óákv-d so fs no-g so
30 frá fyrstu tíð ao ffs lo-d no-d	94 þegar <illa> árar ao ffs <ao> so
31 frá fæðingu ao ffs no-d	
32 frá gamalli tíð ao ffs lo-d no-d	

Figur 1. Adverbiale flerordslemnaer: Leksikalske og grammatiske mønstre

Verdien av den grammatiske markeringen av flerordslemnaene i ordnettet kommer best til syne i verbene der de utgjør et sorteringsgrunnlag for ulike konstruksjonsmønstre og trekker frem karakteristiske syntaktiske egenskaper til de enkelte verb. Selv om mangfoldigheten er mindre blant adverbialene, viser de også en rik leksikalsk og syntaktisk variasjon. Det gjelder bl.a. skillet mellom ulike kasus og markeringen av ledd med varierende innhold. Den grammatiske taggingen viser en stor konstruksjonsmessig variasjon, og visse kombinasjoner skiller seg ut med en spesiell høy frekvens, f.eks. preposisjonsledd i ulike variasjoner der preposisjonen styrer dativ. Som eksempel dekker taggen fs lo-d no-d (preposisjon + adjektiv i dativ + substantiv i dativ) drøyt 700 lemnaer, bl.a.:

- að minnsta kosti ‘i hvert fall’
- af ráðnum hug ‘med overlegg’
- á skammri stundu ‘på kort tid’
- frá blautu barnsbeini ‘fra barnsbein av’
- fyrir opnum tjöldum ‘for åpen scene’
- í fljótu bragði ‘ved første øyekast’
- með glöðu geði ‘med glede’
- undir berum himni ‘under åpen himmel’

Den grammatiske og alfabetiske sorteringen vil i mange tilfeller underbygge semantisk gruppering og klassifisering av lemmaene. Den kan synliggjøre og skjerpe semantiske relasjoner og på den måte bidra til oppbyggingen av synonymgrupper av den type som vises i figur 2. Ved markeringen av synonymrelasjoner blant flerordslemmaene må det tas strukturelle hensyn, slik at en parallell argumentstruktur blir et klart kriterium for synonymgrupperingen.

**med overlegg**

SAMHEITI / SKYLDHEITI

að yfirlögðu ráði ao af ásetningi ao af ásettu ao  
af ásettu ráði ao af forsettu ao af forsettu ráði ao  
af forsettum ásetningi ao af ráðnum hug ao  
af ráðnum huga ao af settu ráði ao af tilsettu ráði ao  
eftir upplögðu ráði ao með ásetningi ao með ásettu ao  
með ásettu ráði ao með ráðnum huga ao  
með upplögðu ráði ao viljandi ao vísvitandi ao

**hele livet**

SAMHEITI / SKYLDHEITI

alla sína lífdaga ao alla sína lífsdaga ao  
alla sína lífstíð ao alla sína ævi ao  
alla ævi ao allt sitt líf ao

**overdrevent**

SAMHEITI / SKYLDHEITI

fram úr hófi ao hóflaust ao í óhófi ao meir en góðu hófi gegnir ao  
meir en hófi gegnir ao meir en hófi hæfir ao meira en góðu hófi gegnir ao  
oftar en góðu hófi gegnir ao óhóflega ao úr hófi ao úr hófi fram ao

**øyeblikkelig**

SAMHEITI / SKYLDHEITI

á auga lífandi bili ao á auga lífandi bragði ao á augabragði ao  
á lífandi bragði ao á samri stundu ao á skammri stundu ao  
á stundinni ao á svipstundu ao áður en <maður> veit af ao  
eins og skot ao í einni svipan ao í einu vetfangi ao í hvelli ao  
óðara ao strax ao undir eins ao

Figur 2. Synonymgruppering av adverbiale flerordslemmaer

## 6. Sluttord

Denne korte gjennomgangen av adverbenes og adverbialenes plassering og posisjon som betydningsbærende leksikografiske enheter og deres status i en tradisjonell lemmamodell sammenliknet med en utvidet modell der betydningsbærende flerordsenheter er inkludert i lemmalisten, gir oss anledning til å tenke over alle de muligheter og utfordringer den elektroniske ordboken gir. Et mer aktivt samspill mellom de enheter og elementer som er gjenstand for en ordboksbeskrivelse, er en viktig faktor i den forbindelse. Og der inntar adverbene og adverbialene en betydelig plass. Arbeidet med det islandske ordnettet er ment å bidra til å gi disse enhetene en sterkere leksikografisk fokusering i fremtiden.

## LITTERATUR

- DDO = Den danske ordbog 1–6. København 2003–2005. <http://ordnet.dk/ddo/>  
 Íslensk orðabók. 4. útgáfa. Reykjavík 2007.  
 Íslenskt orðanet: [www.ordanet.is](http://www.ordanet.is)  
 Jónsson, Jón Hilmar, 2009: Lemmatisation of Multi-word Lexical Units: Motivation and Benefits. Henning Bergenholtz, Sandro Nielsen & Sven Tarp (eds). *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. S. 165–194. Bern: Peter Lang.  
 Jónsson, Jón Hilmar og Þórdís Úlfarsdóttir, 2011: Íslenskt orðanet: Et skritt mot en allmennspråklig onomasiologisk ordbok. I: *LexicoNordica* 18: 87–108.  
 Metaordboka. Norsk ordbok 2014. <http://no2014.uio.no/tekster/sok/index.html>  
 Orðaheimur. Reykjavík 2002.  
 Orðastaður. 2. útgáfa. Reykjavík 2001.  
 Retskrivningsordbogen. 3. udgave. København. 2001.  
 Svensk ordbok utgiven av Svenska Akademien. Stockholm 2009.  
 Svenska Akademiens ordlista. 13. uppl. Stockholm 2006.

*Jón Hilmar Jónsson*

Stofnun Árna Magnússonar í íslenskum fræðum - Háskóli Íslands, Reykjavík  
 jhj@hi.is

# Ordbogsmanuskripter og et historisk ordbogsarbejde

*Guðrún Kvaran*

In this article the author discusses the importance of using unpublished dictionaries in manuscripts, especially to collect examples from the spoken language that have not been excerpted for published dictionaries. Three collectors are mentioned as examples, Jón Ólafsson from Grunnavík, Dr Hallgrímur Scheving and Dr Björn M. Ólsen. In the 18th century the first one wrote a large manuscript that is extremely important for the history of the Icelandic language. An examination of the vocabulary shows that a great deal of the words have their roots in the spoken language.

In the early 19th century Scheving collected two different manuscripts (Lbs. 220 8vo and Lbs. 283–285 4to) of words that were intended for an Icelandic dictionary, but he never finished. In these manuscripts there are a great many words from the spoken language with no examples in printed dictionaries, probably with Scheving's pupils, who came from all parts of the country, as the source. Scheving's third manuscript is 'Florilegium', with Danish and Low-German loan words in Icelandic, probably intended for a book on language purism as well as for a dictionary of loan words.

The third collector mentioned is Björn M. Ólsen who collected words he heard on his journeys around the country, originally intended for a special dictionary of the spoken language.

The author emphasizes that lexicographers working on historical dictionaries cannot ignore these unprinted sources of an important part of the vocabulary.

*Nögleord:* historisk leksikografi, ordforrådet, talesprog, fremmedord

## 1. Indledning

En historisk ordbog, trykt eller på webben, må lægge vægt på at behandle så stor del af ordforrådet som muligt, såvel skrift- som talesproget. Den må forsøge at behandle det lokale ordforråd så vidt som det er muligt. Oplysninger derom fra ældre tid er som bekendt begrænsede, men de findes dog spredt i ældre ordbogsmanuskripter. I artiklen lægges der vægt på det historiske ordbogsarbejde. Jeg har i sinde at diskutere og vise med eksempler hvorledes manuskripter fra det



18. og 19. århundrede støtter det islandske historiske ordbogsarbejde. Tre manuskripter blev valgt som jeg synes at er vigtige i denne hensigt. Fra det 18. århundrede drejer det sig om det store manuskript af Jón Ólafsson fra Grunnavík (Guðrún Kvaran 2005, 2009) og fra det 19. århundrede vil jeg vælge fire vigtige ordsamlinger af Hallgrímur Scheving (Guðrún Kvaran 2008) og Björn M. Ólsen (Guðrún Kvaran 2003) og lægge vægt på de oplysninger som findes der om det lokale ordforråd. Min hensigt er at vise at uden at tage hensyn til den slags utrykte ordsamlinger ville man gå glip af en god del af det ældre ordforråd.

## 2. Jón Ólafsson fra Grunnavík (1705–1779)

I Árni Magnússon instituttet for islandske studier findes der opbevaret i ni foliobind Jón Ólafssons ordbogsmanuskript med nummeret AM 433 fol. Forfatteren er bedst kendt som håndskriftsamleren Árni Magnússons sekretær og senere stipendiat hos Arne Magnusson fonden i København. Jón Ólafsson arbejdede på ordbogen med pauser fra 1734 og til sin død i året 1779, men han fik aldrig den mulighed at få ordbogen udgivet. Selve manuskriptet er ikke indbydende for dem som ikke er vant til at læse hans håndskrift. Han har på grund af mangel af penge måttet bruge papiret så godt han kunde, han skrev også på margenerne og ind i mellem linierne hvis det var nødvendigt. Alle opslagsordene i manuskriptet er blevet skrevet på sedler samt deres betydning og samt kommentarer, som har gjort det lettere at arbejde med selve ordforrådet. Ialt er sedlerne omkring 50.000, mens enkelte opslagsord er godt 40.300. Opslagsordene findes i en liste på instituttets hjemmeside ([arnastofnun.is](http://arnastofnun.is)).

Manuskriptets ordforråd stammer for det meste fra islandske middelalder skrifter, digtning og sandsynlighvis fra alle de trykte islandske bøger som Jón Ólafsson fik fat i i København. En god del af ord og ordforbindelser er fra det 18. århundredes talesprog, og ganske ofte står som forklaring fra hvilken del af landet han har sine eksempler. Ret mange eksempler er fra Vestfjordene fra hans eget eller hans slægnings talesprog, andre eksempler stammer fra andre informanter, muligvis islandske studenter i København. Sine kilder nævner han dog ikke ved navn. Han skriver f.eks. „Ostfiordensibus [i Ostfjordene], vocula Vestfiordensibus usitata [et ord brugt i Vestfjordene]; loqvendi modus Vestfiordensibus [en måde at tale på i Vestfjordene]; vocula Vestfiordensis [ord fra Vestfjordene]; Vestmanneyensibus [på Vestmanna-øerne]; vocatur Vestmann-eyensibus [nævnt så på Vestmanna-øerne]“ og så videre.

Jeg har gennemgået hele seddelsamlingen og der findes talesprogseksempler fra alle dele af landet. Omkring halvtreds af dem har jeg undersøgt med hensyn

til brug og udbredelse og slået dem op i instituttets seddelsamlinger og i ordbøger. Unægtelig har jeg i nogle tilfælde ikke fundet meget som støtter den lokale brug som Jón Ólafsson nævner, mest fordi ordene er gået ud af brug for lang tid siden, men også fordi de synes at have været kendt fra flere egne end han nævner. Han samlede ikke selv til sin ordbog i Island og måtte støtte sig til de oplysninger han fik fra sine informanter. I andre tilfælde har jeg fundet kilder som støtter lokal brug af ordene helt op til dags dato.

Jeg har valgt ét eksempel fra manuskriptet. Det er *ordet bónstóll*.

### Eksempel

*Bónstóll* er sammensat af *bón* 'bøn' og *stóll* 'stol'. På seddelen står:

cathedra precatória, ad liggja aa boon-stool fyrir manni loqvendi modus Vestfiordensibus, ut puto, tantum in usu, multis precibus aliquem adoriri, multum sollicitare.

D.v.s. Jón Ólafsson synes at kende godt frasen *að liggja á bónstól fyrir einhverjum* fra Vestfjordene i betydningen at 'bestorme én med mange bønner, ulejlige én meget'. Frasen forekommer i manuskriptet også under opslagsordet *stóll* hvor der står: „obnix aliqvem rogare“, d.v.s. 'bede én om noget af al sin kraft'.

I instituttets seddelsamling findes der to eksempler på ordet *bónstóll*. Det ene forekommer i et digt fra det 17. århundrede af en præst i Nordlandet hvor det bruges om en stol, mens det andet er fra Jón Ólafsson selv i et brev som han skrev sin bror og hvor han praler af at han nu ikke længere behøver at „liggja á bónstól frammi fyrir honum“, han behøver ikke længere at 'tigge og bede ham om noget'. Det er klart at han havde frasen på rede hånd.

*Bónstóll* er ikke opslagsord i ordbøger over oldsproget og det findes heller ikke i de ældste islandske ordbøger og det forekommer ikke i Sigfús Blöndals islandsk-danske ordbog fra 1920–1924. Frasen forekommer ikke i islandske samlinger af talemåder. Det ser derfor ud til at den først og fremmest brugtes lokalt i Vestfjordene hvor Jón Ólafsson sagde at den var hyppig. Det ser ud til at frasen er lokalt talesprog, men det kan man ikke bestemme uden flere eksempler. Ordbogen støtter dog og forklarer eksemplet som findes i instituttets seddelsamling.

### 3. Hallgrímur Scheving (1781–1861)

Det er god grund til at have respekt for de idealister som fordybede sig i ord-samling som ofte ikke blev til noget andet end ordlister på sider eller sedler, resultatet var næsten aldrig en trykt ordbog. I Landsbiblioteket findes der

mange sådanne manuskripter, store og små, derunder to som Hallgrímur Scheving, lærer i Bessastaðaskóli, samlede til og har numrene Lbs. 220 8vo og Lbs. 283–285 4to. De blev aldrig trykt, men Sigfús Blöndal (1920–1924) benyttede sig af det sidstnævnte ved sin islandsk-danske ordbog som jeg allerede har nævnt. Grunden til at Scheving begyndte sin ordsamling er sandsynligvis den at han syntes at der var mangel på ord, især fra talesproget, i Björn Halldórssons islandsk-latinsk-danske ordbog fra 1814.

Hvis man først ser på manuskriptet 220 8vo så er det skrevet på 570 sider med latinske definitioner og er øjensynlig en renskrivning, ikke med Schevings håndskrift.

Hvis man sammenligner manuskriptet med Björn Halldórssons ordbog så er det helt klart at det drejer sig i ganske mange tilfælde om tilføjelser. Der findes henvisninger som *sjá, hjá, aliter, apud*, eller *perperam BH*<sup>1</sup>, og nye nummererede betydninger er anført som om de var tilføjelser til selve ordbogen. Som eksempel nævner jeg ordet *lapi*. I Björn Halldórssons ordbog er ordet forklaret først på latin og siden på dansk som ‘*homo sui negligens*, en (især i sin Gang) skjodesløs Person’ (1992:292). I Schevings manuskript står ved ordet *lapi* ‘*exiguuum humerale*, p.e. lítill jökull, aliud est lapi apud BH’. Den tilføjede betydning er ‘gletsjer, bræ’. Kun et eksempel findes i instituttets seddelsamling om en skødesløs person. I den islandske etymologiske ordbog af Ásgeir Blöndal Magnússon (1989: 544–545; herefter ÁBM) findes kun Björn Halldórssons forklaring, mens *bræ* nævnes ikke. Schevings tilføjelse er således noget som man skal tillægge betydning.

Det yngre manuskript Lbs. 283–285 4to er også en renskrivning af en anden. På titelbladet står ‘Orda-Safn úr nýara og daglega málinu tínt saman af skólakennara Dr. H. Scheving’, d.v.s. en ordsamling fra det nyere og daglige sprog samlet af skolelærer Dr. H. Scheving.

Det er en stor fordel at landsdel, undertiden amt, er næsten altid nævnt hvis et ord stammer fra talesproget og ikke fra en trykt kilde. Som eksempel kan nævnes at ved ordet *loddi* står ‘Árnessýslumál’ [brugt i Árnessýssel på Sydlandet] hvis betydningen er ‘grov utvunden uldtråd’, mens der står ‘Múlasýslumál’ [brugt i Múlasýssel på Østlandet] hvis betydningen er ‘vante’. Det er helt klart at Scheving har været særdeles opmærksom på ord fra talesproget, og deres antal i manuskriptet er utrolig stort hvis man tager hensyn til at han så vidt man ved aldrig tog på rejser specielt for at samle til sin ordbog. Men han lyttede til sine elever som kom fra alle egne i landet og spurgte dem om ord og ordforbindelser.

<sup>1</sup> Oversættes: se, hos, anderledes, hos, fejlagtigt BH.

Her nævnes nu to eksempler, et fra hvert manuskript, for at vise hvorledes manuskripterne kan støtte instituttets samling til en historisk ordbog.

## Eksempler

*alivant* 'A.M. víðsvegar, hér og hvar' [vidt og bredt, i alle retninger] er et opslagsord i 220 8vo. A.M. for „austanmál“ betyder at ordet er brugt i den østlige del af landet. Det findes hverken i instituttets samlinger fra skriftsproget eller fra talesproget. I den islandske etymologiske ordbog (ÁBM 1989: 11) er den ældste kilde fra det 19. århundrede, mens der nævnes ikke om ordet kun blev brugt lokalt. Kilden er sandsynligvis fra Scheving og Ásgeir Blöndal Magnússon gætter på at *alivant* er en omdannelse fra dansk *alle vegne*, hvad ikke er usandsynligt.

*alpá* 'A.M. ss. óvíst. Það er alpá.' Betydningen er 'usikkert'. I instituttets samling er der kun et eksempel fra manuskriptet 283–285 4to. Ordet findes ikke i talesprogssamlingen og heller ikke hos Jón Ólafsson fra Grunnavík eller i Björn Halldórssons ordbog. Scheving er derfor instituttets eneste kilde. Substantivet *þá* betyder 'smeltet, ikke frosset jord eller smeltet vand på jorden'. Det er muligt at knytte *alpá* til *þá*. Det er så brugt i en overført betydning om noget som ikke er helt sikkert, lige som det er usikkert at færdes i et terræn når isen er ved at smelte.

Fra Scheving stammer endnu et manuskript som man på Orðabók Háskólans ikke har gennemgået systematisk endnu og jeg kort vil nævne. Det drejer sig om en samling af ord som Scheving har ment at var fremmedord i islandsk (Guðrún Kvaran 2011). Samlingen nævnte han *Florilegium* som er en latinsk oversættelse af det græske ord *antologia*. Schevings samling er opbevaret i syv hæfter på Landsbiblioteket. Hans hensigt er ikke helt klar. På den ene side ser det ud til at han arbejdede på en fremmedordbog. Det viser de lister fra alle bogstaver som han har ladet renskrive. På den anden side kunne han have haft en sprogrensningssamling i sinde. Det viser en del ord fra bogstavet *s* med overskriften „Uppkast til lexicī barbarismorum in lingua Isl“ som synes at være et udkast til færdige ordbogsartikler. De begynder ved ordet *sabbat* og det sidste er *sjá* (i forbindelsen *se til med nogen*). Danske ord er i de fleste tilfælde nævnt i parentes.

Scheving nævner næsten aldrig lokal brug. En undtagelse er dog ordet *bró-dolia*. Der står: „Álptnesingamál – gjöra einhverja bróðolfu = gjöra skandala“, d.v.s. fra Álptanes om at skandalisere. Ordet findes ikke i Björn Halldórssons ordbog, mens Ásgeir Blöndal Magnússon har lemmaet *brotalía* og derunder *brótolia* og *bróðolia* (1989: 84), men markerer dem ikke som lokalt sprog. Der fandtes ingen eksempler i talesprogssamlingen. Ordet er en omdannelse af det

danske ord *badulje* eller dets sideform *bredulje* som er taget til låns i dansk fra fransk *bredouille*.

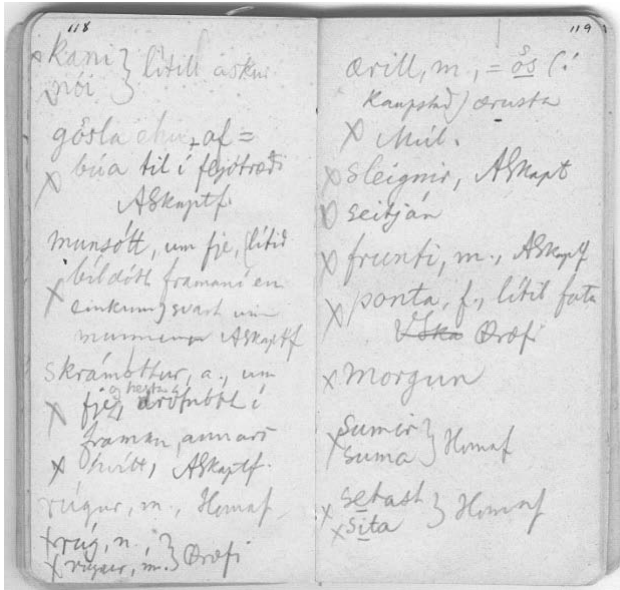
Jeg har undersøgt de ord som begynder på *d* og på *g*. I de fleste tilfælde drejer det sig om ord som er taget til låns fra dansk, fra nedertysk via dansk, eller direkte fra nedertysk. Det er helt klart at *Florilegium* er til stor nytte ved indsamling og bearbejdelse af fremmedord som blev brugt i talesproget i den første halvdel af det 19. århundrede og som ikke så gerne blev brugt i trykte bøger. En del af ordene er ganske vist i Björn Halldórssons ordbog, mens andre er Schevings tilføjelser som er værd at tage hensyn til i beskrivelsen af det historiske ordforråd.

#### 4. Björn M. Ólsen (1850–1919)

Den sidste samler som jeg vil nævne er Björn M. Ólsen, uddannet i sprogvidenskab fra Københavns universitet og Islands universitets første rektor. Efter at han kom hjem fra sit studium havde han i sinde at samle til og skrive en videnskabelig ordbog over det islandske nutidssprog, især det lokale ordforråd i hele landet. Han fik stipendium fra det danske Carlsbergfond og påtog sig nogle rejser. Ordene skrev han i 40 lommebøger som nu er opbevarede i Árni Magnússon instituttet for islandske studier. Björn M. Ólsen kunne ikke fuldende sit formål på grund af sygdom og overlod lommebøgerne til Sigfús Blöndal til brug i den islandsk-danske ordbog som han arbejdede på.

Hvis man undersøger lommebøgerne er det klart at Björn M. Ólsen har været vidt omkring i landet i årene 1885–1892 og skrevet ned ord, kollokationer og fraser som han hørte eller spurgte om. Sigfús Blöndal benyttede sig meget af hans samling. Det kan man f.eks. se af undersøgelse af ord fra Vestfjordene, især ord som ifølge Björn stammede fra Arnarfjord (Guðrún Kvaran 2003: 168–170). Han havde skrevet ned omkring et hundrede ord enten i Arnarfjord eller efter en informant eller informanter fra Arnarfjord. Omkring tredive var ifølge min undersøgelse også kendt i andre dele af landet, d.v.s. de var ikke lokale. I tilfælde af tyve af dem kunne man ikke fastsætte udbredelsen på grund af mangel af eksempler. Omkring fyre ord eller bestemte betydninger af ord var med stor sikkerhed lokale i den nordlige del af Vestlandet, især omkring Arnarfjord og Dýrafjord. Sigfús Blöndal har i de fleste tilfælde markeret ordene „Arnf.“ (Arnarfjord) eller „Vf.“ (Vestfjordene) i ordbogen, og hans forklaringer og bemærkninger antyder at hans kilde er den førnævnte lommebog. Det samme gælder også mine undersøgelser af ord som hos Sigfús Blöndal er markeret fra Østfjordene, Syd-østlandet og Sydlandet. En god del af dem viste sig at være lokale.

Trods det at ordene oftest er markeret en landsdel eller en egn er det slet ikke altid tilfældet og ofte kan man umuligt gætte sig til om ord på den samme side er markeret Nl. for Nordlandet, Sl. for Sydlandet eller Vf. for Vestfjordene eller er slet ikke markeret. Informanterne nævnes næsten aldrig. Undertiden er lommebøgerne svære at bruge da Björn skrev med en blyant og skriften er på mange steder næsten udvidsket.



Figur 1 Fra Björn M. Ólsens lommebog

Selvom Björn M. Ólsen skrev størstedelen ned selv af det han hørte eller fik svar på fra sine informanter findes der dog på nogle steder ordlister fra andre samlere. Som eksempler kan nævnes en fra præsten Jónas Jónasson frá Hrafnagil, *Rangvella, sive idiomata Rangvellensium*. En anden liste stammer fra geologen Þorvaldur Þoroddsen med ord fra talesproget på Østlandet som han skrev ned på sine forskningsrejser og den tredje fra præsten Jón Jónsson i Bjarnarnes med ord fra Sydøstlandet (fig. 1). Alle tre har til fælles ikke at være indfødte på egnen, og de har hørt ord og ordforbindelser hos lokale mennesker som de ikke brugte eller kendte selv. Mine undersøgelser har vist at største delen af det de skrev ned var lokalt talesprog.

## Afslutning

Jeg har nu kort drøftet fem ordbogsmanuskripter uden at have haft mulighed til at behandle dem grundigt. Jeg kunde også have nævnt talrige andre, men min hensigt var jo først og fremmest at lægge vægt på utrykte manuskripters vigtighed ved indsamling til en historisk ordbog. En stor del ord, kollokationer og fraser findes der som ikke forekommer i trykte bøger. Ordsamlerne skrev ned det de hørte eller brugte selv i det daglige sprog, talesproget, og uden deres kendskab og interesse for sproget ville meget materiale være gået tabt. Uden at tage hensyn til disse manuskripter er det efter min mening umuligt at skrive udførlige artikler til en historisk beskrivelse af sproget.

## LITTERATUR

- Ásgeir Blöndal Magnússon, 1989: Íslensk orðsifjabók. Reykjavík.
- Björn Halldórsson, 1814: Lexicon Islandico-Latino-Danicum Biörnnonis Haldorsonii. Biörn Haldorsens islandske Lexikon. I–II. Hafnia.
- Björn Halldórsson, 1992: Björn Halldórssons Orðabók. Íslensk – latnesk – dönsk. Ný útgáfa. 1992. Jón Aðalsteinn Jónsson (red.). Orðfræðirit fyrri alda II. Reykjavík.
- Guðrún Kvaran, 2003: Sigfús Blöndal og vasabækur Björns M. Ólsens. I: Íslenskt mál og almenn málfræði 25. S. 149–172.
- Guðrún Kvaran, 2005: Úr orðabókarhandriti Jóns Ólafssonar úr Grunnavík. I: Íslenskt mál og almenn málfræði 27. S. 201–216.
- Guðrún Kvaran, 2008: Hallgrímur Scheving og staðbundinn orðaforði. I: Íslenskt mál og almenn málfræði 30. S. 153–177.
- Guðrún Kvaran, 2009: Staðbundinn orðaforði í orðabók Jóns Ólafssonar úr Grunnavík. I: Íslenskt mál og almenn málfræði 31. S. 181–195.
- Guðrún Kvaran, 2011: Hallgrímur Scheving og tökuorðin. I: Orð og tunga 13. S. 51–75.
- Sigfús Blöndal, 1920–1924: Íslensk- dönsk orðabók. Reykjavík.

*Guðrún Kvaran*

Stofnun Árna Magnússonar í íslenskum fræðum / Háskóli Íslands  
gkvaran@hi.is

# Varför inte dub(b)lettformer? Om SAOB som normativ ordbok

*Lennart Larsson*

The Swedish Academy Dictionary (SAOB) is nowadays regarded – not least by its editors – as a decidedly descriptive dictionary, with the objectives of registering and describing the factual use of the Swedish language from 1521 to the present day. However, this has not always been the case. When the dictionary was first published at the end of the nineteenth century, it had, in addition to the descriptive aim, an explicit normative purpose that can be seen not only in the many value judgements occurring in its earlier volumes, but also in the shaping of the structure of the entries, as well as in the selection of the information contained therein.

This article provides examples of how the conflict between the descriptive and normative purposes is manifested in SAOB. The main focus is on derivatives with the stressed suffix *ett*, e.g. *dubblett* and *pollett*, a group of words in which language users tend to prefer spellings like *dublett* and *polett* in accordance with the pronunciation, but contrary to the norm established by the Swedish Academy Word List (SAOL). As becomes plain from this article, SAOB has chosen in such cases to follow without exception the norm rather than the use, a choice that at times has resulted in more or less defective or contradictory entries.

**NYCKELORD:** Svenska Akademiens ordbok (SAOB), historisk lexikografi, deskriptiva och normativa ordböcker, stavningsnormering

Att tala om *Svenska Akademiens ordbok* (SAOB) som en normativ ordbok kan tyckas lite märkligt. SAOB brukar ju – inte minst av oss redaktörer – betraktas som en utpräglad deskriptiv ordbok, en som har till uppgift att beskriva det faktiska språkbruket från 1521 och fram till vår tid, inte att slå fast hur detta bruk bör vara eller borde ha varit.

Inte desto mindre finns det normativa inslag i SAOB. Och det har inte bara att göra med det faktum att även den mest deskriptivt inriktade ordbok i viss utsträckning fungerar normerande (se t.ex. Svensén 2004:29 f. liksom diskussionen i LexicoNordica 2002). Det handlar också om att SAOB:s självbild som



en mer eller mindre renodlat deskriptiv ordbok är en ganska modern företeelse; så var det inte alls i ordbokens begynnelse. I den programförklaring som medföljde det första häftet av SAOB 1893 sägs uttryckligen att ”ordboken vill verka för riktig användning af vårt modersmål i tal och skrift”, och denna normativa inriktning kommer tydligt till uttryck i de många värdeomdömen som möter i de tidigare banden. Där finns gott om explicita domar över språkbruket (och språkbrukarna), i stil med att något använts ”äfv[en] hos de bästa förf[attare]”, att något begagnats ”ända till öfvermått”, att man ”hellre” ska använda en viss form än en annan eller att vissa former är ”berättigade” – och ännu under bokstaven S kan man träffa på den klart värderande markören ”i vårdat spr[åk]” (se vidare Larsson 2009:41 f.).

Viktigt i sammanhanget är likaså SAOB:s förhållande till *Svenska Akademiens ordlista* (SAOL), den ordbok som alltsedan sin första upplaga 1874 framträtt med mer eller mindre uttalade anspråk på att utgöra rättesnöret för svenskans morfologi och ortografi. Även om de två ordböckerna numera i stort sett bara har sin huvudman gemensam fanns tidigare mycket intimare kopplingar. Huvudaktören bakom SAOL:s första upplaga, J. E. Rydqvist, betraktade ordlistan som ”ett slags ram till Ordboken” (citerat efter Sigurd 1986:194), och ända fram till den tionde upplagan 1973 tillkom SAOL vid SAOB:s redaktion i Lund, i två av upplagorna med ordbokschefen som huvudredaktör (Sigurd 1986:195 ff.).

Med tanke på de historiska förutsättningarna – såväl SAOB:s tidigare normativa ambitioner som de nära relationerna till SAOL – är det knappast förvånande att SAOB haft vissa svårigheter att hantera sådana fall där SAOL:s norm inte helt överensstämmer med bruket. En god illustration av detta ger stavningen av lånord med det betonade suffixet *-ett*. För den obetonade stavelse som föregår den betonade slutstavelsen föreskriver normen här i vissa fall dubbeltecknad konsonant, i andra fall enkeltecknad: så har SAOL t.ex. *dubblett*, *honnett*, *kassett*, *pollett* och *sufflett* med dubbelteckning, men däremot *palett*, *sonett*, *stafett*, *subrett* och *triplett* med enkelteckning. Vilka morfologiska eller etymologiska överväganden som varit bestämmande i de enskilda fallen är långtifrån alltid så självklart, och det går knappast att komma ifrån att det ibland verkar lite godtyckligt vilken form SAOL kom att fastna för när dessa ord togs med i ordlistan (samtliga av exemplen ovan infördes i antingen den sjätte upplagan 1889 eller den sjunde 1900, då lånorden gavs ett betydligt större utrymme än tidigare). Det återkommande argumentet för att t.ex. stava *dubblett* med <bb> är att det hör samman med *dubbel* (så skriver Språkrådet på sin hemsida att man bör ”[o]bservera att *dubblett* stavas med två b för att visa på samhörigheten med *dubbel*”). Invändningen att *triplett* då borde skrivas med <pp> eftersom det ju är lika tydligt samhörande med *trippel* ligger förstås nära

till hands. (Faktum är också att SAOL i upplagorna 1900 och 1923 hade stavningen *triplett*; det var först i den nionde upplagan 1950 som det i all tyshet<sup>1</sup> ändrades till *triplett*.)

De ord på *-ett* som ställer till problem för språkbrukarna är i första hand de som enligt normen ska ha dubbeltecknad konsonant. Det faktum att stavelserna är obetonade innebär ju att dubbelteckningen kommer att sakna all uttalsrelevans, och det är lätt att konstatera att stavningar som *dublett*, *kasett*, *polett* och *suflett* varken är eller har varit särskilt ovanliga. Exempelvis ger en sökning i Språkbankens konkordanser 67 träffar på ”dublett\*” och 24 på ”dublett\*” medan motsvarande sökning i *PressText* (omfattande en stor mängd tidningstext från 1990 och framåt) ger 286 respektive 146 träffar.<sup>2</sup> Och går man längre tillbaka i tiden är det uppenbart att *dublett* var den dominerande stavningen. I t.ex. databasen *Digitaliserade svenska dagstidningar* (en samling tidningstexter som har sin tyngdpunkt i årtiondena kring förra sekelskiftet) ger en sökning på ”dublett”, ”dubletter”, ”dubletten” och ”dubletterna” sammanlagt 766 träffar medan motsvarande former av *dubblett* bara uppvisar 189 träffar.

Reglerna för hur SAOB – åtminstone numera<sup>3</sup> – ska hantera sådana ord där språket har två levande stavningsformer finns formulerade i *Handbok för redigeringen av Svenska Akademiens ordbok*, utarbetad av den tidigare ordbokschefen Hans Jonsson. Där kan man bl.a. läsa följande (1999:82): ”Varierande form eller stavning av ordet i nusvenska medför obligatoriskt två (eller flera) uppslagsformer i SAOB.” Vad detta innebär i praktiken framgår av följande exempel från 1968 (S 2797) respektive 1985 (S 10010):

**SJAPPA** *ʃap³a²* l. **SCHAPPA** *ʃap³a²*, v. -ade. vbal-sbst. -ANDE, -NING, -ARE (numera bl. tillf., *NKristianst-Bl.* 1870, nr 94, s. 1); jfr **SJAPP**, sbst.<sup>1</sup> (*schappa* 1847 osv. *sjappa* 1872 osv.) [kortform till *chappera* (se *ECHAPPERA*). — Jfr **SJAPPAGE**]

<sup>1</sup> Bland de ortografiska förändringar som Holm 1951 redogör för nämns inte övergången till *triplett*. Dock ligger den helt i linje med ordlistans strävan ”att om möjligt göra stavningen lite lättare [originalets spärning] och att ta hänsyn till vad som kännes naturligt” (s. 55 f.).

<sup>2</sup> Att så pass många exempel på stavningarna *dublett* och *kasett* fortfarande förekommer är desto mer anmärkningsvärt som det förvaldsinställda stavningsprogrammet i Windows inte – som för t.ex. *polett* och *suflett* – nöjer sig med att ge dem röda understrykningar utan korrigerar dem automatiskt.

<sup>3</sup> De tidigare anvisningarna för SAOB-redaktörer, dels Ernst Kocks handbok från 1904 och dels den s.k. Codex Anders från 1970-talet, går överhuvudtaget inte in på frågan.

**SPREJ** *spräj*<sup>4</sup> l. **SPRAY** *spräj*<sup>4</sup> l. (numera föga br.)  
**SPRÄ** *sprä*<sup>4</sup>, r. l. m. (*SD(L)* 1895, nr 323, s. 4, osv.), äv. n.  
(*SvD* 1974, nr 72, s. 15, osv.); best. **-en** resp. **-et**; pl. **-er**  
(*IllSvOrdb.* (1964) osv.) l. **-ar** (HOLMGREN *Öronsj.* 113 (1925)  
osv.). (**spray** 1892 osv. **sprej** 1964 osv. **sprä** 1895–1948)  
[liksom t. o. fr. *spray* av eng. *spray*, till ä. eng. *spray*, stänka o.  
d., av mnl. *spraeyen*, motsv. mht. *spræwen*, *spræjen*; i avljuds-  
förh. till sv. dial. *språ(s)*, gro, skjuta brodd, spricka ut o. d.;  
rotbesläktat med **SPRUTA**, spruta ut o. dyl., o. (avlägsnare) gr.  
*σπειρω*, sår m. m.]

Här finns alltså vid sidan av huvuduppslagsformerna *sjappa* och *sprej* också sido-  
uppslagsformerna *schappa* respektive *spray* utan några bruklighetsinskrän-  
kningar (i den senare artikeln finns dessutom ytterligare en sidouppslagsform  
*sprä* med begränsad bruklighet). Att formerna *schappa* och *spray* är fullt brukliga  
(eller åtminstone var det vid den tidpunkt när artiklarna i fråga tillkom) framgår  
också av de s.k. formparenteserna där årtalet för förstabelägget följs av ett ”osv.”  
(jfr *sprä* som försetts med ett årtal för den sista dokumenterade förekomsten).

Denna beskrivningsmodell med dubbla uppslagsformer förekommer emel-  
lertid inte alls när det gäller orden med suffixet *-ett*, trots att bruket i vissa fall  
rimligtvis borde ha motiverat det. Här har i stället SAOL:s norm fått styra: i  
samtliga de fall som orden på *-ett* förekommer i SAOL är dess form allena-  
rande som uppslagsform i SAOB.

Hur SAOB:s redaktörer i stället har valt att hantera denna ortografiska vack-  
lan varierar, och i huvudsak kan man urskilja tre olika förfaringssätt. Ett sätt att  
undkomma problemet som för en nutida redaktör inte skulle komma på fråga  
är att man helt enkelt inte noterar eller bryr sig om variationen, trots att den  
finns representerad bland språkproven. Paradexemplet här är artikeln CIGAR-  
RETT som trycktes 1904 och vars huvud ser ut på följande sätt (C 157):

**CIGARRETT** *sig<sup>4</sup>arät<sup>4</sup>* l. *si<sup>4</sup>g-*, r. l. m.; best.  
**-en**; pl. **-er**. (*cigarretto SvT* 1852, nr 27, s. 3 (*på tal*  
*om en spanjor*)) [jfr t. *cigarette*, f., eng. *cigarette*, af  
fr. *cigarette*, f., dimin. af *cigare*, motsv. span.-amer.  
*cigarrito*, diminutiv af *cigarro* (se föreg.)]

Här har vi alltså en uppslagsform med dubbeltecknat <rr> och en formparentes  
som inte ger den minsta antydning om en alternativ stavning med bara ett <r>. Tittar man  
närmare på de autentiska språkprov som illustrerar artikeln (C 158 f.) visar det sig  
emellertid att den stora majoriteten av dem har stavningen med <r>: av de 18 gånger  
ordet uppträder i språkproven är det i 14 fall skrivet *cigaret* medan stavningen  
*cigarrett* bara förekommer i 4 fall. *Cigar(r)ett* är för

övrigt det enda av orden med suffixet *-ett* där SAOL numera accepterar alternativa former: sedan den tionde upplagan 1973 träffar man på båda stavningarna i ordlistan.<sup>4</sup> Men dessförinnan, från det att ordet infördes i den sjätte upplagan 1889, var det enbart dubbelteckningen som gällde, och uppenbarligen var det detta som var bestämmande för den SAOB-redaktör som skrev artikeln CIGARRETT.

Det finns ytterligare några exempel på att enkelteckningen finns representerad bland språkproven men utan att denna form överhuvudtaget tas upp i artikeln huvud: det gäller ASSIETT (A 2515) från 1903 (även där har majoriteten av språkproven enkeltecknad konsonant), INHONNETT (I 528) från 1933, och – som sista exempel – NOVELLETT (N 738) från 1947 (med ett språkprov ”Gades noveletter” från 1933).

Att på detta sätt ignorera de stavningar som inte överensstämmer med normen är otänkbart idag; numera är ordboken noga med att all stavningsvariation som inte kan klassificeras som rena skriv- eller tryckfel ska noteras i ordboken. De återstående två förfaringsätten träffar man däremot på ända in i modern tid. I den ena utnyttjas möjligheten att – som det heter i SAOB-sammanhang – *anföra* stavningsvariationen men utan att *redovisa* den, dvs. båda stavningarna förtecknas i formparentesen men utan att dateras var för sig (se Jonsson 1999:83). Det här är en beskrivningsmodell som ofta används i fråga om äldre stavningsvarianter som saknar uttalsrelevans och ”bedöms såsom enbart ortografiska” (1999:84), så den borde ju väl kunna försvara sin plats också när det gäller den här stavningsvariationen – under förutsättningen att variationen inte längre förekommer. Ett exempel är det numera obrukliga ordet *pus(s)ett* i en artikel från 1955 (P 2482):

**PUSSETT**, sbst.1, r. l. f.; best. -en. (**poussette** 1843. **pussett** (pus-) 1845–1885) [jfr eng. *poussette*; diminutivum till ä. fr. *pousse*, stöt (jfr fr. *poussette*, benämning på dels viss barnlek med nålar, dels vissa vagntyper), till fr. *pousser*, stöta m. m. (se **PUSSERA**, v.1)] (†) stöt, puff l. dyl. KNORRING *Förh.* 2: 365 (1843). Han .. får en icke så liten pusett i pannan (av dörren). DENS. *Skizz.* II. 2: 270 (1845). SCHULTHESS (1885).

<sup>4</sup> Enligt Mattsson 1974 togs variantformen *cigarett* med i SAOL ”i anslutning till *cigarill* och med anknytning till ordets stavning i danska, norska, engelska, franska och tyska” (s. 109); redan 1957 hade Nämnden för svensk språkvård här uttalat sig för stavningen *cigarill* ”i överensstämmelse med uttalet och med skrivningen i flertalet europeiska språk” (se Bergman 1962:57 f.). Motiveringarna för enkelteckningen – såväl uttalet som anknytningen till andra språk – skulle med lika stor rätt kunna appliceras på *dublett*.

Här har alltså stavningarna med <s> och <ss> behandlats i klump, ett smidigt och enkelt sätt att få fram variationen. Om det däremot gäller ord där de två stavningsvarianterna fortfarande är i bruk så blir beskrivningsmodellen betydligt mera tveksam. Där borde det ju – som i följande exempel från 1935 (K 693) – egentligen finnas två uppslagsformer:

**KASSETT** *kasät*<sup>4</sup>, r. l. f. l. m. (m. UB 4: 600 (1873)); best. -en; pl. -er. (förr äv. skrivet *ca-cas(s)ette* 1758–1796. *kas(s)ett* (ca-) 1814 osv. *kas(s)etten* (ca-), sg. best. 1712 osv.) [liksom d. *kas-set(te)*, t. *kassette* o. eng. *cassette*, av fr. *cassette*, av it. *cassetta*, dimin. till *cassa*, kista (se KASSA, sbst.1)]

Faktum är att den här artikeln skulle kunna läsas som ett – om än starkt nedtonat – medgivande att stavningen med <s> fortfarande förekommer, inte minst i och med att stavningen med <c> i formparentesens inledning explicit förklaras vara obruklig; dock strider det ju mot regeln att fortfarande brukliga stavningsvarianter också ska resultera i sidouppslagsformer.

Värt att notera är att exemplet KASSETT utgör något av ett undantag; denna beskrivningsmodell är förhållandevis ovanlig när det gäller sådana ord på -ett där normen har dubbelteckning. Däremot förekommer den ofta när enkelteckningen är norm; det gäller t.ex. för samtliga de fem tidigare nämnda exemplen PALETT (P 73, tryckt 1952), SONETT (S 8870, slutredigerad 1982), STAFETT (S 10795 f., 1986), SUBRETT (S 14361, 1997) och – som hittills färskaste exempel – TRIPLETT (T 2522, 2007):

**TRIPLETT** *triplät*<sup>4</sup>, r. l. m. (DALIN (1854) osv.) ((†) n. RP 8: 655 (1641)); best -en (ss. n. -et); pl. -er. (*triplett* (-pp-) 1836 osv. *triplette* 1839–1897. *tripleter*, pl. 1632 osv. *triplettet*, sg. best. 1641) [bildat medelst TRI- i analogi med DUBBLETT]

Förklaringen till denna ojämna fördelning ligger antagligen i att språkbrukarna när det gäller orden på -ett tenderar att föredra enkelteckningen, och om denna också överensstämmer med normen kommer fallen av dubbeltecknad konsonant att uppträda så pass tillfälligt att redaktörerna med relativt gott samvete kan låta dem passera på detta sätt.

I de fall dubbelteckningen är norm är det i stället en annan beskrivningsmodell som dominerar. Denna innebär att båda varianterna *redovisas* i formparentesen, dvs. stavningarna med och utan dubbelteckning av konsonanten dateras var för sig. Även här är saken oproblematisk så länge det rör sig om ord som numera är obrukliga, som i följande artikel från 1986 (S 10804):

**STAFFLETT** (Staflétt WESTE, stafflä'tt DALIN), r. l. m. (LIND (1749), *SvTyHlex.* (1872)) l. n. (LIND (1749)); best. (ss. r. l. m.) -en; pl. -er. (staf- 1734—1841. staff- 1790—1896. staffe- 1817. stav- c. 1755. -let 1734—c. 1755. -lett 1749—1896) [liksom ä. d. *staffelet*, av ä. t. *staffelet*, biform till *staffelei* (se STAFFLI), sannol. uppkommen under påverkan av fr. *chevalet*, staffli] (†) staffli; äv. bildl. SERENIUS Q 2 b (1734). Ett stort Rum i frontispicen skulle emottaga Solens första strålar och hysa mitt Ritbord och min Staflett. LINNERRHIELM *IBr.* 165 (1797). Naturen sjelf sig satt vid hans (*dvs. Runebergs*) staflett / Och i hans hand sin egen pensel gett. FRANZÉN *Skald.* 7: 138 (1841). HALLSTÉN o. LILIUS (1896).

Knepigare blir det när det handlar om ord som fortfarande brukas i båda sina stavningsformer. Alternativet att (som i fallen SJAPPA och SPREJ ovan) i formparentesen ha ett ”osv.” vid båda stavningsvarianterna är ju inte möjligt eftersom SAOB:s regler då skulle kräva dubbla uppslagsformer – vilket alltså inte förekommer bland dessa ord. I stället har lösningen blivit att den stavningsvariant som inte överensstämmer med normen förpassas till historien genom att förses med ett årtal för den sista förekomsten. Så är bl.a. fallet i artikeln DUBBLETT, tryckt 1922, vars huvud ser ut som följer (D 2330):

**DUBBLETT** *dublät*<sup>4</sup>, i bet. I 1—2, 4 o. II äfv. (i sht i lärddt fackspr., särsk. i bet. I 1 *b*) **DUPLETT** *dup-*, sbst. o. adj.; ss. sbst. r. (m. SAHLSTEDT, LUNDELL; f. DALIN (1871)) ((†) n. ÅOXENSTIERNA 2: 408 (1620), *HT 1895*, s. 433 (*i handl. fr. 1645*)); best. -en ((†) ss. n. -et ÅOXENSTIERNA 2: 408 (1620), *HT 1895*, s. 433 (*i handl. fr. 1645*)); pl. -er. (dubbl- 1620 osv.; dubl- 1645—1904; doub(b)l- 1632—1871; dobl- 1613. dupl- 1668 osv.; doupl- 1628—1791. -ett 1620 osv. -et 1613—1888. -ette 1791—1871)

I. [jfr d. *doublet*, t. *doublette*, af fr. *doublet*, af *double* (se DUBBEL, adj.); formen med p beror på direkt anslutning till lat. *duplus* (se DUBBEL, adj.)] sbst.

Här har alltså varianten med enkeltecknad konsonant getts ett slutårtal 1904, grundat på ett i artikeln citerat språkprov från Henrik Schücks *Studier i nordisk litteratur- och religionshistoria*: ”Domar ser ut att vara blott en dublett af föregångaren”. Att denna gränsdragning i tiden, med slutårtalet mindre än 20 år före tryckåret, utgick från normen – *dublett* hade införts i SAOL:s sjunde upplaga 1900 – och inte från bruket är något som framgår också ur SAOB själv. För det första finns redan i artikeln DUBBLETT ett yngre språkprov med enkeltecknad konsonant. Det gäller sammansättningen DUBBLETT-KAMRAT (D

2331), där ordbokens första och enda belägg är hämtat från Olof Högbergs roman *Fribytare* från 1910. I SAOB:s artikel är språkprovet visserligen inte citerat, där finns enbart ett s.k. naket belägg i form av en källangivelse, men går man till källan ser det ut så här: ”Hennes dubblettkamrat – det var en liten trevlig danska” (s. 362). Det kan förstås röra sig om ett redaktionellt förbiseende. En annan möjlighet är dock att det uttrycker en värdering från redaktörens sida; kanske räknades inte Olof Högberg – till skillnad från akademiledamoten Henrik Schück – bland ”de bästa författare” vilkas stavningsvarianter ansågs vara värda att tidfästa och dokumentera.

För det andra kan man konstatera att stavningen *dubblett* utan tvivel var den naturliga också för många inom SAOB:s redaktion. Dels är stickorden på de excerptlappar i SAOB:s arkiv som utgör underlaget för artikeln DUBBLETT påfallande ofta skrivna med ett <b>, även i sådana fall där källan har dubbelteckning (se SAOBs).<sup>5</sup> Dels har stavningen *dubblett* vid åtminstone tre tillfällen smugit sig in i SAOB:s redaktionella text. Det äldsta exemplet träffar man på i definitionen av AFSKRIFT (A 506), tryckt redan 1895, där det talas om ”dubblett af akt l. dokument”. Det andra exemplet återfinns i artikeln BILDNING från 1910, alltså sex år efter det angivna slutårtalet, där man ger hänvisningen DUBLETT-[BILDNING] (B 2563 f.). Det tredje fallet slutligen hör hemma i nyare tid, i en artikel slutredigerad 1986; det gäller definitionen av STAMKORT, som i en av sina betydelser uppges vara ”motsatt: dubblettkort” (S 10913).

Detta förfaringssätt, att ge ett – som det förefaller – mer eller mindre fiktivt sista årtal för förekomsten av den enkeltecknade varianten, är inte begränsat till SAOB:s inledande band utan uppträder också längre fram i alfabetet. Bl.a. gäller det de tre tidigare nämnda HONNETT (H 1132, tryckt 1932), POLLETT (P 1423, tryckt 1953) och SUFFLETT (S 14433 f., slutredigerad 1997):

**HONNETT** *honüt*<sup>4</sup>, adj. **-are**; n. o. adv. =.  
 (honet 1740—1790. honett 1692—1897. honette 1735—1791. honnet 1716—1849. honnete 1784. honnett 1770 osv. honnette 1782) [av fr. *honnête*, ffr. *honeste*, motsv. it. *onesto*, span. o. port. *honesto*; av lat. *honestus*, hederlig, ansedd (se HONEST)]

<sup>5</sup> T.o.m. de citerade språkproven har ibland stavningen med <b> trots att källan har <bb>. Det gäller t.ex. samtliga nio lappar med excerpter från Adolf Noreens *Vårt språk* från 1906 (s. 59 ff.). Medan Noreen genomgående skriver *dubblett* använder sig excerptisten däremot – säkerligen utan att alls tänka på det – lika genomgående av *dubblett*.

**POLLETT** *polät*<sup>4</sup>, äv. *pω-*, sbst.<sup>2</sup>, förr äv. **BOLLET**, r. l. m. (*2RARP* I. 2: 27 (1720) osv.) ((†) n. STIERNMAN *Com.* 1: 890 (1623), *GullbgDomb.* <sup>29</sup>/1 1656); best. **-en**; pl. (äv. att hänföra till sg. **-ette** l. **-äte**) **-er** (HOLMKVIST *BergslHyttsspr.* 73 (*i handl. fr.* 1557) osv.) ((†) **-en** SCHRODERUS *Os.* III. 1: 159 (1635)); förr äv. **POLETTE** l. **POLÁTE**, r. l. m. (l. n.?) med avs. på pl. se ovan. (bollet 1623. bolletter, pl. 1623. palletter, pl. 1845. polet 1680—1766. polett c. 1764—1944, 1952 (*i bet.* 10). polette 1796. poletter, pl. 1627. pollet 1557—1921. pollett 1771 osv. polletter, pl. 1663 osv. poläte 1640. påleten, pl. 1635. påleter, pl. 1625) [av ä. t. *bollett*, *bolette*, *pallet*, assignation, ett slags mynt (av bly), sannol. av ä. fr. *bullet(t)e*, (litet, runt) sigillavtryck, intyg, attest, biljett m. m. (se **BILJETT**). — Jfr **BULLA**, **BULLETIN**]

**SUFFLETT** *sufflät*<sup>4</sup>, äv. *suf-*, r. l. m.; best. **-en**; pl. **-er**. (**soufflet** 1799—1848. **soufflett** (**-ette**) 1808—1854. **soufflett** (**-ette**) 1808 (: Soufflette Vagn)—1826. **sufflet-** i ssg 1877—1971. **sufflett** 1824 osv. **sufflett** 1879 (: sufletthudar)—1958) [av fr. *soufflet*, blåsbälg, bälgliknande anordning, av *souffler*, blåsa (se **SUFFLERA**); stavningen med **-tt** sannol. efter ord som **KADETT**, **LAVETT**, **SERVETT**]

Mest anmärkningsvärt här är fallet POLLETT, där stavningen med enkeltecknad konsonant är belagd bara ett år innan artikeln trycktes – närmare än så kan man ju knappast komma den ”varierande [...] stavning av ordet i nusvenska” (Jonsson 1999:82) som enligt reglerna obligatoriskt ska leda till dubbla uppslagsformer. Att förfaringsättet i artiklarna POLLETT och DUBBLETT är detsamma betyder dock knappast att det handlade om samma slags bevekelsegrunder. Vad gäller DUBBLETT lät sig upphovsmannen förmodligen så till den grad styras av SAOB:s då ännu högst aktuella normativa uppgift att han medvetet valde att bortse från språkprov som bröt mot normen. I fallet POLLETT däremot är det tydligt att redaktören nogsamman redovisade de språkprov som utgjorde hans arbetsmaterial; här var det snarast frågan om ett – inte helt lyckat – försök att inom ramarna för en i en äldre normativ tradition etablerad beskrivningsmodell foga in en språklig mångfald som inte riktigt rymdes där.

Givetvis är denna problematik inte begränsad till orden på *-ett*. Ett annat exempel på hur oförmågan att helt och fullt erkänna stavningsvarianter som inte är godtagna av SAOL leder till att SAOB:s artiklar biter sig själva i svansen är PETIMÄTER (P 764), i likhet med POLLETT tryckt 1953. Här finns variantformen PETIMETER visserligen med som sidouppslagsform – i detta fall



har ju variationen uttalsrelevans – men i formparentesen inskränks dess bruklighet till åren 1760–1952, alltså även här fram till bara ett år innan artikeln gick i tryck.

## KÄLLOR OCH LITTERATUR:

- Bergman, Gösta, 1962: Rätt och fel i språket. (Skrifter utgivna av Nämnden för svensk språkvård 26.) Stockholm.
- Codex Anders = Anvisningar för SAOB-redaktörer sammanställda i sedelform av Anders Sundström, 1970-talet. (I: SAOBs.)
- Digitaliserade svenska dagstidningar: <http://magasin.kb.se/searchinterface/> (augusti 2011)
- Holm, Pelle, 1951: Nytt och gammalt i Svenska Akademiens ordlista. (Skrifter utgivna av Nämnden för svensk språkvård 8.) Stockholm.
- Högberg, Olof, 1910: Fribytare. Förra delen. Stockholm.
- Jonsson, Hans, 1999: Handbok för redigeringen av Svenska Akademiens ordbok. 2 uppl. Lund. (Otryckt manuskript. I: SAOBs.)
- Kock, E. A., 1904: Handbok för redigeringen af Svenska Akademiens ordbok. Tryckt som manuskript. Lund.
- Larsson, Lennart, 2009: Hur märker användaren av SAOB de 110 åren mellan A och TYNA? I: Språk och stil NF 19, s. 35–52.
- LexicoNordica 9, 2002. Red. Henning Bergenholtz & Sven-Göran Malmgren. S. 1–172.
- Mattsson, Gösta, 1974: Ordlistans tionde upplaga – riktlinjer och kommentarer. I: Johannisson, Ture & Mattsson, Gösta: Svenska Akademiens ordlista under 100 år. (Skrifter utgivna av Svenska språknämnden 55.) S. 55–116. Stockholm.
- Noreen, Adolf, 1906: Vårt språk. Nysvensk grammatik i utförlig framställning. Sjunde bandet. Lund.
- PressText Online: <http://www.presstext.se/> (sökning 23/8 2011)
- SAOB = Ordbok över svenska språket utgiven av Svenska Akademien (Svenska Akademiens ordbok), 1893 ff. Lund.
- SAOBs = Samlingar till Ordbok över svenska språket utgiven av Svenska Akademien. Lund.
- SAOL = Svenska Akademiens ordlista över svenska språket, 1874–2006. 1–13 uppl. Stockholm.
- Sigurd, Bengt, 1986: Ordboken, ordlistan och några andra av Svenska Akademiens språkliga insatser under 1900-talet. I: Allén, Sture, Loman, Bengt & Sigurd, Bengt, 1986: Svenska Akademien och svenska språket. Tre studier. S. 143–231. Stockholm.
- Språkbankens konkordanser: <http://spraakbanken.gu.se/konk/> (augusti 2011)
- Språkrådet: [http://www.sprakradet.se/servlet/GetDoc?meta\\_id=1950](http://www.sprakradet.se/servlet/GetDoc?meta_id=1950) (augusti 2011)
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.

*Lennart Larsson*

Institutionen för nordiska språk vid Uppsala universitet

Svenska Akademiens ordbok, Lund

e-postadress: [lennart.larsson@nordiska.uu.se](mailto:lennart.larsson@nordiska.uu.se); [lennart.larsson@svenskaakademien.se](mailto:lennart.larsson@svenskaakademien.se)

# Atypiske lemmaer i Det Norske Akademis Store Ordbok

## Noen betraktninger om deres plass og funksjon i en moderne diakron og deskriptiv ordbok

*Hanne Lauvstad*

*Det Norske Akademis Store Ordbok* ('The Norwegian Academy's Dictionary'), an extended and updated version of *Norsk Riksmålsordbok* ('The Norwegian Riksmål Dictionary', vols I–IV 1937–57, vols V–VI 1995) will be a descriptive, diachronic dictionary covering the Norwegian Bokmål and Riksmål, and, as such, one of two national dictionaries for the Norwegian language. As modern Bokmål has its origins in both Old Norse and Danish, the dictionary's vocabulary poses certain challenges to lemmatization. The dictionary contains several archaisms and dialect words, i.e. diachronic and diatopic variants, among them several atypical lemmas, which have to be adapted and standardized as part of the dictionary's macrostructure. Examples illustrate how this problem may be solved in the light of the basic criteria for the lemmatization: etymology, part of speech, gender and inflection, style nuance and usus.

*Nøkkelord:* lemmatisering, diatopisk og diakron varians, makrostruktur

### Bakgrunn

Det Norske Akademis Store Ordbok (NAOB) skal publiseres digitalt i 2014 og er en utvidelse og modernisering av Norsk riksmålsordbok (NRO). NRO utkom i fire bind i årene 1937–57, pluss to supplementsbind i 1995. NAOB skal være en deskriptiv, diakron ordbok for bokmål og riksmål. Redigeringspråket er moderat bokmål. NRO er hovedsakelig basert på ekserpter fra norsk skjønnlitteratur fra 1814 og fremover. Vi arbeider med å supplere basen med ord fra nyere skjønnlitteratur og sakprosa og dessuten med å øke tilfanget av nyord, bl.a. anglisismer, fagord som er gått inn i allmennspråket og slang. Dessuten skal det eksisterende materialet moderniseres språklig og innholdsmessig der det er behov for det. Endret syn på språket og strengere krav til leksikogra-

fisk metode er også grunner til å revidere det ordboksmaterialet som stammer fra NRO.

NRO har en lemmabestand som bl.a. består av en rekke av det jeg i tittelen har valgt å kalle *atypiske lemmaer*. Vår utfordring i NAOB er å definere lemmatiseringsprinsipper<sup>1</sup> som også passer til disse atypiske lemmaene, og å gjennomføre dem. Hovedprinsippet for lemmatisering i NAOB er etymologi.<sup>2</sup> Men også faktorer som gjelder form (ordklasse og genus) eller semantikk (f.eks. at betydningen er radikalt forskjellig fra den opprinnelige) vektlegges, jf.

**ask**<sup>3</sup> (subst., m.) (av norr. *askr*)

I 1. (bot.): 'løvtreart i oljetrefamilien, med motsatte, ulikefannede blader; lat. *Fraxinus excelsior*';

2. (koll.): 'virke av dette treet'

II 1. (dial.): 'eske'; 2. (om norrøne forh.) hulmål for væske = 4 boller | jf. *bolle* og *juste*

Merk at dette er satt som to forskjellige lemmaer i NAOB. Lemmatiseringssystemet er imidlertid ikke helt konsekvent i NRO. Til tross for ordklasseskifte har det f.eks. vært vanlig å anføre adjektiv- el. adverbavledninger under et tilhørende substantiv, sannsynligvis for å spare plass i papirordboken NRO. Dette er nå endret i NAOB, slik at ny ordklasse gir nytt lemma, jf.

**atlask** (subst., m.) (via mnty. *at(te)las*, *atlask*, fra arab. 'glatt'): 'glinsende stoff, særl. av silke, med en noe løs binding av trådene slik at renningen ses på et lengre stykke'

**atlaskes** (adj.) (avl. av *atlask*): 'som er laget av atlask'

EKSEMPEL: *en atlaskes kjole*

Noe fast system i NRO for når betydningen er så forskjellig at det er grunnlag for å opprette et nytt lemma med egen artikkel, er det heller ikke lett å få øye på. Det er f.eks. uklart hvorfor man har valgt å lemmatisere *bakke* (subst., m.) i betydningene 'lite serveringsbrett' og 'line [til å fiske med]' all den stund det siste skal ha en opprinnelig betydning 'brett som linen er viklet rundt' og derfor etymologisk tilhører samme lemma som det første. Begge ordene er varianter av *bakk* (subst., m.), som er lånt fra nty. *back* el. nederl. *bak* 'fat, trau'; oppr. fra romansk.

<sup>1</sup> Jf. Wolskis term *Lemmaverfabren* (Wolski 1989: 366).

<sup>2</sup> I pakt med Nordisk leksikografisk ordbok (NLO 173f.) bruker jeg her termen *lemmatisering* om det Wolski (med referansen til Wiegand 1983) kaller *Lemmansetzen* (Wolski 1989: 365).

<sup>3</sup> Eksemplene fra NAOB er forkortet og redigert her.

## Problemer – og litt om årsakene til dem

NRO inneholder en rekke lemmaer som avviker fra standardordforrådet. For det første skyldes dette den diakront-deskriptive hensikten med ordboken. Det norske bokmålet har en sammensatt karakter med røtter i så vel norrønt som dansk skriftspråk. Når man vil dokumentere språket diakront, må man derfor ta hensyn til en rekke arkaismer, samtidig som det dukker opp mange dialektale ordformer som må innpasses i ordbokens makrostruktur. Årsaken til atypiskheten er i slike tilfeller altså enten diatopisk eller diakron. For det andre skyldes avvikene metoden. Lemmabestanden i NRO – og dermed også i NAOB – er blitt til ved ekserpering. Dette har medført at man har hatt en viss tendens til å feste seg ved rariteter, dvs. leksikalske enheter som i den litterære riksmåls-/bokmålstradisjonen har skilt seg ut fra det øvrige ordforrådet. Disse leksikalske enhetene er blitt opphøyet til atypiske lemmaer i NRO.

I NRO er mange leksikalske enheter som representerer slik ulik bruk – språkets *parole*, blitt behandlet atomistisk som egne hovedlemmaer, med til dels vilkårlig normering og med liten vekt på deres plass i språkssystemet som helhet, *langue*. Det forekommer ikke så helt sjelden at en leksikalsk enhet er definert med en lydlig variant av seg selv, f.eks. er den dialektale varianten *bløyg* blitt definert med den normerte formen *blei*, jf.

*blei* (subst., m.) (av norr. *bleðgi* 'flis'): 'kile; tynn, smal skive som man skjærer av noe'  
*bløyg* (subst., m.): se *blei*

Begge formene er norske og av samme etymologiske opprinnelse. Merk at lemmatiseringen er endret i NAOB i forhold til i NRO, slik at den første formen er representert med et hovedlemma i NAOB, mens den siste derimot er oppført som et henvisningslemma som er lenket til hovedlemmæet.

Den leksikografiske utfordringen i arbeidet med den moderne og utvidede ordboken NAOB er å ta hensyn til diakron og dialektal variasjon i ordbokens lemmabestand, dvs. *parole*, samtidig som man velger ut hvilken form av en leksikalsk enhet (jf. Zgustas term *lexical unit*, Zgusta 1971: 249 ff.) som skal danne (hoved)lemma i ordboken, altså det som representerer *langue*.

## Diskusjon

### Varianter

Variantene i NAOB kan være a) ortografiske varianter av samme leksikalske enhet (f.eks. *frilans/freelance*, med norvagisert vs. opprinnelig engelsk ortografi), b) (ortografisk gjengivelse av) fonologiske varianter av samme leksikalske enhet

(f. eks. *trykk/trøkk*), der det siste er en dialektal form, som bl.a. hører hjemme i visse nokså faste kollokasjoner av muntlig karakter, som *få seg en trøkk* ('få hard medfart'). Man har også c) lemmaer uten tilsvarende lemmatiserte hovedformer i ordboken, f.eks. *alvskott* (subst., n.), '(i folketro) brudd el. brekk innvendig hos dyr', en leksikalisert sammensetning hvor sisteleddet ikke har noen parallellform 'skudd', som ellers ville være hovedformen i NAOB, men isteden den (sett fra moderat bokmålsstandpunkt) muntlige variantformen *skott*, som henger sammen med bruksområdet. Det finnes derfor ikke noe alternativ til hovedlemma.

### Kriterier for valg av hovedlemma

I arbeidet med NAOB må man ta stilling til når det er grunnlag for å slå sammen etymologisk beslektede varianter av en leksikalsk enhet til ett lemma og hvordan man skal behandle atypiske, enkeltstående varianter i ordboken. Kriteriene for å definere lemmaer som atypiske er enten *frekvens* eller *form*.

Dette er, med inspirasjon fra Bo Svenséns diskusjon av homonymi kontra polysemi (Svensén 2004: 116ff.), NAOBs kriterier for hva som bør være hhv. hovedlemmaer og henvisningslemmaer:

- etymologi
- ordklassetilhørighet
- genus og bøyning
- stilnivå og bruksområde

#### *Eks. 1: Normative og synkrone parallellformer i norsk bokmål med lignende stilvalør*

*ben* (subst., n.) fra da. *ben*; tilsv. norr. *bein* 'knokkel; legg (senere bet.)'

*bein* (subst., n.) av norr. *bein* 'knokkel; legg (senere bet.)'

Dette kan regnes som varianter av samme leksikalske enhet, men de er lemmatisert hver for seg i NRO med delvis overlappende betydninger og samlet under hovedlemmaet *ben*, med *bein* angitt som variant og som eget henvisningslemma i NAOB. Etymologien er parallell; den ene variantformen er nedarvet direkte fra norr. *bein*, den andre lånt inn fra dansk *ben*. I moderne bokmål blir disse variantformene imidlertid oftest oppfattet som likeverdige når det gjelder stilnivå og bruksområder. Den norske bøyningsformen i pl. på *-a* er nå blitt normen, tidligere ble den danske pl. formen *benene* brukt i norsk, men den er nå temmelig sjelden. La oss se nærmere på spørsmålet om stilvalør og normering.

*Eks. 2: Variantformer med ulik stilvalør og ulik vei inn i moderne bokmål*

**hu** (subst., m.) (fra da. *hu*; sa. ord som *hug*), (mest foreld. el. arkais. og i faste uttr.): 1. 'sinn; sinnelag; tanke(r)'

UTTRYKK:

*med berådd, velberådd hu* | se berå og velberådd

(*bli, være, gjøre*) *gram i hu* | se gram, adj.)

*komme i hu* 'komme i tanker om; komme på; huske' | jf. ihukomme

*gå el. rinne noen av hu* (foreld. el. arkais.) 'gå noen av minne, glemmes av noen' | jf.

omhu

2. (særl. om elskovsfølelse) lyst; attrå

**hug** (subst., m.) (av norr. *hugr*; sa. ord som hu):

1. (dial. el. arkais.) 'sinn; sinnelag; tanke(r)' | jf. *hu*;

2. 'lyst; attrå' | jf. *hu*

Formene tilsvarer hverandre etymologisk, begge har betydningene 1. 'sinn; sinnelag; tanke(r)'; 2. 'lyst; attrå', men de har ulik brukshistorie i norsk (som hhv. arveord og lånord) og (noe) ulik stilistisk valør, dessuten ser det ut til at *hu* sjelden bøyes. Ellers virker det ut fra ordbokens litterære belegg som om begge variantformer kan brukes i fast uttrykk som *komme i hu*, mens idiomatiske uttrykk som *med velberådd hu* 'etter moden overveielse', *gram i hu* 'forbitret' og *rinne noen i hu* 'dukke opp i minnet' utelukkende knyttes til den danske formen. Også Bokmålsordboka (BMO), som ellers har valgt *hug* som hovedlemma, oppgir disse uttrykkene utelukkende med formen *hu*. Begge variantformer vil bli beholdt i NAOB som to lemmaer med etymologiopplysninger. Her er også eksemplene blitt stående igjen under hhv. *hug* og *hu* for å skille dem tydelig ut. Praksis for plassering av autentiske eksempler varierer for øvrig fra tilfelle til tilfelle, jf.

**brudd/brott** (subst., n.): Den første formen er lånt fra da. *brud* og dominerer i (moderat) bokmål, den andre er arveord fra norr. *brot*; begge formene er dannet til verbet *bryte*. Etymologi, genus og bøyning stemmer over ens, mens bruksområdene bare delvis overlapper; *brott* har flere dialektale betydninger som mangler belegg i formen *brudd*, bl.a. 'bratt skråning el. bakke', 'brenning (bølge(r) som bryter mot kyst el. skjær)', 'elvestryk', 'brus; knak, brak'; 'avbruket stykke, skår, særl. av glass el. steintøy; bruddstykke', (på klær) 'oppslag, brett el. kant, især som står ut'. Men de fleste betydninger er felles, f.eks. 'det å brytes el. være brutt', 'sted hvor man bryter stein e.l.', 'det å få (et bestående forhold) til å opphøre', 'anstøt (mot), krenkelse (av et løfte, en lov, en sedvane e.l.)', 'forbrytelse' (den siste betydningen er imidlertid sjelden når det gjelder formen 'brudd').

Stilvaløren må også sies å være forskjellig, i moderat bokmål vil *brott* oppfattes som en radikal eller dialektalt preget form. Her er det altså stilistiske og semantiske egenskaper som gjør at man anser *brudd* og *brott* som to forskjellige leksikalske enheter og oppretter dem som to hovedlemmaer til tross for felles etymologi.

*Eks. 3: Former uten stilistisk motstykke i bokmål men med beslektet lemma*

Visse arkaistiske former mangler motstykke i moderne bokmål. Et eksempel er **vovet**: (adjektivisk perfektum partisipp av *vove*, dansk form av *våge*), en variantform som har andre konnotasjoner i moderne norsk enn den mer umarkerte formen *våget*. Her kunne man eventuelt ha opprettet et eget lemma for adjektivet, siden de øvrige av verbets bøyingsformer er nærmest fraværende i moderne norsk. Men som diakron ordbok må NAOB også gi plass til og dokumentere en rekke tilfeller av eldre, litterær bruk av variantformen, som ville bli mindre synlig dersom alle sitater ble flyttet til artikkelen *våge*. Igjen er det stilpreg og brukshistorie som gjør at man, på tross av etymologi, opprettholder et eget lemma for *vove*. Imidlertid vil vi i NAOB, av hensyn til den marginale rollen verbet i denne danske formen spiller i moderne norsk språkbruk, forenkle NROs omfattende artikkel for *vove*, og den adjektiviske perfektum-partisippformen vil bli fremhevet med eget betydningsnummer og behørig belagt med autentiske eksempler, også fra den senere tid.

*Eks. 4: Hapax legomena, unormerte dialektord og andre sjeldne leksikalske enheter*  
La oss se på noen eksempler på *hapax legomena*, unormerte dialektord o.a. sjeldne leksikalske enheter:

**abengslig** (adj.) (avl. av norsk dialektalt *avbendes*, *avbendeleg(e)* 'vanskelig tilgjengelig, bakvendt, avbaklig, uhåndterlig, overvettes'; se *-lig*), brukt med bet. 'urimelig; forkjært' og belagt med et autentisk eksempel fra P.Chr. Asbjørnsens *Huldreeventyr*. Lemmaet blir i NAOB sammenlignet (til slutt i artikkelkroppen) når det gjelder bruk og semantikk med de fonetisk lignende lemmaene *avbragdslig* og *abrenslig*.

**blodge** (vb.) (av norr. *blóðga*; avl. av *blóðugr*; se *blodig*; jf. *blogge*, *blygge* og *bløgge*) brukt dialektalt, med bet. 1. 'gjøre blodig; såre så blodet flyter', 2. 'slakte, stikke (fisk) slik at blodet renner av'. Lemmaet *blodge* forekommer alt i NRO og fremstår som et konstruktlemma, en restituert form i pakt med den etymologiske opprinnelsen, men uten autentiske belegg. I NAOB er dette valgt som hovedlemma. De lydlige, dialektale variantene *blogge*, *blygge* og *bløgge* er bevart som henvisningslemmaer av hensyn til *brukshistorien* (de autentiske eksemplene er bevart i artiklene som hører til henvisningslemmaene for å synliggjøre *usus*) og *brukerrelevansen* (en ordboksbruker kan ikke forventes koble



*blygge* til *blodge*). I NRO var alle disse variantene egne, isolerte (hoved)lemmaer, uten sammenkobling og hensyn til *langue*.

## Sammenfatning

I NAOB har vi følgende grunnlag for våre lemmatiseringsprinsipper:

- *representativitet og normativitet*: Vi velger som hovedlemma den variant som best samsvarer med den ortografiske formen for ordbokens språklige norm, i vårt tilfelle moderat bokmål. Om nødvendig oppretter vi et konstruktlemma. Normen kan imidlertid være vanskelig å avgjøre, jf. eksemplene *ben/bein* og *hul/hug*. I en deskriptiv ordbok som NAOB vil brukshistorie, frekvens og stilnivå tillegges vekt når man lemmatiserer. Etymologisk beslektede former med lignende bruk og stilpreg blir i den grad det er mulig samlet under et hovedlemma, eventuelt med henvisningslemmaer, som i en digital ordbok som NAOB lenkes til hovedlemmaene.
- *autentisitet*: Vi ønsker ikke å overnormere sjeldne varianter; jf. eksempelet *abengslig*. Denne fremgangsmåten bryter naturligvis med foregående prinsipper, men den gjelder vel å merke unntakstilfeller.
- *brukerrelevans*: Vi er opptatt av at brukeren er i stand til å finne den informasjon han eller hun leter etter. I denne hensikt bruker vi bl.a. digital lenking. Dessuten legger vi vekt på at lemmaet skal ha form som er slik at brukeren finner det igjen (jf. *blodge*-eksempelet).

Et problem vi ofte støter på i NAOB, er å avgjøre når vi skal samle litterære belegg (sitater) under hovedlemmaet og når vi bør la dem stå under henvisningslemmaet. Dette avgjøres av hvor og i hvilken grad vi ønsker å dokumentere og synliggjøre variantene. Hvis hovedlemmaet har få eller ingen belegg, blir sitateksempler flyttet dit, mens de gjerne får stå under henvisningslemmaet dersom det blir ansett som viktig å skille ut egne belegg på variantformen. Altså er det skjønn som avgjør hvor interessant det er å dokumentere bruken av en (underordnet) variantform.

## Konklusjon

En diakron, deskriptiv ordbok som NAOB stiller leksikografen overfor visse utfordringer når det gjelder lemmatiseringen, utfordringer som er knyttet til diakron og diatopisk varians, altså ved det jeg omtaler som atypiske lemnaer. Vi har sett at etymologien er det viktigste kriteriet for hvilke av de atypiske lemnaene som blir hovedlemmaer og hvilke som blir henvisningslemmaer. Også ord-

klasse og genus er sentrale lemmatiseringskriterier. Imidlertid finnes det tilfeller hvor vi oppretter egne lemmaer for varianter med samme etymologiske opprinnelse. Da er det bruksmåter og stilnivå som avgjør. Spørsmålet blir hvor store forskjeller man kan godta når det gjelder bruk og stilvalør for å lemmatisere varianter samlet – eller velge å opprette forskjellige lemmaer.

Også språkpolitisk er dette en vanskelig balansegang. Det hersker ofte uenighet om de forskjellige variantenes stilvalør og hvilke som skal regnes som ortografisk hovedform. Eksempelvis vil mer konservative krefter mene at variantformer som *ben* og *bein* ikke er stilistisk likeverdige innenfor alle bruksområder. Også hensikten med ordboken virker inn på lemmatiseringen – i en deskriptiv ordbok som NAOB kan vi tillate oss å ta mer hensyn til diakron brukshistorie og stilistiske særegenheter enn det som er vanlig i en normativ ordbok. Som nevnt er også brukerrelevans viktig for oss. Samtidig er det vårt mål at lemmatiseringen er i pakt med språkssystemet som helhet, slik at etymologisk beslektede ordformer med felles bruksmåter og stilpreg blir samlet, mens lydlige varianter og andre variantformer henviser til en hovedform. Det digitale lenkesystemet blir et hjelpemiddel for å gjøre ordbokens makrostruktur oversiktlig og brukervennlig, siden man ved et tastetrykk kan gå fra henvisningslemmaer til hovedlemmaet, hvor all informasjon er samlet.

Hovedlemmaene representerer på denne måten *langue*, mens *parole* dels blir ivaretatt av henvisningslemmaene, dels av autentiske eksempler i artikkelkroppene. Inkonsekvensene når det gjelder lemmatisering i NRO skyldes antagelig at man der ikke alltid har skjelnet tydelig mellom *langue* og *parole*, noe som ikke bare viser seg i lemmatiseringen, men også på andre nivåer i ordboken, f.eks. når det gjelder utformingen av definisjonene. Men det er en annen historie.

## LITTERATUR

- BMO = Wangensteen, Boye (red.): Bokmålsordboka. Definisjons- og rettskrivningsordbok. 3. utg. Oslo 2005.
- NLO = Bergenholtz, Henning m. fl. (red.): Nordisk leksikografisk ordbok. 1997. Oslo.
- NRO = Norsk riksmålsordbok. B. I–IV 1937–57, B. V–VI 1995. Oslo.
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. Andra, omarbetade och utökade upplagan. Stockholm.
- Wiegand, Herbert Ernst, 1983: Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung. I: Studien zur neuhochdeutschen Lexikographie. B. III. Utg. av Herbert Ernst Wiegand. Hildesheim, Zürich, New York (Germanistische Linguistik 1–4/82. S. 401–474).
- Wolski, Werner, 1989: Das Lemma und die verschiedenen Lemmatypen. Berlin, New York. I: Handbücher zur Sprach- und Kommunikationswissenschaft. B. 5.1. S. 360–371.

Zgusta, Ladislav, 1971: *Manual of Lexicography*. Haag, Paris. I: *Janua linguarum. Series maior* 29.

*Hanne Lauvstad*

Det Norske Akademis Store Ordbok, Kunnskapsforlaget

hanne.lauvstad@kunnskapsforlaget.no

# Ejendomsordbogen fransk/dansk: ny integreret e-ordbog

*Patrick Leroyer & Liselotte Kruse*

This article outlines the theoretical and practical considerations that have guided the planning and compiling of the *French/Danish Real Estate e-Dictionary*, a new online specialized dictionary which is now available and fully operational at Ordbogen.com. The genuine purpose of the dictionary is to assist its intended Danish users in solving foreseen, typical communication and knowledge-related problems in various real estate situations such as searching for and buying, selling, or owning a property in France. The dictionary is innovative in a number of ways as it is deliberately polyfunctional, but has been specifically designed to match the various needs of its intended user groups (experts and semi-experts with various degrees of proficiency in French) thanks to an adaptive data access and presentation. It also features a pragmatic user interface allowing consultation according to the situation of the users in relation to their practical real estate situation: searching, buying, owning or selling a property.

NØGLEORD: Online fagordbog, hybrid ordbog, polyfunktionel ordbog, brugertilpasning, integreret koncept, situationel tilgang, praktisk guide, leksikografisk samarbejde

## 1. Indledning

*Ejendomsordbogen fransk/dansk* – herefter EOB (Kruse og Leroyer: 2012) – er en ny integreret e-ordbog. Efter at have været tilgængelig online som gratis testversion (betaversion) i et halvt år hos Ordbogen.com (2012) er EOB i dag tilgængelig via abonnement med login. Denne korte præsentationsartikel vil give en klar idé om, hvordan vi har tænkt og udviklet en innovativ fagordbog, som på en integreret måde er i stand til at tilgodese flere typer brugere og tackle flere brugersituationer, og som genremsigt er en krydsning mellem fagordbog, leksikon, og praktisk guide. EOB gør brug af den nyeste databaseteknologi og kan konsulteres stationært såvel som *on the move* på en PC eller på en i-phone, en i-pad, eller en i-pod touch, og overgår således de leksikografiske drømme, som de Schryver (2003) formulerede for blot få år siden. Med andre ord integrerer

EOB en række funktionelle, metaleksikografiske innovationer og brugen af flere IT-platforme, deraf navnet integreret e-ordbog.

Sagens kerne er *feriebolig i Frankrig*, som er kilde til manges drømme, men også kilde til en lang række komplekse informationsbehov. Vi vil først kort beskrive den leksikografiske baggrund, som har ført til udvikling af ordbogsprojektet, blandt andet hvilke informationsproblemer, der kan opstå i forskellige situationer, når vi taler feriebolig i Frankrig, og hvordan vi har prøvet at tænke innovativt for at tilfredsstille de forudsete informationsbehov. Da det praktiske formål med ordbogen er handlingsorienteret (Wiegand 2001: 231), er samtlige informationer udvalgte for at hjælpe brugeren med at anskaffe, eje, eller sælge en feriebolig i Frankrig. Vi vil dernæst kort skitsere det leksikografiske samarbejde, der har muliggjort projektet. Samarbejdet mellem flere komplementære partnere har fremskyndet den leksikografiske innovationsproces. Hovedparten af artiklen vil præsentere det leksikografiske koncept med de intendede brugere og brugerprofiler, de leksikografiske funktioner (cf. Bergenholtz og Tarp 2005; Tarp 2006) samt korpus og dataselektion. Artiklen viser til sidst de adaptive datapræsentationer, både brugergænsefladen og kommunikative og kognitive visninger. I konklusionen vil vi præcisere, hvordan integrerede ordbøger adskiller sig fra gængse fagordbøger og kan ligge til grund for effektive leksikografiske værktøjer til løsning af praktiske informationsproblemer og konkrete opgaver.

## 2. Leksikografisk baggrund og forudsete situationer

Som andre nordeuropæere føler danskerne sig tiltrukket af Frankrig, og mange overvejer at købe fast ejendom i nogle af de mest populære regioner, enten som feriebolig, som investeringsobjekt, eller for at bosætte sig ved pensionsalderen. De søger derfor informationer, der kan hjælpe dem til at realisere disse planer. Dette kan ske hjemmefra, hvor man undersøger markedet på nettet, men det kan også ske under ferieophold i Frankrig, hvor man rejser rundt for at finde velegnede beliggenheder. Købsovervejelser afføder en række komplekse informationsbehov om fransk sprog og kultur, det franske ejendomsmarked, samt råd og vejledning vedrørende ejendomshandlen, som fagleksikografien kan tilfredsstille.

Der er ingen dedikerede fagordbøger om ejendomshandel mellem dansk og fransk. Der findes kun online termlister, men disse er ret primitive og mangelfulde. Dem, der kan fransk på højt niveau kan benytte sig af franske, monolingvale fagordbøger (for eksempel DIV 2009, LVI 2012), leksika og encyklopædier. Men alt i alt er der ikke megen hjælp at hente. Den optimale hjælp forud-

sætter, at fagleksikografien ændrer ved den klassiske fagordbogs form og funktioner. Hjælpen skal være praktisk tilrettelagt. Fagordbogen skulle gerne fusionere med en praktisk guide. Dette taler for udvikling af integrerede ordbøger. En sådan ordbog vil kunne tilbydes online og konsulteres fra en PC-arbejdsstation, men det vil på sigt være oplagt at tilpasse grænsefladen til mobile terminaler i form af en ferieboligapplikation.

Udviklingen er sket i samarbejde med Ordbogen.com, den førende udbyder af online ordbøger i Danmark. Konceptet er afledt af en grundig markedsanalyse af de ekstra-leksikografiske situationer og af de leksikografisk relevante informationsbehov, som danskerne typisk oplever som købere, sælgere eller ejere. Vort fokus har været rettet på de problemer, som danskerne erfaringsmæssigt løber ind i, og på den nøje udvalgte information, som bedst vil yde hjælp til løsning af disse problemer.

1. De kommunikative situationer er kendetegnet ved, at mange danskere ikke kan fransk – eller tilstrækkeligt fransk, og derfor er hæmmet i at afsøge det franske ejendomsmarked. De manglende receptive færdigheder vil afholde mange potentielle købere fra at undersøge markedet for relevante boligemner og opsøge de franske ejendomsmæglere. Danskere, der alligevel forsøger sig vil opleve kommunikationsproblemer, der kan sætte en stopper for deres boligdrømmer eller skaber misforståelser under selve købet eller ved et senere salg.
2. De kognitive situationer er kendetegnet ved, at danskerne er uvidende om det franske ejendomsmarked, inklusive annoncering, prisfastsættelse, købsforhandlinger, boligdokumenter, lovgivning, skatter og afgifter, administrative regler og procedurer mm., da ejendomsmarkedet i høj grad er et kulturbundet vidensområde.
3. Konceptet bygger på en graderet opdeling af de intenderede brugere efter L2-kompetence (graden af kendskab til fransk) og fagkompetence (graden af kendskab til fransk ejendomshandel), og inddrager en pragmatisk, situationsstyret datatilgang.

Dette har medført udviklingen af et multisøgbart brugerinterface. Vi vil i det følgende kort redegøre for, hvordan konceptet er blevet til, og vise de løsninger, vi har udviklet for at implementere konceptet.

### 3. Leksikografisk samarbejde i teori og praksis

I praksis er alt ordbogsarbejde teamwork, men dette har også en vigtig teoretisk dimension. Det er almindeligt anerkendt i den metaleksikografiske litteratur, at

udvikling af en god fagordbog forudsætter samarbejde mellem mindst tre slags eksperter, der henholdsvis sidder inde med viden om faget, viden om fagsproget, og viden om selve fagordbogen (Nielsen 2009). Ejendomsordbogen går videre og er bygget op som et samarbejdsprojekt, der inddrager 4 slags eksperter:

1. Eksperter i ejendomshandel i Danmark og i Frankrig, og mellem Danmark og Frankrig – her er det selskabet Sunliving, der udover at drive ejendoms-mæglervirksomhed i Danmark, har specialiseret sig i salget af ferieboliger i Frankrig. De leverer og sikrer den faglige viden. Denne viden suppleres af et netværk af andre eksperter, fx juridiske eksperter og skatteeksperter.
2. Eksperter i ejendomsfagsprog mellem de to sprog, inklusive fagoversættelse. Her er det selskabet DK&France, der som ordbogsredaktør leverer den fagsproglige viden.
3. Eksperter i fagleksikografi, der er tilknyttet Center for leksikografi ved Aarhus Universitet, Business and Social Sciences. De leverer den metaleksikografiske viden.
4. Endelig er der med Ordbogen.com eksperter i databaseteknologi, mobilteknologi og salg og marketing af online og mobile ordbøger. Ordbogen.com er Danmarks største udbyder af online og mobile ordbøger. De leverer databaseløsningen og sikrer den fortsatte udvikling af databasefunktionaliter og applikationer (også kaldet apps).

Projektkoordinering og ordbogsmanagement varetages af ordbogsredaktøren i samarbejde med Ordbogen.com.

#### 4. Integreret koncept – profiler, funktioner, selektion

Det integrerede koncept er frugten af en metaleksikografisk innovationsproces, der understøttes af en flerårig, omfattende markedsanalyse og markedsresearch af de ekstraleksikografiske situationer, og af de leksikografisk relevante informationsbehov. Der er i konceptet defineret 4 prioriterede målgrupper i alt:

1. Danskere, der overvejer at købe eller sælge
2. Danskere, der er midt i en købs/salgsp proces
3. Danskere, der allerede ejer en bolig
4. Danske translatører (L2 > L1)

Konceptet opererer desuden med 2 profiler ud fra L2 kompetence (eller mangel på samme), dvs. danskere med eller uden kendskab til fransk – og ud fra kend-

skabet til faget, dvs. danske eksperter og semi-eksperter (danskere, der udelukkende har viden om ejendomshandel i Danmark).

De 4 primære kommunikative (1 og 2) og kognitive (3 og 4) funktioner omfatter følgende forudsete brugersituationer:

1. Læse og forstå franske ejendomstekster
2. Oversætte franske ejendomstekster til dansk
3. Tjekke viden på dansk og på fransk om fransk ejendomshandel og ejerskab af bolig i forbindelse med en igangværende oversættelsesproces
4. Erhverve sig – enten før, under, eller efter – viden på dansk om fransk ejendomshandel og ejerskab af bolig

og er møntet ud følgende brugerprofiler:

Funktioner 1, 3 og 4	– KØBERE; ADVOKATER; EJENDOMSMÆGLERE
Funktioner 1 og 3	– BOLIGEJERE
Funktioner 1, 2, 3 og 4	– TRANSLATØRER
Funktion 4	– POTENTIELLE KØBERE

Ejendomsordbogen rummer en række datakategorier, der hver især eller i kombination kan opfylde de forudsete informationsbehov i de forskellige situationer. De omfatter:

- franske lemmata og synonymer
- danske lemmata (nøgleord)
- danske ækvivalenter
- illustrationer
- grammatiske oplysninger
- forklaringer på dansk (og fransk), der kan udvides (kort eller lang visning)
- kontrastive kulturnoter
- kollokationer, standardfraser og deres oversættelser
- henvisninger til integrerede byggede, fx ejendomsguiden og boligdokumenter

Konceptet inkluderer også en række integrerede brugs- og funktionsrelaterede byggede:

- En metaleksikografisk præsentation af ordbogen
- En brugervejledning
- En integreret praktisk guide til fransk ejendomshandel
- En samling, integrerede dokumenter, fx annoncer, købsaftale, skøde



- En række links til foretrukne sider

The screenshot shows the website interface for Ordbogen.com. At the top, there are navigation links for 'Køb privat', 'Uddannelse', 'Erhverv', 'Downloads', and 'Support & Kontakt'. The main header features the site logo and a navigation menu with 'Hjem', 'Køb adgang', and 'Ejendomsordbog Fransk / Dansk'. A prominent banner states 'Gratis tilgængelig i May' (Free available in May). Below this, a navigation bar includes 'Introduktion', 'Indhold', 'Brugeren i centrum', 'Kvalitet', 'Guide', 'Redaktør', and 'Kontakt'. The main content area features a headline: 'Ejendomsordbogen fransk-dansk-fransk udgives i samarbejde med en eksternt redaktør. Ordbogen lanceres som beta-version den 3. januar 2011.' This is followed by two paragraphs of text describing the book's focus on helping users in France and the book's structure. A small image of the book cover is shown on the right. At the bottom, a note states that all books from Ordbogen.com are available in a beta version.

Figur 1: ordbogens præsentationsside

Det er især den praktiske guide til ejendomshandel, der adskiller ejendomsordbogen fra konventionelle fagordbøger. Guiden er konciperet til at opfylde kognitive funktioner, og fungerer hermed på sin vis som en faglig indledning. Men den er også designet til praktiske instruktioner, så den på sigt vil kunne opfylde operative funktioner, når køberen på sin smartphone fx skal kunne hente hjælp under selve købs-/salgsprocessen, fx til at forstå og udfylde franske boligdokumenter.

Korpus er opbygget omkring et centralt, fransksproget fagsprogligt korpus, der udelukkende består af autentiske dokumenter, og er støttet af et sammenligneligt dansk korpus indeholdende:

- salgsannoncer og salgsoptillinger
- tekniske attester, rapporter og mærkninger
- juridiske dokumenter, herunder købsaftale og skøde, vedtægter
- administrative dokumenter
- autentisk korrespondance med franske myndigheder

Datasektionen er styret af ordbogens genuine formål: at bistå danskere med erhvervelse eller frasalg af feriebolig i Frankrig. Vi har eksempelvis nedprioriteret selektion af data til informationer om finansiering af feriebolig, da dan-

skerne ofte bruger deres eget pengeinstitut til at finansiere erhvervelse af en feriebolig.

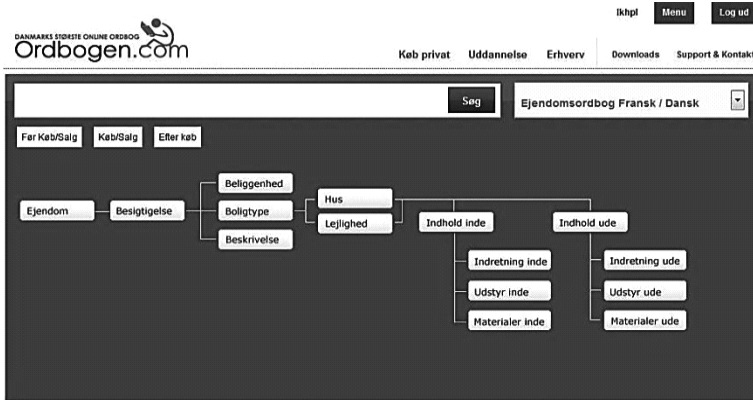
## 5. Pragmatisk datapræsentation og brugergrænseflade

Ejendomsordbogen er udstyret med pragmatisk filtrering til datasøgning og datavisning. Da ordbogen skal opfylde flere funktioner, skal datavisninger også kunne filtreres ud fra funktionelle situationer. En sådan filtrering vil kunne generere op til 5 forskellige, adaptive datasøgninger og – visninger:

1. Data for at læse og forstå franske boligtekster, med bl.a. forklaringer til de franske lemmata og ækvivalentangivelser.
2. Data for at oversætte fra fransk til dansk, med bl.a. ækvivalenter, kollokationer, standardfraser og deres oversættelser.
3. Data for at hurtigt at tjekke viden om franske begreber, fx i forbindelse med en oversættelse, med illustrationer og korte forklaringer på dansk til de franske lemmata.
4. Data for at tilegne sig viden om fransk ejendomshandel. Disse data tilgås ud fra danske lemmata eller nøgleord, der peger på alle de relevante franske indgangslemmata. Disse udstyres med lange forklaringer på dansk samt henvisninger til andre relevante lemmata, til den praktiske ejendomsguide, og til eksterne kilder.
5. Operative data, i form af instruktioner, der kan bruges i konkrete køb/salgssituationer eller ved kontakt til franske myndigheder.

Udover søgefeltet, hvor brugeren kan indtaste de danske og franske ord og udtryk, som søges på, er EOB brugergrænsefladen udstyret med en interaktiv, visuel præsentation af de situationer, som er almindelig forekommende i forbindelse med ejendomshandel. Denne såkaldte gruppesøgning via danske nøgleord er endnu en funktionalitet, der adskiller ejendomsordbogen fra konventionelle fagordbøger. Den situationelle brugergrænseflade skal sikre tilgang til data, ud fra hvilken praktisk situation brugeren befinder sig i. Der er ved start mulighed for at benytte sig af tre hovedtilgange til søgning og indhentning af informationer:

1. Før køb/salg
2. Under køb/salg
3. Efter køb

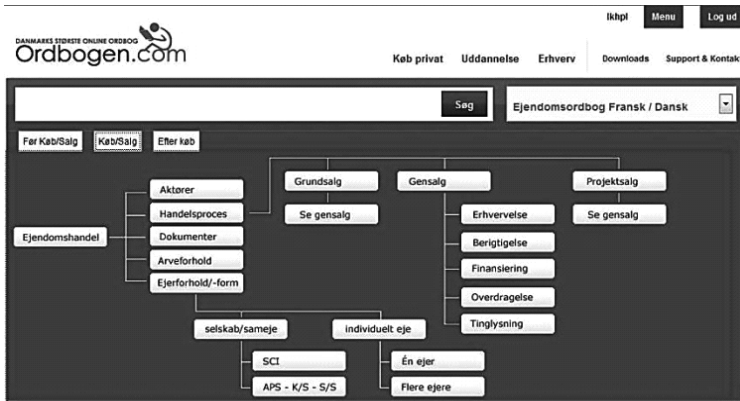


Figur 2: Brugerinterface med pragmatisk 'før', 'under' og 'efter' datafiltrering

Hver af disse situationstilgange ledsages af en detaljeret visuel fremstilling, hvori situationen, som brugeren befinder sig i i forbindelse med at skulle realisere planen om en ejendom/feriebolig både før, under og efter købet, er opdelt i relevante indholdskategorier. Disse relaterer sig til processen med at finde, erhverve og eje en ejendom/feriebolig. Brugeren kan så efterfølgende klikke sig ind i det relevante felt og få direkte adgang til systematiske optegnelser af alle de lemmata, der er tilknyttet den pågældende kategori, fx:

- Implicerede aktører
- Handelsproces
- Dokumenter
- Forhold vedrørende arv
- Ejerforhold, herunder beskatning

I kraft af sin pragmatiske forankring er den situationelle grænseflade specielt velegnet til kognitiv datavisning.



Figur 3: brugerinterface med situationel tilgang til købs-/salgsproces

I kommunikative situationer skal brugerne kunne vælge at få specifikke visninger frem, som er styret af de enkelte kommunikative funktioner, 1. Læs og forstå og 2. Oversæt, samt en tredje visning, der vil kombinere visninger 1 og 2, når en kommunikativ situation afløses af en kognitiv situation. Dette sker typisk under oversættelsesprocessen, hvor brugeren har brug for at stoppe op og hurtigt tjekke specifik viden, inden oversættelsesproces genoptages.

**diagnostic technique immobilier** Søg Ejendomsordbog Fransk / Dansk

**Information**

- Om ordbogen
- Chat med support
- Spørg Sprogcentret

**Seneste opslag**

- diagnostic technique immobilier
- diagnostic
- maison à colombages
- hup
- chaumière
- d'antan

**Værktøjer**

- Tilføj note
- Tilføj ord
- Foretst rettelser
- Foretst ord

**Ejendomsordbog Fransk / Dansk** *(gratis testudgave)*

**Diagnostic Technique Immobilière** substantiv m <Diagnostics Techniques Immobilières>

Synonymer

- diagnostic obligatoire substantiv m <diagnostics obligatoires>
- diagnostic immobilier obligatoire substantiv m <diagnostics immobiliers obligatoires>
- état technique substantiv m <états techniques>
- diagnostic technique substantiv m <diagnostics techniques>
- diagnostic immobilier substantiv m <diagnostics immobiliers>
- D.T.I. forkortelse m <D.T.I.>
- DTI forkortelse m <DTI>
- expertise immobilière substantiv f <expertises immobilières>

Fortklaring

*I Frankrig skal sælger ved ejendomsandel sælge ....*

Oversættelser

- lovpålagt teknisk undersøgelse
- lovpålagt teknisk rapport

Fraser

- Les diagnostics obligatoires sont les suivants : le diagnostic amiante, l'état relatif à la présence de termites, la Loi Carrez (Mesurage), le diagnostic plomb (CREP), la performance énergétique (DPE), les risques naturels et technologiques (ERN/T), l'état De lovpålagte undersøgelser er følgende: ...

Henvisning

- Dossier de Diagnostic Technique

**Skriftstørrelse**

- Større
- Mindre
- Nulstil

**Egne noter**

Ejendomsordbog Fransk / Dansk

- Ingen noter

**Tilfreds med artiklen?**

- Kunne bruges
- Kunne ikke bruges

Figur 4: artikelopbygning for kommunikativ visning

En anden kognitiv datavisning sker gennem den praktiske ejendomsguide. Ejendomsguiden kan konsulteres på 2 måder:

1. Direkte fra ordbogens hovedmenu, som vist her – hvor brugeren kan navigere mellem de enkelte sektioner, eller læse fra ende til ende.
2. Indirekte fra de enkelte artikler, hvor der linkes til de specifikke afsnit, i hvilke informationerne ekspliciteres yderligere eller sættes i kontekst.

The screenshot shows the top part of a website. At the top right, it says 'Gratis tilgængelig i May'. Below that is a search bar and a dropdown menu for 'Se andre produkter'. A navigation menu contains links for 'Introduktion', 'Indhold', 'Brugeren i centrum', 'Kvalitet', 'Guide', 'Redaktion', and 'Kontakt'. The main heading is 'Praktisk Guide Boligkøb i Frankrig'. Below this is the author information: 'DK-France v/ Liselotte Kruse', 'Statsautoriseret translator og tolk i fransk', and the date '01.01.2011'. A paragraph of text follows, explaining the guide's purpose. Below the text is a section titled 'Indholdsfortegnelse' with a list of links: 'Før boligkøb', 'Dømmeboligen', 'Udbuddet', 'Prisniveauet', 'Ejendomsannoncerer', 'Overvejelser', 'Gode råd', 'Boligkøb', and 'Købsproceduren (la procédure d'acquisition)'.

Figur 5: praktisk guide for kognitiv visning

I kognitive situationer, kan brugeren få adgang til informationer via danske nøgleord, som indbyder til navigation. Datavisning vises som lister af relevante lemmata, som er samlet i grupper. Vi kalder det gruppesejning og -visning. Lemmata er linket til et net af relevante henvisninger, både interne og eksterne. Fagforklaringer til de enkelte lemmata er alle redigeret på dansk, og der kan vælges mellem kort eller lang visning. I tilknytning til nogle af de danske nøgleord, hvor forskelle mellem de to lande er store, er der udarbejdet kulturnoter.

The screenshot shows the website 'Ejenomsordbogen Fransk / Dansk'. At the top, there is a navigation bar with links for 'Køb privat', 'Uddannelse', 'Erhverv', 'Downloads', and 'Support & Kontakt'. A search bar contains the word 'hus' and a 'Søg' button. Below the search bar are buttons for 'Før Køb/Salg', 'Køb/Salg', and 'Efter køb'. The main content area is divided into three columns:

- Information:**
  - Om ordbogen
  - Chat med support
  - Spørg Sprogcentret
  - Seneste opslag:**
    - hus
    - chaumière
    - d'antan
    - antan
    - beskrivelse
    - boligtype
  - Værktøjer:**
    - Tilføj note
    - Tilføj ord
    - Foreslå rettelser
    - Foreslå ord
- Ejenomsordbog Fransk / Dansk (gratis testudgave):**
  - Gruppesejning:**
    - Følgende ord blev fundet i gruppen: hus
      - bastide - lille lystgård
      - bastidon - lille landsted
      - bastille - landsted
      - båtsise - stor bygning
      - bergerie - fårehus
      - bien immobilier - (bolig)lemne
      - cave - kælder
      - cellier - vinkælder
      - chalet - chalet
      - chambre d'hôtes - B&B
      - chapelle - kapel
      - charges de copropriété - fællesudgifter
      - chaumière - stråtkøbt ejendom
      - collectif - kollektiv
      - combles - tagetage
      - conciergerie - portnerbolig
      - coopérative d'habitation - andelsboligforening
      - copropriétaire - samejer
      - copropriété - medejerskab
      - corps - hoved
      - coup de cœur - forelskelse
      - création - indretning
      - dépendance - udbygning
      - écurie - hestestald
      - ensemble immobilier - boligkompleks
- Skriftstørrelse:**
  - Sterre
  - Mindre
  - Nulstil
- Egne noter:**
  - Ejenomsordbog Fransk / Dansk:
  - Ingen noter
- Tilfreds med artiklen?:**
  - Kunne bruges
  - Kunne ikke bruges

Figur 6: listepresentation

## 6. Afsluttende bemærkninger

Den integrerede e-ordbog er velkendt i forvejen; den bruges nemlig på hjemmesider, hvor der ofte er behov for definitioner, fagforklaringer eller instruktioner i pragmatiske situationer, fx hos virksomheder i finanssektoren, pension og forsikring, eller varesalg og distribution, der ønsker at hjælpe deres kunder i forbindelse med transaktioner. Den er oftest integreret i selve teksterne og har form af pop-up vinduer, som kommer frem, når man klikker på ord eller flytter markøren over dem. EOB er ligeledes forankret og integreret i forudsete pragmatiske situationer: ejendomshandel. EOB er et ægte leksikografisk informationsværktøj, der kombinerer det bedste fra fagordbogen og leksikonet med en praktisk guide. Det, der adskiller den fra de konventionelle fagleksikografiske værker, er dens fokus på det pragmatiske sigte: den er udelukkende konciperet til at hjælpe brugerne til at løse en konkret, veldefineret opgave. Det er det praktiske sigte, som er EOB's genuine formål. Derfor er den også blevet udstyret med en situationel tilgang, dvs. en tilgang til informationer i de konkrete situationer, som brugerne befinder sig i, i forbindelse med at skulle realisere planen om en ferie-

bolig, både før, under og efter købet. Tilgang og navigation er også bidirektionel – brugeren kan efter behov tage udgangspunkt i danske eller franske søgninger – og datapræsentationer er adaptive, da data bliver filteret afhængigt af den valgte funktion eller situation. På sigt har vi planer om at supplere ordbogens interaktivitet, som p.t. er begrænset til rent leksikografisk feedback fra brugerne. Vi vil integrere muligheden for at oprette og vedligeholde et brugernetværk af abonnenter, hvorved interesserede brugere kan komme i dialog med hinanden, dele viden, reagere til indlæg og støtte hinanden op med praktiske informationer, der er relateret til det franske marked for ferieboliger. Dette kunne meget vel være en integreret del af ordbogsapplikationen til mobil brug, og med mulighed for integration af sociale netværker, fx en dedikeret Facebook side. Hermed bliver det endnu tydeligere, i hvor høj grad (fag)leksikografi er en særlig del af informations-, business og sociale videnskaber.

## LITTERATUR OG KILDER

- Bergenholtz, Henning og Sven Tarp, 2005: Wörterbuchfunktion. Irmhild Barz, Henning Bergenholtz, Jarmo Korhonen (red.), Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch. Frankfurt a.M./Bern/Bruxelles/New York/Oxford/Wien: Peter Lang. S. 11-25.
- DIV = Dictionnaire de l'immobilier et de la presse. 2009. L'agence Innovapresse.
- DK&France, formidling af franske ferie- og vinboliger, undervisning, oversættelse, tolkning, rådgivning. <http://www.dk-france.dk/>
- EOB = Kruse, Liselotte og Patrick Leroyer: Dansk/fransk ejendomsordbog. Ordbogen.com. <http://www.ordbogen.com/?iqmedier=1> (marts 2012)
- de Schryver, Gilles Maurice, 2003: Lexicographers Dreams in the Electronic-dictionary Age. *International Journal of Lexicography* 16.2. S. 143-199.
- LVI = Lesiteimmo.com. Lexique du vocabulaire immobilier, dictionnaire immobilier, dictionnaire juridique du logement. <http://www.lesiteimmo.com/financement/dictionnaire-immobilier/lexique/A#> (marts 2012)
- Ordbogen.com. Danmarks Største Online Ordbog. <http://www.ordbogen.com/?iqmedier=1> (marts 2012)
- Sunliving. Trygt boligkøb i udlandet. <http://www.sunliving.dk/>
- Nielsen, Sandro, 2009: Reviewing printed and electronic dictionaries. A theoretical and practical framework. I: Nielsen, Sandro og Sven Tarp: *Lexicography in the 21st Century*. In honour of Henning Bergenholtz. Amsterdam: John Benjamins. S. 23-41.
- Tarp, Sven, 2006: Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi. Doktorafhandling. Århus: Handelshøjskolen i Århus.
- Wiegand, Herbert Ernst, 2001: Was eigentlich sind Wörterbuchfunktion? Kritische Anmerkungen zur neueren und neuesten Wörterbuchforschung. *Lexicographica. International Annual for Lexicography* 17. S. 217-248.

*Patrick Leroyer*

Center for leksikografi, Institut for Erhvervs kommunikation, Business and Social Sciences, Aarhus Universitet

pl@asb.dk

*Liselotte Kruse*

Institut for Erhvervs kommunikation, Business and Social Sciences, Aarhus Universitet

lkh@asb.dk



# Netordbøger – hvordan finder brugerne frem til dem, og hvad gør de, når de har fundet dem?

*Henrik Lorentzen & Liisa Theilgaard*

In general, user behaviour studies of online dictionaries have focused on user behaviour once the user is on the site. But before a potential user even reaches this stage, he or she must succeed in finding the dictionary on the Web. The purpose of this paper is to investigate users' linguistic search strategies before they enter our dictionary site, ordnet.dk. What kind of search engine queries are successful and why (not)? Similarly, we have studied the queries in the search field of our online dictionaries. Are the search strategies the same as when using a search engine? Taking the no-match searches as a starting point, we have asked ourselves whether our content and search functionality correspond to the search behaviour of the users, i.e., if we can give an answer to the users' query and if data are organized and presented in an appropriate way. Given the results of these analyses, we decided to make several changes to the site in order to optimize user access and attract new users. We present these changes and the results and finally some ideas about future development. The paper is based on empirical data from Google Analytics and our log files.

*Nøgleord:* netordbøger, brugeradfærd, søgestrategier, webstatistik, søgemaskineoptimering.

De fleste undersøgelser af brugeradfærd i forbindelse med netordbøger har koncentreret sig om adfærden, når brugeren er kommet ind på ordbogssitet. Men inden en potentiel bruger overhovedet er nået dertil, må han eller hun finde frem til ordbogen på nettet.

I denne artikel undersøger vi de søgestrategier, brugerne anvender, når de leder efter sproglig viden ved hjælp af en søgemaskine. Med resultaterne af disse undersøgelser i hånden har vi stillet os selv spørgsmålet: Hvordan kan vi optimere Det Danske Sprog- og Litteraturselskabs netordbøger på ordnet.dk, så de tilgodeser denne søgeadfærd bedst muligt og hjælper flest mulige brugere frem

til ordbogen og dermed til et svar? Vi præsenterer nogle af de ændringer, vi har foretaget, og svarer på, om de har givet resultat i form af flere nye brugere, flere oplag med videre.

På tilsvarende vis har vi undersøgt, hvad brugeren søger på i netordbøgernes søgefelt. Er søgestrategierne de samme som i søgemaskinerne, og er brugeren overhovedet bevidst om, hvad Den Danske Ordbog og Ordbog over det danske Sprog er for en slags ordbøger? Med udgangspunkt i de søgninger, der ikke har givet resultat, har vi spurgt os selv, om vores indhold og søgefunktion stemmer overens med brugernes søgeadfærd, dvs. om vi kan svare på brugerens søgning, og om data er organiseret og præsenteret hensigtsmæssigt. Det giver desuden anledning til at overveje, hvilke veje vi skal gå i fremtiden for at supplere netordbøgerne med nyt og andet indhold.

## Metode, termer og redskaber

Webstedet ordnet.dk hviler i øjeblikket på tre søjler: Den Danske Ordbog (DDO), Ordbog over det danske Sprog (ODS) og KorpusDK (se Fjeld og Malmgren 2010). Den første netversion af ODS blev lanceret i november 2005, og derefter fulgte DDO i november 2009. I november 2010 lanceredes anden version af ODS, hvor hele, nye artikler fra supplementet er inkluderet. KorpusDK muliggør søgninger i to korpuser over moderne dansk almentsprog og er til en vis grad integreret med ordbogsressurserne, men vil i øvrigt ikke blive berørt i denne artikel.

I undersøgelserne af brugernes adfærd har vi primært benyttet to redskaber, dels Google Analytics, dels vores egne logfiler. Google Analytics er et online webanalyseværktøj (se Clifton 2010), der leverer en lang række statistikker over de besøgendes adfærd. Disse statistikker oplyser fx om, hvor mange besøgende siden har pr. dag, og om, hvordan de besøgende er kommet frem til siden. Det er muligt at sammenligne forskellige perioder med hinanden, fx den samme måned i to forskellige år, eller se, hvor længe en bruger opholder sig på siden.

Samtidig holder vi også selv øje med brugerne i kraft af vores egne logfiler (jf. bl.a. Almind 2008 og Nygaard og Fjeld 2008), nærmere betegnet via de såkaldte no match-lister, dvs. lister over, hvad brugeren *ikke* finder ved sine søgninger på ordnet.dk. De undersøgte data gælder DDO, dels fordi det er den ordbøgerne, der har langt flest besøgende, dels fordi det er den ordbogsbase, vi primært opdaterer, men det er muligt at foretage de samme undersøgelser for ODS.

I undersøgelsen opererer vi med et “før” og “efter”. “Før” er september 2010, og “efter” er marts 2011. I den mellemliggende periode har vi foretaget

en række ændringer i webstedets funktionalitet, og det er derfor muligt at se, om disse ændringer har haft nogen indvirkning på besøgstal og brugeradfærd.

Når vi taler om “besøg”, er der tale om antallet af individuelle sessioner, som en bruger har startet. Hvis brugeren har været inaktiv i 30 minutter eller derover, regnes ny aktivitet som et nyt besøg.

## Hvordan finder brugerne frem til ordbogen?

Der er tre forskellige adgangsveje til ordbogen:

- 1) Brugeren kan søge i en søgemaskine og klikke på et link til ordbogen i søgeresultatet
- 2) Brugeren kan komme direkte til ordbogen ved at klikke på et bogmærke gemt i browseren eller ved at taste webadressen direkte i browserens adresselinje
- 3) Brugeren kan komme via en linkhenvisning på et andet websted

I september 2010 kom halvdelen (49 %) til DDO ved at søge i en søgemaskine – og heraf står Google for 99 % – en tredjedel (33 %) kom direkte til DDO, og de sidste 18 % kom via en henvisning på et andet websted, fx sproget.dk. Henvisningsbrugere bruger kortest tid pr. besøg (5:31 min.) og ser færrest sider pr. besøg (3,53 sider). De flittigste kommer direkte og bruger 7:21 min. og ser 4,1 side pr. besøg, og det undrer ikke, da de kender webstedet og ordbogen i forvejen.

Brugerne foretrækker altså at søge i en søgemaskine, og dermed bliver det meget vigtigt at figurere som søgeresultat i Google – og helst på den første søgeresultatside – hvis en potentiel bruger skal blive opmærksom på ordbogen.

## Hvordan søger brugerne i en søgemaskine?

For at undersøge brugernes søgeadfærd i en søgemaskine har vi undersøgt alle de søgeudtryk, som brugerne har søgt på i en søgemaskine i september 2010, og som har resulteret i et klik videre til ordbogen. Undersøgelsen siger altså *ikke* noget om, hvilke søgeudtryk der mislykkedes, dvs. ikke ledte frem til ordbogen, eller hvilke søgeudtryk brugeren eventuelt har forsøgt sig med inden.

Først har vi set på antallet af søgeord i søgeudtrykket, defineret som en streng uden mellemrum. Her viser det sig, at 60 % af brugerne har skrevet ét søgeord, 23 % to søgeord og 17 % tre eller flere ord, dvs. at hele 40 % søger på mere end ét ord.

Dernæst har vi grupperet alle søgeudtryk i tre kategorier, alt efter om det betegner et emne (metaord, fx *ordnet*, *dansk ordbog*), et sprogligt udtryk (materialiterord, fx *slesk*, *få blod på tanden*) eller en kombination af disse to (fx *hvilken ordklasse er ordet efter* eller *nonchalant betydning*, hvor brugeren søger efter betydningen af ordet *nonchalant*):

	Antal søgeudtryk	Antal besøg	Besøg i %	Eksempler
Metaord	1.160	28.355	64 %	<i>ordnet</i> , <i>dansk ordbog</i> , <i>ddo</i>
Materialiter	2.220	13.290	30 %	<i>slesk</i> , <i>selvom selv om</i>
Kombination	1.251	2.780	6 %	<i>nonchalant betydning</i> , <i>snyde synonym</i>
I alt	4.631	44.425	100 %	

Figur 1: Fordeling af søgeudtryk via en søgemaskine i september 2010

Søgeudtrykkene fordeler sig med 2.220 materialitersøgeudtryk, 1.160 metasøgeudtryk og 1.251 søgeudtryk, hvor metaord er kombineret med materialiterord. Disse 4.631 forskellige søgeudtryk har resulteret i 44.425 besøg, dvs. i gennemsnit 10 besøg pr. søgeudtryk. Men der er stor forskel på effektiviteten af søgeudtrykkene. Således har 2.530 søgeudtryk kun genereret ét besøg hver, svarende til 5,7 % af besøgene. Heroverfor står de godt 1.160 metaord, der har genereret 64 % af besøgene, hvoraf de fem mest effektive søgeudtryk *ordbog*, *dansk ordbog*, *den danske ordbog*, *ordnet* og *ordnet.dk* tegner sig for hele 39 % af alle besøg. Det er altså i høj grad søgeudtryk med metaord, der har givet resultat, og det hænger sammen med, hvad der er blevet indekseret af webcrawlerne.

Ved lanceringen af *ordnet.dk* blev det af frygt for manglende serverkapacitet besluttet udelukkende at tillade indeksering af forside, hjælpesider og omtekster, dvs. omkring 25 sider fra DDO og omkring 65 sider i alt på *ordnet.dk*. Der var således ingen ordbogssider, der blev indekseret automatisk, men enkelte ordbogssider er alligevel blevet indekseret, fordi andre websteder har linket til en bestemt ordbogsside. Denne praksis harmonerer åbenbart dårligt med brugernes faktiske søgeadfærd, så der ligger et stort potentiale i at kunne tilgodese materialitersøgning fra en søgemaskine.

## Hvordan søger brugeren i søgefeltet?

Sammenligner vi brugernes søgeadfærd i en søgemaskine med søgeadfærden, når brugeren er kommet ind på DDO og søger i søgefeltet, tegner der sig et ganske andet billede.

Til denne delundersøgelse har vi måttet bruge data fra marts 2011, fordi logningen af søgeudtryk fra søgefeltet ikke var slået til i Google Analytics i september 2010. Her viser det sig, at 96 % søger på ét ord, mens 2,4 % søger på to ord og 1,6 % på tre eller flere ord. Set i forhold til de 60 % étordssøgninger i en søgemaskine er det tydeligt, at brugeren har en anden adfærd, når det gælder søgninger i søgefeltet.

Eftersom søgninger via søgefeltet i marts 2011 udgjorde så meget som 641.140 forskellige søgeudtryk, har vi ikke kunnet behandle dem kvantitativt ved en systematisk klassifikation, men et blik ned gennem søgeudtrykkene viser, at de fleste søger på et materialiterord. Søgning med jokertegn er sjældent brugt, idet kun 0,3 % af søgeudtrykkene indeholder et jokertegn (fx *\*skib, ???tøj*). Vi har ellers gjort meget ud af, at man kan søge på både udtryk og flerordsforbindelser samt med jokertegn, men tilsyneladende har brugerne ikke lagt mærke til det eller følt det som et behov.

Meta- og kombinationssøgninger forekommer heller ikke særlig ofte, fx optræder strengen *ordbog* kun 144 gange, *synonym* 56 gange og *betydel betydning* 46. Brugeren opfatter altså søgefeltet som et opslagsfelt til ordbogsdelen alene, og sådan fungerer søgefeltet også normalt i elektroniske ordbøger i modsætning til andre websteder, hvor det bruges til fritekstsøgning på hele hjemmesiden.

## Hvad finder brugeren ikke?

Ved hjælp af vores logfiler har vi undersøgt samtlige registrerede søgninger via DDO's søgefelt i september 2010, og heraf var der 20,6 %, der ikke gav noget resultat. Dette tal er lavt sammenlignet med to tidligere undersøgelser, hvor ikke-vellykkede søgninger i Bokmålsordboken udgjorde 31 % (Nygaard og Fjeld 2008: 59) og i Lexin online Svenska ord udgjorde 41 % (Hult 2008: 78), hvilket i det sidste tilfælde måske hænger sammen med, at ordbogen indeholder så få opslagsord, og at ordbogen har brugere med svensk som andetsprog som hovedmålgruppe.

Man skal være opmærksom på, at dette tal kun dækker søgninger uden direkte resultat i form af en (ordbogs)side, og at det derfor ikke siger noget om, hvor mange af disse søgninger der resulterede i et meningsfuldt mente du-for-slag. Vi har nemlig truffet det valg, at brugeren kun ledes direkte til en ordbogs-side, når der er tale om bøjningsformer og ortografiske varianter, der er registre-

ret i ordbogsartiklerne. Ellers præsenteres brugeren for en række mente du-forslag, der dels bygger på den søgealgoritme, vi anvender<sup>1</sup>, dels på en liste med kendte fejlformer.

Herefter har vi kategoriseret de 1.000 mest frekvente ikke-vellykkede søgeudtryk i en række typer. Disse 1.000 søgeudtryk repræsenterer 12,6 % af samtlige ikke-vellykkede søgeudtryk. Et søgeudtryk kan optræde i flere kategorier, fx både som en stavfejl og som en bøjningsform. Den mest almindelige ikke-vellykkede søgetype var stave- og slåfejl med 30 % (fx *unskyld*, *priotere*, *sjangre* for *genre*), og dernæst kom bøjningsformer med 20 % (fx *dominerende*, *terroristernes*, *bibringes*, *skrid*, *fornemste*). Det overraskede os, at der blev søgt på så mange bøjningsformer. En mulig forklaring er, at vi tilbyder en opensearch-funktion. Med den installeret i sin browser (Explorer eller Firefox) kan man bruge DDO som søgetjeneste og slå et ord op direkte i ordbogen. Man markerer et ord på en hvilken som helst webside og vælger "Søg med Den Danske Ordbog", og så sendes det markerede ord af sted som søgning til ordbogen og vises på et nyt faneblad. Vi har dog ikke nogen tal for, hvor mange der benytter denne funktion, så vi ved det ikke. Under alle omstændigheder fik det os til at tænke over, hvilke bøjningsformer vi ikke havde taget højde for i vores data.

Resten fordeler sig især på typer som fremmedord (9 %), proprier (7 %), sammensætninger (7 %), nye ord (6 %), flerordsudtryk (4 %), forkortelser (3 %), afledninger (2 %) samt flere andre, og de siger selvfølgelig noget om, hvad brugerne leder forgæves efter i DDO (se mere nedenfor om fremtidsperspektiver).

## Ændringer

På baggrund af vores undersøgelser besluttede vi at foretage en række ændringer. Først og fremmest ønskede vi større synlighed i søgemaskinernes søgeresultater – og helst på første resultatside – og derfor åbnede vi for, at webcrawlere kunne indekserer alle de enkelte ordbogssider. Så hvor der før var under 100 indekserede sider, er der nu ca. 100.000 sider for DDO og 350.000 for hele ordnet.dk.

Dernæst skulle der gøres noget ved de mange bøjningsformer, som vi ikke var klar over, at der blev søgt på. Vi har derfor tilføjet præsens participium som skjult bøjningsform til verbernes bøjning. Der er sikkert mange, der søger på

---

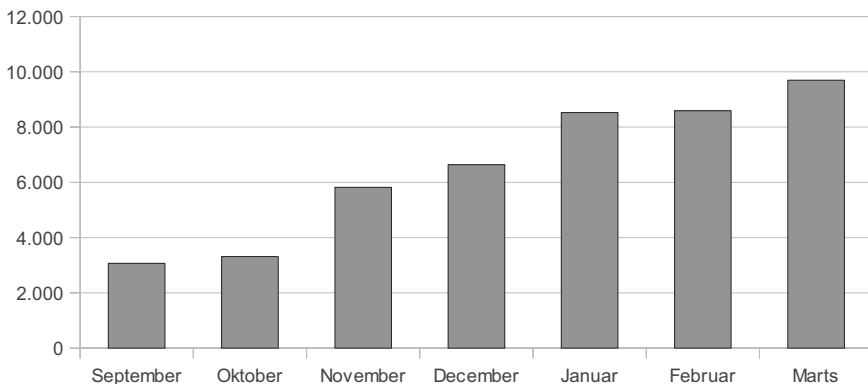
<sup>1</sup> Det drejer sig om stavekontrolprogrammet *Aspell* i kombination med *Levenshtein*-distancen, som er et mål for forskellen mellem to strenge, og som vi bruger til at sortere de dårligste forslag fra.

denne bøjningsform, fordi de bruges som adjektiver, men i DDO beskrives de ofte i verbalartiklen og ikke i en selvstændig artikel. Så nu kommer man automatisk til artiklen *dominere*, når man søger på *dominerende*. På samme måde har vi genereret skjulte genitivformer til alle substantiver, så man nu automatisk kommer til artiklen *trommehinde*, når man søger på *trommehindens*. – Til gengæld har vi ikke gjort noget for at tilføje imperativ og passiv, da disse former ofte genkendes af søgealgoritmen bag mente du-funktionen. Adjektivernes komparationsformer er medtaget som bøjningsformer i DDO, men bestemt form superlativ (*den/det/de fornemste*) er ikke medtaget, og den har vi heller ikke gjort noget ved endnu.

Endelig har vi gjort mente du-funktionen mere synlig ved også at vise den i midterfeltet og ikke blot under søgefeltet, så der kom mere fokus på de ordforslag, der skal hjælpe brugeren, når en søgning ikke giver resultat. Desuden har vi forbedret mente du-funktionen, så den håndterer sammenskrevne former bedre og viser et henvisningslink, når ens søgeudtryk findes i ODS, men ikke i DDO.

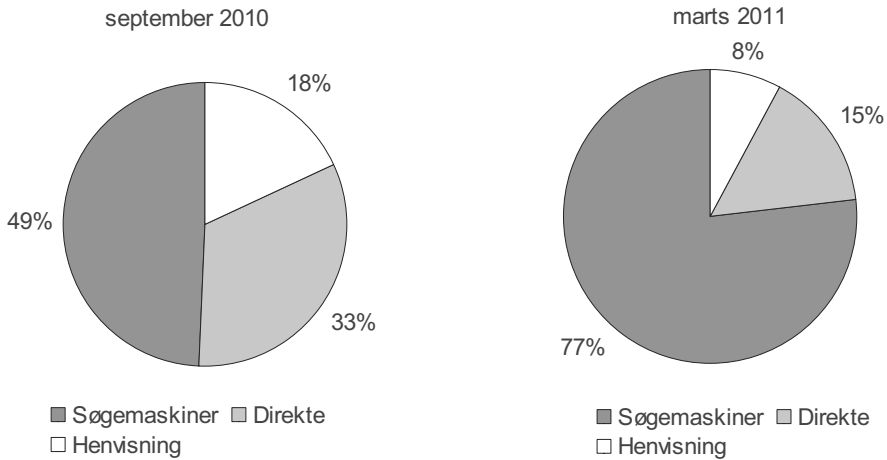
## Resultater

Indekseringen af alle ordbogssider har haft en markant effekt. Antallet af daglige besøg er mere end tredoblet i løbet af det halve år, vi har undersøgt trafikken, og det fortsætter.



Figur 2: Besøg på DDO pr. dag i perioden september 2010 til marts 2011

Dette afspejler sig naturligvis også i fordelingen af brugernes adgangsvej til ordbogen. Hvor det i september 2010 var 49 % besøg, der kom fra en søgemaskine, er det i marts 2011 hele 77 %:



Figur 3: Adgangsveje til DDO i september 2010 og marts 2011

Indekseringen har altså betydet, at vi nu figurerer dér, hvor brugerne søger efter information (også sproglig), nemlig i søgemaskinerne (Google) – og ofte på første søgeresultatside. Det gælder både de nye og de tilbagevendende brugere, idet 82,3 % af de nye brugere i marts 2011 kommer via en søgemaskine, men også 72,6 % af de tilbagevendende. Samtidig bidrager søgemaskinernes egne søgealgoritmer til at guide brugerne frem til et resultat, for eksempel i forhold til stavfejl, bøjningsformer og lignende. Dertil kommer, at brugerne nu henvises direkte til svaret i form af en ordbogsside og ikke bare til forsiden. Og endelig bør det vel nævnes, at denne synlighed ikke har kostet en krone til webmarkedsføring eller lignende, og at webcrawlnes indeksering ikke har betydet længere svartider.

Andelen af nye brugere er også steget, så fordelingen mellem nye og tilbagevendende besøgende i marts 2011 er 44 % nye over for 56 % tilbagevendende, mens det i september 2010 var 34 % nye over for 66 %. Det er dog stadig mere end halvdelen, der er tilbagevendende besøgende, og de bruger stadig længst tid på hvert besøg. Opgørelserne viser, at det altså både er lykkedes at tiltrække nye brugere og at gøre en del af dem til trofaste brugere.

Ser vi på de søgeudtryk, som har ledt brugerne frem til ordbogen via en søgemaskine, er der nu fem gange så mange, der kommer til ordbogen via et søgeresultat, men samtidig er der blevet 18-19 gange så mange forskellige vellykkede søgeudtryk, nemlig 85.601 i forhold til septembers 4.631, dvs. at der er en langt større spredning, når det gælder søgeudtryk. Det hænger sammen med, at alle ordbogssiderne nu bliver indekseret, så der er chance for at få resultat med et langt større antal søgeudtryk. Det kan illustreres ved, at der i september



2010 var 2.530 forskellige søgeudtryk, der resulterede i hver 1 besøg, svarende til 55 % af alle søgeudtryk, mens det i marts 2011 var 60.384 svarende til 70,5 %. Noget tilsvarende ses i fordelingen på søgeudtrykstyper: Hvor det før i høj grad var metasøgeudtryk som *ordbog*, der gav resultat, er der nu meget større sandsynlighed for at få match med et materialitersøgeudtryk, fordi opslagsordet indgår i navnet på de enkelte websider med ordbogsartikler. Alt i alt har webcrawlernes indeksering af ordbogssiderne været den vigtigste enkeltfaktor, når det gælder om at tilgode se brugernes søgeadfærd: De søger sproglig information via en søgemaskine, og de søger ofte med materialitersøgeudtryk.

Tilføjelsen af yderligere bøjningsformer har også givet resultat. Af de 1.000 mest frekvente ikke-vellykkede søgninger fra september 2010 kan alle bøjningsformer undtagen passivformen *fåes* og imperativformen *ryd* håndteres, enten via de tilføjede bøjningsformer eller via mente du-funktionen, idet søgealgoritmen kan foreslå det rigtige ord, når det gælder mange af stave- og slåfejlene. Samtidig kan vi henvise til ODS, hvor der ikke er resultat i DDO. ODS har 225.000 opslagsord og dækker en ældre periode, så der henvises for eksempel, når der er tale om ældre ord som *møgle*, sammensætninger som *fødselsoverskud* og fremmedord som *sekularisme*. Det betyder, at af alle de 1.000 mest frekvente ikke-vellykkede søgninger fra september 2010 kan vi nu håndtere de 45 %, hvoraf de 10 % er henvisninger til ODS. Hvis vi antager, at resultatet for de frekvente kan overføres til også at gælde alle ikke-vellykkede søgninger, kan vi altså håndtere 45 % eller næsten halvdelen af de i alt 20 % ikke-vellykkede søgninger, dvs. at kun ca. 10 % af samtlige søgninger ikke kan håndteres, og det synes vi er et meget tilfredsstillende resultat.

## Fremtidsperspektiver

Undersøgelserne af logfilerne og statistikkerne fra Google Analytics giver os mange idéer til fremtidigt arbejde, dels med hensyn til indhold, dels med hensyn til funktionalitet: Hvad kan og skal brugeren gøre på siderne?

Fremmedord er en efterspurgt kategori, jf. følgende liste over ord, brugerne ikke kan slå op: *denotativ, emergent, residual, ebrieret, dopamin, intentionalitet, eksplorativ, inkrementel, semiotisk, nominal, elaborere*. Her kan redaktionen med fordel supplere. DDO er lidt underforsynet, hvad angår lidt sjældnere fremmedord på grund af de relativt strenge krav om korpusbaseret, som redaktionen har fulgt. Mange ord af denne type var ikke særlig frekvente eller overhovedet med i DDO's korpus. Dertil kommer, at mange brugere næppe slår udbredte dagligdags ord op, men mere ord, de ikke kender. Det skal dog tilfø-

jes, at man i DDO godt kan finde forklaring på flere af de nævnte ord via et beslægtet opslagsord, fx *semiotik* og *intentionel*.

En anden efterspurgt type er proprier. De er principielt slet ikke med i DDO, undtagen hvis de indgår i faste udtryk som *fra Herodes til Pilatus*. Der kan være flere grunde til, at proprier slås op: Der kan være problemer med stavningen: *Ålborg*, *Taastrup* (å eller aa), *Svejts*, *Egypten* (kan også staves *Schweiz* og *Ægypten*), eller der kan være tale om et vidensbehov, som vi kun kan gætte på: *Orion* (stjernebilledet?), *Luna* (?), *Nato* (hvad er det, og kan det staves *NATO*?), *Anders* (hvad betyder navnet?).

Sammensætninger er ikke overraskende også almindelige på listen over ikke-vellykkede søgninger: *bilreparation*, *bryllupsår*, *juletræstøppe*, *sigøjnerbande*, *koncessionsaftale*, *lømmelpakke*, *svenskerfrisyre*, *københavnskort*. Dansk, svensk og norsk er sprog, der kan danne i princippet uendeligt mange sammensætninger, og mange brugere vil måske bare slå et ord op for at se, om det findes, dvs. er registreret. Ord som de nævnte afspejler, hvad der tales om i tiden, fx *bryllupsår* og *sigøjnerbande*, og er som sådan ad hoc-dannelser, mens andre er udtryk for leksikalisering, fx *juletræstøppe*, der betegner en helt bestemt genstand. I *lømmelpakke* er ordet *pakke* leksikaliseret i en nyere betydning om et politisk tiltag, her et, der giver politiet udvidede beføjelser over for potentielt kriminelle. *Svenskerfrisyre*, der åbenbart indtastes sådan (korrekt stavning er *frisure*), er sammen med *svenskerhår* etableret på dansk som betegnelse for en bestemt frisure til mænd (kort hår fortil og langt bagtil; kaldes også *bundesligahår*).

Listerne byder også på helt nye ord, dvs. ord fra de sidste 10-15 år, som fx *slush ice*, *vintage*, *avatar*, *kvindechauvinisme*, *wasabi*, *zumba*, *simkort*, og af disse er de tre sidste da også kommet med i DDO inden for det sidste halve år.

Forkortelser kan man også godt forstå, at brugere vil slå op, og her kan der også suppleres med nyt stof: *LOL* (laughing out loud), *ifm* (i forbindelse med), *CI* (cochlear implantat), *ATP* (Arbejdsmarkedets Tillægspension). DDO indeholder allerede en del forkortelser, men området er naturligvis ikke udtømt, og der kommer nyt til, fx fra sms- og chatsproget. Mange forkortelser er proprier som fx *FN*, *EU*, *SNG*, *SS*, og også her kan redaktionen vælge at medtage denne type fremover, ikke mindst fordi de ofte er flertydige som *SS*: *Schutzstaffel*, *dampskib* (steamship) eller chatforkortelse for "smiler sødt".

Når det gælder synonymmer, må det siges, at DDO opererer med et snævert synonymbegreb, dvs. at den generelt ikke indeholder så mange synonymmer som fx Nudansk Ordbog og slet ikke så mange som en regulær synonymordbog, men brugernes indtastninger tyder på, at de ønsker flere synonymmer. På DSL arbejdes i disse år med en ny begrebsordbog for dansk, baseret på DDO og på det danske wordnet DanNet, og det er vores plan engang i fremtiden at kunne

tilbyde flere synonymer og beslægtede ord på grundlag af data fra begrebsordbogen (se Nimb 2011 og Trap-Jensen 2011).

Det er også vigtigt og oplagt, at redaktionen arbejder videre med at forbedre søgning på ét eller flere ord og på mente du-funktionen, der hjælper brugere, som begår mindre taste- eller stavefejl. I øjeblikket får man ikke hjælp, hvis man indtaster en bøjningsform af en almindelig stavefejl som fx *paranteser*, mens man bliver ledt på rette vej ved indtastning af singularisformen *parantes* (for *parentes*). Her er det oplagt at udvide bøjningsleksikonet med bøjningsformer til stavefejl. Andre indtastninger ligger for langt fra det rigtige til, at de kan opfanges af mente du-funktionen, selvom stavemåden *angsement* afspejler normal dansk udtale af *arrangement*, og her vil vi gerne på længere sigt give mulighed for fonetisk søgning, sådan som det allerede længe har været kendt i flere udenlandske ordbøger (fx den franske TLFi). Vi kan også overveje at tilføje bøjningsformer som passiv og imperativ; no match-listerne indeholder som nævnt mange bøjningsformer, hvilket tyder på, at brugerne enten kopierer en ordform ind i søgefeltet eller bruger de muligheder for direkte opslag, som vi tilbyder (den tidligere omtalte opensearch-funktion).

Andre fremtidsperspektiver går på at udvide og forbedre hjemmesidens funktionalitet. I øjeblikket er idiommer og andre faste udtryk søgbare, så de kommer frem i søgeresultatet, hvis man indtaster ét eller flere af de ord, der indgår i udtrykket. Det gælder dog kun de faste udtryk, der internt kaldes sublemmaer; de står i en særlig sektion af artiklen og er forsynet med mange af de samme oplysninger som hovedopslagsordene. DDO indeholder imidlertid også et stort antal ukommenterede ordforbindelser (kollokationer i bred forstand), der blot er nævnt under de relevante betydninger. De findes ikke frem med de nuværende søgemuligheder, og det vil være en oplagt forbedring, hvis de kommer til det. Brugere kan nemlig ikke vide, om en given ordforbindelse står som fast udtryk eller kollokation – og kender nok heller ikke forskel. I øjeblikket finder man fx ikke følgende ordforbindelser: *i forbindelse med, heller ikke, akademisk uddannelse, familiens sorte får, gøre honnør, veje for meget*, og det er da ærgerligt, når de faktisk er taget med i ordbogen.

Endelig er der perspektiver i at se mere på brugernes metasøgeudtryk, fx *synonym, andet ord for*. Det tyder – som nævnt – på, at brugerne ofte er interesserede i synonymer, og hvis de er det, bør vi give synonymer en mere fremtrædende plads i præsentationen og også gerne tilbyde flere synonymer. Et søgeudtryk som *ord med p* kunne antyde, at hjælpeteksterne bør være bedre til at lære de besøgende at bruge jokertegn, eller at vi skal kunne håndtere natursproglig søgning i et eller andet omfang, fx så man kan stille et spørgsmål.

## Konklusion

Vi kan således drage flere positive konklusioner på vores foreløbige erfaringer med brugerundersøgelser: For os har det været en succes at tillade webcrawlerne at indeksere alle vores sider, så vi kan kun opfordre andre til at prøve det. Besøgstallene på ordnet.dk er mere end tredoblet, siden vi begyndte i november 2010, og de er stadig stigende pr. september 2011, uden at vi har oplevet længere svartider eller nedbrud. Andre har før os konstateret, at logfiler giver værdifuldt input, og det er meget vigtigt at få dette både redaktionelle og funktionelle input. Helt sikkert er det i hvert fald, at der er meget arbejde at gøre fremover, og en væsentlig del af opgaven bliver at prioritere indsatsen.

## KILDER OG LITTERATUR

- Almind, Richard, 2008: Søgemønstre i logfiler. I: *LexicoNordica* 15. S. 33–55.
- Becker-Christensen, Christian m.fl., 2010: *Politikens Nudansk Ordbog*, 21. udg. København.
- Clifton, Brian, 2010: *Advanced Web Metrics with Google Analytics*, 2. udg., Wiley Publishing, Inc. Indianapolis.
- DanNet = DanNet 2009-. København: Det Danske Sprog- og Litteraturselskab og Center for Sprogteknologi, Københavns Universitet: [www.wordnet.dk](http://www.wordnet.dk) (oktober 2011)
- DDO = Hjorth, Ebba, Kjeld Kristensen m.fl.: *Den Danske Ordbog 2003-*. København: Det Danske Sprog- og Litteraturselskab. <http://ordnet.dk/ddo> (oktober 2011)
- Fjeld, Ruth Vatvedt og Sven-Göran Malmgren, 2010: Vård ett besök – om DSL:s nya webbsida ordnet.dk. I: *LexicoNordica* 17. S. 297–311.
- Hult, Ann-Kristin, 2008: Användarna bakom loggfilerna – redovisning av en webbenkät i Lexin online Svenska ord. I: *LexicoNordica* 15. S. 73–91.
- Nimb, Sanni, 2011: Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger. I: *LexicoNordica* 18. S. 135–155.
- Nygaard, Lars og Ruth Vatvedt Fjeld, 2008: Analyse av søkelogger for bedre søkmu-ligheter i elektroniske ordbøger. I: *LexicoNordica* 15. S. 57–72.
- ODS = Dahlerup, Verner m.fl.: *Ordbog over det danske Sprog inkl. supplement 1918-2005*. København: Det Danske Sprog- og Litteraturselskab. <http://ordnet.dk/ods> (oktober 2011)
- sproget.dk: <http://sproget.dk> (oktober 2011)
- TLFi = Le Trésor de la Langue Française informatisé. <http://zeus.atilf.fr> (oktober 2011)
- Trap-Jensen, Lars, 2011: Dansk begrebsordbog: baggrund og metode. I: *LexicoNordica* 18. S. 183–197.

*Henrik Lorentzen*

Det Danske Sprog- og Litteraturselskab  
hl@dsl.dk

*Liisa Theilgaard*

Det Danske Sprog- og Litteraturselskab  
lt@dsl.dk

# Hapax i Olof Verelius ordbok från 1681

*Anton Lundqvist*

Olof Verelius was the custodian of national monuments (*riksantikvarie*) in Sweden during the second half of the 17th century. He translated the Icelandic sagas into Swedish and published them. He also produced a dictionary of primarily Icelandic, but also Old Swedish words where the headwords are translated into Swedish and Latin. Among the equivalents in the Swedish translations, words occur that have not been documented elsewhere (hapax legomena). In the article I try to find out where these words come from and how they are formed. My conclusion is that the vast majority of them are formed as loan translations according to the Icelandic model. Many of the equivalents have simply been translated to the etymologically corresponding Swedish morphemes while others have been interpreted part by part. Direct loanwords also occur but are not so common. I also found, however, that words that only occurred in the dictionary were documented in the collections of the Swedish dialects. This shows that these were not the creations of Verelius, but must have been used in vernacular language.

*Nyckelord:* Verelius, Index, hapax, lånord, översättningslån, oegentlig dubbelöversättning, dialekt

## Olof Verelius

Olof Verelius innehade en personlig professur i fäderneslandets antikviteter i Uppsala och var riksantikvarie på 1660- och 1670-talen. 1600-talets nya stormakt Sverige hade behov av att visa på en lång och ärorik historia och i det syftet dokumenterade man fornlämningar samt studerade fornsvenska lagar och isländska texter. Verelius fick i uppgift att översätta isländska sagor som berörde svenska förhållanden. 1664 utkom ”Göthreks och Rolfs Wesgötha kongars historia”, 1666 ”Herrauds och Bosa saga” och slutligen 1672 ”Hervarar saga”, vilka alla utspelar sig i Sverige. I översättningarna införs flera isländska begrepp i svenskan såsom *ättestupa* (SAOBs: ättestupa), *bautasten* (Haugen 2007: 37 f.) och *bärsärk* (Hirvonen 1987: 113). De gamla texterna var inte lätta att förstå, så för att underlätta studiet fick Verelius även uppdraget att ”förfärdiga ett dictionarium öfwer alla gamble swenske ordh som både i dhe gamla laghar och

andra gamla skrifter fins på dhet man i framtijden till alla sådanne sakers interpretation måtte en godh anledning och underwijsning hafwa” (Schück 1933a: 14).

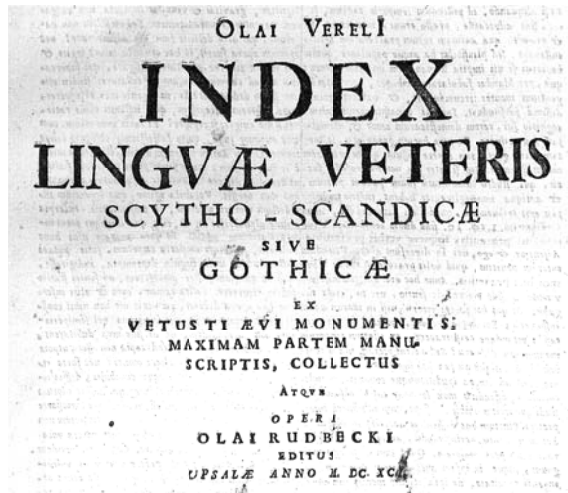
## Index lingvæ veteris scythoscandicæ sive gothicæ

Verelius uppgift och syfte var således att producera en ordbok och där förklara svårförståeliga ord, dels i de gamla fornsvenska lagtexterna, dels i andra gamla skrifter, dit tydligen också de nämnda isländska sagorna räknades.

Det svenska språket kan grovt delas in i en runsvensk, en fornsvensk och en nysvensk period. Den fornsvenska perioden sträcker sig från ca 1225

till ca 1526 varefter den nysvenska perioden vidtar. SAOB omfattar nysvenskt skriftspråk från 1521 och framåt. Det är tydligt att skillnaden mellan det äldre språkstadiet, fornsvenskan, och språket vid mitten av 1600-talet var så stor att behov fanns för översättning av vissa äldre ord och begrepp.

Resultatet blev ”Index lingvæ veteris scythoscandicæ sive gothicæ”, hädanefter kort och gott kallad Index. Verelius hade för eget bruk gjort en lista över isländska ord under sitt arbete med Olafssagan, med översättning till latin, och den listan kom att utgöra grunden till Index (Schück 1933b: 131). Majoriteten av uppslagsorden är därför isländska men bland ordbokens uppslagsord förekommer även ord på fornsvenska och gotiska. 1681 var Index så gott som färdig men Verelius avled 1682 utan att den hade publicerats. Hans gode vän Olof Rudbeck färdigställde den och 1691 kunde den tryckas.



## Hapax i SAOB och Index

Index är en ofta anlitad källa i SAOB. Från A till och med TYNA i SAOB är den citerad över 1 600 gånger. I Index svenska definitioner förekommer ord som endast är påträffade där och sålunda utgör s.k. hapax. Begreppet hapax är

grekiska och betyder ”det en gång sagda”. Med det avses ord eller uttryck som bara har påträffats en gång i ett äldre språkmaterial. Ofta medför det att innebörden av ordet blir osäker, då man inte kan jämföra flera olika belägg och författare. Problematiken har tidigare belysts av Anki Mattisson (Mattisson 2010). För att inte riskera att misstolka innebörden av ett ord krävs det i SAOB, sedan år 2000, minst två av varandra oberoende belägg på ett ord för att det skall tas med. Det medför sålunda att hapax inte längre tas med i SAOB. Eftersom SAOB inte täcker fornsvenska får dock ensamstående belägg i tidig nysvenska med förankring i fornsvenska medtas. De blir visserligen hapax i SAOB men är det ju inte i språket som helhet.

I denna undersökning utgår jag från att hapax i SAOB ur Index sannolikt även innebär hapax i Index. Det kan tyckas märkligt att Index innehåller hapax då syftet med ordboken var att förklara svårförståeliga ord. Tidigare okända ord bör ju ha ett begränsat förklaringsvärde i en ordbok. Syftet med min undersökning var därför att ta reda på vilka dessa hapax är och hur de har tillkommit eller bildats.

## Materialet

Materialet till undersökningen består av ord från Index som har citerats i SAOB och som utgör hapax i SAOB. Detta material kan ju inte betraktas som fullständigt, men kan väl ändå utgöra ett tämligen gott underlag för en diskussion. Materialet tog Språkdata i Göteborg fram åt mig. De har möjlighet att göra sökningar i ”SAOB på nätet” med fler parametrar än vad man normalt kan vid fritextsökning, vilket gjorde att de kunde få fram ett bättre grundmaterial. Sökningen gav ca 300 träffar och i den listades ställen i SAOB där Index utgjorde enda belägget. Efter att för undersökningen ointressanta poster hade tagits bort återstod 137 träffar. Orsaker till strykning var t ex att källan var felaktig, att momentet var levande eller att Index inte var enda belägg. I ett femtiotal av de strukna träffarna hade SAOB bedömt att betydelsen i Index var en annan än den hos övriga belägg varför de fått egna moment. Eftersom det i dessa fall endast är betydelsen som är udda, inte ordet i sig, har jag inte räknat dem som hapax och därmed inte tagit med dem i undersökningen.

## Utformning av artiklarna i Index

Artiklarna i Index består av ett uppslagsord följt av källhänvisning samt ekvivalenter på eller översättningar till nysvenska och latin. Ekvivalenterna åtskiljs i allmänhet endast med kommatecken. Någon enstaka gång används istället *och*



eller *eller*. Ordningen på källhänvisning och översättningar varierar. Ibland tillkommer ett språkprov där lemmat ingår och det översätts till de båda målspråken. Ibland förekommer kommentarer på målspråken. De olika språken är tryckta i olika stilar. För enkelhetens skull betraktar jag även fornsvenska som källspråk i Index, även om fornsvenska och nysvenska naturligtvis egentligen är olika stadier av samma språk.

Enligt Svensén (2004: 314 f.) kan ekvivalenterna i en ordbok utifrån sin utformning indelas i dels översättningsekvivalenter, dvs. ekvivalenter som går att sätta in i löpande text på målspråket, dels förklaringsequivallenter, som utan att vara insättbara ger ”användaren en uppfattning om det källspråkliga teckeninnehållet”. Utöver dessa finns encyklopediska förklaringar, som ofta används då målspråket saknar en användbar ekvivalent.

Utifrån denna terminologi kan ekvivalenterna i Index delas in i tre olika typfall:

1. Översättningsekvivalent(er).

*Illgerda flockur, Kong. S. pag. 224. Manipulus facinorosorum, Ilgerningz-hoopen.*

*Isl.* Illgerda flockur ... *sv.* Ilgerningz-hoopen.

2. Översättningsekvivalent(er) med en följande förklaringsequivallent.

*Sialfrædi, Ol. S. cap. 38. so. Friiheet / godtycke / sielfrådande att göra och låta / Arbitrium, liberium, facultas agendi.*

*Isl.* Sialfrædi ... *sv.* Friiheet, godtycke, sielfrådande att göra och låta.

3. Enbart förklaringsequivallent eller encyklopedisk förklaring.

*Slimuseta, Yöll. Manhel. cap. 29. Sicia i veitz læna slimuseta, Når man sitter så længe qwar i Gæstebodet / liſta som han med rumpan wore sömader eller beckader qwar på bæncken / Intempestive nimis in convivio adesse; abitum nescire.*

*Isl.* Slimuseta ... *sv.* När man sitter så länge qwar i Gästebodet, lijkja som han med rumpan wore sömader eller beckader qwar på bäncken.

Av de 137 träffarna i mitt material utgörs ca 115 stycken av översättningsekvivalenter, vilket därmed är den klart dominerande varianten.

## Lexikala novationer

Vid översättning, i synnerhet mellan närliggande språk, är det lätt att källspråket påverkar utformningen av översättningen. Anna Helga Hannesdóttir menar att ”språkens likhet kan förvillia översättaren så att dennes känsla för målspråket sätts ur spel” (Hannesdóttir 2001: 127). Det kan leda till att ord, efter viss anpassning, används som om de vore svenska. Det kan även gälla ordvalet, så att ett etymologiskt närliggande ord som är mindre vanligt eller folkspråkligt i målspråket, väljs framför ett mer brukligt (Edlund & Hene 1996: 92 ff.). För Verelius – precis som tycks ha varit fallet hos Biörner (se Hirvonen 1987: 132) – finns även möjligheten att översättaren hade ett politiskt intresse av att markera språkens gemensamma ursprung och därför medvetet valde att översätta med etymologiska motsvarigheter, ord av samma stam eller till och med etymologiskt skilda ord med liknande stavning.

Hapax är ofta tillfälliga lexikala novationer. Lexikala novationer skapas t ex då målspråket saknar ett ord för en företeelse i källspråket, dvs. då en semantisk lucka föreligger. Ofta använda lexikala novationsstrategier är då ordlån, blandlån, begreppslån eller användning av förklaringskvivalenter. (Edlund & Hene 1996: 33 ff.)

För att skilja översättningar där en av ekvivalenterna utgörs av ett ordlån, som endast nödtorftigt har anpassats till målspråkets morfologi, från sådana där samtliga ekvivalenter har ett förklaringsvärde, använder Ilkka Hirvonen begreppen ”oegentlig dubbelöversättning” respektive ”egentlig dubbelöversättning”. Hirvonen har jämfört Biörners Nordiska kämpadater från 1737 med Verelius sagaöversättningar och då kommit fram till att de oegentliga dubbelöversättningarna är betydligt vanligare hos Biörner än hos Verelius (Hirvonen 1987: 122 f., Haugen 2007: 104). Jag kommer också att begagna mig av denna terminologi.

## Olika typer av hapax i Index

### Ordlån

Verelius introducerade isländska ord i sina sagaöversättningar, oftast då orden rörde fornnordiska företeelser och fyllde ett behov i svenskan (Hirvonen 1987: 138–145). I Index verkar han dock ha blivit mer restriktiv med att använda ordlån och morfemlån, vilket är naturligt då syftet med ordboken ju var att förklara svårförståeliga ord. Enligt förordet är han försiktig med att dra paralleller till och att använda ord ur andra språk över huvud taget, något han förklarar med att: ”eftersom dessa små bäckar är från vår källa, har jag inte velat återföra ett

vatten som strömmar i fjärran till sitt källsprång” (Verelius 1681: Verelius företal, s. 2; översättning från latin: Gunnar Grauman). Några exempel finns dock:

Gron (SAOB G 961). Artikeln *sprotta* i Index lyder: *isl.* Sprotta ... *sv.* Spricka ut, spritta up. *isl.* Ey mun gron sprottinn ... *sv.* Skegge har inthet sprucki ut. *lat.* inde proverb. *Sv.* Gron näsa, Trum tryne. (Verelius 1681: 241). ”Inde proverb.” betyder ’därifrån ordspråket’. Samtliga ord i det anförda ordspråket betyder ’näsa, mule’ i någon form. *Gron* är en u-omljudd form av ”gran” som förekommer på västnordiskt område i betydelsen ’mule, tryne’ (Widmark 1959: 52). På östnordiskt område tycks Verelius vara ensam om ordet. Det finns inte heller belagt i OSDs. Det verkar därför rimligt att med Widmark konstatera att ordet inte bör ha ingått i Verelius egen vokabulär utan utgör ett lånord. I Index finns *gron* också med som egen artikel under lemmat *isl.* Graun & gran med betydelse ’mule, läpp; haka; skägg; näsa’ (Verelius 1681: 100).

Rate (SAOB R 401). *Isl.* Kampe ... *sv.* Gammal Kampe, Gamble raten (Verelius 1681: 137). *Rate*, motsvarande *isl.* *hrati* ’skräp’, återfinns som *råte* ’svag stackare; fänig person’ i Dalarna, Gästrikland och Härjedalen. Verelius stavar ordet med *-a-* som i isländskan, och inte med *-å-* vilket kan tyda på att han inte var bekant med det dialektala ordet. *Hrati* förekommer inte som uppslagsord i Index. Parallellordet *kampe* i betydelsen ’gammal, svag person’ är utöver i Index endast belagt i Glossarium sveo-gothicum från 1712 av Haqvin Spegel. Då Spegel har ordagrant samma definition på latin som Verelius är det sannolikt att han har skrivit av Index. Det fornisländska ordet *kampi* betyder enligt bland andra Cleasby & Vigfusson (1874: 330) och Kahle (1910: 170) ’person med skägg eller mustach’ och bör vara bildat till *kampr* ’skägg, mustasch’. Ordet används i Islendinga saga som nedsättande benämning på en gammal biskop. *Rate* och *kampe* används i samma betydelse och kan utgöra parallella ordlån motsvarande *isl.* *hrati* och *kampi*. SAOB för istället det svenska *kampe* (känt från 1616 i sammansättningen *hov-kampe*) till *kamp* ’strid’ och jämför med *isl.* *kappi* ’kämpe’ (SAOB K 305) vilket alltså etymologiskt är ett helt annat ord. *Rate* är sålunda sannolikt ett lånord medan *Kampe* antingen är ett lånord eller ett för Verelius känt inhemskt ord, homonymt med det isländska.

Det förekommer även hapax som utgörs av lånord från andra språk än något av källspråken:

Koppgåld (SAOB K 2350). *Isl.* Nefgildi ... *sv.* Näseskatt ... Mantalzpenningar, Koppgelt, Rompeskatt (Verelius 1681: 180). *Koppgelt* betyder ’huvudskatt, skatt erlagd av varje huvud’ och är ett lån av ett lågtyskt *koppgeld*.

Bärnhyttarearbete (SAOB B4832). *Isl.* Vesalmanlegt verk ... *sv.* Bernhyttarearbete (Verelius 1681: 288). Ordet används i betydelsen ’fuskverk’ som utgår

från det tyska uttrycket *auf der bärenhaut liegen* 'ligga på björnhuden' dvs. 'vara lat'. *Bernhyttare* är bildat efter tyska *bärenhäuter* till *bärenhaut*, 'björnhud'.

Kasa (SAOB K 652). *Isl.* Huskytra ... *sv.* Koite, Koja, Kasa (Verelius 1681: 129). *Kasa*, i betydelsen 'hydda, koja', är bildat av latin *casa*, 'hydda'.

I dessa senare fall bör orden ha funnits naturligt i Verelius språk. Med tanke på att han enligt citatet ovan inte ville använda ord från främmande språk kan man anta att de var någorlunda väl integrerade i svenskan, trots att de inte har blivit excerperade mer än en gång vardera i SAOB.

## Översättningslån

I Index finns flera exempel på hapax i form av en nysvensk ekvivalent som verkar vara påverkad av den isländska eller fornsvenska förlagan. I synnerhet vad gäller ekvivalenter med fornsvenska förlagor är det svårt att avgöra om orden var levande men inte nedtecknades annat än i Index eller om Verelius inspirerades av de fornsvenska lagtexterna och tog upp orden som nysvenska. Även då källordet är isländskt och den svenska ekvivalenten består av på 1600-talet fullt levande ordled är det svårt att avgöra om det är Verelius egen skapelse eller om det var ett gångbart ord. De flesta orden har även andra ekvivalenter eller förklaringar vilket gör att de kan ses som översättningslån som ingår i oegentliga dubbelöversättningar, det vill säga att källordet först ges en svensk gestaltning och sedan får en kompletterande översättning för att förklara vad det faktiskt betyder. Jag tar upp några exempel på översättningsekvivalenter som utgör hapax nedan.

Från isländska:

Boklära (SAOB B 3743). *Isl.* Bokfræde ... *sv.* Boklära (Verelius 1681: 38).

Bugnäsot (SAOB B 4551). *Isl.* Biug-nefiadur ... *sv.* Kroknäsot, eller Bugnäsot (Verelius 1681: 35). Förleden är väl en anpassning till svenska (jfr fsv. *bugh* 'krökning') medan *-nefiadur* har substituerats med det svenska *-näsot*.

Falsklös (SAOB F 204). *Isl.* Fals-laús ... *sv.* Falsklöös, trogen (Verelius 1681: 63). *Isl.* *fals* 'falskhet' har översatts med *falsk*, ett numera utdött substantiv med samma betydelse och *laus* med *lös*.

Illgärningshop (SAOB I 179). *Isl.* Illgerda flokkur ... *sv.* Ilgerningz-hoopen (Verelius 1681: 133). *Illgärning* var och är ett fullt levande ord och det finns en parallell bildning i *illgärningsfolk*. *Flokk* är substituerat med *hop*.

Illäte (SAOB I 182). *Isl.* Ill-læte ... *sv.* Illäte, jämmer, weklagan (Verelius 1681: 133). *Illäte* finns även belagt i fornsvenskan i betydelsen 'jämmer, klagan' (Söderwall 1: 590). SAOB tar upp flera sammansättningar på *ill-* som är belagda på 1600-talet och en del av dem kan vara inspirerade av isländskan, t ex

*illbragd* och *illråde*. Huruvida *illäte* i Index skall ses som ett inhemskt ord eller ett översättningslån är alltså osäkert.

Krigsklädnad (SAOB K 2737). *Isl.* Herskrud ... *sv.* Krigzklädnat, Krigzwapn (Verelius 1681: 117).

Landsöda (SAOB L 206). *Isl.* Landaudn ... *sv.* Landzöda, och förstörande (Verelius 1681: 152).

Lättfoter (SAOB L 1774). *Isl.* Lettfoetr ... *sv.* letfoter, snab på föttren (Verelius 1681: 158). Stavningen *let-*, 'lätt-', var levande på 1600-talet och senare leden *-foter* förekommer i fornsvenskan och var på 1600-talet levande i sammansättningen *snarfoter* (SAOB S 7975) med samma betydelse.

Nyverke (SAOB N 843). *Isl.* Nyvirki ... *sv.* Nywerke, arbete som är nyligen gjort (Verelius 1681: 183).

Olönt (SAOB O 664). *Isl.* Uleynt ... *sv.* Olönt (Verelius 1681: 296). Betydelsen är här 'inte dold'. *Löna* i betydelsen 'dölja, hemlighålla' var levande in på 1800-talet i svenskan, så den betydelsen bör inte ha känts så främmande för Verelius som den gör idag. Den vanliga betydelsen av *olönt* var och är dock 'obelönad, som inte får någon lön', så det verkar troligt att källspråket har påverkat översättningen.

Rättskjutande, substantiv (SAOB R 4103). *Isl.* Beinskyri ... *sv.* Rettskiutande (Verelius 1681: 32). *-Skeyri* är rimligen fel för *-skeyti*. Det isländska ordet är bildat av *beinn*, 'rät, rak, i en rak linje', och *skeyti*, 'något som skjuts' och betyder enligt Cleasby-Vigfusson 1874: 56 'a straight-shooting, a good shot'. *Isl. beinn* översattes med 'rätt'. *Isl. skeyti* översattes med en avledning av *skjuta*. *Prickskytte*, som företeelsen kallas idag, är inte belagt förrän 1944.

Skeppsöm (SAOB S 3779). *Isl.* Skipsaumr ... *sv.* Skepspijk, skepsöm, skepsz-nagell (Verelius 1681: 228).

Skälmskott (SAOB S 5805). *Isl.* Geigarskot ... *sv.* Skälms-skott som träffar dem som snarare borde hielpas (Verelius 1681: 89).

Stålbyxor (SAOB S 13807): *Isl.* Jarnhosor ... *sv.* Ståålböxor (Verelius 1681: 132). *Stål-* tycks ha varit den förled som tidigast föredrogs i fråga om utrustningsdetaljer av järn, t ex i *-handske* (1561), *-hatt* (1558), *-hjälm* (1541) och *-krage* (1558). Att Verelius byter ut förleden *järn-* till *stål-* kan innebära att han inlemmar ordet i ett känt mönster.

Sällhetsrum (SAOB S 16160). *Isl.* Sælldarstador ... *sv.* sälheetz rum (Verelius 1681: 212).

I några fall uppträder hapax i form av svenska islandismer som ekvivalenter till andra ord än den isländska förlagan.

Harskalös (SAOB H 476). *Isl.* Ottalaust ... *sv.* Harskalöst, tryggeligit (Verelius 1681: 195). *Harskalöst* är enligt SAOB bildat efter isl. *háskalauss*, ett ord

som inte förekommer i artikeln men i Index finns lemmat *Haskalaust* som översätts med 'vthan fara'. Verelius var alltså väl bekant med det isländska ordet då han använde *harskalöst* som ekvivalent.

Nästånge (SAOB N 1127). *Isl.* Fit ... *sv.* Slät plas ... Nästånge, Vdd som löper uth i Siön doch slät och jämn (Verelius 1681: 72). *Nästånge* är enligt SAOB bildat efter *isl. nestangi*. I Index finns en artikel *isl.* Nestangr ... *sv.* Näs, udd, wid Hafwet (Verelius 1681: 181).

Från fornsvenska:

Några hapax utgörs av ord som har använts i översättning till nysvenska från fornsvenska och kan därmed möjligen betraktas som översättningsekvivalenter.

Husrån (SAOB H 1474): *fsv.* Bos ran ... *sv.* Huusråhn och Röfwande (Verelius 1681: 38).

Husöda (SAOB H 1480): *fsv.* Bol-özla ... *sv.* Huus-öda, hus-röta (Verelius 1681: 39).

Rånsböte (SAOB R 3646): *Fsv.* Ransbot ... *sv.* Rånsböte (Verelius 1681: 201). *Rånsböte* betyder detsamma som *rånsbot*, 'bot för ett rån', ett ord som användes sista gången 1726 enligt SAOB, dock i ett referat av Kristofers landslag från 1442.

Rånsverk (SAOB R 3647): *Fsv.* Ransverk ... *sv.* Rånande, Rånswärk (Verelius 1681: 202).

I Index har som framgår av exemplen ovan källspråkets ord översatts både med hjälp av motsvarande etymologiska morfem och genom översättning av betydelsen hos de skilda morfemen. Flera av exemplen syns påverkade av isländska och dessutom kompletteras de av ytterligare ekvivalenter med ett större förklaringsvärde. De kan tolkas som översättningslån ingående i oegentliga dubbelöversättningar.

## Folkspråkliga ord

Bland orden i mitt material finns det sådana som, trots att de för övrigt inte har hamnat i SAOB eller SAOBs, har levt ett liv i de svenska dialekterna. Exempel på det är:

Hamn (SAOB H 223). *Fornnorska* Hamna ... *sv.* Hugg, oc hamna, Thet är Bya allmenning på hwilken flera intressera uthi Skogzhugge och Mulebete. Sådanne hamner hafwa the uthi Wåla härad (i Västmanland). Hwarest de och kalla Hamnarör eller Hamnarå och Hamnarätt (Verelius 1681: 107). *Hamn, hamna* 'betesmark' återfinns i dialektuppteckningar från Bohuslän, Dalsland

och Värmland och även i norskan. Verelius består oss även med de enda beläggen i SAOB på *hamnarör*, *hamnarå* och *hamnarätt*.

Rubbot (SAOB R 2723): *Isl.* Rubbungligur ... *sv.* Som har fuult och stygt Ansichte, Rubbuter (Verelius 1681: 210). *Rubbot* med betydelserna 'med skrovlig, ojämn yta; grov i ansiktet' är i SAOBs endast dokumenterat en gång men i OSDs förekommer det i formerna *rubbot* och *rubbug* i Bohuslän, Jämtland och Ångermanland.

Rävköp (SAOB R 4235). *Isl.* Kauprof ... *sv.* Kiöpruff, Räfköp (Verelius 1681: 139). *Rävköp* är etymologiskt identiskt med det tidigast 1698 belagda *revköp*, 'köp som rivs upp eller annulleras' (SAOB R 1604). En annan variant är det i SAOB endast 1637 belagda *rivköp* (SAOB R 2264). Samtliga varianter finns även belagda i OSDs även om *revköp* är den vanligaste. *Rävköp* är i OSDs endast belagt från Hulterstad sn på sydöstra Öland. Ordformen berodde enligt meddelaren på att vederbörande var "rävakti" (Lindroth 1926: 217). Lindroth funderar över om det är en individuell folketymologisk omtolkning av ordet, men eftersom formen även föreligger hos Verelius tycks den ha haft viss spridning.

Fala (SAOB F 104). *Isl.* Auskufaulr ... *sv.* Swart bleck, Askefärgat som falan på Askan (Verelius 1681: 24). *Fala* i betydelsen 'falaska' är egentligen oblikt kasus av ett svagt maskulinum på -e (SAOB F 104). I OSDs finns belägg från Dalarna på form på -a medan formen *fale* finns belagd i Blekinge, Bohuslän, Dalarna, Dalsland, Jämtland, Värmland och Västergötland.

Gryntäljare (SAOB G 1092): *Isl.* Gloggur ... *sv.* Nisker, Granräknat, Gryntäljare (Verelius 1681: 95). *Gryntäljare* har jag inte funnit några fler belägg på, däremot bistår OSDs med den närliggande varianten *grynräknare*, 'mycket snål person', från Blekinge. Bilden av en snål person som någon som noggrant räknar sina gryn har alltså funnits och blivit dokumenterad i vart fall två gånger.

Vissa ord som i SAOB uppträder som hapax har som några exempel ovan visar varit levande i dialekterna och utgör folkspråkliga inslag i Index.

## Sammanfattning

Verelius uppdrag med Index var att i en ordbok förklara svårbegripliga ord. Trots det är en del av de ord han använder i definitionerna endast påträffade i Index och utgör därmed hapax. Jag har med utgångspunkt från citat i SAOB studerat dessa hapax, bland annat för att se hur de har tillkommit eller bildats. Rena ordlån förekommer men är mindre vanliga. En stor del av orden i materialet utgörs av översättningslån, antingen översatta med motsvarande etymologiska morfem eller med ekvivalenter som inte är etymologiskt besläktade. De

kompletteras för det mesta med ytterligare ekvivalenter eller förklaringar. De kan därför ses som isländska ord i svensk gestaltning ingående i oegentliga dubbelöversättningar. Det är ibland svårt att avgöra om ett visst hapax är en nybildning av Verelius. Några ord som är hapax i SAOB har visat sig förekomma i dialektuppteckningar och kan därför ha varit kända av Verelius.

## KÄLLOR OCH LITTERATUR

- Cleasby, Richard & Vigfusson, Gudbrand, 1874: An icelandic-english dictionary. Oxford.
- Edlund, Lars-Erik, & Hene, Birgitta, 1996: Lånord i svenskan. Om språkförändringar i tid och rum. 2. uppl. Stockholm.
- Hannedóttir, Anna, Helga, 2001: Ekvivalent och ekvivalent – det beror på vad man menar. I: Gäller stam, suffix och ord. Festskrift till Martin Gellerstam den 15 oktober 2001. Göteborg.
- Haugen, Susanne, 2007: Från bautasten till bautastor. Studier över fornvästnordiska *bautasteinn* och svenska ord bildade med *bauta(-)*. Umeå.
- Hirvonen, Ilkka, 1987: ”Fornälskarom till tjänst” – om språket i E. J. Björners Nordiska kämpadater. En jämförelse med Olof Verelius sagaöversättningar. I: Diverse. Vänskrift till Bertel Fortelius 3.8.1987. Åbo.
- Kahle, B, 1910: Die altwestnordischen Beinamen bis etwa zum Jahre 1400. I: Arkiv för nordisk filologi 26, s. 142-202. Lund.
- Lindroth, Hjalmar, 1926: Ölands folkmål. Göteborg.
- Mattisson, Anki, 2010: Hapax i SAOB. I: Från sjö till hav. Namn- och ordstudier tillägnade Birgit Falck-Kjällquist. Göteborg.
- OSDs = Samlingarna till Ordbok över Sveriges dialekter. Dialekt- och folkminnesarkivet i Uppsala.
- SAOB = Ordbok över svenska språket, utgiven av Svenska Akademien, 1893 ff. Lund.
- SAOBs = Samlingar till Ordbok över svenska språket. Lund.
- Schück, Henrik, 1933a: Kgl. Vitterhets- historie- och antikvitetsakademien. 2. Antikvitetskollegiet 1. Stockholm.
- Schück, Henrik, 1933b: Kgl. Vitterhets- historie- och antikvitetsakademien. 2. Antikvitetskollegiet 2. Stockholm.
- Spegel, Haqvin, 1712: Glossarium-sveo-gothicum eller Swensk-ordabook, inrättat them til en wällmeent anledning, som om thet härliga språket willia begynna någon kunskap inhämta.
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. Stockholm.
- Söderwall, Knut, 1884–1918: Ordbok öfver svenska medeltids-språket. Lund.
- Verelius, Olof, 1664: Göthreks och Rolfs Wsegötha kongars historia på gammal göthiska fordom beskrefwen och nu med ny uttolkning utgången. Uppsala
- Verelius, Olof, 1666: Herrauds och Bosa saga. Med en ny vtolkning iämppte gamla götskan. Uppsala.



- Verelius, Olof, 1672: *Hervarar saga på gammal götska med Olai Vereli vttolkning och notis*. Uppsala.
- Verelius, Olof, 1681 (postumt utg. 1691): *Index lingvæ veteris scythoscandicæ sive gothicæ*. Uppsala.
- Widmark, Gun, 1959: *Det nordiska u-omljudet. En dialektgeografisk undersökning. I. Text*. Uppsala.

*Anton Lundqvist*

Svenska Akademiens ordbok i Lund  
anton.lundqvist@svenskaakademien.se

# Lexikografisk beskrivning av ortnamn

*Jonas Ljöfström*

Although the relative lexical frequency of proper nouns is 10 to 15 per cent of the vocabulary in most texts, only a limited number of them are listed in language dictionaries, and with very little linguistic information.

Toponyms constitute, besides anthroponyms, the main representative of proper names. They deserve a specific linguistic description, all the more as complex toponyms may be formed with common nouns, adjectives and grammatical words, a fact which blurs the boundary between proper names and others.

Simple as well as complex toponyms (including both endonyms and exonyms) require a formal description concerning e.g. gender, number and use of determiners. A functional analysis approach taking into account, on the one hand, appellative and proprial elements and, on the other, their generic and specific functions, aims to modelize the structures of toponyms and will pave the way to relevant contrastive analyses of toponyms in different languages, and will also help formalising alphabetic order principles.

Last but not least, pragmatic information, based on corpus studies, is essential for a realistic description of the use of variant forms.

Man kan observera en tendens att behandla ortnamn styvmoderligt både i grammatiker och i ordböcker. Denna svaga ställning stämmer inte med ortnamnens generellt sett relativt höga lexikaliska frekvens. Man kan även konstatera att de språkliga uppgifterna om ortnamnen inte alltid är fullständiga. Detta är problematiskt eftersom viss information, som genustillhörighet och kongruensböjning, inte systematiskt överensstämmer med det övriga ordförrådets och man inte säkert kan sluta sig till dem utifrån allmänspråklig kompetens. Ortnamnen, liksom egennamn i allmänhet, hänvisas oftast till uppslagsverk och de språkliga uppgifterna där är sällan utförligare. Även i Svenséns (2009: 73ff.) utförliga handbok nämns endast personnamnen och ortnamnens komplexitet och speciella problem lyser med sin frånvaro.

Problemet med ofullständig språklig beskrivning av ortnamnen är speciellt akut när det gäller tvåspråkiga ordböcker där redogörelsen traditionellt huvudsakligen är begränsad till viktiga orter i de områden där de berörda språken

talas. En fråga är hur välmotiverad denna geografiska begränsning är; en annan vilken språklig information som ges och kunde ges ordboksanvändaren.

En inte obetydlig andel av ortnamnen är komplexa och innehåller lexikaliska element från allmänspråket. Redan detta faktum motiverar en generösare behandling av såväl endonymer som exonymer, då dessa ofta blir föremål för eller är resultatet av översättningar. Eftersom det rör sig om ett specialområde är vokabulären inte heller alltid helt förutsebar utifrån allmänspråket.

## Egennamn och transparens

I Svenska Akademiens grammatik finns en detaljerad genomgång av egennamnens struktur och funktion samt deras referenttyper. Likheten med substantiven exemplifieras med att de prototypiskt har en individuativ referent och syntaktiskt fungerar som huvudord i en nominalfras som definitionsmässigt har definit innebörd (SAG 2: 117). Övergångar mellan de två ordklasserna är inte ovanliga, genom att egennamnen kan vara lexikaliseringar av en form av ett substantiv respektive att egennamnen substantiveras vid metaforisk användning. SAG konstaterar också beträffande ordklassen egennamn att gränsen mellan svenska och främmande språk ofta är "oklar, eftersom utländska referenter normalt bibehåller det egennamn som de fått i respektive språkmiljö: *Los Angeles, Beethoven.*" (SAG 2: 117). Egennamnen och inte minst ortnamnen befinner sig i praktiken i skärningspunkten mellan olika språk och kulturer.

Denna särställning beror inte bara på att ett utländskt egennamn används i (till exempel) svenskan utan också på att egennamnens, och då särskilt ortnamnens, struktur kan vara komplex och innehålla ord från olika ordklasser. Även om ortnamnen (proto)typiskt inte har någon betydelse, så är de komplexa ortnamnen i sig ofta möjliga att tolka i det givna språket med avseende på de ingående komponenterna. Dessa ortnamn blir därmed i någon mening transparenta, vilket öppnar mot ett möjligt överförande av ursprungsspråkets endonym till ett exonym i det mottagande språket (exonymisation). I den mån ortnamnet innehåller ord som betecknar geografiska eller administrativa begrepp motsvarar det inte helt den typiska egenskapen hos egennamnet, uttryckt sålunda i SAG: "Det prototypiska egennamnet ... säger oss i sig inget annat om namnets bärare än att den bär just det namnet." (SAG 2: 116). Motsägelsen är oftast skenbar eftersom man ofta har att göra med sekundära (användningar av) ortnamn.

## Synkroni, språkperspektiv och komplexitet

För en analys som behandlar ortnamnen och deras plats i det svenska språksystemet krävs ett synkroniskt perspektiv och en utgångspunkt i svensk språkkompetens. Vid analysen görs skillnad mellan enkla och komplexa enheter. Utifrån svensk språkkompetens behandlas alltså utländska enheter som oanalyserbara. Enkla ortnamn på svenska är till exempel *Hjo* och *Partille*; de är synkroniskt oanalyserbara svenska endonymer. De utländska ortnamnen *Paris*, *Belleville* och *Kriegsheim* är ur det svenska perspektivet också enkla ortnamn.

Komplexa ortnamn kan delas upp i olika kategorier. Sammansättningarna (SAG 2: 129ff.), som uppvisar olika betoningmönster, utgör en stor grupp i svenskan, med exempel som *Östersjön*, *Storstockholm*, *Karlstad* och *Mariestad*. Lexikaliserade ordgrupper (SAG 2: 133) kan också utgöra ortnamn: *Lilla Edet*, *Västra Frölunda* som har ordgruppsbetoning och är utan framförstådd bestämd artikel.

För de tre nämnda typerna gäller att de kan vara lexikaliserade böjningsformer av substantiv (SAG 2: 132ff.): *Vänern*, *Viskan*; *Klarälven*, *Säveån*; *Lilla Edet*, *Ume älv*, *Göta kanal*. Lexikaliserade böjningsformer finner vi även i de svenska exonymerna: *Nilen*, *Tibern*, *Themsen*; *Kongofloden*, *Uralbergen*, *Pyrenéerna*, *Mellanöstern*, *Godahoppssudden*; *Främre Orienten*, *Döda havet*. Staters officiella namn kan ofta uppvisa rätt komplexa ordgrupper: *Amerikas förenta stater*, *Socialistiska folkliga libyska arabiska Jamahiriya* är några exempel. Till detta skall läggas de i många språk vanliga ortnamnen med prepositions konstruktion. I svenskan verkar de endast finnas i exonymer översatta från ursprungsspråket (eller ett förmedlande språk): *Öarna under vinden*. Den nyligen myntade svenska endonymen *Lidköping vid Vänern* står tämligen isolerad. Man kan vid en undersökning av flera språk konstatera förekomsten av flera olika typer av komplexitet och inre struktur (se Löfström & Schnabel-Le Corre 2011).

De utländska komplexa ortnamnen är i princip ogenomskinliga och fungerar som oanalyserbara strängar: *Costa del Sol*, *Frankfurt am Main*, *Aix-en-Provence*, *Stratford-on-Avon*. I den mån dylika spanska, tyska, franska och engelska endonymer skulle förekomma i en svensk ordbok är det en öppen fråga vilken språklig information läsaren skulle behöva.

## Lexikalisk frekvens

Ordklassernas olika frekvens i löpande text varierar av stilistiska och innehållsliga skäl beroende på texttyp. Det är inte helt lätt att finna tillförlitliga uppgifter om just ortnamnens frekvens eftersom frekvensundersökningar snarast gör en ordklassindelning av ordförrådet på en nivå som avser enskilda ord (på grafords-

nivå) och där behandlas alla egennamn bara som proprier oavsett om de är förnamn, efternamn, ortnamn, institutionsnamn etc. Även om andelen proprier varierar från texttyp till texttyp, finner man generellt en inte obetydlig del proprier. Vi exemplifierar här med resultaten från två frekvensundersökningar gällande svenska tidningar respektive franska romaner (Allén 1971 och Engwall 1984). I dessa utgör proprierna 4,45 respektive 2,44 procent av orden textuellt medan den relativa lexikaliska frekvensen är hela 15,10 respektive 10,40. Denna relativa lexikaliska frekvens kan jämföras med substantivens (55,48 respektive 49,00), adjektivens (10,45 respektive 14,25) och verbens (6,53 respektive 18,20). I de svenska tidningstexterna utgör alltså proprierna den tredje största ordklassen lexikaliskt och i de franska romanerna den fjärde; ordförrådet av adjektiv och verb i litterär text är inte förvånande mer varierat. Att en ordklass som är den tredje eller fjärde största beroende på typ av text borde vara värd en detaljerad lexikalisk beskrivning förefaller uppenbart.

När det gäller att avgöra ortnamnens andel av proprierna kan man gå igenom materialet som ligger till grund för frekvensberäkningarna på ordnivå. Tillgängliga data är basvokabulären, d.v.s. ord med en god spridning över olika texttyper och tidningar (med en modifierad frekvens på mer än 4 i tidningarna (i en korpus på 1 miljon löpande ord)). Det ger ett resultat på ungefär 200 ortnamn, 500 personnamn och 40 andra namn. Men metodologiskt viktigare är att ta hänsyn till det faktum att ortnamn, kanske mer än andra egennamn, utgörs av varierande kombinationer av ord. I själva verket kan i stort sett alla ordtyper ingå i en syntagm som bildar ett ortnamn. En del av de ord som på ordnivå klassificerats som till exempel adjektiv, substantiv, preposition, verb, utländsk enhet eller proprium ingår i den underliggande texten i en (ort) namnsyntagm. Detta innebär som nämnts, å ena sidan, att frekvenserna för proprier på ordnivå bara delvis motsvarar ortnamnsenheter. Å andra sidan tillkommer på frasnivå ett inte obetydligt antal namnsyntagmer, varav givetvis en stor del utgörs av ortnamn. Dessa kan inte direkt utläsas av de publicerade frekvenserna; de på frasnivå redovisade namnen är bara de som har en god spridning över texttyper och därmed bortfaller alla enstaka förekomster. Av både strukturella och frekvensrelaterade skäl kvarstår det faktum att ortnamnen utgör ett viktigt undersökningsobjekt för lexikografen.

Svenska tidningar				Franska romaner					
Klass	Ftext	Flex		Klass	Ftext	Flex			
	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	
Adjektiv	77k	7,75	7441	10,45	Adjektiv	22k	4,52	2308	14,25
Subst.	221k	22,17	39486	55,48	Subst.	86k	17,24	7934	49,00

<u>Proprium</u>	44k	4,45	10746	15,10	<u>Proprium</u>	12k	2,44	1684	10,40
Verb	172k	17,22	4649	6,53	Verb	96k	19,33	2947	18,20
---					---				
Totalt	1M	100,00	71178	100,00	Totalt	500k	100,00	16193	100,00

## Bruk av artikel i vissa språk, genus och numerus

I de enspråkiga ordböckerna har ortnamnen sällan sin plats. I uppslagsverk däremot är de talrika och får en encyklopedisk behandling. Någon annan språklig information än stavningen ges inte för de inhemska ortnamnen. Däremot ges vid behov uttalsangivelse för utländska namn och då den encyklopediska informationen om uttalet på ortens språk. Det utländska endonymets plats i det svenska språkssystemet berörs av naturliga skäl inte. Indirekt hänvisas läsaren till språkkänslan eller den svenska grammatiken.

Exempel på behandlingen av ortnamn i uppslagsverk (Bra Böckers Lexikon):

- Aachen [fonetik] på franska Aix-la-Chapelle, ...
- Appalacherna [fonetik], en 2600 km lång bergskedja...
- Appenzell [fonetik], kanton i nordöstra Schweiz...
- L'Aquila [fonetik], huvudstad i provinsen Abruzzi...
- Arabien eller Arabiska halvön, halvö i ...
- Arabiska republiken Egypten, det officiella namnet...
- Arabiska viken eller Persiska viken mellan Iran, ...
- Ålborg, dansk stad...
- Öxnevalla, församling i Marks kommun...

I franska ordböcker och uppslagsverk finns det en tradition att ange ortnamnens genus (Petit Larousse illustré, Petit Robert des noms propres). Exempel:

Appalaches	n.m.pl. ...	/n.f.pl.
Apuseni (monts)	- ...	
Arabes unis [Emirats]	→ Emirats arabes unis (Etats des)	
Arabe unie [République]	→ République arabe unie	
Arabie	n.f. ...	
Arabie Saoudite	n.f. ...	
Arabique (golfe)	- ...	
Arabique (désert)	- ou désert Oriental...	
Xian ou Si-ngan	V(ille) de Chine...	
Yamuna ou Jamna	n.f. Rivière de ...	
Zwundrecht	Comm(une) de Belgique ...	

Genusangivelsen är här till för att ange att ortnamnet används med bestämd artikel i kontext. När ett substantiv inom parentes anges efter uppslagsordet anges inte genus eftersom ett substantiv från det allmänna ordförrådet förutsätts vara känt av läsaren. De första raderna i exemplen ovan tolkas alltså: *les Appalaches*, *les monts Apuseni*, (*les Etats des*) *Emirats arabes unis*, etc. När det gäller *le desert Arabique*, staden ('ville') *Xian* och floden ('rivière') *Yamuna* anges alternativa former. Staden *Xian* och kommunen *Zwundrecht* redovisas som övriga ord med encyklopediska uppgifter men däremot utan genusangivelse. Detta får tolkas så att de inte tar bestämd artikel.

Principen att uppgift om genus endast används för att ange bruk av bestämd artikel riskerar att ge en snedvriden bild av ordförrådets egenskaper. Detta kan ha att göra med den osäkerhet som observeras i hur kongruensen fungerar i kontext. Jämför ordet *Appalaches* ovan där två encyklopedier anger olika genus, maskulinum pluralis respektive femininum pluralis. Man kan också i bruket, såväl i franska som i svenska, konstatera att ett enkelt personligt pronomen undviks för att hänvisa till en stad och att man föredrar uttryck som 'den gröna idyllen', 'vår huvudstad'. Dessa ersätter ett enkelt 'det'. Däremot fungerar kongruensen vid adjektiv; i svenskan är tätorter neutrer, 'det glada Hudik', 'Malmö är vackert beläget vid Öresund'. I franskan verkar vacklan i valet av genus vara större än i svenskan. Den bestämda artikelns status i franskan är inte sällan oklar. Kongruens med genus och numerus kan skilja sig från bruket vid motsvarande vanliga substantiv och adjektiv. Så kan även vara fallet i svenskan. Principerna är delvis olika beroende på språk. Ur lexikografisk synvinkel bör dessa förhållanden speglas i ordboksartiklarna. Det förhåller sig, som man kan vänta sig, delvis annorlunda i de tvåspråkiga ordböckerna.

I tvåspråkiga ordböcker blir informationen om genus och bruk av artikel viktigare än i de enspråkiga. För den svenska publiken har följande presentation valts för Fransk ordbok. I den fransk-svenska delen anges den franska toponymen med uttal och genusuppgift. Dessutom ges toponymen utskrivet med artikel samt den svenska översättningen utan någon ytterligare upplysning, t.ex. Irak [uttal] m, *l'Irak* – Irak. På motsvarande sätt görs i den svensk-franska delen men utan uttalsangivelse: Irak – *l'Irak* m. Informationen om artikelbruket är mer explicit än i de franska citerade uppslagsverken. För de fall där ingen artikel används markeras detta genom att genusangivelse utelämnas: Malte [uttal] – Malta, respektive Malta – Malte.

För svenskan och andra nordiska språk är problemet med artikelns vara eller icke vara marginellt eftersom artikeln är ett suffix och när den förekommer är den alltså lexikaliserad och närvarande redan i uppslagsformen.

## Motsvarigheter i tvåspråkiga ordböcker

Ovan behandlades ett sätt att i tvåspråkiga ordböcker redogöra för artikelbruket. På ord- och ordgruppsnivå tillkommer problemet att på ett välavvägt sätt presentera enheter med sammansättningar eller variantformer, i synnerhet när de två språken, som franskan och svenskan, inte har samma inre struktur. Det blir ett fall där presentationen i bokstavsordning inte är given.

### Fransk – Svensk

<b>azur</b> [uttal] m <i>la côte d’A~</i>	himmelsblått, azurblått; Franska Rivieran
<b>azuré, -e</b> [uttal] <b>Afrique</b> [uttal] f. l’ <i>Afrique</i> -l’ <i>Afrique australe</i> -l’ <i>Afrique du Nord</i> -l’ <i>Afrique du Sud</i>	himmelsblå, azurblå Afrika södra Afrika Nordafrika Sydafrika
<b>Maghreb</b> [uttal], <i>le M~</i> <b>Sahel</b> [uttal], <i>le ~</i> geogr.	Maghreb, Nordafrika Sahel
<b>Baltique</b> [uttal] I <i>la mer ~</i> II <i>la ~</i>	Östersjön Östersjön
<b>Scandinavie</b> [uttal], <i>la ~</i> –	Skandinavien
<b>balte</b> [uttal] <i>adj</i> <i>les pays ~s</i>	baltisk Baltikum, de baltiska staterna

### Svensk – Fransk

<b>azur</b> s <b>Rivieran</b>	azur <i>m</i> franska la Côte d’Azur; italienska la Riviera
<b>azurblå</b> <i>adj</i> <b>Afrika</b> - <i>södra Afrika</i> <b>Nordafrika</b> <b>Sydafrika</b>	azuré, bleu d’azur (oböjl.) l’Afrique f. l’Afrique australe l’Afrique du Nord l’Afrique du Sud
– – <b>Östersjön</b>	– – la [mer] Baltique
<b>Skandinavien</b> <b>Skåne</b> <b>Öresund</b> <b>baltisk</b> <i>adj</i> ; <i>de ~ staterna</i> <b>Baltikum</b>	la Scandinavie la Scanie le Sund balte, baltique les pays <i>mpl</i> baltes les pays <i>mpl</i> baltes

Man konstaterar för till exempel *Côte d’Azur* och *pays Baltes* att de presenteras under ett allmänt uppslagsord, substantiv eller adjektiv, *azur* m. respektive *balte* adj. I den fransk-svenska ordboken presenteras komplexa toponymer som *Afrique australe*, *Afrique du Nord* under det enkla uppslagsordet *Afrique*. Motsvarande presentation har inte valts i den svensk-franska delen utom för *södra Afrika*, där *Afrika* uppträder som självständigt graford. Vid sammansättningarna *Nordafrika* och *Sydafrika* har strikt bokstavsordning valts.

För övrigt kan man gällande dessa toponymer som avser topografiska objekt inom fransk respektive svensk kultursfär eller geografisk sfär konstatera att de inte alltid är ömsesidigt närvarande; *la Scanie* och *le Sund* finns inte i den fransk-svenska delen och *le Maghreb* och *le Sahel* inte i den svensk-franska delen. Trots att *Maghreb* återges med *Maghreb*, *Nordafrika* på svenska och att uppslagsformen *Nordafrika* (översatt med *Afrique du Nord*) finns i den svensk-franska delen ingen ingång till *Maghreb*. Samma sak gäller *Sahel*. I vad mån detta är följden av medvetna val eller ett resultat av blandade eller motstridiga principer är inte lätt att avgöra. Situationen illustrerar i alla fall de principiella problem som toponymerna ställer lexikografen inför. Några andra exempel följer nedan.



## Fransk – Svensk

Amérique [uttal] f. l'Amérique

Se även *Etats-Unis*  
 -l'Amérique centrale  
 -l'Amérique latine  
 -l'Amérique du Nord  
 -l'Amérique du Sud

Emirats Arabes unis [] mpl  
les ~Etats-Unis [] mpl, les *Etats-Unis*

Grande-Bretagne [],  
 la *Grande-Bretagne*  
 Royaume-Uni [], le ~

Amerika

Centralamerika  
 Latinamerika  
 Nordamerika  
 Sydamerika

Förenade Arab-  
emiraten

Förenta staterna

Storbritannien  
 Förenade kungariket

## Svensk – Fransk

**Amerika** l'Amérique f;  
 ~s **förenta stater** les Etats-Unis  
 [d'Amérique]

Se även *Förenta staterna*

**Centralamerika** l'Amérique f centrale  
**Latinamerika** l'Amérique f latine  
**Nordamerika** l'Amérique du Nord  
**Sydamerika** l'Amérique du Sud

**Förenade Arabemiraten**  
 les Emirats mpl Arabes unis  
**Förenta staterna** les Etats-Unis m

**Storbritannien** la Gde-Bretagne

Ovanstående exempel redovisar, med utgångspunkt i några franska uppslagsord i bokstavsordning, de svenska motsvarigheterna samt, i särskilda kolumner, dessa svenska exonymer som uppslagsord i den svensk-franska delen av ordboken med deras franska exonymer. Överensstämmelsen är som synes mycket god, med ett undantag, *le Royaume-Uni – Förenade kungariket*, som saknas i den svenska delen (se nedan under Variantformer). Man konstaterar här att de två språkens olika struktur gör, när det gäller toponymer med ingående *Amérique* respektive *Amerika*, att för franskan alla komplexa toponymer presenteras under bokstaven A, medan man i svenskan, som har bestämningsledet framförställt, får leta efter de olika toponymerna på olika ställen för att få den överblick som presenteras här ovan. Frågan diskuteras nedan under rubriken Val av uppslagsform.

## Variantformer

Ovan visades att den engelska endonymen *United Kingdom* motsvaras av den franska exonymen *Royaume-Uni* och den svenska *Förenade kungariket*. Detta är den officiella svenska översättningen med varianten *Förenade konungariket Storbritannien med Nordirland*. Man ser så gott som aldrig *Förenade kungariket* i media och få svenskar verkar känna till denna svenska exonym. I svenska texter använder man i stället t.ex. *Storbritannien*. Det vore inte orimligt att då åtminstone under rubriken *Storbritannien* i den svensk-franska ordboken införa *Grande Bretagne* och *Royaume-Uni* ('*Förenade kungariket*'). Det är inte heller helt orimligt att överväga att presentera en rad variantformer med uppgift om användningsområde. Sålunda kunde *Royaume Uni – Royaume uni de Grande-Bretagne et de l'Irlande du Nord – Grande-Bretagne – îles Britanniques* och kanske oegentligt även *England* sammanföras eller få korshänvisningar. Motsvarande

svenska exonymer är *Förenade kungariket – Förenade konungariket Storbritannien med Nordirland – Storbritannien – Brittiska öarna – England*. Problemet med variantformer illustrerades också ovan med *Maghreb* och *Afrique du Nord*. Man har observerat (se t.ex. Cislaru 2006, 2008) när det gäller namn på länder att man bör skilja på definitionsdomän och referensdomän. Ett namn på ett land kan avse en geografisk enhet, en politisk eller kulturell enhet och följaktligen är de variantformer som finns inte fritt utbytbara utan mer eller mindre bundna till vissa kontexter. Toponymernas komplexa referens på en skala mellan lokativ och institutionell syftning är en utmaning för lexikografin.

## Val av uppslagsform, funktionell analys

I vissa sammanhang har man på senare tid tillämpat en strikt bokstavsordning varvid nya enheter inne i artiklarna undviks. Man kan fråga sig om det är lösningen för de många komplexa toponymerna. Det är inte självklart för läsaren att förstå enligt vilken princip enheterna är ordnade. Som vi kunnat se ovan varierar det. En funktionell analys av toponymer underlättar jämförelsen mellan olika språk (se Lofström & Schnabel-Le Corre 2011 och Stani-Fertl 2001). Den funktionella analysen bygger på två dikotomier, å ena sidan skillnaden mellan appellativer (A) och proprier (P), egennamn eller avledningar därav, å andra sidan skillnaden mellan generisk (g) och specifik (s) komponent.

Franska	Svenska	Engelska
<i>Cap Nord</i>	<i>Nord kap</i>	<i>North Cape</i>
Ag + As	As + Ag	As + Ag
<i>Nouvelle Guinée</i>	<i>Stor stockholm</i>	<i>New Hampshire</i>
As+ Pg	As + Pg	As + Pg
<i>Pays Basque</i>	<i>Tysk land</i>	<i>Swasi land</i>
Ag + Ps	Ps + Ag	Ps +Ag
<i>Guyane française</i>	<i>Franska Guyana</i>	<i>French Guyana</i>
Pg + Ps	Ps +Pg	Ps +Pg
–	<i>Dals Ed</i>	<i>Land's End</i>
–	As+KASUS+Ag	As+KASUS+Ag
<i>Golfé de Gascogne</i>	<i>Öarna under vinden</i>	<i>Stratford-on-Avon</i>
Ag+PREP+Ps	Ag+PREP+As	Pg+PREP+Ps

I uppställningen framgår tydligt de allmänna tendenserna i de två germanska språken engelska och svenska i förhållande till det romanska språket franska. Den specificerande, determinerande komponenten är framförställd i de germanska språken medan den i franskan är efterställd (med ett lexikaliskt undantag, adjektivet *nouvelle* i *Nouvelle Guinée*). Samtliga språk har samma struktur på konstruktioner med preposition, där definitionsmässigt den determinerande komponenten är efterställd.

Huruvida användandet av den funktionella analysen också kan vara vägledande för den lexikografiska presentationen av endonymer och exonymer kan vara värt en undersökning. En första approximering kunde vara att uppslagsordet, som det verkar förhålla sig i många av de exempel vi sett, ska vara den specificerande komponenten (*Nord (cap)*, *Nord|kap*) utom i de fall där där det finns ett generiskt proprium (*Guyane française*, *French Guyana*). Det är inte omöjligt att ett systematiskt utnyttjande av den funktionella analysen, med nödvändiga detaljanvisningar, kan hjälpa till att precisera och samordna principerna.

## Perspektiv

En intressant observation är att ordböcker oftare tar upp substantiv eller adjektiv avledda av eller sammansatta med ett ortnamn än ortnamnet självt. I betydelseangivelsen anges det aktuella ortnamnet, men givetvis utan formella uppgifter. Man kunde vända på saken och säga att ortnamnets morfologi kan ligga till grund för den typ av avledning eller sammansättning som kan göras. I fall som *Mälaren*, *Vänern* eller *Vättern* borde en ordboksartikel, förutom genus och numerus, kunna ange den inre strukturen sålunda *Mälar|en*, varav slutsatsen kan dras att ortnamnet används i sammansättningar som *Mälar|pirater*.

Eftersom ortnamnen ofta är lexikaliserade böjningsformer av substantiv eller nominalfraser uppstår motsättningar mellan olika kongruensprinciper. När det gäller kongruens skulle man kunna önska sig att det för morfologiskt plurala ortnamn anges om de tar pluralis eller singularis och om detta är beroende på i vilken bemärkelse de används, pluralis för det topografiska objektet, *de vackra Färöarna*, men singularis för den administrativa, politiska enheten, *ett självständigt Färöarna*.

Man observerar att genustillhörighet ibland är osäker men starka tendenser i svenskan gör att de olika typerna av toponymer har ett visst genus. Detta är beroende dels på vilket topografiskt eller institutionellt objekt de avser (tätorter är neutrum t.ex.), dels genus på slutledet (*Östersjön*, *Vänern* utrum, men *Långvattnet*, *Ishavet* neutrum).

I själva verket ligger dessa frågor alltså inte förvånande i skärningspunkten mellan grammatik och lexikon. Utarbetad av principer för ortnamns beskrivning i ordböcker innebär sålunda avvägningar i förhållande till grammatiskt beskrivbara regelbundenheter och hänsynstagande till de speciella problem som särskilt tvåspråkiga ordböcker ställer.

## LITTERATUR

- Allén, Sture, 1971: Nusvensk frekvensordbok baserad på tidningstext. 2 Lemman. Data Linguistica 4. Almqvist & Wiksell International. Stockholm.
- Bra Böckers Lexikon. Bokförlaget Bra Böcker. 1983-1990. Höganäs.
- Cislaru, Georgeta, 2006: Propriétés catégorielles des noms de pays. I : *Onoma* 41. S. 83-113.
- Cislaru, Georgeta, 2008: Le nom de pays comme outil de représentation sociale. I: Boyer, Henri & Paveau, Marie-Anne (red.) *Toponymes, instruments et enjeux, / Mots. Les langages du politique/* 86. S. 53-64.
- Engwall, Gunnel, 1984: Vocabulaire du roman français (1962 – 1968). Dictionnaire des fréquences. Data Linguistica 17. Almqvist & Wiksell International. Stockholm.
- Fransk ordbok. Fransk-svensk. Norstedts. 1991. Stockholm.
- Fransk ordbok. Svensk-fransk. Norstedts. 1991. Stockholm.
- Löfström, Jonas & Betina Schnabel-Le Corre, 2011: Comment analyser et comparer les toponymes de différentes langues dans une perspective synchronique. I : *Nouvelle Revue d'Onomastique (NRO)* 52:291-318.
- Petit Larousse illustré. Larousse. 2002. Paris.
- Petit Robert des noms propres. Dictionnaires Le Robert. 2007, 2009. Paris.
- SAG = Teleman, Ulf, Hellberg, Staffan & Andersson, Erik, Svenska Akademiens grammatik, band 1-4. 1999. Stockholm: Norstedts.
- Stani-Fertl, Roman, 2001: Exonyme und Kartographie. *Weltweites Register deutscher geographischer Namen, klassifiziert nach Gebräuchlichkeit, und ihrer ortsüblichen Entsprechungen.* Wien: Wiener Schriften zur Geographie und Kartographie, Band 14.
- Svensén, Bo, 2009: *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge University Press.
- Utrikes namnbok. Utrikesdepartementet. 1998. Stockholm.

*Jonas Löfström*

TopoLing, Université Rennes 2

jonas.lofstrom@uhb.fr

# Från Svenska ord (Lexin) 3 till Svenska ord 4

*Sven-Göran Malmgren*

In the late 1970s, the Swedish Board of Education started the LEXIN project (LEXIN=Lexikon för invandrare, ‘Dictionaries for immigrants’). The purpose was to produce lots of Swedish-L2 dictionaries, where L2 represents immigrant languages for which no commercial publishing houses could be expected to engage in dictionary-making. These bilingual dictionaries were to be based on a common Swedish dictionary and database, *Svenska ord* (‘Swedish words’). This monolingual dictionary, with slightly less than 20,000 lemmas, was compiled at the Department of Computational Linguistics at the University of Gothenburg, under the direction of Martin Gellerstam. Its first edition was published in 1984.

The LEXIN project was a great success; today, more than 20 immigrant languages are provided with bilingual dictionaries based on *Svenska ord*. The project has inspired similar projects in Norway, Denmark, and Iceland. In the 1990s, an Internet version of all the LEXIN dictionaries was published, as well as extended editions of *Svenska ord*.

In 2008–2010, an updated and extended version of *Svenska ord* (Svenska ord 4) was developed at the Centre for Lexicology and Lexicography at the University of Gothenburg. It was published in 2011, connected to the website of the Swedish Board of Language Planning, now responsible for the LEXIN project.

In the paper, the improvements of *Svenska ord 4* are described, not least in relation to the intense metalexicographic discussion of *lexicographic functions* in recent years. In particular, the excellent search possibilities of Svenska ord 4 are underlined (made possible thanks to work performed at KTH, the Royal Institute of Technology in Stockholm).

*Nyckelord:* Inlärningsordbok, Produktion, Reception, Definitionsvokabulär (begränsad), Transparenta sammansättningar, Språkprov, Idiom, Partikelverb (behandling av)

## Inledning

Det svenska ordboksprojektet *Lexikon för invandrare* (LEXIN, härefter Lexin) startade i slutet av 1970-talet på initiativ av Skolöverstyrelsen (senare Skolver-

ket). Bakgrunden var att Sverige mer och mer hade blivit ett mångkulturellt land, och det behövdes därför tvåspråkiga ordböcker mellan svenska och en rad minoritetsspråk. I många fall var antalet personer som talade dessa språk för litet för att sådana ordböcker skulle bära sig ekonomiskt. Det var därför nödvändigt med särskilda insatser från samhällets sida.

Vetenskapligt ansvarig för Lexinprojektet var Martin Gellerstam vid dåvarande Institutionen för språkvetenskaplig databehandling (Språkdata) vid Göteborgs universitet. Som ett första steg utarbetades en enspråkig svensk ordbok med ungefär 17 000 uppslagsord med lättlästa beskrivningar (inlärningsordboken *Svenska ord*, SO 1, 1984). Denna ordbok skulle utgöra gemensamt underlag för de tvåspråkiga ordböckerna.

Lexinprojektet blev en stor framgång: sedan 1984 har, förutom Svenska ord, Lexinordböcker mellan svenska och mer än 20 minoritetsspråk publicerats. På 1990-talet publicerades två nya, utökade upplagor av Svenska ord, enligt omslagstexten nu med ca 25 000 uppslagsord. (Den tredje upplagan avviker dock bara obetydligt från den andra.) Sedan slutet av 1990-talet är samtliga Lexinordböcker tillgängliga på nätet (<http://lexin2.nada.kth.se>). Lexinprojektet har också exporterats till flera andra nordiska länder. (Se vidare Gellerstam 1999.)

Lexinprojektet sorterar numera under Språkrådet (tidigare Svenska språknämnden). Som ett led i en planerad uppdatering av de tvåspråkiga Lexinordböckerna gav Språkrådet år 2008 Lexikaliska institutet vid Institutionen för svenska språket vid Göteborgs universitet i uppdrag att bearbeta tredje (eller andra) upplagan av Svenska ord. Resultatet av denna bearbetning, här kallat *Svenska ord 4* (SO 4), föreligger nu som nätordbok. Förutom projektledaren, författaren till denna artikel, har Daniel Berg, Sture Berg, Ann-Kristin Hult, Louise Holmer, Christian Sjögreen, Emma Sköldberg och Maria Toporowska Gronostaj deltagit i projektet, och även Anja Allwood, Anna Jansson och Joel Olofsson har lämnat bidrag i anslutning till en lexikografikurs med inriktning på inlärningsordböcker. Arbetet med webbversionen har utförts på KTH i Stockholm under ledning av Viggo Kann. Ansvarig på Språkrådet har varit Christian Mattsson.

I fortsättningen ska jag ge en översikt över de förändringar som har gjorts från Svenska ord 2/3 (SO 2/3) till Svenska ord 4 (SO 4), och tankarna bakom dem. Tonvikten ligger på kvalitativa förändringar, jag bortser t.ex. från att ungefär 3 000 nya uppslagsord har tillkommit. Dispositionen är densamma som i en ordinär ordboksartikel, dvs. framställningen börjar med lemmatecknet och avslutas i stort sett med idiomdelen. Men allra först ska jag kort redovisa några allmänna överväganden som gjordes i början av projektet. (För en tidigare, men

numera inte helt aktuell, översikt, riktad till en internationell publik, se Hult et al. 2010.)

Som framhålls t.ex. av Pálfi & Tarp (2009) är Lexinordböckerna, däribland SO 2/3, på många sätt beundransvärda verk. Projektet som helhet torde vara i det närmaste unikt, särskilt om man räknar med dess avknoppningar i bl.a. Norge (se t.ex. Bjørneset 1999 och <http://www.lexin.no>). Ändå var det tydligt att det på vissa punkter fanns en potential för förbättring av SO 2/3, inte minst mot bakgrund av den intensiva metalexikografiska diskussion som förts under de senaste 20 åren (se t.ex. Tarp 2008, en teoretisk översikt med tillämpning på inlärningsordböcker).

För det första borde det, menade vi, vara möjligt att ytterligare skärpa SO:s karaktär av just *inlärningsordbok*. En del av informationen kunde redovisas på ett explicitare sätt, och viss svårbegriplig grammatisk terminologi kunde kanske utelämnas eller översättas till klarspråk. Och man kunde kanske ta ett steg i riktning mot en begränsad definitionsvokabulär, något som rekommenderas av Pálfi & Tarp (2009), eller i varje fall kraftigt reducera antalet ord i definitionerna som inte själva förklarades i ordboken.

För det andra borde man bestämma sig för om SO i första hand ska vara en receptions- eller produktionsordbok. Nu är det uppenbart att den i viss utsträckning är bådadera (jfr nedan), men det finns tillfällen när receptions- och produktionsaspekten faktiskt kan komma i konflikt (jfr Malmgren 2009). I sådana fall, menade vi, borde receptionsaspekten ha företräde. Den enkla tanken var att man måste förstå ett ord innan man kan använda det aktivt.

Givetvis är många Lexinanvändare nöjda med att slå upp svenska ekvivalenter i en lämplig tvåspråkig Lexinordbok. Med tanke på dem kan vissa av de följande resonemangen möjligen synas mindre relevanta, t.ex. diskussionerna kring definitioner och språkexempel. Men även om Svenska ord:s främsta uppgift kanske är att utgöra underlag till de tvåspråkiga ordböckerna, är den självklart också en enspråkig svensk ordbok i sin egen rätt. Och man får aldrig glömma att många minoritetsspråk i Sverige fortfarande saknar tvåspråkiga Lexinordböcker.

## Förändringarna – från lemmatecken till idiom

Vi inleder varje underavsnitt med exempel från ordböckerna, först den gamla versionen från SO 2/3, sedan den nya från SO 4. (Layouten i exemplen kan avvika något från ordböckernas.)

## Lemmatecknet

SO 2/3:

järn|väg  
omorganiserar

SO 4:

järn|väg  
om|organiserar

I SO 2/3 är alla äkta sammansättningar, t.ex. *järnväg*, försedda med lodstreck som ger användaren värdefulla upplysningar om deras struktur. Däremot saknas lodstreck vid sammansättningar med adverb eller preposition som förled (*omorganiserar*). Vi menar att lodstrecket ger lika värdefull strukturinformation i dessa fall. Det är annars inte alls säkert att en användare med svaga kunskaper i svenska ser att *omorganiserar* är nära besläktat med *organiserar*.

I förbigående sagt framgår det av exemplet **omorganiserar** att SO 4 liksom SO 2/3 har verbets presensform, inte dess infinitivform, som uppslagsform. Mer om detta nedan.

## Uttalsangivelser

Uttalsangivelserna var gjorda på ett utmärkt sätt i tidigare upplagor av SO. Exempelvis var SO 1 på 1980-talet ensam bland svenska definitionsordböcker om att ange musikalisk accent eller tonaccent. Det fanns dock en liten svaghet i uttalsangivelserna i de tidigare upplagorna av SO: det gavs ingen information vid böjningsformer med avvikande uttal (jfr Malmgren 1999: 83 f.). I SO 4 är alla böjningsformer med oförutsägbart uttal, t.ex. *doktorer* med accentförskjutning jämfört med *doktor*, försedda med uttalsmarkering. Dessutom ges i handledningen till SO 4 några enkla regler om förändring av tonaccenten i böjningsformerna, så att användaren med utgångspunkt i grundformens uttal i princip kan komma fram till rätt uttal av alla böjningsformer.

## Böjningsangivelser

SO 2/3:

kastar *kastade kastat kasta(!)*  
regnar *regnade regnat regna*  
film *filmen filmer*



SO 4:

**kastar** att kasta, kastade, har kastat, är kastad, kasta!**regnar** att regna, regnade, har regnat, –, –**film** filmen, filmer, filmerna

SO 2/3 hade som synes ett ytterst ekonomiskt sätt att ange verbböjning. Efter uppslagsformen i presens gavs preteritumformen, supinumformen och till sist, där så var möjligt, imperativ- och infinitivformerna tillsammans (tack vare utropstecknet inom parentes efter *kasta*). Vi tror dock att det kan vara fördelaktigt med en explicitare notation, som därtill är försedd med småord som indikerar vilken böjningsform det är fråga om (t.ex. att före *kasta*). Vi har också lagt till perfektparticipformen och anger alltid explicit när den formen och/eller imperativformen saknas (som vid verbet **regnar**).

Vid substantiven anger SO 4 alltid, i motsats till SO 2/3, bestämd form pluralis.

## Definitioner

SO 2/3:

**kastar** .. få något att fara genom luften**färgglad** .. som har många glada färger

SO 4:

**kastar** .. får något att fara genom luften med en svängande rörelse med armen**färgglad** .. som har starkt lysande färger (1)

Exemplet **kastar** påminner oss åter om att verbens uppslagsform är presensformen. Men det första verbet i definitionen SO 2/3 står i infinitiv, vilket innebär en viss (logisk) inkongruens. Kanske är det därför bättre att välja presensform även i definitionen, så som det är gjort i SO 4. Inom parentes kan tilläggas, att definitionen i SO 2/3 passar lika bra till verbet **skjuter** och därför har kompletterats något i SO 4.

Exemplet **färgglad** illustrerar ett annat problem. Ordet *färg(er)* i definitionen är starkt polysemt och en inlärare kan ha svårt att räkna ut vilken av Svenska ord:s tre betydelser det är fråga om. I Svenska ord 4 anges detta explicit i det här fallet och många liknande fall (dock inte alltid).

Vi övergår till en fråga av annan typ. Betrakta följande exempel:

SO 2/3:

**sektor ..**

- del av en cirkel som begränsas av två radier och en cirkelbåge allmänt avgränsat område

*den offentliga sektorn*

SO 4:

**sektor ..**

- del av en cirkel som begränsas av två radier och en cirkelbåge
- även bildligt del (t.ex. av ett samhälle)

*den offentliga sektorn*

Även här är beskrivningen i SO 2/3 mycket kompakt. Problemet är att användaren inte kan veta till vilken av de båda betydelseerna exemplet *den offentliga sektorn* ska föras.

I SO 4 har vi infört underavdelningar av betydelsemoment och har som krav att det alltid ska framgå vilken betydelse ett exempel illustrerar. I artikeln **sektor** är det helt tydligt att frasen *den offentliga sektorn* illustrerar den bildliga betydelsen.

Vi återkommer till frågan om begränsad definitionsvokabulär längre fram.

## Morfologiska språkprov (sammansättningar)

SO 2/3 (**gata**):

gatu|belysning  
gatu|kök  
gat|hörn  
affärs|gata

SO 4 (**gata**):

gatu|belysning  
gat|hörn  
affärs|gata

Språkprov i form av sammansättningar spelar en viktig roll i Svenska ord. De visar hur grundordet förändras som förled i sammansättningar (*gata* > *gat-*, *gatu-*), och de illustrerar betydelsebeskrivningen. Enligt vår mening bör de dock vara helt transparenta, annars måste de förklaras. I SO 2/3 är *gatukök* ('korv-

stånd') ett exempel på en icke-transparent sammansättning som blir eget uppslagsord i SO 4.

Observera lodstrecken som ger användaren god service. Strängt taget är de inte nödvändiga vid efterledssammansättningar (*affärsgata*), men i explicithe- tens namn sätts de ut även här i SO 4.

## Syntaktiska språkprov

SO 2/3:

**orientering** ... *ge en orientering om läget*  
**naturligtvis** ... *du har naturligtvis rätt*  
**slutsats** ... *dra förhastade slutsatser; komma till en viss slutsats*

SO 4:

**orientering** ... *chefen fick en orientering om läget*  
**naturligtvis** ... *naturligtvis var världsmästaren helt överlägsen*  
**slutsats** ... *de hade 50.000 kronor och bilen kostade 100.000 – slutsatsen var att de måste låna 50.000*

Tillsammans med sammansättningsexemplen spelar de syntaktiska språkproven i en inlärningsordbok en utomordentligt viktig roll för att stödja och konkretisera definitionerna. Många gånger kan ett bra exempel bidra mer till förståelsen av ett ord än en kanske abstrakt definition. I SO 4 finns i genomsnitt nästan ett syntaktiskt exempel per uppslagsord, betydligt fler än i SO 2/3.

Vi tror sålunda att det är fördelaktigt om de syntaktiska exemplen är så konkretiserande som möjligt. Därför bör de i allmänhet utgöras av hela satser, inte infinitivfraser (jfr **orientering**). De bör vidare i möjligaste mån bidra till betydelsebeskrivningen; användaren ska t.ex. förstå varför just ordet *naturligtvis* är lämpligt i det givna exemplet (vilket inte är fallet i SO 2/3). Exemplen under **slutsats** illustrerar slutligen en konflikt mellan receptions- och produktionsaspekten. SO 2/3:s exempel är utmärkta i en produktionsordbok, men de förutsätter att användaren har förstått ordet *slutsats*. I det här fallet måste ansvaret för betydelsebeskrivningen huvudsakligen vila på exemplen. Exemplet i SO 4 ersätter i praktiken den definition som är omöjlig att utforma på ett lättbegripligt sätt.

**Konstruktionsuppgifter (valenser)****SO 2/3 (älskar, glad):**

A älskar B/x/att+INF  
glad (över x)

**SO 4 (älskar, glad):**

ngn älskar ngn/ngt/att+VERB  
ngn är glad (för/över ngt/SATS)

Konstruktions- eller valensuppgifterna är det mest påtagligt produktionsinriktade inslaget i Svenska ord. De var med redan i första upplagan och var resultatet av en verklig pionjärsats av Martin Gellerstam (jfr Malmgren & Toporowska Gronostaj 2009: 183). Valensuppgifterna har uppskattats i varje fall av lite mer avancerade användare, men notationssystemet har ibland uppfattats som en smula tekniskt (i SO 2/3-exemplen ovan står A och B för mänskliga referenter, x (och y) för sakliga eller abstrakta referenter, och parenteser betecknar fakultativ bestämning). I samråd med Språkrådet beslöt vi att införa de kanske mindre tekniska ”variablerna” *ngn* och *ngt*, och dessutom ersatte vi INF (initiv) med det mer välkända VERB. Slutligen byggde vi ut uppgifterna vid adjektiven i analogi med verbuppgifterna. Kanske bör vi slopa förkortningarna *ngn* och *ngt* och skriva *någon* resp. *något* i klartext.

**Idiom****SO 2/3:**

**kastar** ... *kasta yxan i sjön* ge upp  
**yxan** ... *kasta yxan i sjön* ge upp

**SO 4:**

**kastar** ... *kastar yxan i sjön* ger upp  
**sjön** ... *kastar yxan i sjön* ger upp  
**yxan** ... *kastar yxan i sjön* ger upp

Platsen för redovisning av idiom har diskuterats livligt de senaste åren (t.ex. Sköldberg 2006). Vanligt i pappersordböcker är att ta upp idiommet på det lemma som motsvarar det första substantivet, så som det är gjort i exemplet både i SO 2/3 och SO 4. Så görs också t.ex. i DDO och Svensk ordbok utgiven

av Svenska Akademien, och ofta ges hänvisningar från de andra ”tunga” orden i idiomet (t.ex. från **sjö** till **yxa** i exemplet ovan).

I nätdordböcker finns det dock knappast någon anledning att inte ta upp idiomet på alla de ”tunga” orden, i exemplet från SO 4 sålunda på alla tre lemanana **yxa**, **kastar** och **sjö**. Det är ett ytterst enkelt sätt att bespara användarna många fruktlösa uppslagningar. Givetvis bör definitionerna vara ordagrant lika på alla de ställen där ett och samma idiom tas upp (jfr Sköldberg 2006); i SO 4 är definitionerna bokstavligen kopierade från det första stället till alla övriga.

Verbala idiom som *kasta(r) yxan i sjön* får i SO 4 (men inte i SO 2/3) uppslagsformer och förklaringar ”i presens”. Från Lexinordböckernas utgångspunkter är detta konsekvent om man ser idiom som självständiga lexikala enheter, nästan på nivå med lemmen.

## Ytterligare ett problem: partikelverb

De svenska partikelverben är en komplicerad historia. Många av dem uppträder i två skepnader, i s.k. fast sammansättning (*utkommer*) och s.k. lös förbindelse eller lös sammansättning (*kommer ut*), med samma betydelse men ofta olika stilnivå. Enligt vår mening måste sambandet mellan verb som *utkommer* och *kommer ut* framgå klart i en inlärningsordbok. Så är dock inte alltid fallet i SO 2/3, och när sambandet anges sker det gärna med hjälp av svårtillgänglig grammatisk terminologi. I SO 4 har vi försökt få fram sambandet på ett mer explicit sätt:

SO 2/3:

**utkommer** .. publiceras  
**kommer ut** .. bli publicerad

**hyr ut** -> **uthyr**  
**uthyr** .. mest i lös sms ..

SO 4:

**ut|kommer** .. blir publicerad, kommer ut (1) ..  
**kommer ut** .. 1) blir publicerad, utkommer .. 2) berättar öppet om sin (homo)sexuella läggning

**hyr ut** .. låter disponera mot betalning

I SO 2/3 finns ingen länk mellan **utkommer** och **kommer ut**, men väl mellan **hyr ut** och **uthyr**. Dels i form av en hänvisning (från **hyr ut**), dels genom en grammatisk angivelse under **uthyr**. Frågan är dock om användaren kan tolka

den grammatiska terminologin rätt och med ledning av den bilda den mycket vanligare variantformen **hyr ut**.

Idén bakom SO 4:s notationssystem framgår av exemplet *utkommer/kommer ut*, som numera är mer komplicerat än på 1990-talet. Variantformen anges på samma sätt som synonymer, t.ex. *utkommer* direkt efter den egentliga definitionen av *kommer ut*. På så sätt skapas en länk mellan den fasta sammansättningen och den lösa förbindelsen, och dessutom framgår det för vilka betydelser varianterna är utbytbara. Sålunda kan användaren se att det går att säga *boken utkommer/kommer ut i höst*, men bara *han har kommit ut* (inte *\*utkommit*) *som homosexuell*.

De löst sammansatta partikelverben orsakar inlärare stora svårigheter inte minst genom att partikeln ofta kan hamna långt ifrån verbet: *boken kommer kanske inte ut i höst*. I en något mer utvecklad version av SO 4 bör man kunna klicka sig direkt från ett ord i en digital text till motsvarande lemma i ordboken. Men inte bara det; så snart det finns ett diskontinuerligt partikelverb (t.ex. *kommer ... ut*) i en mening i texten bör sökprogrammet kunna hitta det och, om så önskas, skicka användaren till motsvarande ordboksartikel.

## Mer om definitioner: om begränsade definitionsvokabulärer och om navigerbarhet

I SO 2/3 finns ett stort antal ord – flera tusen – i definitionerna som inte själva är definierade i ordboken. I de flesta fall rör det sig om transparenta sammansättningar och avledningar av definierade ord, men även sådana ord kan vara svåra att tolka för många användare. Idealet bör vara att alla definitionsord själva hittas som uppslagsord, allra helst genom en klickning (något som i stort sett genomförts i SO 4). Ännu bättre är det om antalet olika ord som används i definitionerna kan hållas relativt lågt, och att de huvudsakligen återfinns i basordförrådet. En sådan begränsad definitionsvokabulär har särskilt använts i engelska inlärningsordböcker, bl.a. den utmärkta *Macmillan English Dictionary* (2002). Vi återkommer till detta mot slutet.

I SO 4 har vi inte helt kunnat förverkliga ambitionen att alla definitionsord själva ska vara definierade, än mindre idén med en begränsad definitionsvokabulär. Vissa sammansättningar i definitionerna saknar fortfarande formellt definitioner, men kravet är då att de åtminstone återfinns som morfologiska exempel. Exempelvis definieras *mat* 'näringssämnen som man äter, föda', varvid *näringsämne* inte är uppslagsord men står som språkprov under **näring**. Eftersom vi har att göra med en transparent sammansättning får detta anses ersätta en regelrätt definition. Användaren kan hitta från definitionen av **mat** till arti-

keln **näring** genom att skriva ”näringsämnen” eller genom att klicka på detta ord. I de allra flesta fall kan man klicka på vilket ord som helst i definitionerna, oavsett böjningsform, och komma till motsvarande uppslagsord. Dessa navigeringsmöjligheter innebär stora förbättringar jämfört med nätversionen av SO 2/3.

Låt oss till sist se lite närmare på Macmillan och något jämföra förutsättningarna för begränsade definitionsvokabulärer i engelska och svenska ordböcker. I inledningen till Macmillan läser man: ”All the definitions in this dictionary are written using a carefully chosen list of under 2500 words.” Det låter bra, men tyvärr är det inte riktigt sant. Ordboken använder många andra definitionsord än de föregivna 2 500 orden, och dessa ord är satta i kapital. En liten stickprovsundersökning ger vid handen att antalet olika ”kapitalord” kan ligga runt 4 000. Med andra ord består definitionsvokabulären i Macmillan inte av 2 500 ord utan troligen av minst 6 000 ord.

Men i ärlighetens namn är även detta antal klart lägre än antalet olika definitionsord i SO 4. Nu har vi inte gjort något allvarligt försök att radikalt begränsa detta antal, men även om vi hade gjort det hade det varit svårt att komma ner till Macmillans siffor. Orsaken är närmast självklar: tendensen att bilda sammansättningar i svenskan där engelskan föredrar nominalfraser. Engelskans *university course* och *atom of oxygene* motsvaras av svenskans *universitetskurs* och *syreatom*. Eftersom *university*, *course*, *atom* och *oxygene* redan fanns i Macmillans definitionsvokabulär, innebär *university course* och *atom of oxygene* inget ytterligare tillskott. Men det skulle uppenbarligen de svenska sammansättningarna *universitetskurs* och *syreatom* göra. Det enda sättet att hålla jämna steg med Macmillan är då att göra omskrivningar till sammansättningarna: *kurs på universitetet* resp. *atom av syre*. Men då blir i stället definitionsspråket onaturligare.

Slutsatsen är att det – i varje fall för svenskans del – knappast är meningsfullt att försöka slå rekord i grenen ”definitions-vokabulärens litenhet”. Det viktiga är att det går att hitta förklaringar till alla definitionsord, antingen direkt, som i de allra flesta fall i SO 4, eller indirekt, som i fallet *näringsämne* i samma ordbok. Därtill kommer givetvis kravet att definitionsorden om möjligt ska vara enklare än uppslagsordet.

## LITTERATUR

Bjørneset, Tove, 1999: Ordboksprojektet NORDLEXIN-N. I: *LexicoNordica* 6. S. 35–45.

DDO = Den danske ordbog, 2003–2005, Köpenhamn.

- Gellerstam, Martin, 1999: LEXIN – lexikon för invandrare. I: *LexicoNordica* 6. S. 3–17.
- Hult, Ann-Kristin, Sven-Gran Malmgren & Emma Sköldberg, 2010. *Lexin – a report from a recycling lexicographic project in the North*. I: Dykstra, Anne & Tanneke Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress, Leeuwarden 6–10 July 2010*. Ljouwert. S. 800–809.
- Malmgren, Sven-Göran, 1999: De svenska LEXIN-ordböckerna ur ett (ställföreträdande) användarperspektiv. I: *LexicoNordica* 6. S. 79–90.
- Malmgren, Sven-Göran, 2009: On production-oriented information in Swedish monolingual defining dictionaries. I: Nielsen, Sandro & Sven Tarp (red.), *Lexicography in the 21st Century*. In honour of Henning Bergenholtz. Amsterdam/Philadelphia. S. 93–102.
- Malmgren, Sven-Göran & Maria Toporowska Gronostaj, 2009: Valensbeskrivningar i svenska ordböcker – och några andra. I: *LexicoNordica* 16. S. 181–196.
- Macmillan = Macmillan English Dictionary for Advanced Learners, 2002, London.
- Pálfi, Lorand-Levente & Sven Tarp, 2009: Lernerlexikographie in Skandinavien – Entwicklung, Kritik und Vorschläge. I: *Lexicographica* 25/2009. S. 135–154.
- Sköldberg, Emma, 2006: *Idiom och semantisk variation – i autentiskt språkbruk och i ordböcker*. I: Lorentzen, Henrik & Lars Trap-Jensen (red.), *Nordiske studier i Leksikografi* 8. Rapport fra konference om leksikografi i Norden, Sønderborg 24.–28. Maj 2005. Köpenhamn. S. 359–371.
- SO 1 = Svenska ord, 1. uppl., 1984. Stockholm.
- SO 2 = Svenska ord, 2. uppl., 1992. Stockholm.
- SO 3 = Svenska ord, 3. uppl., 1995. Stockholm.
- SO 4 = Svenska ord, 4. uppl., 2011, <http://lexin2.nada.kth.se/lexin/>.
- Svensk ordbok utgiven av Svenska Akademien, 2009, Stockholm.
- Tarp, Sven, 2008: *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen. (*Lexicographica Series Maior*. 134.)

*Sven-Göran Malmgren*

Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet  
malmgren@svenska.gu.se



# Farväl till en stor grupp adverb

## En breddad prepositionskategori i ordboksartiklar?

*Nina Martola*

In the comprehensive Norwegian grammar, *Norsk referansegrammatikk* (NRG), a new categorization of prepositions and adverbs is introduced. Several types of words, which we have been used to regarding as adverbs, are classified in the NRG as prepositions, more precisely as intransitive prepositions. In this article I discuss the implications such a recategorization would have for Swedish lexicography, both monolingual and bilingual. My conclusion is that it does not necessarily lead to increased transparency of dictionary items for the categories, and that it is wiser to wait with reforms of this kind until a new categorization is well established within general grammar.

*Nyckelord:* ordklass, preposition, adverb, kategorisering, enspråkiga ordböcker, tvåspråkiga ordböcker, adresseringsstruktur.

### 1. Inledning

Prepositioner är besvärliga att hantera som uppslagsord i ordböcker. Betydelseinnehållet är vagt och därmed vanskligt att definiera i definitionsordböcker. Att ange motsvarigheter i andra språk kan vara bedrägligt eftersom prepositionsvalet i ganska hög grad bestäms av andra faktorer än betydelse (*på bordet* – *on the table* men *titta på X* – *look at X*; *höra på X* – *listen to X*; *priset på X* – *the price of X*). Ännu mer komplicerat kan det bli vid språk som saknar prepositioner (*på bordet* – finska *pöydällä* < *pöytä* + adessivändelsen *-lla*<sup>1</sup>). Adverb har överlag mer betydelseinnehåll och är därmed lite lättare att behandla, men kategorin adverb är i gengäld mycket brokig.

---

<sup>1</sup> Det saknas inte prepositioner helt och hållet i finska men gruppen har en mycket begränsad användning jämfört med i germanska språk. Postpositioner är betydligt vanligare. I den stora finska grammatiken (ISK) används f.ö. främst den övergripande termen *adpositio* 'adposition'.

I Norsk referansegrammatikk (i fortsättningen NRG) görs en annorlunda kategorisering av adverb och preposition än vi varit vana vid inom nordisk lexicografi: en stor grupp adverb hänförs till kategorin prepositioner. I en recension av band VIII av Norsk Ordbok i LexcioNordica 17 föreslås det att ordboken borde följa NRG (Faarlund 2010). Att börja tillämpa nya syntaktiska principer mitt under pågående utgivning är dock knappast att rekommendera. Däremot kan nya principer vara värda att överväga om man börjar på ett nytt ordboksprojekt eller gör en grundligare revidering. Jag ska i det följande försöka reda ut de lexikografiska implikationerna av den föreslagna omkategoriseringen.

Som bakgrund diskuterar jag kortfattat NRG:s kategorisering och ställer den i relation till den indelning som görs i systemverket Svenska Akademiens grammatik (härefter SAG). Jag går också igenom hur kategoriseringen är gjord i två moderna, enspråkiga svenska ordböcker. Efter det undersöker jag ett antal prepositions- och adverbartiklar i olika ordböcker med svenska som källspråk för att ta reda på vilka konsekvenser en omkategorisering skulle få.

Det går att komma med syntaktiska motargument mot omfördelningen mellan adverb och prepositioner men här ägnar jag mig bara åt de lexikografiska för- och nackdelarna av en eventuell grammatisk reform.

## 2. De syntaktiska kategorierna preposition och adverb

NRG och SAG skiljer sig i fråga om definitionerna av preposition och adverb, och jag ska härnäst i korthet gå igenom och jämföra de två grammatikverkens synsätt. Därefter redovisar jag hur prepositions- och adverbartiklar ser ut i svenska ordböcker.

### 2.1 Kategoriseringen enligt Norsk referansegrammatikk

En *preposition* är enligt NRG (s. 23, 411–412) oböjliga ord som utgör kärna i en bestämning efter verb, adjektiv och substantiv. Prepositionerna styr vanligen en substantivfras men kan också uppträda utan efterföljande bestämning (*Hun klatret opp stigen* resp. *Hun klatret opp*). Prepositionerna kan med andra ord precis som verb uppträda transitivt eller intransitivt. Rent intransitiva prepositioner, som aldrig kan ta fyllnadsled är t.ex. *bort*, *fram*, *heim*, *vekk*; *borte*, *framme*, *heime*, *vekke* (NRG s. 414).

I den svenska återgivningen av prepositionsdefinitionen ovan har jag använt det övergripande *bestämning*, men i själva verket är definitionen i NRG lite motsägelsefull. På ett ställe heter det ”kjerne i en utfylling etter verb, adjektiver og substantiver” (NRG s. 23) och på ett annat står det ”de danner kjerne i fraser

som kan fungere som adledd till verb, adjektiv, substantiv eller pronomen” (NRG s. 411).

Enligt NRG är *adledd* en övergripande term och avser de ord i en (grammatisk) fras som inte är kärna. I nominalfrasen *de yngsta barnen i familjen* t.ex. är substantivet *barnen* kärnan och *de yngsta* och *i familjen* är bestämmingar (adledd). Bestämningar av typen *utfylling* definieras som ”(etterstilte) adledd som er forutsatt eller forventet ut fra innholdet i kjernen”. (NRG s. 31, 45.) Exempel på ”utfylling” till adjektiv är *bjørnen* respektive *av deg* i meningarna *Ho var redd bjørnen* och *Jeg er stolt av deg*. I verbfraser är generellt sett alla valensbundna led ”utfyllinger”. (NRG s. 403–404, 661–663.) Eftersom prepositionsfraser kan vara bestämmingar till verb<sup>2</sup> utan att vara ”utfyllinger” (*vi åt middag ute, barna leiker i hagen*) bör väl *adledd* vara det adekvata ordvalet i definitionen.

Adverb definieras i NRG (s. 24) som en grupp oböjliga ord vars funktion det är att stå som bestämning utan att vara ”utfylling” eller att stå som modifiering före ett satsled, en fras eller ett ord. Adverben placeras mycket ofta i satsens mittfält (*Du ska bare gjøre som jag sier*), men de kan också stå först i substantiv- och adjektivfraser (*bare en liten bit, nesten ferdig*). Adverben *nok* och *også* står sist i frasen. Adverben är alltså en mycket smalare kategori än i traditionell grammatik, och en mängd ord som vi vant oss vid att betrakta som adverb räknas som prepositioner i NRG.

## 2.2 Kategoriseringen enligt Svenska Akademiens grammatik

En preposition är enligt SAG (2: 721) huvudord i en prepositionsfras och bestäms av ett nominalt led, rektionen. Men många prepositioner kan stå utan rektion i så kallad absolut funktion och utgör då bundet eller fritt adverbial, partikeladverbial, efterställt attribut eller predikativ. I rektionslös användning är avgränsningen av preposition mot adverb inte klar, eftersom prepositionen ensam fungerar som adverbial (*käppen gick av; han ville inte sätta sig i soffan utan ställde sig bredvid; jag behöver en tidning att läsa i*). I SAG klassificeras *av, bredvid, i* i de här fallen som prepositioner. Också SAG är inne på tanken om intransitiv användning av prepositioner. (SAG 2: 726.) Men man är alltså inte lika radikal som NRG utan räknar ord av typen *bort, hem, fram, in* osv. till adverben. Det sägs explicit att verbpartiklar (i SAG benämnda *partikeladverbial*) till sin ordklassstillhörighet är antingen adverb (*lägga in, fara hem*) eller preposition (*gå av, hitta på*). (SAG 3: 417–420.)

<sup>2</sup> Man kan förstås diskutera om fria adverbial bestämmer verbet, verbfrasen eller rentav hela satsen.

I vissa fall är det befogat att räkna med homonymi mellan preposition och adverb, och det avgörande är att betydelsen skiljer sig. I *sedan kriget* uttrycker prepositionen varaktighet medan adverbet i *sedan tände han en cigarett* inte gör det. (SAG 2: 681–682.)

### 2.3 Kategoriseringen i moderna enspråkiga, svenska ordböcker

Innan jag börjar diskutera utfallet av en eventuell grammatisk reform i ordböcker ska jag ge en översikt över hur prepositioner och adverb är behandlade för närvarande i ordboksartiklar. Jag har valt de två enspråkiga ordböckerna *Svenska Akademiens ordlista* och *Svensk ordbok utgiven av Svenska Akademien* (SO). Tvåspråkiga ordböcker med svenska som källspråk kan i hög grad antas redigeras med dessa två tungt vägande ordböcker som förlaga. Vid analysen har jag utnyttjat SAOL i elektronisk utgåva (*SAOL Plus*, i fortsättningen SAOL+) som innehållsligt är identisk med SAOL 13.

En sökning på strängen *prep* i artikeltext ger 172 träffar i SAOL+. Av dessa 172 uppslagsord förekommer 56 (en dryg tredjedel) enbart som preposition. För 93 lemmarna är ordklassangivelsen *preposition och adverb* och i 23 fall finns det en artikel för ett homonymt adverb. I SO görs homonymindelning alltid utifrån ordklass tillhörighet och av de undersökta prepositionsartiklarna har över hälften en adverbial homonym (tabell 1).<sup>3</sup> Det faktum att över hälften av prepositionerna i ordböckerna har en adverbial homonym pekar mot ett visst kategoriseringsproblem och kunde tala för en klassificering enligt NRG.

	SAOL+		SO	
	N	%	N	%
Enbart prep.	56	32,6	66	46,8
Prep. och adv.	93	54,1		
Adv.homonym	23	13,4	75	53,2
Σ	172		141	

Tabell 1: Prepositions-kategorin i SAOL+ och SO.

För de 145 lemmarna som är gemensamma för de båda ordböckerna stämmer kategoriseringen överens till drygt åttio procent, dvs. orden anges vara antingen enbart preposition eller både preposition och adverb i både SAOL+ och SO.

<sup>3</sup> SO kan innehålla fler än 141 prepositionsartiklar. Eftersom SO inte är elektroniskt sökbar har jag sett bara på de prepositioner som SO har gemensamt med och SAOL+.

Det betyder att den återstående femtedelen (27 stycken) har fått olika klassificering. Största delen (23) är ord som angetts vara enbart preposition i SO men både preposition och adverb i SAOL+. (Se närmare tabell 2.)

Kategorisering	N	%	%	Exempel
Bara prep. i båda ordböckerna	43	29,7	81,4	alltsedan, bland, hos, inemot, längsefter
Prep. och adv. i båda ordböckerna	75	51,7		av, bortifrån, bredvid, intill, ovanpå, över
Bara prep. i SAOL+, annat i SO	3	2,1	18,6	c/o, likmätigt, uppemot
Bara prep. i SO, homonymer i SAOL+	23	15,9		exklusive, framom, från, invid, mittför
Bara adv. i SO, homonymer i SAOL+	1	0,7		an
Σ	145	100,0	100,0	

Tabell 2: Kategoriseringsöverensstämmelse i SAOL+ och SO.

Antalet adverb i SAOL+ är 1387<sup>4</sup> och av dessa är bara en bråkdel (116 eller drygt 8 %) homonyma med prepositioner. Adverben är av flera olika typer. Nästan en femtedel utgörs av sammansättningar på *-ligen* och *-vis*. Som adverb klassificeras också en viss typ av kraftuttryck (*himla, jäkla, nedrans, sjuhelsikes*) och andra jämförbara uttryck (*gudskelov*). Vidare räknas en stor mängd uttryck av utländskt ursprung till adverben (*accelerando, ex officio, fifty-fifty, loco, scherzando*) och likaså ord som ingår i mer eller mindre frasartade uttryck (*bom – skjuta bom, fatt – hinna fatt ngn, barbacka, grensle, huvudstupa, pyton – må pyton, sönder*). Likaså anförs ett antal prepositionsfraser och prepositionssammansättningar som adverbiala uppslagsord *i ffol, i fred, i kväll, medsols, medurs, mothårs, till fullo*.

Två stora grupper adverb som i den nya indelningen skulle föras över till prepositionskategorin är sammansättningar på *här-, där-, hit-, dit-, var-, vart-* (154 stycken) och sammansättningar med väderstrecksangivelse som förled (54). Därtill kommer ett betydande antal sammansättningar med preposition (*-av, -efter, -(i)från* osv.) utöver de som ingår i de två föregående grupperna.

<sup>4</sup> En sökning på strängen *adv* i artikeltext ger bara 288 träffar. Antalet 1387 har jag kommit fram till genom att söka strängarna *a\**, *b\**, *c\** osv. i uppslagsord med avgränsningen *Sök i adverb* och summera. Bokstavvis sökning med ordklassfilter går inte att göra för prepositionerna eftersom de ingår i *Övriga ordklasser* i SAOL+.

### 3. Ordboksartiklar med den nya prepositions-kategorin

Den kategorisering av prepositioner och adverb som föreslås i NRG skulle naturligtvis få följderna för artiklarna i de ordböcker som innehåller grammatisk etikettering. I all synnerhet de uppslagsord som i nuvarande ordböcker är homonymindelade utifrån en klassificering i adverb och preposition skulle kräva radikalare ingrepp för att bli överskådliga.

Jag ska i det följande redovisa vilka konsekvenser en nykategorisering skulle få för de enspråkiga ordböckerna SAOL och SO och för några tvåspråkiga ordböcker med svenska som källspråk.

#### 3.1 Nykategorisering i SAOL

För SAOL behöver följderna av en nykategorisering inte bli särskilt radikala. Antalet prepositioner i SAOL skulle stiga från 172 till flera hundra och den mycket vanliga etiketteringen *preposition och adverb* kunde ersättas med enbart *preposition*. Eftersom SAOL:s artiklar är mycket kortfattade och inte tar upp konstruktioner är en indelning i transitiva och intransitiva prepositioner knappast ens nödvändig (se illustration 1). Men det är naturligtvis också möjligt att ange vilka prepositioner som är enbart transitiva och vilka som är rent intransitiva.

Det stora problemet med en omkategorisering för SAOL:s del skulle snarast vara att skapa medvetenhet om den hos användare som lärt sig definiera prepositioner på ett snävare sätt. En del användare har antagligen ganska vaga begrepp om grammatiska kategorier och för dem spelar klassificeringen ingen större roll.

SAOL nu	SAOL efter omkategorisering
<b>bland</b> preposition <b>innanför</b> preposition och adverb <sup>1</sup> <b>intill</b> [-til'] preposition • <i>intill väggen</i> <sup>2</sup> <b>intill</b> [-til'] adverb • <i>se bilden här intill</i> <sup>1</sup> <b>hit</b> adverb • <i>hit bort, fram, in, ned, upp</i> äv. vard. <i>opp, ut, över</i> hopskr. äv. <b>upp</b> adverb • <i>det är upp till dig</i> <vard.> kommer an på; <i>vara ngn upp i dagen; vända ngt upp och ned</i> el. <i>ner</i> ; jfr <i>uppochmedvänd</i>	<b>bland</b> preposition <b>innanför</b> preposition <b>intill</b> [-til'] preposition • <i>intill väggen; se bilden här intill</i> <sup>1</sup> <b>hit</b> preposition • <i>hit bort, fram, in, ned, upp</i> äv. vard. <i>opp, ut, över</i> hopskr. äv. <b>upp</b> preposition • <i>det är upp till dig</i> <vard.> kommer an på; <i>vara ngn upp i dagen; vända ngt upp och ned</i> el. <i>ner</i> ; jfr <i>uppochmedvänd</i>

Illustration 1. Artiklar i SAOL före och efter omklassificering av prepositionerna.

### 3.2 Nykategorisering i SO

När det gäller SO, som är en definitionsordbok och som också ger konstruktionsuppgifter, är en indelning i transitiv och intransitiv preposition mer påkallad. För prepositioner som *bland* och *hos*, som inte kan stå utan rektion, skulle ordklassangivelsen vara *transitiv preposition*, förkortad på lämpligt sätt. På motsvarande sätt skulle ord som *hit* och *upp* i stället för *adverb* få markeringen *intransitiv preposition*. I övrigt behöver artiklarna inte nödvändigtvis revideras på grund av den nya klassificeringen.

I artiklar med sammansatta prepositioner/adverb är integreringen också förhållandevis enkel att genomföra. För t.ex. *intill* och *bortöver* är betydelseangivelsen densamma, så det är bara att förena artiklarna och ge exemplen efter varandra. En explicit konstruktionsuppgift kunde eventuellt komplettera exemplet: *KONSTR.: ~ (ngt)*. En komplikation innebär dock angivelsen om första belägg, som är olika för preposition och adverb i bägge fallen.

Den stora utmaningen för en ordbok som SO skulle vara att skapa överskådliga artiklar för de enkla prepositionerna som i ordboken nu har homonyma adverbartiklar. De här artiklarna är överlag omfattande i SO. För tio kontrollerade homonymer (*av, efter, för, i, med, om, på, till, åt, över*) är det genomsnittliga antalet huvudbetydelser för prepositioner 6,5 och för adverb 4,7. Underbetydelseerna är i genomsnitt 8 respektive 3.

Jag ska låta artikeln *till* illustrera utmaningarna vid en eventuell revidering. Prepositionsartikeln är indelad i nio huvudbetydelser och sammanlagt tretton underbetydelser. Motsvarande antal för adverbartikeln är sex och fem. I illustration 2 återfinns de två artiklarna i en kraftigt förkortad uppställning; alla underbetydelser är inte noterade och idiomerna är uteslutna.

<p><b>till</b> prep.</p> <p>1 med (riktning mot och) slutpunkten i el. på el. vid el. hos; <i>han reste från Göteborg - Stockholm</i>;</p> <p>° spec. för att ange målet för överlämnande <i>han skickade boken - henne</i>;</p> <p>° i fråga om tidsförh. och mer abstrakta förh., spec. för att ange gräns för ngt: <i>hon stannar ända - jul</i>;</p> <p>° spec. äv. i fråga om uppnående av nytt tillstånd: <i>frågan måste bringas - en lösning</i>;</p> <p>2 (ofta i konstruktion med subst. i genitiv) med läge vid el. på el. i: <i>sitta - bords; sitta - häst</i>;</p> <p>3 i samhörighet med i fråga om konkreta el. abstrakta företeelser som på ett naturligt sätt hör ihop: <i>locket - burken; nyckeln - dörren</i>;</p> <p>4 med syfte eller användning bestående i: <i>han använde bilen bara - (att göra) söndagsutflykter</i>;</p> <p>5 med resultat bestående i: <i>varje krona blev - nytta</i>;</p> <p>6 med avseende på: <i>hon är gladlynt - sitt väsen</i>;</p> <p>7 med ett värde utgörande: <i>sträckan uppmättes - 50 m</i>;</p> <p>8 (såsom ytterligare en faktor) vid sidan av: <i>- detta kommer ...</i></p> <p>9 (i några uttr.) med (plötsligt) användande av: <i>ta - kniven; ta/gripa - vapen</i></p>	<p><b>till</b> adv.</p> <p>1 utöver vad som redan förekommit eller är förhanden: <i>en gång -; 5 kronor -</i>;</p> <p>° spec. för att uttrycka att ngt utgör bidrag e.d.: <i>skjuta - en slant; hjälpa -</i>;</p> <p>2 med plötslig eller kraftig verkan i fråga om inledning av förlopp e.d.: <i>elden flammade -; han hostade -</i>;</p> <p>3 i påvisbar närvaro eller existens: <i>att leva, att vara -; ursäkta att jag finns -</i></p> <p>4 med (riktning mot och) slutpunkt i ngt som vanl. är underförstått: <i>folk började strömma -; sillen går -; båten lade - (vid bryggan)</i></p> <p>5 med visst resultat ofta med förbleknad betydelse, ibl. endast för att understryka fullbordandet av verbhandlingen: <i>skära - fina skivor; laga - en vickning; trassla - det för sig; hur gick det -?</i></p> <p>° spec. för att uttrycka tillslutning e.d.: <i>knäppa - kragen</i></p> <p>6 i läge för att vara synlig vanl. för talaren: <i>han vände den andra sidan -</i></p>
--	--

Illustration 2. Artiklarna för *till* i SO (starkt förkortade).

Den enklaste lösningen på problemet för hur den nya artikeln ska ställas upp är naturligtvis att klassificera *till* som preposition och därefter ange transitiv och intransitiv användning utan att vidare integrera de två delarna, exempelvis: *till* prep. A) trans. 1–9; B) intr. 10–15.

Använder man betydelseangivelser som utgångspunkt, vilket är kutymen i SO, ser vi att samma betydelse återkommer i moment 1 i prepositionsartikeln och moment 4 i adverbartikeln, och likaså överensstämmer betydelsen i moment 5 i båda artiklarna. För moment 9 i prepositionsartikeln och moment 2 i adverbartikeln är den gemensamma nämnaren plötslighet. De här adverbmomenten kunde integreras i motsvarande prepositionsmoment och göras till



underbetydelser. Med en lätt modifiering av betydelseangivelsen i moment 1 kunde också moment 6 ur adverbartikeln passa in där.

I moment 5 i den nya prepositionsartikeln kunde det efter de nuvarande betydelsenyanserna (4 till antalet) följa ytterligare en: ° *förbleknad betydelse, intr.* Då återstår två huvudbetydelser som representerar enbart intransitiva användningar (momenten 1 och 3 ur adverbartikeln). Om dessa läggs in i prepositionsartikeln skulle den alltså sammanlagt bestå av 11 huvudmoment.

Men av detta uppstår frågan om man ska betrakta transitiv användning som normalvärde och ange bara intransitiva användningar explicit eller om man för varje betydelse (och rentav betydelsenyans) borde ange om det handlar om en transitiv eller intransitiv användning av prepositionen (med hjälp av antingen etikett eller konstruktionsuppgift). Utifrån de tio kontrollerade homonymerna verkar den transitiva användningen vara den vanligaste, men ser man till hela det nya prepositionsbeståndet är de rent transitiva prepositionerna i minoritet, och transitivitet skulle därmed bli den markerade egenskapen hos prepositioner (i SAOL: 56 transitiva, 126 transitiva/intransitiva, över 200 nya intransitiva, jfr avsnitt 2.3). En möjlighet är att man inte använder etiketter alls utan låter exemplen i betydelsemomenten tala för sig själva, men det är knappast förenligt med praxis i en mer omfattande ordbok.

En radikalare reform av *till*-artikeln skulle vara att se mer till funktion och inte i så hög grad utgå från betydelsen, eftersom vi har att göra med ett funktionsord. Det är ju först i kombination med andra ord som en prepositions ”betydelse” utkristalliseras. Ordet *till* i sig har inte alla de betydelser som räknas upp i artiklarna i illustration 2. Inte ens kärnbetydelsen ’med riktning mot och slutpunkt’ i är verboberoende, utan betydelsen uppstår i kombination mellan ett rörelseverb och prepositionen. (Jfr t.ex. Norén & Linell 2006.) Också i kombination med en rektion som entydigt kan tolkas som platsangivande får prepositionen konkret riktningsinnebörd (t.ex. *till skogen*), men mer kontext kan ändra på det (*kärlek till skogen*).

Utifrån en korpusbaserad undersökning jag utfört är användningen av *till* i texter verbdominerad. Drygt 70 procent av alla förekomster utgörs av valensförbindelser i vid bemärkelse, medan bara 12 procent är attributiva (*resan till Stockholm*). I knappt 4 procent handlar det om *till* som verbpartikel. Av kombinationerna med verb består 20 procent av konkreta målanvändningar (*åka till Stockholm*) medan hela 35 procent är abstrakta, konstruktions- eller frasliknande förbindelser (*förvandlas till X, döma ngn till X*). Man kunde tala om ett slags metaforiska mål. De tre vanligaste verben i kombination med prepositionen *till* är *gå, komma* och *leda*. (Martola 2009.)

Momentindelningen kunde med andra ord utgå mer från kombinationstyp och beskrivningarna formuleras så att det framgår att det är de kombinationer där *till* ingår som har de angivna betydelseerna, inte *till* ensamt.

Noteras kan att momenten 2–5 i adverbartikeln beskriver användning av *till* som verbpartikel. I konstruktions- och frasordboken *Svenskt språkbruk* är verbpartiklarna anförda i egna kortfattade artiklar: *till* partikel vid verb, t.ex. *bli till*, *finnas till*, *göra sig till* osv. SO kunde eventuellt låna en etikettering av det slaget för ett block inne i prepositionsartiklarna.

### 3.3 Nykategorisering i tvåspråkiga ordböcker

Indelningen av prepositioner och adverb i tvåspråkiga ordböcker följer i mycket hög grad de enspråkiga ordböckernas kategorisering. Jag har undersökt sammanlagt sex ordböcker eller ordboksdelar med svenska som källspråk. Två är svensk-engelska, dels den ordbok som ingår i ne.se (NE/S-E i tabellen), dels den som Norstedts tillhandahåller gratis över webben (Nor/S-E). Därtill har jag sett på en svensk-tysk ordbok (Nor/S-T), en svensk-fransk (Nor/S-Fr), en svensk-finsk ordbok (S-Fi) samt en svensk-estnisk (S-Est). Kategoriseringen framgår av tabell 3.

Lemma	NE/S-E	Nor/S-E	Nor/S-T	Nor/S-Fr	S-Fi	S-Est
<b>bland</b>	prep	prep	prep	prep	prep	prep
<b>hit</b>	adv	adv	adv	adv	adv	adv
<b>innanför</b>	ø	prep adv	prep adv	prep adv	prep adv	prep adv
<b>intill</b>	prep adv	prep adv	prep adv	prep adv	prep adv	prep adv
<b>till</b>	prep adv	prep adv	prep adv	prep adv	prep adv	prep adv
<b>upp</b>	adv	adv	adv	adv	ø	adv

Tabell 3: Kategoriseringen av prepositioner och adverb i tvåspråkiga ordböcker.

De moderna tvåspråkiga ordböckerna (alla relativt stora ordböcker) anger nästan genomgående ordklassstillhörighet. En skillnad gentemot SO är dock att de tvåspråkiga inte hanterar adverb- och prepositionspar som homonymi utan som polysemi. Ordklassindelningen finns inne i artikeln som två huvudblock med egna betydelsebeskrivningar. En bidragande orsak till den stora överensstämmelsen mellan ordböckerna kan förstås vara att tre av sex är från samma förlag.

Större ordböcker ger sannolikt också i högre grad grammatiska upplysningar än mindre – jag har inte granskat mindre ordböcker för denna studie.

Tack vare polysemiutförandet kan etiketterna *preposition* och *adverb* vid en eventuell ordklassreform enkelt bytas ut mot *transitiv preposition* och *intransitiv preposition* utan större ingrepp i artiklarna. Men revideras SO påverkar det sannolikt de tvåspråkiga ordböckerna på sikt.

Emellertid tillkommer problemet med ekvivalenternas kategoritillhörighet i de tvåspråkiga ordböckerna. Det här hänger samman med adresseringsstruktur och räckvidd (jfr Svensén 2004:100–102). Artikeln *hit* ur Nor/S-E får tjäna som exempel (illustration 3).

<p><b>hit</b></p> <p><i>adverb</i></p> <p>allm. <b>here</b>; åt det här hållet <b>this way, in this direction</b>; dit el. så långt <b>thus far</b></p> <p>Exempel:</p> <p>kom hit (hit ner etc.)! <b>come (come down etc.) here!</b></p> <p>[...]</p>
--

Illustration 3. Artikeln *hit* ur Nor/S-E.

Kan man förutsätta att användare överlag inser att angivelsen *adverb* syftar till vänster på *hit* och inte också till höger på de engelska ekvivalenterna? Och är det faktiskt så? I en ordbok där målspråket är det främmande språket borde ju angivelsen snarast gälla ekvivalenterna, som ju är den kategori användarna i första hand behöver information om. Eftersom kategorimotsvarigheten är förhållandevis stor är adresseringsriktningen inte ett så stort bekymmer i ordböcker mellan germanska språk. Men om angivelsen för *hit* i Nor/S-E skulle vara *intransitiv preposition* i stället för *adverb* blir en tolkning av etiketten som högeradresserad direkt vilseledande, så länge man inom engelsk grammatik klassificerar *here* som adverb. (I OED anges exempelvis *here* vara adverb.) Det borde därmed strängt taget finnas etikettering för både lemma och ekvivalent, t.ex.: *hit* (itr.prep; (adv.) *here*).

I den svensk-finska och den svensk-estniska ordboken är adresseringsproblematiken inte lika stor. Eftersom källspråket är det främmande språket för användarna är vänsteradressering naturlig; det är ingångsspråket användarna behöver upplysningar om. Målspråket kan de, och de har därför en klar uppfattning om hur ekvivalenterna konstrueras (också om de inte nödvändigtvis vet vad konstruktionssätten kallas). Finska användare förleds knappast att tro att vare sig adverbena *tänne*, *tähän* 'hit' eller postpositionerna *joukossa*, *keskellä* 'bland' ska konstrueras som prepositioner. Men för de användare som lärt sig

skilja mellan preposition och adverb enligt den traditionella indelningen blir angivelsen *intransitiv preposition* förvirrande.

#### 4. Avslutande resonemang

En omdefiniering av ordklassen preposition skulle få följder inom svensk lexikografi. Det skulle inte leda till några oöverstigliga svårigheter, men å andra sidan är det långt ifrån säkert att det utan vidare skulle leda till klarare och mer lättolkade ordboksartiklar, vilket gärna borde vara målet för lexikografiska reformer. Det nuvarande systemet, att kalla ord som *av, på, till* osv. för adverb i fråga om verbpartiklar är kanske inte optimalt, även om det har en lång tradition inom lexikografien. Det går i vart fall stick i stäv med SAG, den överlägset mest omfattande deskriptiva grammatiken för modern svenska, som explicit säger att prepositioner i absolut användning (alltså rektionslös) är prepositioner och inte adverb.

Men det är tveksamt om den bästa lösningen är att i stället räkna exempelvis *hit* och *hemma* som preposition. Grammatiska teorier kommer och går och lexikografien gör kanske klokt i att vara lite konservativ och inte gå i bräsch för moderna strömningar som eventuellt aldrig får tillräckligt genomslag. De grammatiska etiketternas funktion i svenska ordböcker är ju att vara till hjälp för de användare som har kännedom om svensk grammatik. För övriga användare är de inte till någon nytta. Så länge största delen av de grammatikkunniga användarna inte internaliserat en ny ordklassindelning skapar en reformerad etikettering sannolikt mer förvirring än klarhet. Men om en ny indelning börjar vinna allmänt gehör och anammas i skolornas grammatikundervisning, då blir det dags för en reform också inom lexikografien.

#### LITTERATUR

##### *Ordböcker*

NE/S-E = Nationalencyklopedin. Engelsk ordbok. [www.ne.se](http://www.ne.se) (september 2011)

Nor/S-E = Norstedts engelska ord. Engelska gratis. [www.ord.se](http://www.ord.se) (september 2011)

Nor/S-F = Norstedts stora franska ordbok: fransk-svensk, svensk-fransk. Projektledare: Håkan Nygren; redaktion: Wandrille Micaux, Mathias Thiel (datalogist), Mona Wiman. Stockholm: Norstedts akademiska förlag, 2009.

Nor/S-T = Norstedts stora tyska ordbok: tysk-svensk, svensk-tysk. Huvudredaktör: Irmgard Lindestam. Stockholm: Norstedts akademiska förlag, 2008.

OED = OED. Oxford English Dictionary. [www.oed.com](http://www.oed.com) (september 2011)

SAOL 13 = Svenska Akademiens ordlista över svenska språket. 13. uppl. Stockholm 2006.

- SAOL+ = SAOL Plus. Svenska Akademiens ordlista över svenska språket. Svenska Akademien & Oriibi AB & Språkdata vid Göteborgs universitet, 2007.
- S-Est = Rootsi-eesti sõnaraamat. Svensk-estnisk ordbok. Huvudredaktör: Virve Raag, Tallinn: Valgus, 2004.
- S-Fi = Cantell, Ilse & Anu Hietala, Jenni Hurri, Johanna Månsus & Anja Sarantola. Ruotsi–suomi-suursanakirja. Stora svensk-finska ordboken. Helsingfors: WSOY 2007.
- SO = Svensk ordbok utgiven av Svenska Akademien. Utarbetad vid redaktionen för Svenska Akademiens samtidsordböcker, Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet. Svenska Akademien & Norstedts 2009.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser. Utarbetad av Svenska språknämnden. Stockholm 2003.

*Annan litteratur*

- Faarlund, Jan Terje, 2010: Norsk Ordbok, band VIII. I: LexicoNordica 17. S. 313–320.
- ISK = Hakulinen, Auli & Maria Villkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen och Irja Alho: Ison suomen kielioppi. 'Stora finska grammatiken'. Kotimaisten kielten tutkimuskeskuksen verkkojulkaisuja 5. scripta.kotus.fi/visk (september 2011)
- Martola, Nina, 2009: Prepositionerna *till* och *åt* i ett jämförande perspektiv. I: Lotta Collin & Saara Haapamäki (red): Svenskan i Finland 11. Skrifter från Svenska institutionen vid Åbo Akademi Nr 7, Åbo. S. 115–124.
- Norén, Kerstin & Linell, Per, 2006: Meningspotentialer i den språkliga praktiken. MISS 54. Meddelanden från Institutionen för svenska språket. Göteborgs universitet.
- NRG = Faarlund, Jan Terje & Svein Lie & Kjell Ivar Vannebo, 1997: Norsk referansegrammatikk. Oslo: Universitetsforlaget.
- SAG = Teleman, Ulf & Staffan Hellberg & Erik Andersson, 1999: Svenska Akademiens grammatik. Stockholm: Svenska Akademien & Norstedts ordbok.
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. Stockholm: Norstedts Akademiska Förlag.

*Nina Martola*

Institutet för de inhemska språken  
nina.martola@kotus.fi

# Språkteoriens avtryck i SAOB

*Pär Nilsson*

The first volume of the Swedish Academy Dictionary (SAOB) was published in 1898 and the latest in 2009. A lot has happened during this time within lexicographic theory as well as in general linguistics, and to SAOB editors today the view of language and linguistics is to a large extent unlike the one the editors had in the late nineteenth century. Comparative linguistic research dominated in the nineteenth century, and the Neogrammarian impact on SAOB is obvious. Since the beginning of the SAOB project the form and structure have been essentially fixed in terms of the layout of articles, the definition language etc. In this study I investigate whether it is possible to trace a change of overarching linguistic theories in the dictionary text. The study is based on a comparison of current and older manuals and set of rules and instructions that are available to the editors.

## 1. Inledning

Frågan som jag ställer mig i föreliggande studie är om förändringar som har att göra med övergripande språkvetenskapliga teorier och synsätt kan spåras i Svenska Akademiens ordbok, SAOB.

SAOB är tydligt präglad av junggrammatiska tankar och ideal. Riktlinjerna drogs upp i slutet av 1800-talet med de samtida tyska och engelska motsvarigheterna som förebild. Ordbokens historiska perspektiv är väl förankrat i tidens anda: nyckeln till förståelsen för dagens språk var studier av gårdagens. Vidare står betydelseutvecklingen i centrum. Man hade nu börjat se språklig utveckling som något positivt. SAOB:s presentationsordning av ordens betydelsemoment försöker följaktligen fånga sådan utveckling med utgångspunkt, i regel, i den etymologiskt äldsta betydelsen och inte sällan med upplysning om hur en betydelse utvecklats ur en annan.<sup>1</sup> Ordboken är också i mycket sitt material. Ytterst stor vikt läggs vid att samla in så mycket material som möjligt, och alla detaljer är lika viktiga. Tillsammans ska alla de olika formerna, stavningarna, hänvis-

---

<sup>1</sup> För en ingående redogörelse av SAOB:s principer för momentordning se Eaker (2006).

ningarna till ordens ursprung samt de belysande språkproven i princip tala för sig själva (detta således i kontrast till att låta resultatet av en undersökning av materialet få ta mest plats). Något förenklat: genom att lägga fram allt material som går att få tag i får läsaren den mest kompletta bilden av ordets natur.

Jag letar i min studie efter något som krockar med det här betraktelse- och förhållningssättet.

## 2. Metod och material

Utgångspunkten är de handböcker och regelsamlingar och dyl. som står till buds för personalen på SAOB-redaktionen (presenterade i tabell 1). Dessa har jag läst (kronologiskt) och sökt efter ändringar i anvisningar som explicit eller implicit vittnar om förändringar som ligger på ett överordnat, principiellt plan. Jag söker efter något som tyder på att inte bara språket i sig förändrats, eller att en äldre regel är klumpig eller oekonomisk – utan något som rör synen på hur man ska beskriva eller tolka språket som sådant. Härefter har jag testat resultatet på ordbokstexten, dvs. sökt i artiklarna utifrån regeländringar. Principen här är egentligen att jag vill undersöka om det går att ta sig vidare härifrån, och ungefär hur man skulle kunna gå till väga. Studien ska alltså ses som en förstudie eller en pilotundersökning för en eventuell framtida, större undersökning.

*Tabell 1. Material (i kronologisk ordning)*

**Handböcker, handledningar och anvisningssamlingar:**

Handbok för redigeringen af SAOB (Kock 1904)  
 Kodex Anders, anvisningar i form av lappregister (Sundqvist ca 1950-1990)  
 Vägledning till SAOB (Ekbo o. Loman 1971)  
 Handledning till SAOB (Lundbladh 1992)  
 Handbok för redigeringen av SAOB (Jonsson 1993)  
 Handbok för redigeringen av SAOB, 2 uppl. (Jonsson 1999)  
 Pigge, databas med anvisningar (Svensson m. fl. 1993-)

**Artiklar och fördjupningslitteratur:**

Om betydelse, betydelseanalys, och betydelsebeskrivning (Eriksson 1975)  
 Om historisk lexikografi typ SAOB (Jonsson 1983)  
 Koloss på lerfötter (Eriksson 1993)

### 3. Resultat och analys

#### 3.1 Hänvisning i sammansättningsramsor

Ett exempel på SAOB:s bakåtblickar och rikedom på detaljer finner man i de etymologiska parenteserna. Här leds ordet om möjligt till sin indoeuropeiska rot och eventuella långgivande motsvarigheter i andra språk räknas upp. Motsvarande eller nära besläktade varianter av ord i andra språk kan också anges efter ett ”jfr” (se figur 1). Den här typen av hänvisningar kan rimligtvis tolkas på två olika sätt. Antingen har det jämförda ordet sannolikt eller möjligen påverkat sin svenska motsvarighet (med avseende på form eller viss betydelse) eller förmedlat ordet till svenskan: ett ord är exempelvis påträffat tidigare i danskan, och det är rimligt att anta att detta har varit kanalen in till det svenska språket. Man kan även tänka sig att man har en generellare bild för ögonen, med tanken på ett ursprungligt ords utveckling eller förgrening i olika riktningar: ordet *x* förekommer i svenskan, detta ord kan hänföras etymologiskt till ett visst annat ord, eller en viss rot, som också gett upphov till ordet *y* i danskan.

Figur 1. Etymologi i sammansättningsramsor

#### **BOK-FÖRA.**

- [jfr d. *bogføre*]  
anteckna (i bok); i sht i fråga om räkenskaper;  
äfv. mer l. mindre bildl.

#### **DANS-MUSIK.** <sup>-02</sup>

- [jfr d. *dansemusik*, t. *tanzmusik*, eng. *dancemusic*, *dancing-music*]  
1) musik att dansa efter.

I SAOB:s redigeringsanvisningar hittar man ändringar eller förtydligande som berör just sådana hänvisningar i artiklarnas sammansättningsramsor. I den första handboken (Kock 1904), som i princip har karaktären av en exempelsamling och ofta saknar resonemang, kommenteras inte hänvisningsmodellen med ”jfr” på annat sätt än att exempel ges på i fornsvenskan förekommande ord, inhemskt eller lånat, enkelt eller sammansatt, där hänvisning med ”jfr” påbjuds. I den senare (och ersättande) anvisningssamlingen, *Kodex Anders*, påträffas däremot följande regel: ”utländska motsvarigheter medtas i etymologierna i *s(amman)s(ättnin)gsramsor* endast om de äro av speciellt intresse, dvs. om man kan anta, att det svenska ordet bildats efter mönster av *(el)l(er)* i sin användning påverkats av den utländska motsvarigheten.” (Sundqvist ca 1950-1970: etymo-



logi 5). Av formuleringen kan man misstänka att utländska motsvarigheter medtagits i ordboken i högre grad än motiverat. Regeln gäller fortfarande med tillägget att det hänvisade ordet också måste vara tidigare belagt än det svenska. (Svensson m. fl. 1993-: 209, senast ändrad 2010).

Alltså, enligt de tidigaste reglerna finns möjligheten (eller åtminstone inget i skrift stadgat hinder emot) att i en sammansättningsramsa hänvisa till utländsk motsvarighet med ett ”jfr”, utan särskild tanke på att det jämförda ordet direkt har påverkat det svenska. Detta faktum kan tolkas som ett utslag av en junggrammatisk tanke: ju mer desto bättre. Mer material, fler exempel ger ökad förståelse. Sådana hänvisningar idag blir besvärliga eftersom man inte när man läser sammansättningens etymologi, kan avgöra om den utländska motsvarigheten har påverkat det svenska ordet eller inte, och för att de – om det inte är fråga om påverkan – beskriver utvecklingen av ett ord i ett annat språk (vilket inte bör vara uppgiften för en svensk ordbok). Ett annat problem idag är förstås att hänvisningarna kräver utrymme. Perioden då regeln antecknades i *Kodex Anders* är emellertid inte en tid då man i första hand tänkte på att strama åt materialet.

För att testa om regeln har betytt något för ordbokstexten har jag jämfört några sammansättningsramsor från olika perioder med varandra, och räknat hänvisningar med ”jfr” (se tabell 2). Förlederna har valts ut slumpvis från äldre band och sedan jämförts med ord ur ungefär samma sfär i senare band. Det går givetvis inte att på ett helt rättvist sätt jämföra ett ord med ett annat. Varje ord och varje enskild sammansättning har egna förutsättningar – men en jämförelse av det här slaget kan nog ändå säga något om sättet att skriva hänvisningar och sammansättningsramsor. Jämförelsen är gjord i den tryckta ordboken.

Tabell 2. Hänvisningar med ”jfr” i sammansättningsramsa

Ramsa	Tryckår	Antal hänvisn.	Ssgr totalt	%
BERG ssgr D, bergs-	1904	31	226 (första)	13,7
TRYCK ssgr A, TRYCKA ssgr A	2005	2	226	0,9
DANS ssgr A.	1907	39	128	30,5
SÅNG ssgr A	2001	3	128 (första)	2,3
DEL ssgr A	1909	14	51	27,5
SIDA ssgr A, sid-	1967	2	51 (första)	3,9
ENKEL ssgr A	1922	4	60	6,7
SVÅR ssgr A	1999	0	60 (första)	0

Hänvisning med ”jfr” förekommer i samtliga fall som jag studerat, oftare i de tidigare banden av SAOB. Jag har endast tittat på ett litet antal ramsor (8 st.),

och resultatet kan inte representera hela SAOB, men det visar att det finns skillnader. Resultatet menar jag går att tolka som en enskild punkt där den grundläggande synen på språket förändrats, vilket fått konsekvenser för den faktiska ordbokstexten.

### 3.2 Bruklighetsmarkörer

Under mitten och slutet av 1900-talet börjar de strukturalistiska tankegångarna (med rötterna i Saussure) diskuteras och etablera sig i Sverige. Nu ligger fokus på hur det språkliga systemet hänger ihop och hur enheterna, eller de enskilda byggstenarna i detta system, påverkar varandra. För strukturalisterna är uppmärksammandet och upprätthållandet av skillnaderna mellan det synkrona och diakrona språkstudiet ett välkänt kännetecken.

För SAOB:s del är distinktionen här emellan relevant eftersom uppgiften är att i samma verk både beskriva samtida, nusvenskt språk och äldre språk (egentligen flera äldre språkskeden). I SAOB:s handböcker påträffas även regeländringar som har att göra med ett av redaktörernas verktyg för att navigera mellan olika språkskeden, nämligen bruklighetsmarkörerna.

En SAOB-redaktör har en stor uppsättning bruklighetsuppgifter till sitt förfogande, och dessa preciseras och kommenteras i handböckerna. Uppgifterna kan modifieras på olika sätt, men gemensamt för de allra flesta av dem är att de kan kombineras med ordet *numera*: (*numera* mindre br.), (*numera* bl. tillf.), (*numera* bl. i fackspr.), (*numera* i sht i bygdemålsfärgat spr.) osv. Men hur ska detta *numera* tolkas? I Kocks handbok är innebörden entydig: ”numera mindre br(*ukligt*) (*dvs.*) förr vanligt, men nu jämförelsevis sällan före(*ommande*) uttr(*yck*) för begreppet i fråga (i det att andra uttr(*yck*) numera äro vanligare). Mindre br(*ukligt*) (*dvs.*) både förr o. nu jämförelsevis sällan före(*ommande*).” (Kock 1904: bruklighet § 25.) Mot bakgrund av SAOB:s kronologiska metod – och rent intuitivt – är innebörden klar: så här förhåller det sig nu (i kontrast till hur det var tidigare).

I de senare anvisningarna, Kodex Anders, är innebörden emellertid en annan: ”Den i ... (*Kocks handbok*) angivna skillnaden mellan numera mindre br. – mindre br. ... iakttages icke. I regel användes formuleringen med numera. Formuleringen utan numera användes bl(*ott*) i fall då man kan anta att ifrågakavande användning icke alls förekommit i äldre tid (då endast belägg från 1900-talet finnas och ordet är okänt i dialekterna)” (Sundqvist ca 1950-1970: bruklighet 1). Samma regel gäller idag, men gränsen för äldre tid har (i tillägg till 2 uppl. av Jonssons handbok) flyttats fram till ca 1960.

Den nya regeln innebär att möjligheten att ange bruklighetsgraden för äldre språkskeden inte längre finns. Detta i sin tur innebär att ord som idag är mindre

brukliga, men som förr antingen var fullt brukliga eller också då mindre brukliga, samsas under uppgiften (numera mindre br.). *Numera* ska således enligt den nu gällande regeln tolkas så: vid denna tid (då ordet skrivs). Hur brukligheten var under olika tidigare skeden uttalas inget om. Redaktörens perspektiv är här synkront nutida. Detta perspektiv medför att den lexikografiska beskrivningen i praktiken blir mer korrekt, eller egentligen innehåller färre felaktigt dragna slutsatser. Det är svårt eller omöjligt för oss idag att uttala oss om hur brukligheten var förr, såvida inte allt existerande språkmateriel från alla aktuella tidsskeden finns tillgängligt, alternativt tidigare lexikografiska bruklighetsuppgifter för alla dessa tider, vilket det i praktiken givetvis inte gör.

Problematiskt för SAOB-texten är att man som läsare inte kan uppfatta att bruklighetsangivelserna ändrat innebörd. En titt på artiklarna i figur 2 vittnar om de svårigheter som föreligger för SAOB-läsaren: bruklighetsuppgifterna är identiska, men ska i de två fallen tolkas olika. Jag menar att också regeländringen av bruklighetsuppgifterna kan tolkas som en förändring som ligger på principplanet.

Figur 2. *Bruklighetsuppgiften (numera mindre br.)*

**BERÄTTELSE.** (tryckår: 1904)

- **3)** ... *a)* (numera mindre br.) abstr.: handling(en) att berätta, berättande. ÅGREN *Gell.* 5 (1757). Utom viktigare taleämnen, kommer man också på berättelsen af historier och anecdoter. TÖRNEROS *Bref* 1: 358 (1833).

**TENNIS-INSTRUKTRIS.** (tryckår 2003)

- (numera mindre br.) kvinnlig tennisinstruktör. SDS 29/6 1946, s. 6. SAOL (1973).

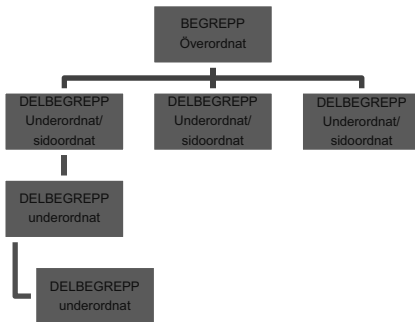
### 3.3 Den hierarkiska strukturen

I SAOB-redaktörernas handböcker och anvisningsmaterial påträffas också explicit resonemang om strukturalistiska betraktelsesätt och metoder. I den första upplagan av Hans Jonssons *Handbok* (1993) introduceras ett kortare avsnitt i kapitlet om betydelsebeskrivning där begrepp som struktur, lexem och opposition inom den postsaussureska strukturella semantiken diskuteras. Avsnittet bygger på en artikel av Jonsson (1982) som samtliga redaktörer på SAOB-redaktionen förväntas ha läst. Tidigare SAOB-redaktören Ulrik Eriksson har skrivit två artiklar med likartat tema, vilka räknas som baslitteratur för SAOB-redaktörer. I de här artiklarna diskuteras semantisk analys och betydelsebeskrivning ur ett strukturalistiskt perspektiv, med schematiska modeller för hur det

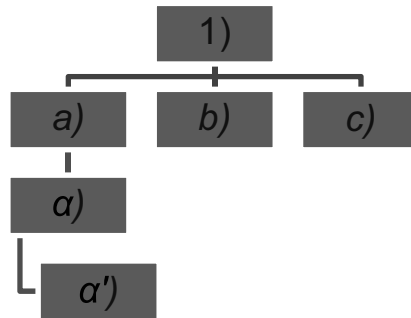
språkliga systemet är uppbyggt. I Erikssons texter förs en kritisk principdiskussion om det analytiska arbetet, medan Jonsson diskuterar de specifika förhållandena för ordböcker som SAOB.

Anvisningarna i handboken och den teoretiska diskussionen i artiklarna utgör alltså (en del av) en kunskaplig bas eller en språkteoretisk infallsvinkel som en redaktör på SAOB har idag (och har haft sedan texterna skrevs), och som skiljer sig mot den som en medarbetare på redaktionen vid arbetets början eller inledande skeden hade. Gemensamt för de nämnda texterna är bl. a. den centrala strukturalistiska tanken om semantisk hierarki: ett överordnat begrepp, i vid mening, delar upp olika delbegrepp under sig, vilka i sin tur kan vara under- eller sidoordnade varandra (se figur 3 a). Jag har gjort en smärre undersökning av detta fenomen och resultatet utgör studiens avslutande del. För SAOB:s rent grafiska struktur är också uppbyggd hierarkiskt, med huvudmoment, undermoment och undermoment till undermoment osv. och med strukturalistisk teori för ögonen kan ordbokens struktur utnyttjas väl vid utformningen av artiklarna (se figur 3 b).

Figur 3 a. Semantisk hierarki



Figur 3 b. Grafisk hierarki i SAOB



Jag ställer mig frågan om det går att undersöka i ordbokstexten om SAOB:s grafiska struktur vid olika tidpunkter används olika. Min utgångspunkt är redaktörens val inför dispositionen av en ordboksartikel (med relativt komplext betydelseomfång och rikt material). Hur ska det presenteras? Ska en framträdande betydelse få bilda ett undermoment till en annan betydelse, eller själv få utgöra ett huvudmoment? Valet är slutligen godtyckligt.

För att undersöka strukturen i SAOB har jag i ordbokens nätversion sökt på frekvensen av  $\gamma'$ . Dvs. ord med den grafiska strukturen huvudmoment (siffra) – undermoment (kursiv) – undermoment till undermoment (grekisk kursiv) – undermoment till undermoment till undermoment (grekisk kursiv prim-

tecken).  $\gamma$  innebär också att momentet befinner sig en bit in i strukturen (föregånget av  $\alpha$ ,  $\beta$ ). Hur många gånger används den här relativt fulla strukturen? Jag har jämfört två intervaller av ordbokstext, motsvarande vardera 20 år. Den första perioden sträcker sig ifrån 1967 till 1986, från ordet *segmentär* till *stassig*. Detta är en period då Hans Jonsson och/eller Ulrik Eriksson arbetade på redaktionen, och då strukturalistiska teorier diskuterades ganska flitigt i Sverige. Den andra perioden sträcker sig ifrån 1937 till 1956, från ordet *knävel* till *rem*, dvs. ett tidigare skede och utan nämnda artikelförfattare. Resultatet presenteras i tabell 3.B

Tabell 3. Utvidgad grafisk struktur i SAOB

ord	år	antal $\gamma'$	ökning (%)
KNÄVEL – REM	(1937 – 1956)	101	
SEGMENTÄR – STASSIG	(1967 – 1986)	193	

91

Det visar sig att möjligheten att disponera ordboksartikeln i en utvidgad hierarkisk struktur, (från huvudmoment ner till undermoment till undermoment till undermoment) har utnyttjats betydligt oftare under den senare perioden, då också de artiklar som refererats till ovan skrevs. Vi ser en ökning med över 90 %. Detta kan ha flera orsaker. Först och främst, de enskilda orden uppför sig olika, och det är som sades ovan, inte helt problemfritt att jämföra olika ord med varandra. Det är också så att SAOB:s material under de här aktuella åren har ökat kontinuerligt, vilket kan ha påverkat resultatet. Ett rikare material kan ge upphov till fler undermoment och en rikare struktur. Jag har bara tagit hänsyn till kvantiteten här, för att dra slutsatser måste materialet givetvis även studeras kvalitativt, för att se hur strukturen används. Men frånsatt detta, är det intressant att den mer komplexa grafiska strukturen är mer frekvent under den senare intervallen, och det kan nog säga något om synen på ords betydelser.

#### 4. Sammanfattning

Jag har tittat på tendenser till motsättningar i SAOB mellan olika föreställningar om vad språket är och hur det ska beskrivas. Först ur ett mikroperspektiv: det finns detaljer i ordbokens etymologiparenteser och bruklighetsangivelser som går att tolka som en krock med äldre synsätt, och tecken på nya. Med tanke på att ordboksartiklarna ska samsas i samma verk, är det nog också på detaljplanet som förändringarna tydligast kan uppmärksammas; inga revolutionerande

ändringar är att vänta. Ur ett makroperspektiv tittade jag sedan på hur ordbokens grafiska struktur använts. Under den senare av de två 20-årsperioder jag undersökte förekommer den utvidgade uppställningen betydligt oftare. Under samma period tillkommer anvisningar för ordboksartiklarnas utformande, där redaktörerna uppmanas att ta hänsyn till och analysera ordens semantiska hierarkiska struktur.

## KÄLLOR OCH LITTERATUR:

- Eaker, Birgit, 2006: Historisk lexikografi i praktiken. Om momentordningen i SAOB. I: *LexicoNordica* 13. Oslo.
- Ekbo, Sven, Loman, Bengt, 1971: Vägledning till Svenska Akademiens ordbok. I: *Skrifter utgivna av Nämnden för svensk språkvård*, 32. Läromedelsförlagen, Stockholm.
- Eriksson, Ulrik, 1993: Koloss på lerfötter. I: *Ord och lexikon*. Festskrift till Hans Jonsson. Stockholm: Norstedts.
- Eriksson, Ulrik, 1975: Om betydelse, betydelseanalys och betydelsebeskrivning. I: *Kring en ordbok*. Festskrift till Sven Ekbo. Stockholm: Norstedts.
- Jonsson, Hans, 1993: *Handbok för redigeringen av Svenska Akademiens ordbok*. SAOBs.
- Jonsson, Hans, 1999: *Handbok för redigeringen av Svenska Akademiens ordbok*. 2 uppl. SAOBs.
- Jonsson, Hans, 1983: Om historisk lexikografi typ SAOB. I: *Nysvenska studier* 63. Stockholm.
- Kock, Ernst Albin, 1904: *Handbok för redigeringen af SAOB*. SAOBs.
- Lundbladh, Carl-Erik, 1992: *Handledning till Svenska Akademiens ordbok*. Stockholm: Norstedts.
- Svensson, Lars m. fl., 1993-: *Pigge, databas med redigeringsanvisningar för SAOB*. SAOBs.
- SAOB = *Ordbok över svenska språket*, utg. av Svenska Akademien, 1893 ff. Lund. OSA = Svenska Akademiens ordbok i databasform. 1997 ff. <http://g3.spraakdata.gu.se/saob/>.
- SAOBs = *Samlingar till Ordbok över svenska språket*. Lund.
- Sundquist, Anders m. fl., ca 1950-1990: "Kodex Anders", redigeringsanvisningar för SAOB. SAOBs.

*Pär Nilsson*

Svenska Akademiens ordboksredaktion  
par.nilsson@svenskaakademien.se

# Nettordbøker og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok

*Christian-Emil Ore*

Printed dictionaries have been the way of disseminating systematized information about the use and meaning of words. During the last three decades electronic dictionaries have been common. Most large online dictionaries are based on retro-digitization and are electronic replica of the originals. To defend the cost of converting the printed text to electronic text, the result should be more than an online reproduction. What exactly this 'more' should be is unclear. As a background some dictionaries are examined to show how the digital medium is exploited.

In a paper in 1999 I described a meta-dictionary linking the entries with the source material as a future strategy for NO thus making the dictionary into an electronic scientific publication with reproducible results. In 2002 the meta-dictionary was redesigned to become a pivot in the new combined source database, text corpus and editing system for NO. Vols 5–12 are edited in the new system.

The meta-dictionary and the linked sources have become a valuable repository. The bidirectional linking in the system makes each headword in a source an entry point to NO and the entire system, thus giving dialect users a unique opportunity to see their dialect in the larger context.

The main objective of the NO project is the printed dictionary. However, the system is well prepared to meet the challenge when the 'web' and 'app' age will soon make the online database the main access point.

*Nøkkelord:* Elektroniske ordbøker, digitalisering, nettpublisering, redigeringsystemer, XML, tekstkoding, vitenskapelighet, formidling

Trykte ordbøker har en lang tradisjon som enerådende teknologi for å formidle systematisert informasjon om bruk og betydning av ord. I løpet av de siste 25 årene har elektronisk publisering blitt mer og mer vanlig. Termer som *nettordbok* og *elektronisk ordbok* blir brukt om slik publisering. Hva som skal til før noe kan sies å være en nettordbok er ennå ikke fastlagt. De skandinaviske språkenes term "ordbok" gir assosiasjoner til et fysisk objekt. Under NFLs 2005 symposium *Elektroniske ordbøker i Norden* var det diskusjon om hvilke krav en skal

sette til en elektronisk tjeneste for at den skal kunne kalles en elektronisk *ordbok*. Termen *ordbok* så ut til å binde flere av deltakerne mentalt til noe som minnet om den tradisjonelle papirordboken. I andre språk brukes termer som er uavhengig av bokformatet, for eksempel *dictionarium*, *glossarium* eller russisk *slovar*. Dette er termer som i en basalbetydning viser til samlinger av ord. I denne artikkelen vil jeg bruke *nettordbok* i vid forstand om alle slags opplysningstjenester for informasjon om ord på nettet og uten noen binding til det tradisjonelle ordboksformatet.

*Norsk Ordbok* fremstår som en tradisjonell, trykt ordbok. Den er planlagt i 12 bind og skal være ferdig i 2014. Under den synsvinkelen har Norsk Ordbok svært lite med nettordbøker å gjøre. Men bak denne trykte fasaden foregår redigeringen av Norsk Ordbok i et heldigitalt redigeringsverksted. De trykte ordboksbindene er en presentasjon i tradisjonelt trykkformat av bare en del av informasjonen i databasene. Våren 2012 ble de nyredigerte delene av Norsk Ordbok gjort tilgjengelig på internett. Denne første versjonen ligger nær opp til den trykte ordboken slik nettversjonen av *Oxford English Dictionary* (OED) ligger nær opp til sin trykte versjon. I skrivende stund (våren 2012) er det også planer om å publisere alt kildematerialet som finnes i Norsk Ordboks databaser sammen med hovedteksten i ett nettsted. Dette vil være en svært omfattende dokumentasjon av norsk språk diakront og geografisk og ikke minst støtte opp under Norsk Ordboks mål om å gi en uttømmende og vitenskapelig beskrivelse av ordtilfanget i norske dialekter og det nynorske skriftspråket.

I denne artikkelen vil jeg gå gjennom hvordan Norsk Ordboks fremtidige elektroniske publisering passer inn i det større bildet av nettordbøker og andre elektroniske ordbøker. Jeg tar også utgangspunkt i en artikkel jeg skrev for 12 år siden om mulighetene for en alternativ publisering av Norsk Ordbok og dens kilder. Artikkelen het *Metaordboken – et rammeverk for Norsk Ordbok* (Ore 1999) og var basert på et foredrag på konferansen til *Nordisk forening for leksikografi* i Göteborg i 1999. Som nå opplevde en den gangen en rivende teknologisk utvikling. I 1990 foregikk elektronisk publisering på disketter og CD-ROM. På slutten av 1990-tallet var det ikke svært vanlig å publisere ordbøker på nettet eller elektronisk på andre måter. De fleste ordbøkene kom bare i trykte utgaver, noen på CD-ROM og et fåtall som nettordbøker; Bokmålsordboka og Nynorskordboka begge 1991, OED 1996, SAOB 1997, for å nevne noen. Siden 1999 har utviklingen fortsatt. I dag er det irritasjon å spore blant våre brukere av Bokmålsordboka og Nynorskordboka for at disse ordbøkene ikke er tilgjengelig via app-er på smarttelefonene deres.

Da de trykte papirordbøkene var enerådende, handlet en viktig del av leksikografien om å få mest mulig informasjon inn på minst mulig plass, jf. førsteutgaven av Svenséns utmerkede håndbok (Svensén 1987). Det elektroniske



mediet har endret det. En har ubegrenset med plass og store pc-skjermer til å vise teksten. Så alt kan skrives fullt ut. Men i de siste årene har vi så fått smarttelefoner med en relativt sett liten skjerm, og dermed har en igjen behov for et mer stenografisk format. Slik vil det antagelig bølge frem og tilbake. Det eneste som er sikkert, er at det til enhver tid vil være komme nye tekniske løsninger for å slå opp informasjon om ord.

Denne erkjennelsen gjør at det er helt nødvendig for en ordboksredaksjon å foreta en grundig analyse av hvordan informasjonen i ordboken skal organiseres. Her er det selvfølgelig ikke snakk om overflatestruktur som kursiv kontra rettskrift, men om å sette opp en formell modell for hvordan elementene henger sammen. En ordbok skal etter min mening ikke sees som en tekst, men en database, som for eksempel Wordnet (Wordnet). Kunsten er å få informasjonen i denne over til et leselig format – en tekst. Når det gjelder eksisterende trykte ordbøker, er det en pågående diskusjon innenfor humanistisk informatikk om de er tekster eller formelle strukturer (Tasovac 2010).

De trykte ordbøkene fungerer jo utmerket. Det har likevel blitt gjennomført mange såkalte retrodigitaliseringsprosjekter for eldre trykte ordbøker. Målet har vært å gjøre dem om til elektroniske ordbøker, eller nettordbøker om en vil. Inntil relativt nylig var digital lagringsplass svært dyr og internett hadde relativt dårlig overføringshastighet. Den mest farbare veien på 1990-tallet var å skrive av ordbøkene og lagre dem som elektronisk tekst. I løpet av de siste ti årene har datalagring blitt så billig at hvem som helst kan lage digitale faksimiler av boksider. Ordet *digitalisering* har skiftet primærbetydning fra 'konvertering til digital tekst' til 'produksjon av digitale faksimiler'. Det er mye billigere å lage en faksimileutgave av en ordbok enn å taste inn teksten, selv i Kina. Det vesentlige må være hva som er det faktiske formålet med den digitale versjonen. Om en kun ønsker å gi brukerne tilgang gjennom den alfabetiske listen av oppslagsordene, trenger en ikke gjøre annet en å skanne inn sidene, taste inn oppslagsordene og lage lenker fra denne listen til faksimilene. En slik løsning er jo valgt for Otto Kalkars ordbok (KO). Kalkar-løsningen er billig, effektiv og ikke minst tro mot originalen. Dersom ren tilgjengeliggjøring er formålet, kan faksimile-løsningen altså være den beste. Som historikeren Manfred Thaller har formulert det: "A million-page collection with catalogue metadata is, to be precise, considerably more useful than a collection of ten volumes with complete transcriptions encoded according to an elaborate markup scheme developed for the occasion" (Thaller 2001).

Dette er lik tanken bak Google Books, som dessverre ikke har gode katalogdata. Så dersom en ønsker å taste inn hele ordboksteksten, må det være fordi en vil noe mer. Hva dette mer egentlig er, er ikke alltid like klart. Et formål kan være å etablere en elektronisk tekst som skal danne grunnlag for en videre revi-

sjon av ordboken. Det første store eksempelet på dette er digitaliseringen av Oxford English Dictionary (OED) i andre halvpart av 1980-tallet. Siden har slik inntasting blitt alminnelig. Formålet er ikke nødvendigvis elektronisk publisering, men nye trykte utgaver. Men tar en seg råd til å konvertere til elektronisk tekst, har en et hendig og kompakt format med en rekke anvendelsesområder.

De fleste såkalte nettordbøker er i dag basert på elektronisk tekst, enten "born digital" eller retrodigitaliserte trykte utgaver. Det er veldig ujevnt i hvilken grad utgiverne har utnyttet teksten til annet en ren (re)produksjon av "pen" tekst. Mange nettordbøker utnytter ikke den elektroniske teksten til noe mer, og leverer derfor egentlig bare en variant av den enkle "Kalkar"-løsningen til Statsarkivet i Bergen. Hverken Bokmålsordboka, Nynorskordboka eller J. Fritzers norrøne ordbok på nett gir ikke så mye mer enn det. Det samme gjelder for så vidt de 60 ordbøkene i Ordbokshotellet til Norsk Ordbok eller Kunnskapsforlagets "ordnett.no". For de ordbøkene jeg har jobbet med skyldes den enkle nettpresentasjonen flere forhold: manglende kjennskap til mulighetene og tilhørende fravær av ønsker hos redaktørene, manglende ressurser til å dypanalisere den elektroniske teksten og at en rekke av ordbøkene ikke har en stringens som ville gjøre det verdt bryet. Men det finnes eksempler, også i Norden, der en har utnyttet mulighetene på en forbilledlig måte. *Den danske ordbog* har en godt strukturert manuskriptdatabase i bunnen og en har evnet å utnytte den til å lenke inn både ordnett, korpus og interne henvisninger. En her har hatt den fordel at en tenkte strukturert fra begynnelsen (DANLEX 1987).

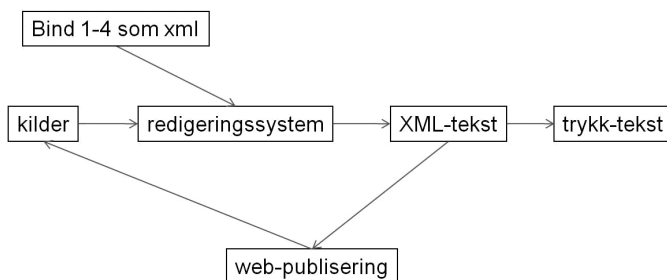
For de såkalte retrodigitaliserte ordbøkene er utgangspunktet mer kompleks. Ordboksredaktørene på 1800-tallet organiserte ikke sine manuskripter etter databasemodeller eller XML-skjema. Men mange eldre ordboksverk er skrevet med en imponerende konsistent artikkelstruktur. Like fullt kreves det et avansert intellektuelt arbeid for å analysere og kode retrodigitalisert ordbøker så langt det går i henhold til en predefinert struktur. Det er ikke uten grunn at TEI Guidelines P5 (TEI P5) for ordbokskoding også har et XML-element "entry-Free" der delementene kan komme i vilkårlig rekkefølge. Da OED ble digitalisert og kodet på slutten av åttitallet, ga en faktisk opp å lage en særlig avansert DTD fordi artiklenes format var for uensartet. Men denne noe flate strukturen har blitt utnyttet godt i den elektroniske utgaven. Den elektroniske OED har en layout som gir assosiasjoner til den originale satsen, men er likevel ganske forskjellig.

En annen stor og viktig retrodigitalisert ordbok er brødrene Grimms *Deutsche Wörterbuch*, *DWB* (DWB). Det digitaliserte manuskriptet ble gitt en detaljert XML-koding i henhold til TEI Guidelines. CD-ROM utgaven fra 2004 bruker kodingen til minste detalj, mens nettversjonen nøyer seg med

strukturkoding i høyre marg og viser originalteksten temmelig kompakt i midtfeltet. Siden er enkel, men ikke elegant som DDOs side. DWB er en del av et tysk ordboksnett [www.wortebuchnet.de](http://www.wortebuchnet.de).

De nordiske svarene på OED og DWB på nettet må sies å være *Ordbok over det danske sprog (ODS)* og *Svenska Akademiens Store Ordbok (SAOB)*. SAOB ble for en stor del konvertert med OCR på nittitallet og kodet og har mange feil, særlig i formatet. Dagens SAOB på nettet hadde kanskje vært bedre om en brukte den elektroniske teksten til å søke i og i tillegg hatt en faksimilevisning av de trykte sidene. Etter sigende skal SAOB skrives inn på nytt slik ODS har blitt det. ODS ble dobbelttastet manuelt i Kina og gitt en enklere XML-koding. Det gjør det i dagens nettutgave mulig å søke på oppslagsord og se enkeltartikler med en delikat layout som minner om den trykte utgaven. Man må kunne anta at utgiverne vil arbeide videre med XML-teksten og knytte den nærmere opp mot DDO og det danske ordnettet.

Norsk Ordbok (NO) er det norske svaret på de store ordbøkene nevnt over. NO er planlagt og designet som en tradisjonell trykt ordbok i 12 bind. Bindene 1 og 2 ble redigert manuelt (skrivemaskin og tradisjonell sats). Manuskriptteksten til bindene 3, 4 og første del av bind 5 ble som en del av redigeringsarbeidet skrevet inn i henhold til et enkelt felt- og kodesystem utviklet på 1980-tallet. Bind 3, 4 og halve bind 5 ble satt elektronisk på grunnlag av den kodete teksten. Det nye redigeringsystemet ble tatt i bruk fra og med bokstaven *i*. Trykksatsen lages i det. NO2014 og EDD har analysert den elektroniske teksten (a-h), gitt den en XML-koding og lastet den inn i redigeringsystemet. Norsk Ordbok redigeres ved hjelp av digital teknologi, men er ennå ikke fullt ut publisert som en digital ordbok, (Ore&Tvedt 2005).



Figur 1 Struktur i Norsk Ordbok-systemet

Selv om Norsk Ordbok ikke har vært planlagt som en digital ordbok og bestillingen fra det norske kulturdepartementet er en trykt ordbok i 12 bind, ligger forholdene godt til rette for å lage en nettversjon av ordboken. I 2011 har det vært arbeidet med å utvikle en slik nettversjon. En prøveversjon ble lansert

våren 2012 (no2014.uio.no). Betrachtingene i den siste delen av denne artikkelen står for min regning og er ikke nødvendigvis NO2014-prosjektets planer eller syn på saken.

Som nevnt ligger manuskriptet til Norsk Ordbok i en database med en struktur som antydet i figur 1. Når teksten skal vise på nettet, skal en da vise det på ODS' maner, det vil si nærmest mulig opp mot det originale satsbildet? Skal en løse opp strukturen og bruke strukturen i XML-teksten for det den er verd? I sistnevnte tilfelle vil vi få en presentasjon av innholdet i Norsk Ordbok som ikke minner så mye om det grafiske utseendet til de trykte bindene. Det er flere forhold en må ta med i betrachtingen når en skal republisere et verk som Norsk Ordbok i nye medier:

1. Norsk Ordbok var og er et prosjekt som skal lage en tradisjonelt utseende ordbok trykt på papir.
2. Redaktørene i Norsk Ordbok er blitt ansatt for å produsere en trykktekst. Kan det for noen medføre at de lager en annen tekst om de skulle ha produsert en nettordbok med dette mediets mange flere muligheter?
3. Er det uetisk eller bryter det med regelverket å lage en nettutgave av Norsk Ordbok som grafisk overhodet ikke minner om originalen? Bryter vi rettighetshaveres ideelle rettigheter tilsvarende Duke Ellingtons versjon av Griegs Peer Gynt suite som ble oppfattet som krenkende av Norsk komponistforening i 1960 og der platen ble forbudt?

På den ene siden er det godt mulig at eldre redaktører vil være fornøyd med et nytt format. På den andre siden kan en vise en viss pietet med hensyn på originalutseende, kanskje ved å gi ut originalversjonen også. Desto eldre et verk en republisere på nett er, desto mindre tenderer en til å tenke over slik hensyn. En god og klar nettpublisering vil øke tilgjengeligheten til innholdet i Norsk Ordbok og dermed bidra til å formidle resultatene av års intellektuelt arbeid. Men det bør nok vurderes om en i tillegg også bør publisere ordboken i et nettformat som svarer til det originale. Men Norsk Ordbok er et prosjekt som har et svært ambisiøst mål:

Norsk Ordbok skal gje ei uttømmande vitskapleg framstilling av ordtilfanget i dei norske dialektane frå 1600 og til i dag og av det nynorske skriftmålet. Hovudvekta skal ligge på grundige tydingsanalysar med klåre definisjonar, kjeldefesta døme på bruk i litteraturen og i målføra, opplysningar om grammatiske eigenskapar og om dialektformer med heimfesting. (www.no2014.uio.no)

Hva som menes med "vitenskapelig" er ikke nærmere beskrevet. Termen "uttømmande" kan implisere at det rett og slett ikke finnes mer informasjon om

ordforrådet enn det som er står Norsk Ordbok. Heldigvis beskrives betydningen av "uttømmende" noe svakere som "tilstrekkeleg, fullgod, dekkjande" i Nynorskordboka og andre norske ordbøker. "Vitskapleg" er mer interessant. "Vitskap" er i følge Nynorskordboka "metodisk innsamling, ordning og etterprøving av kunnskapar og hypotesar etter allment godkjende reglar og krav (om intersubjektiv gildskap [gyldigheit], klarleik osv)". Det er vel stor grad av enighet om at når en beskriver et ords betydninger slik det gjøres i ordbøker eller i leksikalsk semantikk i sin alminnelighet kan det ikke sies å være noen form for hypotesetesting. Det er vel heller i sin beste form en systematisk observasjon og betydningsbeskrivelsene er resultatet av en tilhørende godt begrunnet induksjon. Det vitenskapelige ligger altså i at det er mulig å se det faktiske, empiriske grunnlaget for (induksjons)slutningen. Hvorvidt betydningsbeskrivelsen er skrevet som tekst beregnet for mennesker eller i et mer formelt skjemaspråk er ikke avgjørende for den vitenskapelige gehalten. Et forsøk på å uttrykke betydningen fullstendig analytisk eller i en formell struktur som for eksempel i det generative leksikonformatet er ikke nødvendigvis spesielt vitenskapelig eller forståelig:

One of the several conclusions that may be drawn from this view of lexical meaning [as a set of experiences] is that any attempt to write a completely analytic definition of any common word in natural language is absurd. Experience is far too diverse for that. What a good dictionary offers instead is a typification: the dictionary definition summarizes what the lexicographer finds to be the most common features, in his experience, of the use, context, and collocations of the word. (Hanks 1979)

Jeg vil si det viktigste kravet en kan stille til en vitenskapelig ordbok er at brukerne kan få innsyn i det materialet redaktøren hadde til disposisjon da artikkelen ble skrevet. Det er også en forutsetning at dette grunnlagsmaterialets proveniens som for eksempel kilde, sted og tidfesting er kjent. Det er viktig hvis vi ønsker å kunne trekke pålitelige slutninger fra ordboken med hensyn på utbredelse etc.

Det er ingen tradisjon for at store verkene som ODS, OED, DWB og SAOB skal gi oss mulighet til å etterprøve redaktøren; definitivt ikke i den trykte utgaven, men heller ikke i den digitale utgaven. Litteraturreferansene gir bare kilde for de oppgitte brukseksemplene. Et interessant mellomtilfelle er den delvis redigerte *Ordbog over det norrøne prosaspråk (ONP)*. Nettstedet gir brukeren mulighet til å søke i de ferdig redigerte artiklene (a – em). For resten av alfabetet finnes en database med det digitaliserte kildematerialet og en tilhørende lemmaliste utviklet etter modell av Norsk Ordboks metaordbok. Det er å håpe at redigeringen av ONP blir fullført, og det ville være flott om en kunne lenke opp resten av kildematerialet mot ordboksmanuskriptet. Et uavhengig eksem-

pel er Dictionnaire Étymologique de l'Ancien Français (DEAF) som redigeres i Heidelberg. Her har en digitalisert seddelarkivet og lenket det mot artiklene som redigeres. De eldre delene av ordboken vises som faksimiler av original-sidene. I tillegg er oppslagene i DEAF lenket opp mot tilsvarende prosjekter andre steder i Europa.

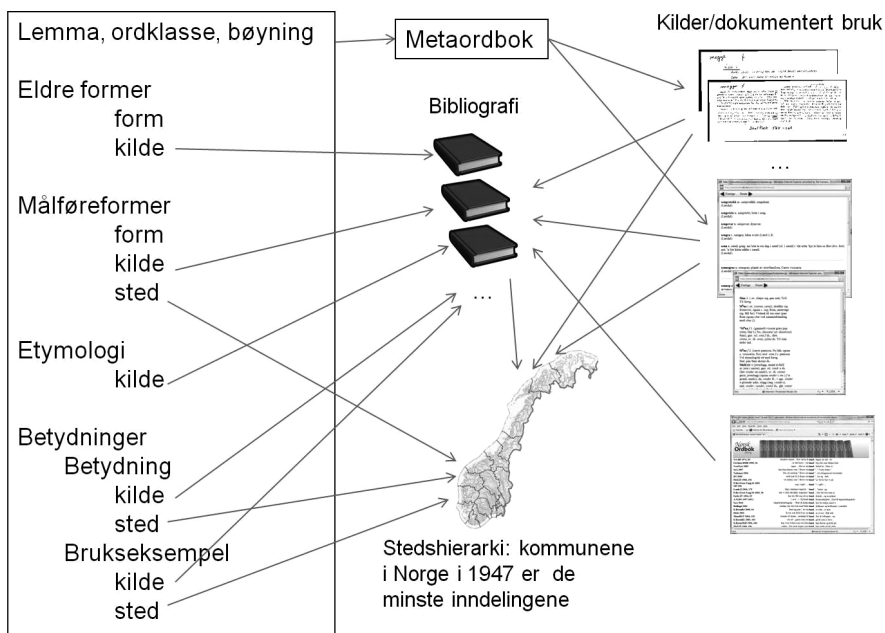
Moderne leksikografiske verktøy som for eksempel Sketch Engine (SkE) eller Tshwanelex inviterer til å lenke korpusanalyse til artiklene i en korpusbasert ordbok. Redigeringsprogrammet Onoma utviklet i 1998–2000 for *Woordeboek van die Afrikaanse Taal (WAT)* av Lexilogik AB er et tidlig eksempel. Redigeringsystemet til Norsk Ordbok (1998 – 2005) er enda et eksempel. Grunnlaget for å lage vitenskapelig etterprøvbare ordbøker er til stede både metodisk og i form av redigeringsprogrammer. Men det har åpenbart ikke blitt gjennomført i noen nettordbøker. Det kan skyldes mange forhold. For mindre, kommersielle ordbøker er det ikke rimelig å forvente noen ambisjoner om vitenskapelighet. Slike ordbøker er praktiske forlagsprodukter og ikke vitenskapelige verk. For de større ordbøkene spiller vel også økonomi inn. Men innen academia er det publikasjonen som teller. Når et vitenskapelig arbeid er fullført og dokumentert gjennom en publikasjon, er det tradisjonelt sett liten vilje til å ta vare på grunnlagsmaterialet. Man kan hevde at det er ingen som kan etterprøve selve innsamlingen og dermed har det ingen verdi å ta vare på det. Det er så, men når grunnlagsmaterialet er borte, kan en heller ikke etterprøve forskerens resonnement fra grunnlagsmaterialet til konklusjonen. Slik sett er resultatene i svært mange empirisk baserte vitenskapelige artikler ikke etterprøvbare og dermed er de i streng vitenskapelig forstand verdiløse.

Norsk Ordboks ambisjon om å være en vitenskapelig ordbok forutsetter dermed at kildematerialet som var tilgjengelig for redaktøren da artikkelen ble redigert, også er tilgjengelig for oss brukere. I en triviell forstand er det tilgjengelig siden seddelsamlingen står i redaksjonens lokaler og alle de andre skriftlige kildene finnes i håndbiblioteket eller i tekstkorpuset. Spørsmålet er om vi kan forlange noe mer? I min artikkel fra Göteborg 1999 viste jeg til eksempler der forskningsprosjekter publiserte det empiriske materialet sammen med de bearbejdede resultatene som en samlet elektronisk utgivelse. At det skulle bli en allmenn trend var i overkant optimistisk. Det har ikke holdt stikk av ulike årsaker, både fordi publikasjonen er det endelige mål og fordi det ikke er noen tradisjon for å dele slike ressurser.

I 1999 hadde Norsk Ordbok 2060 som antatt sluttdato. Det var på det grunnlaget jeg foreslo at modellen med grunnlagsmaterialet og de bearbejdede resultater som en samlet enhet kunne brukes som et substitutt for den ferdige Norsk Ordbok. En slik modell var ikke urealistisk siden mesteparten av bak-

grunnsmateriale allerede var digitalisert. At dette er blitt modellen også for ONP kan en bare notere seg.

Det er også et annet aspekt med en slik modell brukt for Norsk Ordbok. I den mest puristiske korpusbaserte leksikografien er det bare empiri i form av et tekstkorpus som gjelder. Men for verk som Norsk Ordbok må en også støtte seg på et veldefinert sett av utvalgte kilder i tillegg. Slike er for eksempel spesialiserte fagordbøker og dialektordbøker. En fagordbok gir ofte klare og dekkende definisjoner av (allmenn)ord brukt som tekniske termer. Redaktørene av (gode) dialektordbøker har ofte en nærhet til den lokale tradisjonen og kan vie større plass til de lokale tradisjonsordene, det vil si ord som betegner lokale skikker, redskap, matretter og arbeidsprosesser. I Norge er Njål Veres ordbok *Listamålet* (Vere 1993) med sitt lokale tradisjonsstoff et slikt eksempel. I Danmark har en den mye større *Ømålsordbogen* (ØMO) med et rikt etnografisk materiale. Disse to ordbøkene inkorporerer i stor grad det en med rimelighet kan kalle encyklopedisk informasjon. Norsk Ordbok er en allmennordbok, har begrenset plass og må ikke minst se ordens bruk og i et større landsdekkende perspektiv. Det er derfor naturlig at beskrivelsene en finner i disse spesialiserte ordbøkene kan være klarere og gi mer detaljert informasjon enn det som er mulig i henhold til allmennordboksformatet til Norsk Ordbok.



Figur 2 Sammenhengene i Norsk Ordbok-systemet

Den vitenskapelige ideen bak *metaordboken* er altså at det digitaliserte kildematerialet til Norsk Ordbok kunne bli knyttet til ordboksartiklene i Norsk Ordbok slik at brukerne av systemet kunne studere kildene redaktørene hadde til disposisjon når de skrev ordboksartiklene. Men for den allmenne bruker er det kanskje like viktig at lenkingen gir brukerne tilgang til en rik samling av mindre ordbøker og andre kilder som gir brukerne mer detaljerte opplysninger om for eksempel lokale forhold. Omvendt vil artiklene i Norsk Ordbok bli knyttet til metaordboken på linje med de andre kildene. Koblingen går altså begge veier. Det betyr at en bruker av en digital dialektordbok knyttet til metaordboken kan få se hvilken informasjon redaktørene Norsk Ordbok har redigert for det gitte ordet. Hva mer er; oppslagsordene i dialektordbøkene kan brukes som alternativ søkeinngang til Norsk Ordbok. Det kan ofte være stor avstand mellom de normerte oppslagsordene i Norsk Ordbok og de ulike dialektrealiseringene. Sturla Berg-Olsen og Åse Wetås gjorde rede for muligheter og problemer knyttet til dette i sitt fordrag i Lund (Wetås & Berg-Olsen 2011).

Etter disse generelle betraktningene er det riktig å se litt nærmere på arkitekturen i Norsk Ordbok-systemet for å se i hvilken grad den understøtter en vitenskapelig etterprøvbarehet (se Grønvik og Tvedt 2006 for en beskrivelse av redigeringsystemet). En artikkel i Norsk Ordbok består av tre hoveddeler: oppslaget (lemma, ordklasse og bøyingsinformasjon), en beskrivelsesdel ("hakeparentesen") for eldre kilder, avvikende skriftformer, dialektrealiseringer og etymologi, samt den sedvanlige artikkeldelen med betydningsbeskrivelser og brukseksempler. Informasjonen i oppslagsdelen er i henhold til 1938-rettskrivningen og er normativ. Den trenger derfor ingen begrunnelse. De ulike delene av hakeparentesen vil på trykk kunne bestå av ordformer med tilhørende kilde- og stedfesting. Betydningsdelen kan være en enkel liste eller et hierarki; informasjonsteknisk sett kan vi kalle den en trestruktur. Markerte betydninger tradisjonelt angitt i trykk med tall eller bokstaver svarer til en node i et slikt tre. Til hver node finnes det i databasen en informasjonspakke som kan inneholde definisjonstekst, krysshenvisninger, redaksjonelle brukseksempler og sitater fra kildematerialet med mer. Definisjoner og brukseksempler kan ha tilknyttet sted og kildefesting, mens sitatene selvfølgelig må ha det. Stedfestingen er oftest en generalisering til et større område basert på de stedene som er angitt i kildematerialet. Henvisningene til de litterære kildene og sitatene, vil være en delmengde av det som finnes i kildematerialet. Utvalget og generaliseringene er et resultat av redaktørens intellektuelle arbeid. Dette er standard for deskriptive ordbøker som Norsk Ordbok. Men det gir altså ingen direkte tilgang til bakgrunns materialet. En standard digital utgave av Norsk Ordbok vil ikke endre på det. En slik digital utgave bare kunne hjelpe brukeren til å skjønne de stenografiske steds- og litteraturhenvisningene og eventuelt også gjøre det mulig å søke



geografisk eller bibliografisk. For Norsk Ordbok er listen over de 6 000 litterære kildene lagret i en standard bibliografisk format (MARC). For stedfestingen er norske kommuner i 1947 den høyeste nøyaktigheten, og er byggeklossene i stedfestingene i ordboken. Kommunegrensene er blitt digitaliserte, og det åpner for kartbasert søk og visning (se Grønvik og Holmen 2011).

Norsk Ordbok har i tillegg kobling til kildene via metaordboken. Hver artikkel er dermed knyttet til de kildene som finnes for oppslagsordet. Redigering gjør det også mulig å koble kilder, som sedler, andre ordbøker og resultatet av korpussøk direkte til (del)betydningene og til de tre kategoriene eldre kjelder, avvikende skriftformer og målføreformer. Dette er ikke brukt i så stor grad på grunn av høyt arbeidspress. Så en må nøye seg med å sjekke det samlede grunnlagsmaterialet for hele artikkelen. Som nevnt over, kan resultatet av korpussøk lagres i metaordboken i form av kwic-konkordanser i metaordboken. Dermed kan en sjekke hvilket utvalg fra korpus som er brukt. Denne mekanismen er heller ikke brukt i noen særlig grad. Det kan også skyldes at det er tungt å redigere ned en resultatkonkordanse til de linjene en faktisk har brukt. Men dersom noe fra korpus brukes som sitater må det lages sedler som lenkes i metaordboken. Norsk Ordbok bruker IMS Corpus Workbench (CW) som korpusverktøy. Det har en relativt enkel måte å vise tilslag. En kunne i fremtiden tenke seg å bruke for eksempel Sketch Engine (SkE) som leverer ferdigere analyser av korpusmaterialet. Disse skissene kan lett lagres i metordboksystemet for senere sjekk. Denne lille gjennomgangen viser at det er om ikke huller i systemet for å etterprøve redaktørene, så er det visse svakheter. Men om en går tilbake til forslaget til metaordbok fra Göteborg 1999 har en kommet veldig langt mot en fullstendig realisering av en vitenskapelig etterprøvbar ordbok.

Den første versjonen av Norsk ordbok som nettordbok, *Norsk nettordbok*, ble publisert våren 2012 og er en tradisjonell nettordbok med funksjonalitet som OED eller DWB. Den vil i løpet av 2012 etter hvert med kartfunksjonalitet. Det neste skrittet er å integrere kildematerialet på en god måte. Det er en utfordrende oppgave å formidle et slikt kompleks med hensyn på klarhet for den vanlige bruker. Men det vil gi en enestående inngang til kunnskap om norsk språk og synliggjøre den formidable innsatsen til redaktørene av Norsk Ordbok.

## LITTERATUR

DANLEX = Descriptive tools for electronic processing of dictionary data: studies in computational lexicography. The DANLEX-group, Ebba Hjort (et al.). *Lexicographica*, Series Maior 1987. Tübingen: Niemeyer.

- DEAF = Dictionnaire Étymologique de l'Ancien Français, [www.deaf-page.de](http://www.deaf-page.de) (sjekket 05-11-2011)
- DWB = Deutsche Wörterbuch [www.dwb.uni-trier.de](http://www.dwb.uni-trier.de) (sjekket 05-11-2011)
- CW IMS = [www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/](http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/) (sjekket 05-11-2011)
- Grønvik, Oddrun og Tvedt, Lars Jørgen, 2006: Norsk Ordbok 2014 – presentasjon av eit komplekst leksikografisk verktøy. I: Nordiske Studier i Leksikografi nr 8, København.
- Holmen, Jon og Grønvik, Oddrun, 2012: Norsk dialektatlas i digital utgåve og språkdata på kart. I: Nordiska studier i lexicografi 11. Lund. (Skrifter utgivna av Nordiska föreningen för lexicografi 12.)
- Hanks, Peter, 1979: To what extent does a dictionary definition define? I: R. R. K. Hartmann (ed.), Papers from the 1978 B.A.A.L. Seminar on Lexicography. Exeter Papers in Lexicography.
- OED = Oxford English Dictionary: [www.oed.com/](http://www.oed.com/) (sjekket 05-11-2011)
- Ore, Christian-Emil, 1999: Metaordboken – et rammeverk for Norsk Ordbok. I: Nordiska studier i lexicografi 5. Göteborg.
- Ore, Christian-Emil, Tvedt, Lars Jørgen, 2006: Digital sats eller digital satsing. I: Nordiske Studier i Leksikografi nr 8. København.
- SkE Sketch Engine: [www.sketchengine.co.uk](http://www.sketchengine.co.uk) (sjekket 05-11-2011)
- KO = Kalkar, Otto, 1881–1918: Ordbok til det ældre danske sprog (1300-1700). [www.hist.uib.no/kalkar/](http://www.hist.uib.no/kalkar/)
- Svensén, Bo, 1987: Handbok i lexicografi : principer och metoder i ordboksarbetet. Esselte studium; Tekniska nomenklaturcentralen. Stockholm.
- Tasovac, Toma, 2010: Reimagining the Dictionary, or Why Lexicography Needs Digital Humanities, Digital Humanities 2010, Kings College, London, UK
- TEI P5 = Text Encoding Initiative P5: [www.tei-c.org](http://www.tei-c.org) (sjekket 05-11-2011)
- Thaller, Manfred, 2001: From the Digitized to the Digital Library, D-Lib Magazine, February 2001, Volume 7 Number 2, ISSN 1082-9873.
- Tshwanelex: [tshwanelex.com/tshwanelex](http://tshwanelex.com/tshwanelex) (sjekket 05-11-2011)
- Vere, Njål, 1993-1995: Grammatikk, ordbok, stadnamn og folkeminne I-III, Vanse.
- Wetås, Åse og Berg-Olsen, Sturla, 2012: Norsk Ordbok på nett: Utfordringer og løysingar . I: Nordiska studier i lexicografi 11. Lund (Skrifter utgivna av Nordiska föreningen för lexicografi 12.)
- WAT = Woordeboek van die Afrikaanse Taal. <http://www.wat.co.za> (sjekket 05-11-2011)
- Wordnet = [wordnet.princeton.edu](http://wordnet.princeton.edu) (sjekket 05-11-2011)
- ØMO = [dialekt.ku.dk/dialekter/oemaalsordbogen/](http://dialekt.ku.dk/dialekter/oemaalsordbogen/) (sjekket 05-11-2011)

*Christian-Emil Smith Ore*

Institutt for lingvistikk og nordiske studier, Det humanistiske fakultet, Universitetet i Oslo.

[c.e.s.ore@iln.uio.no](mailto:c.e.s.ore@iln.uio.no)

# Dansk leksikografi med polsk: *past, present and future*

*Marcin Overgaard Ptaszynski*

The aim of this paper is to outline the development of Danish-Polish lexicography from its beginnings to the present. The paper adopts a functional perspective. Therefore, its primary focus is on situations which have triggered the need for communication between Danes and Poles, and on the extent to which these are reflected in the structure and content of the dictionaries examined. Each of the investigated dictionaries is examined with a view to determining two things: firstly, its functionality in the particular communicative needs it is meant to satisfy, and, secondly, its capacity as a source of inspiration for the development of Danish-Polish lexicography. Finally, on the basis of the investigations the paper offers several desiderata for future dictionaries between Danish and Polish.

*Nøgleord:* Dansk, polsk, ordbog, leksikografisk funktion, tekstreception, tekstproduktion, kommunikationsbehov

## 1. Indledning

Selv om kontakterne mellem Danmark og Polen strækker sig over 1000 år og har været meget forskelligartede, er det først i forbindelse med den polske indvandring til Danmark i 1890'erne at de første ordbøger mellem dansk og polsk bliver til. Flere ordbøger udkommer senere, da behovet for kommunikation mellem de to lande vokser i følge med dansk investeringsaktivitet i Polen efter 1989 og især indvandring af polsk arbejdskraft efter Polens indtræden i EU i 2004.

I dette bidrag skitseres den historiske udvikling af ordbøger mellem dansk og polsk ud fra et funktionelt perspektiv. Den grundlæggende tese er at leksikografien afspejler konkrete typer informationsbehov som opstår i samfundet (jf. Tarp 2006). Dermed bygger analysen af hver af de undersøgte ordbøger på tre spørgsmål:

1. Hvad er ordbogens funktioner, dvs. hvem skal bruge den og hvad skal den bruges til?
2. Hvor godt passer de i ordbogen implementerede løsninger til ordbogens funktioner?
3. Hvad kan man lære af ordbogen med henblik på at udvikle nye ordbøger?

Undersøgelsen slutter med en redegørelse for de nuværende leksikografiske behov med dansk og polsk.

I denne korte artikel er det naturligtvis umuligt at skabe overblik over alle relevante ordbøger og relaterede, leksikografiske informationsværktøj der findes. Derfor er et særligt fokus rettet mod almenordbøger i trykt format, hvoraf kun et repræsentativt udvalg omtales i detaljer.

## 2. Tidslinjen: Behov for kommunikation og ordbøger

For at kunne analysere den dansk-polske leksikografis historiske udvikling fra en funktional vinkel er det nødvendigt at undersøge hvilke typer situationer der har affødt behov for kommunikation mellem danskere og polakker, og dermed dannet grunden til at lave ordbøger. Her kan nævnes fire overordnede faktorer, hvoraf den der har skabt det unægtelig største kommunikationsbehov, har været polsk indvandring til Danmark.

Indvandringen har foregået i flere bølger over de sidste 120 år (jf. Kolstrup 2010). Fra tidligt i 1890'erne til 1930 kom der adskillige tusinde polakker som sæsonarbejdere til danske roemarker og teglværker. Det er svært at angive nøjagtige tal for hele den lange indvandringsbølge. Kolstrup (2010: 23-24) mener dog at bølgen nåede sit højdepunkt i 1914 med over 14.000 polske arbejdere, og anslår at over 5.000 polakker blev tilbage i Danmark efter indvandringsstoppet i 1930. Dertil skal man lægge omkring 2.500-3.000 jøder, som udvandrede i perioden 1905-1920 fra storbyer i den russisk besatte del af Polen for at bosætte sig i København.<sup>1</sup> Senere kom der polske krigsflygtninge (op til 9.000 i perioden 1945-1949, hvoraf de fleste blev hjemsendt senest i 1949), jødiske tvangsemigranter fra det kommunistiske Polen i slutningen af 1960'erne (omkring 3.000) og politiske flygtninge under krigsretstilstanden i 1981-1988 (2.500-3.000). I løbet af hele kommunistperioden (1945-1989) så man også en beskednen men regelmæssig bølge af afhoppere fra det polske militær og efterretningstjeneste. I 2004 blev Polen medlem af EU, og siden da har man oplevet massiv tilflugt af polsk arbejdskraft, især håndværkere og medicinsk personale,

---

<sup>1</sup> Medmindre andet er angivet, stammer alle tal i dette afsnit fra Kolstrup (2010).

til Danmark (op mod 14.000).<sup>2</sup> For at kunne begå sig i Danmark blev alle polske indvandrere nødt til at tilegne sig det danske sprog, og dermed opstod der behov for ordbøger mellem polsk og dansk.

Ud over indvandringen har følgende faktorer skabt kommunikationsbehov mellem danskere og polakker, og dermed påvirket udviklingen af den dansk-polske leksikografi: undervisning i dansk hhv. polsk som fremmedsprog, turisme og erhvervsaktiviteter. Muligheden for at lære polsk på universiteter i Danmark og dansk i Polen eksisterede allerede i 1930'erne (Stemann 1945: iii), og findes i dag på ét universitet i Danmark (Københavns Universitet) og fire universiteter i Polen (Poznań, Gdańsk, Warszawa og Wrocław) samt på en række sprogskoler i begge lande. Turismen mellem Danmark og Polen har regelmæssigt fundet sted, i hvert fald siden begyndelsen af 1900'tallet – målgruppen af den første parlør mellem dansk og polsk (Nessnejw 1904) talte bl.a. danskere, der skulle besøge Polen. Alene i 2006 har man registreret i alt 242.000 danske overnatninger i Polen og 58.000 polske overnatninger i Danmark.<sup>3</sup> Danske erhvervsaktiviteter i Polen blev først for alvor skudt i gang efter kommunismens fald, men har sidenhen været stigende og meget intensiv: I 2007 var Danmark den 12. største investor i Polen, med andel i flere industrigrene, bl.a. produktion af fødevarer, tekstiler, maskiner og medicin.<sup>4</sup>

De ovennævnte kommunikationsbehov har affødt behov for adskillige typer leksikografiske produkter, for det meste ordbøger og parlører. De udkom løbende siden 1904, dog har det overvejende flertal set dagens lys gennem de sidste 20 år. De fleste værker er almenordbøger til brug ved sprogindlæring. Dertil hører bl.a. Stemann (1945 og 1949), Frank-Oborzyńska (1970), Sørensen og Widding (1989), Nowak (1995), Rajnik og Zgółkowa (2000), og Kuciewicz [uden årstal]<sup>5</sup>. Turismen har resulteret i udarbejdelsen af en række parlører, f.eks. Nessnejw (1904), Stemann (1936), Rajnik og Wójcik (1985), og Michalska og Król (2007). Behov for fagkommunikation, både forretningsmæssig og fritidsrelateret, er indtil videre udmundet i produktionen af to fagordbøger: den juridisk-merkantile ordbog ved Okoniewski (1995) og *Lille Jagtordbog* (2002).

<sup>2</sup> Kilder: *Indvandreere i Danmark 2010* og *Nyidamark.dk*.

<sup>3</sup> Kilde: *Turismen. Regionalt, nationalt og internationalt*.

<sup>4</sup> Kilde: *Inwestycje duńskie w Polsce*.

<sup>5</sup> Udgivelsesåret er ikke angivet, men ordbogen kan næppe være udkommet før årtusindskiftet. På omslaget henvises der nemlig til forlagets hjemmeside, hvilket peger på at værket først dukkede op da Internettet var udbredt.

### 3. Past

#### 3.1 Nessnejw (1904): *Polsk-Dansk Tolk med Ordbog*

*Polsk-Dansk Tolk med Ordbog* er det første leksikografiske værk med dansk og polsk. Det består af to dele: en parlør med ca. 700 ord og udtryk samt en ordbog med ca. 750 opslagsord. Ifølge forordet bestod værkets målgruppe af polske arbejdere i roemarket og teglværkerne, deres danske arbejdsgivere, og danske turister i polsktalende områder. Hvilke typer situationer bruges af *Polsk-Dansk Tolk med Ordbog* var tiltænkt, er ikke angivet. Dens målgruppe taget i betragtning kan man dog med rimelighed antage at parløren og ordbogen skulle anvendes som hjælpemiddel ved tekstreception og (især mundtlig) tekstproduktion på dansk hhv. polsk som fremmedsprog.

Ordbogsdelen består af to simple ordlister, én i hver retning, hvor hvert opslagsord efterfølges af en eller flere ækvivalenter. Andre data angives ikke. Sådan et udvalg af datatyper er i princippet godt i ordbøger til tekstreception (jf. Tarp 2006). For at hjælpe brugeren effektivt med receptionsproblemer burde lemnaudvalget dog være mere omfattende. Da både dansk og især polsk er rige i bøjningsendelser, ville det være oplagt at lemmatisere udvalgte bøjede former af opslagsordene, f.eks. uregelmæssige verber. Man savner også flerordsforbindelser og de faglige ord og udtryk som er relevante for målgruppen (de angives dog i parlørdelen). Mangel på nødvendige data – både faglige ord, men også udtaleangivelser og grammatiske oplysninger – begrænser også ordlisternes brugbarhed ved tekstproduktion. Derudover kommer det at ækvivalenterne til opslagsordene angives uden betydningsadskillelse. Det er et problem da disse ikke nødvendigvis er synonyme. En dansk bruger kunne således komme galt af sted ved at bruge ordlisten til f.eks. at udtrykke ”gaffel” på polsk. Opslagsordet *gaffel* efterfølges af to ækvivalenter, *widelec* og *widły*, men man får ikke at vide at den første af dem betegner en spisegaffel, og den anden en greb.

Parlørdelen består af tematisk arrangerede lister over ord og udtryk. Temaerne omfatter bl.a. mad og drikke, roemarksarbejde, teglværk, og rejsen med dampskib. Hver liste er opdelt i to spalter: den venstre med danske ord og udtryk, og den højre med deres ækvivalenter på polsk. Andre datatyper, først og fremmest synonyme og udtaleangivelser, forekommer sporadisk. Både udvalget af dataene og det at parløren er et værk skabt til envejskommunikation, gør at værket ikke forekommer særlig velegnet til det formål det skulle opfylde. Til tekstreception er parløren måske anvendelig i det omfang brugeren har svært ved at forstå enkelte ord, som vedkommende ser på skrift. Derimod er de mange lemmatiserede sætninger og flerordsudtryk, som er relevante i målgruppens fagkontekst, men som hovedsageligt bruges i tale, ikke umiddelbart tilgængelige, medmindre man kender de ortoepiske regler som gælder fremmed-

sproget. Den hjælp til tekstproduktion, som parløren kan tilbyde er også ret beskedene. Lige som i ordbogsdelen savner man også her nødvendige data, først og fremmest grammatiske oplysninger. Udtaleangivelserne er spredte og vilkårlige. Hvis man ud fra deres meget selektive tilstedeværelse prøver at rekonstruere den tiltænkte brugers informationsbehov, kan resultatet blive (ufrivilligt) morsomt. For eksempel anføres den polske udtale i sektionen om mad og drikke på side 17 ved kun tre ord og udtryk: *picie* ('drikke'), *kieliszek wódki* ('en snaps') og *z maselem* ('med smør'). Udtalen af disse ord er imidlertid hverken sværere eller mindre forudsigeligt for en dansker end det er tilfældet med de øvrige ord inden for samme sektion.

På et mere generelt plan er der flere ting man kan lære af *Polsk-Dansk Tolk med Ordbog*. Der er tale om en rettidig udgivelse – den var et pionerværk og den udkom præcis, mens der var brug for den, hvilket man desværre ikke kan sige om alle ordbøger mellem dansk og polsk. Det samme gælder den nøjagtige specifikation af værkets målgruppe, som også hører til sjældenheder i dansk-polsk leksikografi. Det, som står i skarp kontrast dertil, er værkets manglende tilpasning til de tiltænkte brugerprofiler. Målgruppen omfattede bl.a. polske arbejdere i roemarket, men en anseelig del af dem kunne hverken læse eller skrive (Kolstrup 2010: 39-42). De kunne derfor heller ikke sætte pris på ordbogen.

### 3.2 Stemann (1945): *Polsk-Dansk Lommeleksikon* og Stemann (1949): *Dansk-Polsk Lommeleksikon*

De første egentlige ordbøger mellem dansk og polsk er Stemanns polsk-danske ordbog fra 1945 og dens dansk-polske pendant fra 1949. Siden 1970 har de to værker været udgivet i et samlet bind. Hver af ordbøgerne har en lemmabestand på ca. 6.000.

Målgruppen er ikke udtrykkeligt specificeret i værkerne. Ud fra udvalget og præsentationen af dataene i *Polsk-Dansk Lommeleksikon* kan man dog antage, at ordbogen er beregnet for danskere som er i færd med at lære polsk. Ordbogens omtekster er på dansk og indeholder afsnit om polsk udtale og polsk grammatik; i artiklerne angives der ordklasseangivelser kun ved (udvalgte) polske ord; de metasproglige data i artiklerne er på dansk. Alle disse data ville enten være overflødige eller uhensigtsmæssigt præsenterede for en polsk bruger. Den dansk-polske ordbog synes derimod at være lavet til både polakker, som er ved at lære dansk, og danskere, som er ved at lære polsk. Omteksterne, der er på polsk, indeholder et afsnit om dansk grammatik. I artiklerne angives der bøjningsformer ved danske substantiver og verber, samt oplysninger om danske verbtyper. Disse data er uden tvivl beregnet for polske brugere. For danske bru-

gere anføres til gengæld genus af polske substantiver og semantiske specifikationer vedrørende polske ækvivalenter af de danske opslagsord.

Da der intet står i ordbøgerne om hvad de skulle bruges til, er det svært at vurdere hvor godt de passer til deres formål. Både udvalget af dataene og deres præsentation i *Polsk-Dansk Lommeleksikon* tyder dog på at ordbogen ville være mest behjælpelig ved reception af polske tekster. Artiklerne består nemlig af det polske opslagsord og dets danske ækvivalent(er); kun hvis opslagordet er et substantiv, anføres der også genus, og ved udvalgte verber angives der også perfektive præfikser. Lemmaselektionen er betydeligt mere effektiv end hos Nessnejw (1904), da opslagsordene tæller både grundformer og udvalgte bøjede former, f.eks. *bielszy* ('hvidere'), samt flerordsforbindelser, f.eks. *aparāt fotograficzny* ('fotografiapparat'). Genusangivelserne virker derimod overflødige. De ville være brugbare i en ordbog beregnet til tekstproduktion, men sådan en burde også indeholde andre data, som udebliver fra Stemanns polsk-danske ordbog, så som kollokationer og bøjningsendelser ved verber og substantiver.

Artiklerne i *Dansk-Polsk Lommeleksikon* indeholder lidt flere datatyper end man ser i den polsk-danske ordbog, men det synes at afspejle ordbogens brede målgruppe snarere end en konkret situation, som værket menes at skulle bruges i. En typisk artikel består af det danske opslagsord, eventuelt angivelse af dets genus og pluralis (ved substantiver) eller verbtype (ved verber), samt dets polske ækvivalent(er) og genus- og perfektformangivelser af disse (ved hhv. substantiver og verber). Disse data er ikke tilstrækkelige til, at ordbogen kan yde optimal hjælp ved tekstproduktion, uanset om denne skulle være på dansk eller på polsk. Man ville især savne kollokationer og eksempler, som illustrerer opslagsordenes brug. Samtidig er mange af dataene ikke særlig relevante i en ordbog til tekstreception.

Stemanns værker ligner eksempler på one-size-fits-all ordbøger, hvor ideen kunne være blevet gennemført mere effektivt. Uklarheden om ordbøgernes målgruppe og om deres genuine formål afspejles i tvivlsom dataselektion. Ordbøgerne indeholder for få data, som brugeren vil efterlyse ved tekstproduktion, men samtidig finder man data, som er overflødige ved tekstreception, især i den dansk-polske ordbog.

### 3.3 Sørensen og Widding (1989): *Dansk-Polsk Ordbog*

*Dansk-Polsk Ordbog*, med ca. 12.000 opslagsord, er indtil videre den største af slagsen. Den var en del af en serie af ordbøger udviklet af Dansk Flygtningehjælp, og var beregnet for polske flygtninge i Danmark, som havde brug for et hjælpemiddel i sprogundervisningen. Der er næppe noget tvivl om, at værket hovedsageligt henvender sig til flygtninge og indvandrere. Ordbogens første



artikler er nemlig: *a-bombe*, *a-indkomst*, *a-kasse*, *a-skat* og *a-understøttelse*. I artiklen *absolut* finder man også et af de første eksempler i ordbogen: *Det er absolut nødvendigt at kunne dansk*.

Ordbogens funktion er specificeret i forordet i indirekte (og måske lidt vidtløftige) vendinger:

Flygtninge og andre brugere har mulighed for at forberede sig hjemme. Dermed kan man undgå at spille tiden i klassen på lange og komplicerede ordforklaringer fra lærerens side og i stedet hellige sig oparbejdelsen af en aktiv sprogkompetence (s. 3.)

Det kan diskuteres om en ordbog vitterligt kan erstatte en del af lærerarbejdet, og om intensiv læsning af en ordbog er den mest effektive vej til systematisk indlæring af et aktivt ordforråd. På baggrund af ovenstående kan man dog formode at *Dansk-Polsk Ordbog* skulle hjælpe brugeren med reception og produktion af danske tekster.

Efter udvalget af dataene at dømme, synes ordbogen at være et effektivt værktøj ved tekstreception på dansk. Lemmabestanden tæller grundformer, men også bøjningsformer af uregelmæssige verber (f.eks. *adlød* og *opgav*), substantiver (f.eks. *høns* og *kræfter*) og adjektiver (f.eks. *mindre*), samt leksikaliserede komposita (f.eks. *mindretalsregering*) og flerordsforbindelser, f.eks. (*minister(en) for kulturelle anliggender*). Ved opslagsord som betegner kulturbundne begreber, f.eks. *hygge* og *hvidtøl*, angives der passende forklaringer ud over tilsvarende funktionelle ækvivalenter på polsk. Der anføres ordklasse og bøjningsendelser ved hvert opslagord, hvilket måske ikke er direkte relevant ved tekstreception, men disse data er til gengæld meget brugbare for at ordbogen kan opfylde sin anden funktion, dvs. hjælpe brugeren med tekstproduktion. Det samme gælder de mange eksempler som illustrerer bruget af opslagsordene. Hvor effektiv en dansk-polsk ordbog i det hele taget kan være som et hjælpemiddel til tekstproduktion er dog begrænset. Ideelt set skulle man, i hvert fald i sprogindlæringens begyndelsesfase, have en bilingval ordbog i den modsatte retning til dette formål.

Der er flere måder hvorpå Sørensen og Widdings værk kan bidrage til udviklingen af dansk leksikografi med polsk. For det første er ordbogens målgruppe og intendede anvendelse tydeligt specificeret. For det andet er *Dansk-Polsk Ordbog* – i hvert fald med hensyn til dataselektionen – et eksempel på en velgennemtænkt ordbog til tekstreception. For det tredje er ordbogen baseret på passende kildemateriale. Lemmalisten blev nemlig udarbejdet på baggrund af de (dengang) mest anvendte lærebøger i dansk for udlændinge.

Man kan også lære at undgå et par af ordbogens uhensigtsmæssigheder. Her kan bl.a. nævnes, at ordbogens omtekster ikke er tilpasset brugerprofilen da de kun er på dansk. Man kan også indvende, at udgivelsen måske kom en smule

for sent. I 1989 var flygtningebølgen fra Polen allerede forbi. De polakker, der ankom tidligere, kan forventes at have lært dansk inden da.

### 3.4 Nowak (1995): *Polsk-Dansk Ordbog*

Udgivet som en del af Gyldendals Store Røde Ordbøger-serien er Nowaks polsk-danske ordbog med sine ca. 70.000 opslagsord den største ordbog for sprogparrer dansk og polsk. Det vides ikke med sikkerhed hvem der skal bruge ordbogen og til hvilket formål, men ud fra indholdet at dømme henvender den sig til danskere som er i gang med at lære polsk. Alle omteksterne i ordbogen er nemlig udelukkende på dansk, og de indeholder bl.a. afsnit om polsk grammatik og polsk udtale. Inde i artiklerne angives flere slags grammatiske data om opslagsordene, så som genus, verbtype og udvalgte bøjningsendelser. De vil naturligvis kun være til gavn for en dansk bruger, mens en polsk bruger ville savne tilsvarende data vedrørende dansk. Effektiv sprogindlæring forudsætter erhvervelsen af læse-, skrive- og talekompetencer. Derfor er det kun rimeligt at vurdere, hvor bevendt *Polsk-Dansk Ordbog* kan være ved tekstreception og tekstproduktion på polsk.

Det polske ordforråd som er lemmatiseret i *Polsk-Dansk Ordbog* er særdeles omfattende, og dermed er værket en af de største bilingvale ordbøger med polsk der findes. Dette er dog ikke ensbetydende med at ordbogen kan yde optimal hjælp til at forstå polske tekster. Man kan nævne tre hovedårsager til at *Polsk-Dansk Ordbog* ikke er den perfekte ordbog til tekstreception.

For det første er tilgangen til data i ordbogen langt fra optimal. For at finde frem til den danske ækvivalent må man nemlig ofte læse sig gennem data der både er overflødige ved tekstreception og svære at afkode uden hele tiden at kigge i forkortelseslisten og grammatik-guiden. Et eksempel derpå er nedenstående artikel *pichcić*:

*pichcić* -ę, -isz *ipf. VI (pf. u-)* (Acc) T lave mad, tilberede mad, kokkerere.

Før man når til de danske ækvivalenter, skal man forbi opslagsordets bøjningsendelser i 1. og 2. person singularis præsens, forbi oplysningen om at ordets aspektform er imperfektiv, at ordet hører til sjette konjugationstype, at dets perfekte præfiks er *u-*, at det tager objekt i akkusativform, og til sidst at det hører til daglig tale. Dette er klart en lang og besværlig vej til at finde løsning på et problem med tekstforståelse.

For det andet er valget af danske ækvivalenter af de polske ord og udtryk i ordbogen ret uigennemskueligt. I det ovenstående eksempel *pichcić* er der tale om et spøgefuldt ord for at lave mad. Det er svært at finde en fuldstændig ækvi-

valent på dansk, men *kokkerere* er givetvis det ord, som med hensyn til både denotation og konnotation ligger tættest på det polske opslagsord. Det er ikke tilfældet med de to andre ækvivalenter, *lave mad* og *tilberede mad*. Problemer med både tilgangen til data og valget af ækvivalenter ser man især i forbindelse med kollokationer. Dels er det ikke til at forudsige i hvilken artikel man skal slå op for at finde en konkret kollokation, og dels kan det ske at man finder forskellige ækvivalenter hvis kollokationen angives i flere artikler. For eksempel oversættes det polske *martwy inwentarz* med *bohaver* i artiklen *martwy*, men med *bygninger og redskaber* i artiklen *inwentarz*.

For det tredje er lemmaselektionen ikke optimal. Mange opslagsord er forældede (f.eks. *świszczycała* = 'en letsindig person'), de bruges meget sjældent (f.eks. *nastrojowiec* = 'stemningsmenneske') eller hører til fagsprog (f.eks. *rajzbręt* som er en faglig (og forældet) betegnelse for tegnebræt). Samtidig mangler der mere tidssvarende ordforråd, bl.a. indenfor edb og politik. Årsagen til problemet er givetvis, at det anvendte kildemateriale var temmelig forældet da ordbogen udkom. Det gælder især monolingvale ordbøger over polsk, f.eks. Doroszewski (1963) og Auderska et al. (1973).

Det ovennævnte problem med lemmaselektionen påvirker også ordbogens brugbarhed ved tekstproduktion på polsk. Det er svært at forestille sig, at en dansker som lærer polsk ville have brug for aktivt at anvende forældede, litterære, sjældne eller meget fagspecifikke ord på polsk. Ellers indeholder ordbogen mange data der er nyttige ved tekstproduktion på polsk. Der angives nemlig talrige kollokationer, eksempler og udførlige grammatiske data om hvert polsk opslagsord, og omteksterne indeholder grundige grammatik- og udtaleguides. Dette findes ikke i andre ordbøger mellem dansk og polsk. Ordbogen kan nok bedst bruges af danskere der kan en hel del polsk i forvejen.

Der er flere ting som man kan lære af *Polsk-Dansk Ordbog*. Det mest lærerige synes dog at være, at én og samme ordbog til både tekstproduktion og tekstreception måske ikke er nogen god idé, hvis den skal kunne bruges lige effektivt til begge formål. To separate ordbøger vil være bedst. Ellers kunne man overveje at differentiere behandlingen af opslagsordene. Artikler, der handler om forældede eller sjældne ord burde måske kun laves med henblik på at hjælpe brugeren med tekstreception. De behøver dermed ikke indeholde alle de grammatiske oplysninger, som man har brug for når man vil producere en tekst.

#### 4. Present

De nyeste almenordbøger mellem dansk og polsk er for det meste nye udgaver af ældre ordbøger. Disse har ikke forandret sig nævneværdigt, siden de så dagens

lys for første gang. De seneste udgaver af værkerne ved Stemann (fra 2009), Nowak (fra 2010) og Sørensen og Widding (fra 2009) er uændrede genoptryk af de oprindelige udgaver. Ordbogen ved Frank-Oborzyńska fra 2008, især dens polsk-danske del, har oplevet nogle sporadiske ændringer siden dens første udgave i 1970. Disse synes dog snarere at afspejle tidernes ånd end brugernes reelle behov. En sammenligning af et tilfældigt udvalg på 20 sider fra ordbogens første og sidste udgave viser nemlig følgende forskelle: I artiklen *dyktatura* ('diktatur') blev eksemplet *dyktatura proletariatu* ('proletariatets diktatur') fjernet, og på samme side blev artiklen *dyskietka* ('diskette') tilføjet; i artiklen *karta* ('kort') blev eksemplet *karta kredytowa* ('kreditkort') tilføjet; artiklen *radziecki* ('sovjetisk') blev fjernet, og på samme side blev artiklen *rajstopy* ('strømpebukser') tilføjet; eksemplet *walka o pokój* ('kampen for fred') er erstattet i den sidste udgave med *walka o byt* ('kampen for tilværelsen').

Blandt de få nye ordbøger mellem dansk og polsk, som er blevet lavet fra bunden af, er Kuciewicz's bidirektionelle ordbog den største i Polen. Begge ordbogsdele indeholder ca. 14.000 opslagsord tilsammen. Ifølge forordet er ordbogen beregnet for polakker, der vil lære dansk, og blandt forteksterne finder man en kortfattet beskrivelse af dansk udtale, det danske alfabet, og en meget kort kommentar om de danske substantivers genus. Disse data ville måske være til gavn for de brugere der har behov for hjælp til tekstproduktion på dansk. Løsningen kan dog kun virke efter hensigten, hvis de oplysninger som findes i ordbogsartiklerne er korrekte og hensigtsmæssigt præsenterede. Dette synes umiddelbart ikke at være tilfældet. Både opbygningen og indholdet af ordbogsartiklerne er nemlig ret betænkelige. På hver ordbogsside finder man tre parallelle lister. Den første består af opslagordene, i den midterste angives deres danske ækvivalenter – kun én for hvert opslagsord, mens den sidste liste viser de danske ords udtale i en slags forenklet udtaleskrift. Et kort uddrag fra side 40, som illustrerer ordbogens største skavanker vises nedenfor:

hańba	en vanære	[ˈvanere]
haracz	en tribut	[ˈtriˈbut]
harcerz	en spejder	[ˈspajlo]
hasło	et løsen	[ˈlyysen]
hasło słownikowe	et ordbogløsen	[ˈorbowlysen]
hazard	en hasard	[haˈzard]

Ud over at ordbogen foregiver et mærkeligt indtryk om at der eksisterer 1:1-ækvivalens mellem danske og polske ord, er nogle af ækvivalenterne meget tvivlsomme. *Haracz* betegner – i hvert fald på nutidspolsk – beskyttelsespenge afkrævet af mafia og lignende organisationer, altså ingen *tribut*, som man plejede at betale til en legitim hersker. *Hasło słownikowe* betegner et opslagsord i en

ordbog, men gengives her som (sic!) *ordbogløsen*. Nogle udtaleangivelser er også tankevækkende. Uanset hvor forenklet en lydskrift man anvender, ligger [spajlo] meget langt væk fra en forståelig udtale af *spejder*. Sidst men ikke mindst er *hasard* ikke tælleligt, så genusangivelsen *en* umiddelbart foran ordet er direkte misvisende. Ordbogen kan derfor næppe bruges som et effektivt hjælpemiddel ved sprogindlæring.

## 5. Slutbemærkning: Future

En effektiv, funktional ordbog mellem dansk og polsk hører stadigvæk fremtiden til. Behov for sådan en ordbog kan mærkes hos flere slags brugere. En ny, tidssvarende ordbog til almensproglig kommunikation kunne især være til gavn for studerende af dansk i Polen og polsk i Danmark, men også for de nyindvandrede polske håndværkere og medicinsk personale, der deltager i danskundersøgelse betalt af det offentlige. De to sidstnævnte grupper har også brug for fagordbøger, lige som danske forretningsfolk i Polen og jurister og oversættere i begge lande har det.

Uanset hvad det næste dansk-polske ordbogsprojekt bliver, kan man lære af de løsninger som blev anvendt i de eksisterende ordbøger. Den funktionelle undersøgelse af dansk leksikografi med polsk foretaget i denne artikel har afdækket flere desiderata, som bør tages med i overvejelserne i fremtidige ordbogsprojekter. Ordbogens målgruppe og dens genuine formål må angives tydeligt. Udvalget af data og tilgang til dem må også tilpasses mere effektivt til de tilrænktes brugeres behov. Sidst, men ikke mindst, hvis ordbogen skal være polyfunktional, må funktionerne prioriteres, og dette må afspejles af udvalget af data i ordbogen. En ordbog, især i trykt format, som forsøger at tilfredsstille forskellige slags behov i forskellige slags situationer, kan ikke være lige effektivt i alle situationer, og er næppe en ideel løsning.

## KILDER OG LITTERATUR

- Auderska, Halina, Zofia Łempicka og Stanisław Skorupka, 1973: *Mały słownik języka polskiego*. Warszawa.
- Doroszewski, Witold, 1963: *Wielki słownik języka polskiego* i 11 bind. Warszawa.
- Frank-Oborzyńska, Elżbieta, 1970: *Słownik minimum duńsko-polski i polsko-duński*. (7. udg. i 2008). Warszawa.
- Indvandrerne i Danmark 2010. Udgivet af Danmarks Statistik: <http://www.dst.dk/publikation.aspx?cid=14846> (juli 2011).

- Inwestycje duńskie w Polsce. WPIH Kopenhaga. Ambasada Rzeczypospolitej Polskiej w Kopenhadze: [http://copenhagen.trade.gov.pl/pl/denmark/article/detail,604,Inwestycje\\_dunskie\\_w\\_Polsce.html](http://copenhagen.trade.gov.pl/pl/denmark/article/detail,604,Inwestycje_dunskie_w_Polsce.html) (juli 2011).
- Kolstrup, Søren, 2010: Polske stemmer. Polske indvandringsbølger 1892-2008. København.
- Kuciewicz, Magdalena [uden årstal]: Słownik polsko-duński i duńsko-polski. Kraków. Lille Jagtordbog Dansk-Polsk. Mały słownik myśliwski polsko-duński. Svendborg. 2002.
- Michalska, Urszula og Maja Król, 2007: Rozmówki polsko-duńskie. Warszawa.
- Nessnejw, Jul M., 1904: Polsk-Dansk Tolk med Ordbog. København.
- Nowak, Jan A., 1995: Polsk-Dansk Ordbog. (2. udg. i 2010). København.
- Okoniewski, Grzegorz, 1999: Dansk-Polsk Juridisk-Merkantil Ordbog. København.
- Nyidamark.dk – den officielle portal for udlændinge og integration: <http://www.nyidamark.dk> (juli 2011).
- Rajnik, Eugeniusz og Alina Wójcik, 1985: Mini rozmówki duńskie. Warszawa.
- Rajnik, Eugeniusz og Halina Zgółkowska, 2000: Słownik minimum polsko-duński. Polsk-dansk miniordbog. Poznań.
- Stemann, Ingeborg, 1936: Polsko-duńskie rozmówki. Polsk-dansk parlør. København.
- Stemann, Ingeborg, 1945: Polsk-Dansk Lommeleksikon. Słownik kieszonkowy polsko-duński. København.
- Stemann, Ingeborg, 1949: Dansk-Polsk Lommeleksikon. Słownik kieszonkowy duńsko-polski. København.
- Stemann, Ingeborg, 1970: Polsk-Dansk, Dansk-Polsk Ordbog. (15. opl. i 2009). København.
- Sørensen, Wanda Strange og Lili Widding, 1989: Dansk-Polsk Ordbog. (7. opl. i 2009). Herning.
- Tarp, Sven, 2006: Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi. Århus.
- Turismen. Regionalt, nationalt og internationalt. Udgivet af Danmarks Statistik. [www.dst.dk/publikation.aspx?cid=11676&ci=true&pti=3](http://www.dst.dk/publikation.aspx?cid=11676&ci=true&pti=3) (juni 2011).

*Marcin Overgaard Ptaszynski*  
 Alssundgymnasiet Sønderborg  
 m.o.ptaszynski@gmail.com

# Ein-, to- eller fleirspråkleg ordbok?

*Margunn Rauset, Anna Helga Hannesdóttir  
& Aldís Sigurðardóttir*

This article explores possibilities and challenges that electronic media represent in a collection of three dictionaries with Icelandic as a source language and Danish, Norwegian and Swedish as target languages, i.e. the *ISLEX* dictionaries. This new medium allows dictionary users to conduct searches in many ways, both in lemmas and in equivalents. The dictionaries could therefore be considered as both bi-directional and multi-functional, but it is important to emphasize that they were not developed with this in mind. This article uses this as a point of reference to discuss whether *ISLEX* should be categorized as a mono-, bi- or multi-lingual dictionary.

*Nøkkelord:* fleirspråkleg nettordbok, komplekse brukargrupper, fleksible søkjemoglegheiter, ekvivalentlakuner, multifunksjonalitet, omsette bruksdøme og frasar

## 1. Innleiing

Denne artikkelen løftar fram ulike aspekt både ved arbeidet i tre av fire mål-språksprosjekt i det nordiske samarbeidsprosjektet *ISLEX* og ved produkta – som er gratis tilgjengelege på nettet. Artikkelen gjev ein kort presentasjon av prosjekta som inngår i *ISLEX*, teoretiske og praktiske perspektiv på mål-språks-arbeidet og til slutt nokre oppsummerande konklusjonar.

## 2. *ISLEX*-prosjektet

*ISLEX* er eit fleirspråkleg, digitalt ordboksverk med islandsk som kjeldespråk og frå starten svensk, norsk (bokmål og nynorsk) og dansk som målspråk. Fire institusjonar har arbeidd med dei ulike delprosjekta sidan 2005; dette er Árni Magnússon-instituttet for islandske studiar i Reykjavík, Institutionen för svenska språket ved Göteborgs universitet, Institutt for lingvistiske, litterære og estetiske studium ved Universitetet i Bergen og Det Danske Sprog- og Litteratur-

selskab i København. I tillegg kom Fróðskaparsetur Føroya inn i prosjektet og starta det redaksjonelle arbeidet i februar 2011.

Den islandske redaksjonen har ansvar for handsaminga av kjeldespråket og den tekniske biten med utvikling av databasen og redaksjons- og brukargrensesnittet til ISLEX (jf. Jónsdóttir og Úlfarsdóttir 2008). Ordbøkene inneheld sirka 50 000 oppslagsord og tek utgangspunkt i moderne islandsk. Det blir lagt vekt på å vise eit variert utval av idiom, kollokasjonar og bruksdøme. I tillegg utnyttar ein det digitale mediet med mellom anna mange illustrasjonar og lyd, og alle lemma er utrusta med koplingar til fulle bøyingsparadigme på nettsida *Beygingarlýsing íslensks nútímamáls*, som også er utforma ved Árni Magnússon-instituttet. Kjeldespråkmaterialet er tenkt å skulle kunne fungere med kva som helst målspråk.

Målspråksredaksjonane har ansvar for kvar sitt språk. Artikkane blir utrusta med ekvivalentar, parafraserande forklaringar og omsetjingar. I november 2011 opna den første versjonen av ordboka.

### 3. Teoretiske perspektiv

Denne delen av artikkelen tek føre seg *ISLEX*-ordbøkene i eit teoretisk perspektiv: først funksjonen til ordbøkene og vidare innhaldet og statusen deira.

#### 3.1 Funksjon

Eit teoretisk grunnleggjande utgangspunkt for innhaldet og strukturen i ei ordbok er funksjonane ordboka er meint å fylle. I dei skandinaviske landa er det relativt få som lærer seg islandsk, og den islandske marknaden har heller ikkje talrike brukargrupper. Det er difor nødvendig å leggje til rette for fleire brukargrupper samstundes (Sigurðardóttir o.a. 2008: 779 f). Formålet med *ISLEX*-ordbøkene er at dei skal vere så brukbare som mogleg for så mange som mogleg: Morsmål (L1) > framandspråk (L2) for islendingar som skal produsere tekst på eit av målspråka, respektive framandspråk (L2) > morsmål (L1) for skandinav som skal avkode islandsk tekst.

Det digitale mediet gjer i tillegg at det ikkje berre er mogleg å søkje i makrostrukturen, altså dei islandske oppslagsorda. Eit søk på eit islandsk oppslagsord gjev sjølvstøtt den relevante målspråkseininga for det eller dei målspråka brukaren har valt. Søkjer ein brukar på *þorskur*, får han opp ekvivalenten *torsk*. Dessutan kan ein søkje på dei skandinaviske ekvivalentane, og ved eit fritekstsøk kan ein få tilslag på ord og ordformer i alle delar av ordboksteksten. Ved eit søk på til dømes det svenske ordet *lagom*, får ein ei liste over artikkane der *lagom* føre-



kjem som ekvivalent, det vil seie dei islandske lemma *hæfilega* og *mátulega*. Utvidar ein søket til eit fritekstsøk på *lagom*, får ein opp heile 13 artiklar der det svenske *lagom* førekjem – ikkje som ekvivalent til oppslagsordet, men der det inngår i omsetjinga av døme av ulike slag, og ein kan dermed få eit vidare bilete av ulike kontekstar ordet førekjem i.

Dette er sjølvstøtt søkjemoglegheiter som brukarane kjem til å nytte seg av. ISLEX kan dermed bli oppfatta ikkje berre som ei samling av tospråklege ordbøker, men også som ei tospråkleg, bidireksjonal og multifunksjonell ordbok: for islendingar både i produksjon og avkoding av respektive skandinaviske språk og for skandinavar både når dei skal avkode og produsere islandsk tekst. På den måten skulle alle involverte språk kunne fungere både som kjeldespråk og målspråk. Trass i at brukarane kan ha slike forventningar så lenge søkjemoglegheitene er der, har vi i arbeidet likevel vore nøydd til å i første omgang sjå islandskan som det einaste kjeldespråket og dei andre språka berre som målspråk.

### 3.2 Innhald

Presentasjonen av det islandske materialet går ut frå ein einspråkleg kjeldespråksanalyse både når det gjeld makrostrukturen og mikrostrukturen. Den einspråkleg lekseminndelinga skal kunne fungere for fleire målspråk, ikkje berre dei skandinaviske. Lemmaoppsettet er islandsk, og det er også den semantiske og strukturelle analysen som ligg til grunn for mikrostrukturen. På denne måten kan ISLEX minne om det britiske Dante-prosjektet (sjå Atkins o.a. 2010).

På kjeldespråkssida merkar ein det kontrastive perspektivet som er så sentralt i tospråkleg leksikografi, først og fremst i makrostrukturen. Det betyr mellom anna at det er teke med mange samansetjingar, til dømes er det med over 60 samansetjingar med *bill*. Ein del er gjennomsluttede morfologiske kognatar med alle dei skandinaviske ekvivalentane (som *bildekk*, *bilakaup*), medan andre, som *bilastæði* ('parkeringsplass') og *bildrusla* ('skramlekasse'), er vanskelegare å føresjå.

I mikrostrukturen kjem det kontrastive perspektivet først til syne i arbeidet i målspråksredaksjonane. På den måten kan ein snakke om ei tospråkleg ordbok mellom islandsk og det respektive målspråket. Den einspråklege kjeldespråksanalysen som målspråksredaksjonane arbeider ut frå, er naturleg nok i varierende grad tilpassa dei enskildte målspråka. Til dømes kan den semantiske analysen av islandskan frå eit målspråksperspektiv til tider resultere i ei overdriven oppdeling i ulike tydingar. Regelbundne semantiske relasjonar blir illustrerte på ein måte som ikkje alltid blir opplevde som motiverte ut frå eit enkelt målspråk, men som kan vere høgst relevante ut frå perspektivet til eit anna.

Lemmaet *hæll* kan illustrere dette. Det er delt inn i fire tydingar 1) 'hæl' (på fot) 2) 'hæl' (på sokk) og 3) 'hæl' (på sko) og 4) 'plugg' eller 'kile'. Dei tre første av desse ser ikkje heilt motiverte ut frå eit dansk og norsk perspektiv. Når det gjeld svensken, er også det tredje tydingsmomentet motivert: på svensk finn ein hæl på føter og sokkar, men ikkje på sko. Då snakkar ein om 'klack'. Den fjerde tydinga er motivert ut frå alle dei skandinaviske målspåka. Hadde *ISLEX* vore ei tradisjonell tospråkleg ordbok, hadde det kontrastive perspektivet gjort at ein kunne slått saman ulike tydingar, men her gjer formatet at alle gjer likeins.

Det som ut frå perspektivet til eitt av målspåka ser ut til å vere ein for detaljert analyse, kan vere adekvat i relasjon til eit anna målspåk. Då er det ikkje berre kjeldespråket og den kontrastive relasjonen mellom kjeldespråk og målspåk som formar mikrostrukturen i artikkelen, men også omsynet til andre målspåk. På den måten ber kvar og ei av *ISLEX*-ordbøkene preg av å ha blitt til i ein fleirspåkleg samanheng.

Også i bruksdøma og konstruksjonseksempla som det finst så mange av i kjeldespråkmaterialet, kan det einspråklege kjeldespråksperspektivet vere mindre relevant for dei ulike målspåksredaksjonane. I ei einspråkleg ordbok mynta på morsmålsbrukarar, er bruksdøma relativt uproblematisk, men dei er ikkje like enkle å handtere i ei tospråkleg ordbok. Når eit bruksdøme på eit gjeve kjeldespråk skal utformast på eit målspåk, kan det på den eine sida oppstå ein konflikt mellom ulike eigenskapar ved oppslagsordet og på den andre sida mellom oppslagsordet og bruksdømet, som det blir peika på av Hannesdóttir o.a. (2010: 141 f.):

Enligt den gängse lexikografiska traditionen är ett språkprovs viktigaste funktion att illustrera typiska drag hos uppslagsordet genom att det återges i en naturlig och representativ kontext. Det är emellertid inte givet att alla de naturliga och typiska drag som ryms i ett språkprov med enkla medel och på ett rättvisande sätt kan återges i en [mål]språksformulering. Lexikografen ställs därför inför uppgiften att prioritera bland uppslagsordets egenskaper. Valet kan stå mellan att förmedla uppslagsordets semantiska innehåll och att redovisa dess kombinatoriska, pragmatiska eller andra egenskaper.

Det kan dessutan ofte vere andre ord enn oppslagsordet som valdar problem i eit bruksdøme, og det kan drage merksemda vekk frå den illustrerande bruken av lemmaet dømet skulle ha. Eit eksempel er dei høgfrekvente islandske nemningane *frændi* og *frænka* som er vanskelege å omsetje presist til alle dei skandinaviske språka fordi vi ikkje har ei tilsvarende sekkenemning for mannlege og kvinnelege slektningar. *Amma* og *afi* er derimot berre problematiske på svensk sidan ein der alltid presiserer om ei bestemor eller ein bestefar er på mors- eller

farssida. I omsetjinga av bruksdøma blir difor leksikografen sin dugleik som omsetjar ofte sett på prøve (jf. Hellerud 2012).

### 3.3 Status

Som nemnt ovanfor, kjem dei fleksible søkjemoglegheitene til å bli utnytta, og ordbøkene kjem då til å bli brukte i samanhengar dei ikkje er konstruerte for. Eit bruksområde som ikkje inngår i konseptet, er funksjonen som fleirspråkleg ordbok. Det er ikkje noko som hindrar ein norsk brukar som vil finne ut kva noko heiter på svensk, i å bruke *ISLEX* til dette. Eit søk på den nynorske ekvivalenten *skyming* gjev fem islandske lemma: *húm*, *ljósaskipti*, *móska*, *naturhúm* og *rökkur*. Går ein inn i kvar og ein av desse artiklane, kan brukaren også få tilgang til ekvivalentane på dei andre målspåka – ekvivalentar som i forlenginga også (indirekte) er ekvivalentar til det norske ordet ein har søkt på. For brukaren kan dette vere ein interessant ekstrarfunksjon, men han må brukast med varsemd, for artiklane er ikkje produserte og kvalitetssikra for dette formålet.

## 4. Praktiske perspektiv

Denne delen omhandlar ulike sider ved redaksjonsarbeidet som illustrerer ein del av dei teoretiske perspektiva i avsnitt 3. Først kjem nokre døme på fakultative ledd i fleirordseiningar, deretter døme på korleis ordbøkene kjem høvesvis islandske og skandinaviske brukarar i møte og til slutt nokre døme på korleis særspåklege trekk blir handtererte med omsyn til syntaks, grammatisk kjønn og upersonlege og passive konstruksjonar.

### 4.1 Fakultativitet på kjeldespråket

*ISLEX* byr på eit stort utval av fleirordseiningar i form av bruksdøme, kollokasjonar og idiom, og alle desse blir utrusta med omsetjingar og ekvivalentar på målspåka. I presentasjonen av desse fleirordseiningane inngår ofte fakultative ledd, det vil seie ledd som må realiserast, men der det einskilde uttrykket kan varierast. For å markere slike ledd, brukar ein spissparentesar rundt det utskiftelege uttrykket. Det kan leie til konflikt mellom den einspråklege kjeldespråksanalysen og målspåka. Det er ofte ikkje heilt fri variasjon i det fakultative leddet, og der ein kan rekne med at islandsk brukar kjenner atterhalda, kan vi ikkje forvente at ein skandinavisk brukar gjer det.

**búa + til**

objekt: AKK

*búa til* <kerti>

- ■ lave <stearinlys>, støbe <stearinlys>
- ■ B lage <stearinlys>, støype <lys>
- ■ N laga <stearinlys>, støypa <lys>
- ■ tillverka <stearinljus>, stöpa <ljus>

Figur 1: Del av artikkelen *búa*. B=bokmål og N= nynorsk

Det islandske partikkelverbet *búa til* (fig. 1) betyr 'lage', 'framstille', 'stille til' (sv. 'tillverka', 'skapa', 'göra'). Objektet er vanlegvis konkret (*búa til sósu* ('lage saus')). I dette tilfellet består konflikten i at objektsrepresentanten *kerti* ('stearinlys') ikkje kollokerer med den nøytrale ekvivalenten til *búa til* på dansk, norsk og svensk ('lave'/'lage'/'tillverka'). *Lys støyper* ein, om ein ikkje snakkar om i stor skala, då blir det rettare å bruke *produsere*. Og mellom *búa til* og *produsere* er det berre ein marginell ekvivalensrelasjon: er det snakk om produksjon i stor skala, brukar ein ikkje *búa til*. Då står det mellom å gjere vald på islandskan eller målspråka (jf. Hannesdóttir o.a. 2010: 141 f.).

Ein annan variant av dei utskiftelege ledda er ein spissparentes med to fakultative ledd. Fraseologien rundt begge ledda er ikkje nødvendigvis parallelle i islandsk og dei respektive målspråka, og ofte vil ein måtte dele dei opp på målspråka. Kollokasjonen *listi yfir* <*þátttakendur; vöruúrvalið*> (fig. 2) kan realiserast på fleire måtar på alle målspråka, og det er felles at ein gjerne ville brukt ulike kollokasjonar avhengig av kva for eit av dei to orda i spissparentesen som er aktuelt.

**listi yfir** <*þátttakendur; vöruúrvalið*>

- ■ liste over <deltagerne; sortimentet>
- ■ <deltager>liste;
- ■ <vare>katalog
- ■ B liste over <deltakerne; vareutvalget>
- ■ B <deltaker>liste;
- ■ B oversikt over <vareutvalget>
- ■ N liste over <deltakarane; vareutvalet>
- ■ N <deltakar>liste;
- ■ N oversikt over <vareutvalet>
- ■ lista över <deltagarna, sortimentet>;
- ■ <deltagar>lista, <deltagar>förteckning;
- ■ <varu>katalog

Figur 2: del av artikkelen *listi*

I nokre tilfelle finst det kollokasjonar med to spisse parentesar og dermed to utskiftelege ledd:

*kunna* <þessu> <vel>  
 ■■ kunne lide <dette>, bryde sig om <dette> (især med nægtelse eller begrænsende udtryk)  
 ■■E like <dette>  
 ■■N lika <dette>  
 ■■ tycka <bra> om <det här>  
 ■■ gilla <det här>  
*hann kunni því illa hvernig hún talaði um vin hans*  
 ■■ han brød sig ikke om hvordan hun omtalte hans ven  
 ■■E han likte dårlig måten hun snakket om vennen hans på  
 ■■N han likte dårleg måten ho snakka om vennen hans på  
 ■■ han tyckte inte om hur hon talade om hans kompis, hon talade om hans vän på ett sätt som han ogillade

Figur 3: del av artikkelen *kunna*

Ein islandsk brukar vil vite kva som kan setjast inn i spissparentesen, men det vil ikkje vere like enkelt for ein skandinav. I bruksdømet under kollokasjonen får målspråksbrukaren presentert eit alternativ til begge parentesane, nemleg *því* og *illa*, og den islandske brukaren får også opplysningar om plasseringa av negasjon på dansk og svensk.

## 4.2 Metaspråk

Kva metaspråk ordbøkene inneheld, kjem til å variere. Ein brukar som går inn gjennom islex.da, .no eller .se, vil berre få opp sitt eige språk som målspråk, og går han inn gjennom islex.hi.is, får han opp dansk som målspråk. Brukaren må gjere eit aktivt val om han vil ha med fleire av målspråka. På denne måten kan *ISLEX*-brukarane sjølve vere med på å avgjere om dei ønskjer ei to- eller fleirspråkleg ordbok.

Mykje av omteksten endrar seg alt etter kva land sin portal ein går inn gjennom, men dette kan endrast ved eit tastetrykk inne på sida. Dette handlar om kva språk grammatiske termar kjem på, forskjellige diasystematiske markeringar, ulike fagområde eit ord høyrer til og så vidare. I tillegg kan ein snakke om to typar metaspråk i mikrostrukturen. På den eine sida får kjeldespråksbrukaren kommentarar til ein del av ekvivalentane på islandsk, medan lemma ikkje sjeldan blir avgrensa/presiserte for målspråksbrukarane på skandinavisk. For den islandske brukaren blir dermed metateksten vist på islandsk med presisering av skandinavisk ekvivalent. Eit døme er verbet *greina* som i eine tydinga har fått ekvivalentane 'identifisere' og 'diagnostisere (greina sjúkdóm)'. Avgrensinga

*greina sjúkdóm* ('diagnostisere sjukdom') er i første rekkje mynta på ein islending, for ein kan rekne med at skandinavar vil vite kor tid dei skal bruke *identifisere* og i kva samheng ein brukar *diagnostisere*. Det kan ein ikkje forvente at ein islending har kunnskapar om.

For den skandinaviske brukaren vil metateksten bli presentert på eit skandinavisk språk med avgrensing av eller som kommentar til det islandske lemmaet. Ordet *afasystir* tyder 'syster til bestefar' og blir på norsk og dansk uttrykt med *grandtante*. For å vise at det islandske ordet har ei meir avgrensa tyding enn ekvivalentane på dansk og norsk, som kan vere både 'syster til bestemor' og 'syster til bestefar', må vi for dei danske og norske brukarane tilføre ein parentes framfor ekvivalenten: "(syster til bestefar) grandtante". På svensk må ein til med ei parafraserande forklaring: "syster til morfar eller farfar".

Artikkelen til *góðgæti* (fig. 4) viser også ein av fordelane med det digitale mediet der visse opplysningar blir gjevne om ekvivalentane enten på islandske eller eit av målspåka, alt etter kva portal ein brukar eller kva brukaren sjølv ønskjer.

#### góðgæti no hk

→ BEYGING

góð-gæti

- lækkeri (oftast í fleirtölu), lækkerbiskin
- godbit, snadder, lækkerbiskin
- godbit, snadder, lækkerbiskin
- sötaker, godsaker, godis

#### góðgæti sb. neutr.

→ BØJNING

góð-gæti

- lækkeri (især i pluralis), lækkerbiskin
- godbit, snadder, lækkerbiskin
- godbit, snadder, lækkerbiskin
- sötaker, godsaker, godis

Figur 4: artikkelen til *góðgæti* i islandske og dansk utgåve

Dei to utgåvene av artikkelen er henta ut frå *islex.hi.is* og *islex.da*, og vi skal fokusere på det som står i parentes bak danske *lækkeri*. Slike diasystematiske markeringar gjev i *ISLEX* opplysningar om det skandinaviske ordet, til dømes om stilnivå, om det er fagspråk, bunde til ulike teksttypar, eller som her, om grammatiske eigenskapar. Det er den islandske brukaren som først og fremst treng denne opplysninga, dei fleste danskar vil vite at *lækkerier* oftast blir brukt i fleirtal. Samstundes kan vi tenkje oss andre brukargrupper det kan vere nyttig for å få denne opplysninga på dansk.

### 4.3 Særspåklege trekk

Den vidare teksten tek i hovudsak føre seg fire aspekt: syntaks, grammatisk kjønn, upersonlege konstruksjonar og passivkonstruksjonar. I nokre tilfelle er

dei særspåraklege trekka knytt til kjeldespråket, i andre til eitt eller fleire av målspåraka.

Ordstillinga i dei fire språka er som regel identisk i utsegnsetningar som *það er kalt úti* ('det er kaldt ute'). Ordstillinga mellom anna i leddsetningar med negasjon er lik i dei skandinaviske språka, men ikkje på islandsk. Dette bruksdømet i *ISLEX* er eit døme på det: *hann segir að það sé ekki kalt úti* ('han seier at det ikkje er kaldt ute').

Det er mange nyansar i kjeldespråket som kan vere vanskelege å formidle på målspåraka. Eit særspårakleg trekk ved islandsk samanlikna med dei skandinaviske språka, er den samansette morfologien. Eit døme er at 3. person fleirtal av personlege pronomen får ulik form etter kjønn: Er det to eller fleire hankjønn, heiter det *þeir*, to eller fleire hokjønn er *þær* og to eller fleire av begge kjønn er *þau* – og dette er oftast umogleg å få fram i omsetjingane av bruksdøma. Sjølv om det finst slike nyansar i kjeldespråket som vi ikkje får formidla på målspåraka, bør ein også vere varsam med å gje for mange opplysningar som kan bli oppfatta som unødige av store brukargrupper. Sjølv sagt er grenseoppgangane uklare for kor tid det er nødvendig med ekstraopplysningar, og det blir forsterka av at ein prøver å leggje til rette for så mange brukargrupper i ei og same ordbok.

Eit anna særtrekk ved kjeldespråket samanlikna med målspåraka, er at ein på islandsk brukar mange upersonlege konstruksjonar og passivkonstruksjonar som kan vere utfordrande å finne gode ekvivalentar til på dansk, norsk (ikkje minst på nynorsk) og svensk. Desse trekka finn vi i mange av dei islandske bruksdøma, kollokasjonane og idioma.

#### þar mætast stálin stinn

- ■ nogen sætter hårdt mod hårdt
- ■ der sættes hårdt mod hårdt
- ■ man står stejlt over for hinanden
- ■ <de> setter hardt mot hardt
- ■ <dei> set hardt mot hardt
- ■ de sætter hårt mot hårt

Figur 5: del av artikkelen *stál*

Idiom av typen *þar mætast stálin stinn* (fig. 5), skiftar sjeldan mellom aktiv og passiv (Sköldberg 2004: 247 ff.). For at ein brukar skal kunne bruke idiomatisk direkte i ein given kontekst, kan det vere freistande å gje opp ein mindre idiomatisk frase som *\*der blir hardt sett mot hardt*. For den islandske brukaren er det viktigare å få kjennskap til at ein i dei skandinaviske språka idiomatisk uttrykkjer dette med aktiv: *å setje hardt mot hardt*. Om ein set ein aktivkonstruksjon

som ekvivalent til den islandske i passiv, må brukaren ofte sjølv vende og vri litt på det. Dansk løyser dette med å både gje opp ei aktiv og ei passiv formulering.

*það kveður við* <*þrúma*>  
 ■■ man hører <tordenen> <rulle>  
 ■■ det tordner  
 ■■■ det tordner, man kan høre <tordenen> <rulle>  
 ■■■ det torar, ein kan høyra <tor> <rulla>  
 ■■ man hör <åskan> <mullra>

Figur 6: del av artikkelen *kveða* (verb)

*Það kveður við* <*þrúma*> (fig. 6) er eit døme på ein upersonleg konstruksjon, i tillegg til at det igjen viser korleis det fakultative leddet styrer omsetjinga. Å omsetje dette bruksdømet med ei formulering med ubestemt pronomen, er stilistisk meir problematisk på norsk enn på svensk og dansk, difor har ein litt ulike løysingar i målspåka. Språklege særdrag i målspåka vil ofte gjere at det blir slik. *Det torar* er det idiomatiske uttrykket på norsk, men det er ikkje stilistisk ekvivalent med *það kveður við þrúma*.

I omsetjingsteorien snakkar ein om kompensering dersom ein erstattar tap av språklege effektar ein stad i teksten ved å plassere tilsvarende effektar ein annan stad. Når målspåksredaksjonane her har lagt til ein ekstra spissparentes (<rulle>/<mullra>), er det mellom anna eit forsøk på å kompensere for tapet av stildrag i omsetjinga av *kveða* til *høyre*, og den kompetente språkbrukaren får også gjeve opp ein kollokasjon han kanskje kan dra nytte av. På den andre sida fortonar islandske vendingar seg i ein del tilfelle som høgstil for skandinavar, ikkje minst på grunn av dei passive og upersonlege konstruksjonane, og då er faren at ein kan gjere dei skandinaviske omsetjingane stivare enn det islandske utgangspunktet skulle tilseie.

## 5. Konklusjon

Den nye, digitale teknikken fører til nye funksjonar, og nye strukturar fører til nye utfordringar. Trass i at ordbøkene i ISLEX-prosjektet har blitt til både for og ved hjelp av den nye teknologien, er det den tradisjonelle leksikografiske tenkjemåten som set rammene for ordbøkene. Målspåksredaksjonane har tenkt og arbeidd med utgangspunkt i klart avgrensa funksjonar: eit kjeldespråk og separate målspåk. På den andre sida kjem dei fleksible måtane å søkje på til å bli utnytta. Når ein kan søkje på ekvivalentane, får ein tilgang til mikrostrukturen



i målspråka. Dei får på den måten funksjon av eit kjeldespråk, og lista ein då får av ekvivalentar, dannar ein falsk makrostruktur.

Målspråksord som ikkje blir aktualiserte i ordboksteksten, kan ein umogleg finne, dei kjem snarare til å glimre med fráværet sitt. Semantisk tunge ord som per i dag manglar i mikrostrukturen, er til dømes *hierarki*. Ord som *konsistens* og *integritet* blir berre delvis handsama. *Identitet* er med i artikkelen til *sjálfsmynd*, men ikkje den tydinga som har med namnet og eventuelt stillinga til ein person å gjere (som i ”kan du bekrefte identiteten din?” eller ”fastslå identiteten til den avlidne”). Også meir talespråksprega ord som *skraphaug* (om bil) og *knupp* (om hjelpsam person) manglar, på lik linje med slanguttrykk som *gäddhäng* (svensk)/*grevinneheng* eller *bingovinger* (norsk)/*mormorarme* (dansk). Prøver ein brukar å søkje på desse meir eller mindre sentrale skandinaviske orda, får han med andre ord 0 treff. Dermed får vi i det nye digitale mediet det vi kan kalle – ikkje lemmalakuner – men *ekvivalentlakuner*. Dette viser at *ISLEX*-ordbøkene ikkje er utvikla som bidireksjonale ordbøker.

Svaret på spørsmålet i overskrifta må bli at *ISLEX* ikkje er ei einspråkleg, men tospråkleg og kanskje fleirspråkleg ordbok. Blant dei ordbøkene som har blitt lanserte i *ISLEX*-prosjektet, finst det framleis ikkje ei einspråkleg, islandsk ordbok; grunnlaget er lagt, men den semantiske beskrivinga er ufullstendig. Det har heller ikkje vore eit mål at *ISLEX* skulle fungere som ei einspråkleg islandsk ordbok. Tydingsinndelinga er delvis motivert ut frå at eit eller fleire av målspråka kan ha trong for det, men ikkje ut frå ein kontrastiv analyse som er det vanlege i tospråkleg leksikografi. Mellom islandsk og dei respektive målspråka kjem dei tospråklege ordbøkene forhåpentlegvis med tid og stunder til å bli kompletterte – med omsyn til fleire av dei funksjonane og finessane som det digitale mediet byr på. Der er vi ikkje heilt enno.

## KJELDER OG LITTERATUR

- Atkins, B. T. Sue, Adam Kilgarriff and Michael Rundell, 2010: Database of ANalysed Texts of English (DANTE): the NEID database project. I: Anne Dykstra & Tanneke Schoonheim (red.): Proceedings of the XIV Euralex International Congress. Ljouwert. S. 549–556.
- Jónsdóttir, Halldóra og Þórdís Úlfarsdóttir, 2008: *ISLEX: Islandsk-skandinavisk webordbok*. I: Rapport från den 9:e konferensen om lexicografi i Norden. Akureyri 2007. S. 197–208.
- Hannesdóttir, Anna Helga, Jón Hilmar Jónsson och Sofia Tingsell, 2010: Mot en begreppsaserad isländsk och svensk fraseologisk ordbok. Reflektioner kring pragmatiska idiom. I: Nordiska studier i lexicografi 10. Rapport från Konferensen om lexicografi i Norden. Tammerfors 2009. S. 140–149.

- Hellerud, Ylva, 2012: Ensam fras söker kontext. Översättarens hantering av språkprov i en tvåspråkig ordbok. I: Nordiska studier i Lexikografi 11. (Skrifter utgivna av Nordiska föreningen för lexikografi 12.) Lund.
- Sigurðardóttir, Aldís, Anna Helga Hannesdóttir, Halldóra Jónsdóttir, Håkan Jansson, Lars Trap-Jensen and Þórdís Úlfarsdóttir, 2008: ISLEX. An Icelandic-Scandinavian Multilingual Online Dictionary. I: Elisenda Bernal and Janet DeCesaris (red.): Proceedings of the XIII Euralex International Congress. Barcelona 2007. S. 779–789.
- Sköldberg, Emma, 2004: Korten på bordet. Innehålls- och uttrycksmässig variation hos svenska idiom. Institutionen för svenska språket, Göteborgs universitet, Göteborg.

### *Nettsider*

- Beygingarlýsing íslensks nútímamáls [Inflectional database of contemporary Icelandic]. Reykjavík: Kristín Bjarnadóttir (red.), Stofnun Árna Magnússonar í íslenskum fræðum. <http://bin.arnastofnun.is> (október 2011)
- ISLEX: Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum 2005–2011. [www.islex.hi.is](http://www.islex.hi.is) (október 2011)
- ISLEX: København: Det Danske Sprog- og Litteraturselskab 2005–2011. [www.islex.da](http://www.islex.da) (október 2011)
- ISLEX: Göteborg: Institutionen för svenska språket, Göteborgs universitet 2005–2011. [www.islex.se](http://www.islex.se) (október 2011)
- ISLEX: Bergen: Institutt for lingvistiske, litterære og estetiske studium ved Universitetet i Bergen 2005–2011. [www.islex.no](http://www.islex.no) (október 2011)

### *Margunn Rauset*

Institutt for lingvistiske, litterære og estetiske studium, Universitetet i Bergen  
 margunn.rauset@lle.uib.no

### *Anna Helga Hannesdóttir*

Institutionen för svenska språket, Göteborgs universitet  
 anna.hannesdottir@svenska.gu.se

### *Aldís Sigurðardóttir*

Det Danske Sprog- og Litteraturselskab, København  
 as@dsl.dk

# A corpus based method for a diachronic study of the central vocabulary of New Norwegian

*Daniel Ridings & Oddrun Grønvik*

This article describes how a monitor corpus can be created from existing corpora where the texts are annotated with basic bibliographical information.

The focus is on the core vocabulary and how to document diachronic change. The implementation uses simple, readily available techniques from corpus linguistics.

A norm is defined by isolating a small vocabulary unlikely to change from text to text, consisting of the most frequent words. These words are the building blocks of language – function words and some highly frequent verbs and nouns, words essential to producing grammatical sentences. The relative frequencies of these words from one text collection to another will show minor deviations. This slight deviation is used to specify the norm. It is the ‘wobble room’.

Sub-corpora are then created of the corpus sections to be compared. Relative frequencies are produced for every word in both subsets. When the relative frequency of a word from one subset deviates more than the above-mentioned wobble room compared to the same word in the other chronological subset, something has happened. A word may be moving in or out of the language or texts in one collection may deviate drastically from texts in the other, or an event may have raised an otherwise obscure technical term to the front page.

*Key words:* monitor corpus, relative frequencies, central vocabulary, New Norwegian, corpus linguistics

Corpora are now main-stream for any dictionary project with its own editors. The methodology was set out in the COBUILD dictionary project (Sinclair 1987) and has spread throughout the world. If a dictionary project does not have a corpus, it wants one.

Another trend that goes back to John Sinclair and COBUILD is the characterization of a corpus (Sinclair 1996). Corpora have their typologies. They can be domain specific, useful in the creation of terminologies, balanced, reference,

attempting to provide a gold standard for a language, and monitor, documenting the changes in language over time. In order for a corpus to be really useful for an ambitious dictionary project there is one characteristic that subsumes most of the others. The corpus should be big. If the corpus is large in number of tokens or documents, but skewed by containing for instance too many newspaper texts, it is not big enough. It should preferably be augmented by adding other text types, rather than by cutting newspaper texts and thereby diminishing the size. If there are electronic texts easily available and they are not in the corpus, then the corpus probably needs to be bigger. Scandinavian languages are not spoiled by a wealth of digital texts like some other languages, such as English.

In the mid-nineties, corpora tended to land around 30-40 million words (Sinclair 1996). It was a comfortable size. Online concordances could be generated dynamically and the resulting concordance lines handled fairly comfortably. It was certainly easier than working one's way through the texts and creating archives of dictionary slips.

Computers continue to grow in power, storage continues to drop in price, more and more text is available electronically. Corpora grow and grow. They become so large that they cannot be handled comfortably by traditional means. A lexicographer can only digest so many concordance lines. The known facts about language get reinforced by more and more evidence, but the interesting subtle changes and aspects get drowned out in the wealth of material.

## Norsk Ordbok 2014

The beginnings of the New Norwegian corpus go back to the early 90-ies. Two factors played an important role in the beginning. The first factor was an ambitious project at the University of Oslo 1991-1997, the *Documentation Project*. Its goal was to provide an extensive information system for the Norwegian language and culture. It eventually involved several other universities in Norway, but in the present context, it was the project's focus on providing electronic research material for language studies in general and lexicography in particular that makes it important (Ore et al 1998: 89 ff.).

The second factor was a chance opportunity to test out the importance and usefulness of corpora for poorly documented languages with a short history of written literature. In 1992 representatives from the University of Norway, Christian-Emil Smith Ore and Oddrun Grønvik and two representatives from Göteborg University, Martin Gellerstam and Daniel Ridings, combined their efforts in a NUFU and SIDA financed project at the University of Zimbabwe in

Harare. The project was called the *African Languages and Lexicon Project* (ALLEX). Corpora already had a solid position in Swedish lexicography. The methodology spread to ALLEX and, indirectly, to Norsk Ordbok 2014 as from 2002. The Documentation Project had digitalized many texts for Norwegian and at least two invaluable resources for New Norwegian: Garborg's and Vinje's collected works along with some key smaller texts by Ivar Aasen.

In 2002 one of the authors of this paper, Daniel Ridings, joined the staff of the *Unit for Digital Documentation* at the University of Oslo with the explicit task of building up a New Norwegian corpus for the national dictionary project, Norsk Ordbok 2014 (Ridings 2005).

## The norm

The basic idea of this presentation is that there is a core vocabulary that does not change much from text type to text type. The assumption that the core vocabulary is fairly stable, allows us to turn to the vocabulary that actually does change. These changes can be due to a text type or terminological domain, or the change can be due to time, that is, we can investigate the vocabulary diachronically. This is an attempt to describe how Sinclair's monitor corpus can be implemented.

It goes without saying that any vocabulary that does not change significantly from a scientific text to a non-fiction text is not going to consist of content words. Content words will reflect the domain and text type, and the fact that they do change significantly from one domain to another is used in various techniques, from extracting terminology to the automatic production of keywords and summaries. Function words, the small building blocks of language, are the ones that will be found in any text or transcription of speech.

Another pointer to where we can find the words at the very core of a language comes from experience with part of speech tagging. A common tactic in part of speech tagging is to look up each word of a text in a digital lexicon, a lexicon that consists of a word forms and a list of analyses that a given form can have. That is the easy part, if there is a lexicon available. The hard part, the part where most energy is spent, is disambiguating the tags when a word form can have two analyses or more. *Record* can be a noun and *record* can be a verb. The main task of a part of speech tagger is to decide between the alternatives and assign the correct analysis to an ambiguous word based on the surrounding context.

A few words about the digital lexicon – the reference point for part of speech tagging. The more words there are in the lexicon, the better, up to a point. The

Norsk Ordbok 2014 New Norwegian corpus is fairly large, by any standards. At present, the autumn of 2011, it consists of 87,767,084 running tokens. That is the size of the corpus. Of these almost 88 million tokens, 1,863,524 are unique combinations of a word and a part of speech description. The remaining tokens need disambiguation. The token *kasta*, for example, can be a past tense form, a perfect participle, an infinitive or the definite form of *kaste* in the singular, to name a few. In order to have a lexicon to cover all the analyses of the words in the corpus, the lexicon would have to be very large. In practice, lexica for part of speech tagging are kept much smaller by making certain educated guesses. One such would be that all words with a final *-ane* can be assumed to be masculine plural nouns in the definite form until shown to be otherwise, by the context.

Those who have worked with part of speech tagging know that a lexicon does not need to be unreasonably large. They also know that a good lexicon will do most of the work for them. The lexicon available to the Norsk Ordbok 2014 New Norwegian corpus is the New Norwegian section of *Norsk ordbank* (the Norwegian Word Bank), which has close to 100 000 lemma with paradigms.

A lot of effort is put into raising the precision of part of speech tagging from 94% to 96%, but a good lexicon will reach somewhere between 85 – 90% all by itself, without fancy algorithms. Why is this?

As large as the NO 2014 New Norwegian corpus is, a lexicon would only have to contain 53 words in order to tag 40% of the whole corpus. In other words, 53 words are all that is needed in the lexicon in order to tag over 35 million words of the corpus. 172 words in the lexicon will cover 50% of the corpus, a little over 43 million words, and 592 words in the lexicon will cover 60% of the corpus. It feels like a safe hypothesis to assume that the 53 most frequent words do belong to the core vocabulary. If one million running words are added to the corpus, the assumption is that these 53 words will not show much deviation with regard to relative frequencies. The following is a list of what these 53 words were in 2009. The relative frequencies are per 100,000 words. For example, one can expect to find the word *og* 2,940 times per 100,000 words of the corpus.

og	2940.0	med	994.3	har	631.9
i	2401.7	ein	915.1	eg	618.3
det	2204.6	for	875.8	seg	558.7
er	1378.2	var	843.7	men	557.5
som	1362.9	at	837.4	om	471.5
til	1250.3	av	826.0	ho	462.4
han	1226.9	ikkje	811.3	eit	444.0
på	1131.9	å	791.7	hadde	367.1
dei	1001.4	den	662.7	ei	358.8

frå	323.1	skal	237.9	alle	165.5
vi	302.4	ut	229.1	andre	163.7
så	302.3	paa	222.3	meg	162.8
du	277.1	vart	218.8	her	159.1
der	271.2	vil	194.7	noko	152.7
kan	263.3	dette	193.1	mot	152.6
so	251.3	eller	174.0	aa	149.2
no	248.6	berre	173.2	over	147.5
etter	240.5	kom	170.6		

In May 2009 an opportunity arose to test the assumptions made here. The corpus was already quite large with a wide selection of texts from across the spectrum of New Norwegian literature. A whole year volume of *Dag og Tid* from 2006 was acquired and the question arose: what happens when you add one million words from a specific genre?

Before the new texts were added to the corpus, a list was made all the words in the corpus with their relative frequency. This list was then recreated after the one million words of newspaper texts were added. Then the two lists were compared. There is bound to be some deviation between the relative frequencies for a given word in the two lists. The problem was to figure out how much deviation was significant.

It was known that there were some words in the top of the frequency list that would be affected radically. These were words that reflected the spelling norm from the 1800's and from before 1938, when a major spelling reform was introduced. Since the texts being added were all from after the spelling reform, it was reasonable to assume that the older words would display a *worst case scenario* when it came to a deviation of their relative frequencies before and after. *Paa* and *aa* are examples of such words in the list above. The following list is slightly larger than the one above. It contains the top 76 words. This time the numbers do not represent the relative frequencies, but a measure, in per cent, of how much the relative frequencies differ from what they were before the one million words of modern newspaper texts was added. Some, in particular words from earlier orthographical norms, have, as expected, decreased (-) in relative frequency by a large measure. Others have differed by much smaller measures. The words are sorted in descending order according to how many percentage points they differ from the relative frequencies they had before the new addition to the corpus was made.

aa	-	3.514	um	-	3.385	henne	-	2.528
paa	-	3.512	mange	+	2.567	har	+	2.469
so	-	3.421	år	+	2.552	om	+	2.322

sa	-	1.941	på	+	1.018	meg	-	0.427
seier	+	1.917	dette	+	1.013	alle	+	0.420
å	+	1.853	vore	+	0.995	over	+	0.403
frå	+	1.638	i	+	0.945	to	+	0.366
gjekk	-	1.628	for	+	0.937	så	+	0.355
av	+	1.564	ei	+	0.926	slik	+	0.355
ho	-	1.529	som	+	0.867	inn	-	0.310
er	+	1.517	at	+	0.866	opp	+	0.274
hadde	-	1.465	sjølv	+	0.846	med	+	0.251
vera	-	1.444	enn	+	0.828	ikkje	+	0.238
då	+	1.333	den	+	0.790	vart	+	0.216
kjem	+	1.285	seg	-	0.775	skal	-	0.215
når	+	1.281	til	+	0.746	me	+	0.207
han	-	1.267	eg	+	0.746	og	-	0.204
må	+	1.259	eller	+	0.732	etter	-	0.194
du	-	1.254	men	+	0.663	det	+	0.151
kan	+	1.251	alt	-	0.638	fram	-	0.132
også	+	1.236	meir	+	0.626	ved	-	0.085
ein	+	1.220	noko	+	0.585	berre	+	0.041
hans	-	1.202	ha	+	0.573	kunne	+	0.036
kom	-	1.193	vi	+	0.566	mot	+	0.020
eit	+	1.152	andre	+	0.554	dei	+	0.019
denne	+	1.148	fekk	-	0.553	vil	-	0.018
få	+	1.044	der	-	0.551	ut	+	0.009
var	-	1.034	her	-	0.549	blir	+	0.001
kva	+	1.032	no	-	0.536			

As expected, it can be seen that *aa*, *paa*, *so*, and *um* have fallen the most in relative frequency. They all belong to an older orthography and will only be found in modern texts containing citations from older texts. Their frequencies have decreased from 3.3% to 3.5%. On the other end of the spectrum, the spectrum being the 76 most frequent words, there are words such as *blir*, *ut*, *vil* that have hardly changed at all. This is taken as evidence that there is a stable kernel of words in New Norwegian that are not dependent on text type or domain. One million words of newspaper text could be expected to produce a tangible slant on the vocabulary of corpus, but that is not what is found in these words. On the average, the most frequent words displayed a variation of relative frequency of 1.1%.



## Synchronic study

This figure was then used for a synchronic study, to demonstrate how a corpus can be used to identify good candidates for a bilingual or learner's dictionary of New Norwegian. As mentioned above, there are over 87 million words, or more accurately *tokens*, in the corpus. Many tokens are productive compounds, numbers, names and the like. Such words will not automatically become candidates for a dictionary, no matter how frequent they are. A simple frequency lists in descending order will still leave a substantial amount of work for the lexicographers to filter the list down to a manageable vocabulary. What would the principles be? Would they be principles that could be taught to others and principles that could be objectively applied or would they be based on subjective intuition?

It was then demonstrated how one can, by using the figures above, and settling on a maximum variation of 2.5%, isolate a group of words in the whole corpus that display the same amount of stability as the 75 most frequent words.

This criterion filters out the low frequency words. If there was a hapax in the corpus before the newspaper texts were added, and that word was found one more time in the newspapers, its increase in relative frequency would be 100%.

Two lists of words from the corpus were produced. The two lists represented the corpus vocabulary together with relative frequencies from before and after the addition of new texts. A collection of approximately 25,000 word forms (25,597) was isolated. These word forms display the same measure of stability as was defined by the top most frequent words, the absolute core vocabulary. This list would provide a very good starting point for the vocabulary of a basic dictionary of New Norwegian, based on objective criteria.

## Diachronic extension

This method can be used in many circumstances where one can think in terms of a sub-corpus. The corpus prior to the addition of new texts is a sub-corpus of the whole corpus after the texts have been added.

Each text of the New Norwegian corpus has been assigned a number of attributes. A title is one such attribute. An author is another and the date of publication is yet another. Every single text has the year of publication associated with it. This last attribute, the date, provides what is necessary to perform diachronic investigations of the vocabulary.

It is possible to create sub-corpora based on publication date. Frequency data for all texts before a given year, 1970, for example, can be created as a sub-corpus. Then frequency data for the same texts with the addition of those from

1970–1979 can be created and compared to the frequency data for the pre-1970 texts analogous to the comparison that was made above when one million words of newspaper texts were added to an existing corpus. This time, however, the comparison is not made between existing and new texts, but between subsets of the same corpus. This enables a diachronic investigation of the vocabulary without having to add the texts in chronological order.

The New Norwegian corpus is stored in an Oracle database in order to be integrated into the editing and formatting software that has been produced for the lexicographers by the *Unit for Digital Documentation* at the University of Oslo. There are four main tables:

1. One table, GRAPHWORD, for the word-forms found in the corpus.
2. One table, OCCURENCES, that ties a word-form to a position in a text, one record for each position. If a form occurs 25 times, there will be 25 rows for that form in this table.
3. A table for the text, TEI-markup and all.
4. A table for textual features, DOCUMENTS, containing a unique text number, the abbreviation used for a text in concordances, the year of publication and some other features of no interest here.

It is not necessary to create copies of the corpus when one wants to work with various sub-corpora. The following SQL query, written by Christian-Emil Smith Ore, will create a frequency list of all texts before the year 1970:

```
SELECT
    t4.word, t4.id, t3.datum, count(*) z
FROM
    occurrence t1, text t2, document t3, wordtype t4
WHERE
    (t1.t_id = t2.id) and
    t1.w_id = t4.id and
    (t2.document_id = t3.id) and
    (t3.datum < '1970')
GROUP BY
    t4.word,
    t4.id,
    t3.datum
ORDER BY z DESC
```

It takes a few seconds and a new corpus is not created, only frequency information for all texts before 1970. This frequency list is set aside and the relative frequencies are calculated.

A similar query, changing the date in the query to include texts published earlier than 1980, will create a new frequency list. This list will include all the words in the first list in addition to data for all the words from the next decade as well. This is the equivalent of having a corpus of texts published before 1970 and then adding new texts from the following decade. This list is also set aside and the relative frequencies are calculated.

These two data-sets, word lists with relative frequencies, can now be processed using the techniques described for synchronic processing above. However, instead of looking for *stable* words, a diachronic study might be more interested in the words that have *entered* written language, that is, words that display a greater increase in relative frequency than stable words are expected to have. Or one could look for words that appear to be *leaving* the language, that is, words that display a greater negative change in relative frequency than stable words are expected to have. Such a process can be described as monitoring the language. The corpus is the same, but the techniques used to process it create a monitor corpus.

In principle, one could do this processing of the corpus from the earliest texts to the most recent texts in ten year increments and study the changes in vocabulary, spelling norms and inflectional patterns across time. It would be interesting to see how rapidly the various orthographical recommendations took root.

The techniques described here could also be extended for the extraction of objectively based keywords for a text. It could be argued that a text that strongly reflects a certain subject or domain, will inevitably display differences in the vocabulary. Syntactic structures remain relatively stable across domains, but not the vocabulary. One could add a strongly domain biased text to a general language corpus and extract the words that have increased in relative frequency. In such a text, it is not unreasonable to assume that such words would point to what the text is *about*. This technique allows the text itself to declare what it is about. It is notoriously difficult for a number of individuals to assign the same keywords to the same text. The keywords individual experts chose are based on subjective criteria and the subjective opinions of several individuals do not necessarily coincide. The value of the resulting keywords is accordingly.

One could also compare two authors to each other, Garborg and Vinje, for example. Or one could take a text with an unidentified author and compare it to one by a known author. If the identified author has some idiosyncrasies when compared to others and if the unidentified work displays some of the same behavior, then the techniques could be used to point the way towards a closer inspection.

The possibilities are many. We hope that this paper has described the process in enough detail for it to be reapplied in installations with a different infrastructure.

## LITERATURE

- Ore, C.-E. S. og Kristiansen, N. 1998: Sluttrapport 1992-1997. Dokumentasjonsprosjektet. Universitetet i Oslo.
- Ridings, Daniel, 2005: Nynorskorpuset vid Norsk Ordbok 2014: Integreering med redaktions-arbete. In: Nordiske studiar i leksikografi 8, pp. 315-325.
- Sinclair, John (ed.), 1987: Looking up: An account of the COBUILD Project in lexical computing. London and Glasgow.
- Sinclair, John, 1996: EAGLES, Preliminary recommendations on Corpus Typology, <http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html>.

The ALLEX Project 1991 – 2006: See website at <http://www.edd.uio.no/allex/>  
 Norsk Ordbank 1997-: See <http://www.hf.uio.no/iln/om/organisasjon/edd/forskning/norsk-ordbank/>

*Daniel Ridings*  
 daniel@dlridings.se

*Oddrun Grønvik*  
 NO2014, Universitetet i Oslo.  
 oddrun.gronvik@iln.uio.no

# Ord i kombination

## En undersökning av några adjektiv som uttrycker 'hög grad'

*Bodil Rosqvist*

The study presented in this article attempts to investigate the semantic preferences of three Swedish adjectives when used as attributes to express 'a high degree' of a noun headword: *hård* (Eng. *hard* or *severe*), *kraftig* (Eng. *powerful*) and *stark* (Eng. *strong*). In some semantic contexts the three adjectives seem practically interchangeable, while in other contexts only one of them sounds 'correct'. For example, 'a high degree' of the word *vind* (Eng. *wind* or *breeze*) can be expressed by any of the adjectives *hård*, *kraftig* or *stark*, but 'a high degree' of the word *känsla* (Eng. *feeling*) is normally expressed only by the adjective *stark*. Through corpus searches and through organizing the head nouns into *semantic fields* I try to establish the semantic contexts in which the adjectives *hård*, *kraftig* and *stark* can be used to express 'a high degree' of a noun. The present investigation constitutes a pilot study of a larger research project on the use of intensifying Swedish adjectives and adverbs.

*Nyckelord:* adjektiv som uttrycker 'hög grad', kollokationer, semantiska fält

### 1. Inledning

Var och en som ägnar sig åt lexikografiskt arbete vet att ett ords betydelse ofta är avhängig av betydelsen av de ord som förekommer i dess närhet. I ordkombinationen *hård skare* kan ordet *hård* sägas ha sin grundbetydelse; "som icke l. bl. obetydligt ändrar sin form vid yttre tryck; som har en fast konsistens, som gör starkt motstånd mot tryck" (SAOB H 1829).<sup>1</sup> I ordkombinationen *hård vind* har ordet *hård* däremot snarast betydelsen 'mycket' eller 'i hög grad'. 'Hög grad av vind' kan också uttryckas med någon av ordförbindelserna *kraftig vind* eller *stark vind*. Adjektiven *hård*, *kraftig* och *stark* kan alltså alla ge uttryck för myckenhet och tycks ibland vara utbytbara med varandra. Detta framgår också av det

---

<sup>1</sup> Förkortningar i SAOB: l. = eller, bl. = blott (dvs. *endast*).

betydelsemoment i artikeln HÅRD som beskriver den aktuella betydelsen i SAOB (H 1838): ”i fråga om ngt som sker med kraft l. eftertryck l. i fråga om intensiv värksamhet: kraftig, stark”. Här utgör alltså orden *kraftig* och *stark* definitionen för ordet *hård*.

Att de tre här nämnda adjektiven kan ha samma betydelse av grad och förekomma i samma sammanhang betyder dock inte att de är helt synonyma och utbytbara med varandra. Ordförbindelserna *hård kritik*, *kraftig förstärkning* och *starkt skäl* är t.ex. idiomatiska uttrycksätt på svenska, men det låter lite konstigt att säga *hård förstärkning* eller *kraftigt skäl*. I en ordbok brukar vanligt förekommande semantiska sammanhang för ett visst uppslagsord antydast genom att några typiska språkliga exempel anges. Av praktiska skäl är det svårt att tydliggöra vilket ord av ett antal närsynonymer som är ”det rätta” i ett visst sammanhang. I detta arbete går jag utanför lexikografins sedvanliga ramar då jag genom korpusundersökningar och indelning i semantiska fält kartlägger de semantiska sammanhangen för *hård*, *kraftig* och *stark* då de används som adjektivattribut i kollokationer för att uttrycka hög grad av ett substantiviskt huvudord.

Med termen *kollokation* avser jag (i nära anslutning till t.ex. Malmgren 2003 och Svensén 2004) en grammatiskt sammanhängande och välformulerad ordkombination bestående av två innehållsord varav det ena (basen) kan ses som ett semantiskt huvudord och det andra (kollokatoren) som ett biord. I exemplet *hård vind* utgör t.ex. substantivet *vind* kollokationens bas, och adjektivet *hård* dess kollokator. Basen har sin grundbetydelse medan betydelsen av kollokatoren blir tydlig först i samband med huvudordet (basen). I kollokationen *hård vind* har ordet *hård* den primära funktionen att uttrycka ’hög grad’. Kollokatoren kan således ses som ett slags funktionsord med svagt betydelseinnehåll. Kollokatoren har vidare ofta begränsad kombinerbarhet och är oförutsägbar på så sätt att inga grammatiska regler förklarar varför ett visst ord föredras (vid en specifik bas) framför möjliga synonymer (se vidare Rosqvist 2011: 47 f.).

Enligt Svensén (2004: 207), som företräder vad som kan kallas ett systemorienterat synsätt, är kollokationer ”ordförbindelser som inte går att förklara med hjälp av semantiska selektionsregler”. De är oförutsägbara och ”specifika för det enskilda språket” (ibid.). I detta arbete försöker jag trots allt urskilja regelbundenheter och semantiska mönster i kollokationsbildningen. Min målsättning är att kunna redogöra för i vilka sammanhang ordet *hård* används för att uttrycka ’hög grad’ och när något av orden *stark* eller *kraftig* är att föredra. Undersökningen av *hård*, *kraftig* och *stark* utgör en pilotstudie inför en större undersökning av adjektiv och adverb som kan användas med förstärkande betydelse i svenskan.

De ordförbindelser bestående av ett adjektiv och ett efterföljande substantiv som behandlas nedan kan alla sägas vara kollokationer eftersom de innehåller

ett funktionsord som uttrycker hög grad av ett annat ord, men alla skulle kanske inte kallas kollokationer med ett traditionellt systemorienterat synsätt. Flera av ordförbindelserna är varken frekventa eller lexikaliserade. Det är de undersökta adjektivens funktion som står i fokus.

## 2. Några kollokationsstudier som rör funktionen 'hög grad'

En undersökning som bl.a. rör adjektiv som uttrycker hög grad i svenskan är Tingsells magisteruppsats (2000). Tingsell undersöker där vilka gradbestämmande kollokationer som förekommer för 45 frekventa abstrakta substantiv i några olika korpusar. Hon konstaterar att *stor*, *hög* och *stark* är de flitigast förekommande adjektiven (Tingsell 2000: 26).

Jag vill även nämna en engelsk korpusundersökning av ett förstärkande adjektiv. Partington har undersökt *semantic preferences* för det förstärkande adjektivet *sheer* genom att dela in bassubstantiven i semantiska fält (2004: 149 f.). Han urskiljer därvid följande fyra *semantic sets* för de substantiviska basorden: 1) 'magnitude/weight/volume', t.ex. *the sheer volume of reliable information*, 2) 'force/strength/energy', t.ex. *the sheer force of his presence*, 3) 'persistence', t.ex. *sometimes through sheer insistence* och 4) 'strong emotion', t.ex. *sheer joy*. Han jämför sedan användningen av *sheer* med närsynonymerna *complete*, *pure* och *absolute* och konstaterar att närsynonymerna har andra semantiska preferenser och ingår i andra syntaktiska strukturer.

## 3. Semantiska fält

Två svenska forskare som har arbetat med semantisk indelning av substantiv är Ekberg (1989) och Malmgren (2003), båda i samband med undersökningar av kollokationer bestående av verb och substantiv. Ekberg har undersökt *abstrakta övergångsfraser*, som t.ex. *falla i sömn*, medan Malmgren har undersökt verbförbindelser som anger att djupsjektet i en sats 'får till stånd' eller 'genomför' objektet i satsen (t.ex. *fatta beslut*). De har varit intresserade av att kategorisera substantiven för att kunna uttala sig om vilka semantiska sammanhang de undersökta verben används i. Den modell för semantisk indelning av bassubstantiven som Ekberg använder sig av (och som baseras på Viberg 1980) har modifierats av Malmgren (2003: 153) och ser hos honom ut som följer:

- I. Det mentala/fysiologiska huvudfältet
  - a. EMOTION. Ex. *förtvivlan, raseri*
  - b. KOGNITION. Ex. *grubblrier, medvetande*
  - c. FYSIOLOGISKT TILLSTÅND. Ex. *dvala, sömn*

## II. Huvudfältet Verksamhet/Yttre tillstånd

- a. KONFLIKT. Ex. *krig, anfall*
- b. VERBAL KOMMUNIKATION. Ex. *samtal, kritik*
- c. TVÅNG/BEROENDE. Ex. *beroende, trældom*
- d. ARBETE. Ex. *arbete, verksamhet*
- e. SOCIAL KONTAKT. Ex. *kontakt, samarbete*
- f. ATTITYD/UPPTRÄDANDE Ex. *tystnad, hövlighet*

I detta arbete har jag använt mig av ovanstående kategorisering som en första utgångspunkt. För mina ändamål har jag dock funnit det fruktbart att utöka modellen och att göra vissa förändringar. Jag redogör för min indelning i semantiska fält i avsnitt 4.3.

## 4. Material och metod

Den korpus som jag använder mig av är Parole <<http://spraakbanken.gu.se/parole/>>, en ordklassstaggad textdatabas bestående av ca 19,4 miljoner löpande ord som tillhandahålls av Göteborgs universitet. Den största delen av materialet består av dagstidningar från 1976–1997. 4,4 miljoner löpord utgörs av romaner från Bonniers (1976–1981). Vidare är 1 miljon löpord hämtade från Webb-texter och 0,4 miljoner löpord från tidskriften *Forskning och Framsteg*.

### 4.1 Korpussökningar

För att få många relevanta träffar valde jag att formulera en söksträng där aktuellt adjektiv följs av 0–1 ords lucka och vilket substantiv som helst. För adjektivet *hård* användes exempelvis söksträngen [word="hård"][]{}{0,1}[msd="N.\*"], vilket resulterade i 642 träffar. Med hjälp av sorteringsfunktionen i Parole kunde träffarna presenteras så att de ord som påträffades direkt efter det aktuella adjektivet hamnade i bokstavsordning. I de allra flesta fall hamnade de substantiv som jag var intresserad av på denna plats, vilket gav en ganska god överblick av materialet.

För former som kongruensböjts i neutrum sökte jag på aktuellt adjektiv med preciseringen att det måste vara just ett adjektiv, för att på så sätt sortera bort all adverbial användning. För *hårt* användes exempelvis söksträngen [word="hårt" & msd="A.\*"], vilket resulterade i 533 träffar. Jag har ännu inte inkluderat de adjektivformer som används i plural och bestämd form, *hårda*, *kraftiga* och *starka*, i analysen. Dessa träffar, sammanlagt 3 536 stycken har jag bara hastigt gått igenom för att se om några nya semantiska kategorier anträffas för de olika adjektiven. Så är inte fallet vad jag kan se.



## 4.2 Urval av substantiv

En del av träffarna sorterades bort direkt, exempelvis sådana där adjektivet inte stod som bestämning till substantivet, som i exemplet: ”Pressen på gruppen är extremt *hård*. *Frågan* är ... infekterad”. I övrigt gällde det att bedöma huruvida det aktuella adjektivet användes som uttryck för myckenhet av det substantiv som det bestämde och därmed skulle ingå i undersökningen. Jag använde något av uttrycken *mycket*, *i hög grad* eller *hög grad av* som ersättningsord för adjektiven för att urskilja de substantiv som var intressanta. I ordförbindelsen *hård disciplin* är *hård* t.ex. utbytbar med *hög grad av*, vilket betyder att ordet *disciplin* inkluderas i undersökningen. För de substantiv som hade mer än en träff noterade jag antalet träffar. Eftersom sökningen tillät 0–1 ords lucka mellan adjektiv och substantiv föregås orden ibland av ytterligare ett bestämmande adjektiv, t.ex. *hård politisk strid*.

För adjektivet *hård* sorterade jag bl.a. bort orden *arbetsgivare*, *attityd*, *mandelpotatis*, *tolkning* och *tresetare*. En *hård arbetsgivare* är inte *i hög grad* en *arbetsgivare*. Här passar ord som *tuff* eller *krävande* bättre som ersättningsord för adjektivet. *Hård mandelpotatis* är inte heller en *myckenhet av potatis*, utan *potatis* som har egenskapen eller karaktären att vara *hård*.

Gränsdragningen är absolut inte självklar. Är *hård argumentation* detsamma som *en myckenhet av argumentation* eller betyder adjektivet att argumentationen är *hård* och *tuff* till sin karaktär, utan bibetydelse av *grad*? Det aktuella belägget lyder som följer:

... samtalston, mycket sällan är det retorik och **hård argumentation**. – Vi, de olika ...

Jag gjorde här bedömningen att betydelsen av ordet *hård* visserligen i huvudsak är ’tuff, skoningslös’ eller dyligt, men att betydelsen ’hög grad’ också finns med. Även ordförbindelser som *hård örfil* och *hård kram* har tagits med i undersökningsmaterialet. En *hård kram* är *hård* till karaktären, men man kan också säga att den *i hög grad* är en *kram*.

## 4.3 Indelning i semantiska fält

Min semantiska kategoriseringsmodell, som baseras på Ekberg (1989) och Malmgren (2003), ser (för närvarande) ut som följer:

- A. Det mentala/fysiologiska huvudfältet
  1. EMOTION Ex. *besvikelse*, *indignation*
  2. KOGNITION Ex. *minne*, *framtidstro*
  3. FYSIOLOGISKT TILLSTÅND Ex. *abstinens*, *chock*  
sjukdom/skador Ex. *blödning*, *förkylning*

4. EGENSKAP/ATTITYD/UPPTRÄDANDE Ex. *disciplin, töntighet*  
 5. PERSON Ex. *fiende, kommunist*
- B. Verksamhet/Yttre tillstånd**
6. ARBETE/MÅLINRIKTAD HANDLING Ex. *bevakning, gödsling*  
 7. HÄNDELSE Ex. *explosion, misslyckande*  
     ökning/minskning Ex. *nedgång, ökning*  
     rörelse/accelerering Ex. *forcering, rörelse*  
 8. SOCIAL KONTAKT Ex. *sambörighet, vänskap*  
     (verbal) kommunikation Ex. *kritik, polemik*  
     konflikt/konkurrens/angrepp Ex. *attack, konfrontation*  
     politik Ex. *opinion, röst*  
     pålagor/tvång/beroende Ex. *politisk styrning, tidspress*  
 9. (YTTRE) TILLSTÅND/FÖRHÅLLANDE Ex. *belysning, trafik*  
     lukt/smak Ex. *doft, kantarellsmak*  
     väder Ex. *regnskur, vind*  
 10. KONKRET FÖREMÅL Ex. *bilbomb*

Avdelning (A) representerar människans kropp och själ, med PERSON som avslutande kategori och (B) representerar (i förhållande till människan) yttre händelser, verksamheter, tillstånd och föremål. Kategorierna PERSON och KONKRET FÖREMÅL kan betraktas som något perifera. Det krävs att substantivets betydelse innefattar en (verbal) innebörd av verksamhet eller tillstånd för att ett bestämmande adjektiv ska kunna avse 'hög grad', vilket betyder att det inte är så många substantiv avseende person eller konkret föremål som kan komma i fråga. En *kraftig man* är t.ex. inte i hög grad en man. Däremot kan en person vara t.ex. *fiende* eller *kommunist* i hög grad, vilket motiverar att kategorin inkluderar (åtminstone tills vidare).

Som Ekberg (1989: 51) påpekar är nog en viss överlappning mellan de olika kategorierna ofrånkomlig. I mitt undersökningsmaterial kan exempelvis substantiv som *livsåskådning* och *moral* föras både till KOGNITION och till EGENSKAP/ATTITYD/UPPTRÄDANDE. Det är också vad jag har valt att göra (i överensstämmelse med Ekbergs metod). Vidare kan naturligtvis homografa substantiv eller olika betydelser hos ett och samma substantiv tillhöra olika kategorier. Ordet *röst* kan t.ex. antingen vara synonymt med *stämma* och sorteras in under EGENSKAP eller så kan det hänföra sig till möjligheten för en person att rösta i val och sorteras in under "politik". Jag har inte haft möjlighet att skärskåda varanda substantiv i sitt sammanhang, vilket naturligtvis kan medföra en del felaktiga kategoriseringar. Jag hoppas dock att helhetsbilden inte ska påverkas av detta.

För substantiv som förekommer flera gånger anges antalet träffar efter orden.

## 5. Resultat

Ordet *hård/hårt* genererade 1 175 träffar, *kraftig(t)* 665 träffar och *stark(t)* 2 130 träffar i Parole-korpusen, vilket resulterade i totalt 3 970 träffar att ta ställning till.

Eftersom en fullständig resultatredovisning av samtliga kategorier i den semantiska indelningsmodellen vore alltför platskrävande har jag valt att återge träffarna för tre semantiska fält som tydliggör olikheterna i adjektivens användning: 1) fältet EMOTION, 2) underkategorin 'sjukdom/skador' i fältet FYSIOLOGISKT TILLSTÅND och 3) underkategorin 'konflikt/konkurrens/angrepp' i fältet SOCIAL KONTAKT. Träffarna är sorterade i frekvensordning i första hand och i bokstavsordning i andra hand, ibland efter sammansättningsars senare led (så att ord som *miljöstrid*, *polischefsstrid* och *slutstrid* hamnar intill varandra). Substantiv med neutralt genus följer efter övriga substantiv, föregångna av ett semikolon.

Nedan återges de substantiviska huvudord som jag har fört till fältet EMOTION i det mentala/fysiologiska huvudfältet:

### 1. EMOTION

#### **hård/hårt**

*hård förnöjsamhet*

#### **kraftig/kraftigt**

*kraftigt missnöje*

#### **stark/starkt**

*stark känsla* 37, *oro* 15, *önskan* 7, *närvaro* 6, *självkänsla* 5, *lust* 5, *längtan* 4, *formkänsla* 3, *livskänsla* 3, *medkänsla* 3, *närvarokänsla* 3, *vrede* 3, *irritation* 2, *familjekänsla* 2, *lidelse* 2, *ovilja* 2, *rädsla* 2, *spänning* 2, *tillgivenhet* 2, *ångest* 2, *attraktion*, *besvikelse*, *empati*, *entusiasm*, *fascination*, *fruktan*, *förhoppning*, *förtjusning*, *förtrostan*, *förväntan*, *homofobi*, *indignation*, *inlevelse*, *inspiration*, *kluvenhet*, *fysisk känsla*, *psykisk känsla*, *fartkänsla*, *hemkänsla*, *lustkänsla*, *njutningskänsla*, *nukänsla*, *obehagskänsla*, *overklighetskänsla*, *pulskänsla*, *revirkänsla*, *vi-känsla*, *skaparlust*, *frihetslängtan*, *medlidsamhet*, *missnöje*, *optimism*, *prestationsångest*, *sensualism*, *skepsis*, *skräck*, *skuld*, *sympati*, *upphetsning*, *upprördhet*, *moralisk upprördhet*, *uppskattning*, *fransk önskan*; *starkt hopp*

Av sammanställningen framgår att *stark* är det adjektiv som används mest frekvent inom detta semantiska fält, med baser som *känsla*, *önskan*, *längtan*, *vrede*, *lidelse*, *tillgivenhet*, *attraktion*, *empati*, *förhoppning* och *optimism*. *Känsla* är det basord som förekommer mest frekvent, med 37 träffar. Ordet förekommer också som senare led i många sammansättningar, som t.ex. *familjekänsla*, *fartkänsla*, *obehagskänsla* och *självkänsla*.

Ordet *hård* anträffas endast tillsammans med *förnöjsambet*, vilket nog kan sägas vara en tillfällig, idiosynkratisk, ordsammanställning. En Google-sökning på *hård förnöjsambet* ger inte några träffar. Även *kraftig(t)* anträffas bara tillsammans med ett enda ord i detta semantiska fält, nämligen *missnöje*. Ordförbindelsen *kraftigt missnöje* tycks förekomma då och då i svenskan (och ger 455 träffar på Google), men kollokatoren *stark* brukar trots allt föredras även vid denna bas.

I samband med substantiv betecknande känslor använder man alltså helst adjektivet *stark* för att uttrycka myckenhet. *Kraftig* kan ibland komma i fråga, medan *hård* i princip inte används inom detta område.

I kategorin 'sjukdom/skador', som ingår i fältet FYSIOLOGISKT TILLSTÅND, är det i stället ordet *kraftig* som är den helt dominerande kollokatoren:

### 3. FYSIOLOGISKT TILLSTÅND

**sjukdom/skador**

**hård/hårt**

–

**kraftig/kraftigt**

*kraftig blödning 2, förkylning 2, blåtira, hjärnskakning, huvudvärk, influensa, insöndring (i blodet av tillväxthormon), lårkaka; kraftigt näsbensbrott, sår*

**stark/starkt**

–

Av sammanställningen framgår att några exempel på ordkombinationer som förts till detta semantiska fält är *kraftig blödning, förkylning, hjärnskakning, huvudvärk, influensa* och *lårkaka*. Varken *hård* eller *stark* används inom detta semantiska fält.

I kategorin 'konflikt/konkurrens/angrepp', underavdelning till SOCIAL KONTAKT, förekommer däremot alla de tre undersökta adjektiven:

### 8. SOCIAL KONTAKT

**konflikt/konkurrens/angrepp**

**hård/hårt**

*hård konkurrens 39, kamp 17, strid 11, fight 2, uppgörelse 2, vidräkning 2, attack, dansattack, motattack, driftduell, spurtduell, fjät, inre kamp, närkamp, antinazistisk kampanj, konfrontation, patrull ("möte på hård patrull"), politisk strid, miljöstrid, polis chefsstrid, slutstrid, toppstrid, utmaning (mot ngn); hårt motstånd 30, angrepp 20, prov ("sattes på hårt prov")*

**kraftig/kraftigt**

*kraftig beskjutning 3, känga 2 (bildl.), attack; kraftigt spelövertag 4, motstånd 2, övertag 2, bombardemang*

**stark/starkt**

*stark konkurrens* 3, *mediävlar* 3, *internationell konkurrens* 2, *motsättning* 2, *beskjutning*, *konflikt*, *konfrontation*; *starkt motstånd* 18, *försvar* 15, *försvarspel* 5, *eget försvar* 2, *angrepp*, *sjöförsvar*, *totalförsvar*

Av sammanställningen framgår att *hård* är den vanligaste kollokatorn i samband med substantiv som avser konflikt, konkurrens eller angrepp. Några exempel på kollokationer är *hård konkurrens*, *kamp*, *strid*, *uppgörelse*, *attack* och *konfrontation*. Ordkombinationerna *hårt motstånd* resp. *angrepp* har också relativt många träffar.

Kollokatorn *stark* är näst vanligast och förekommer med baser som *konkurrens*, *beskjutning*, *konflikt*, *konfrontation*, *motstånd* och *försvar*. Här sammanfaller alltså flera av baserna. Tillsammans med orden *konkurrens*, *konfrontation* och *motstånd* fungerar både *hård* och *stark* som förstärkande adjektiv, även om *hård* tycks vara något mer frekvent.

Adjektivet *kraftig* fungerar också med flera av de baser som nämnts i de andra exemplen, t.ex. *attack*, *beskjutning* och *motstånd*, men *kraftig* används inte lika ofta som de två andra kollokatorerna. Ett mönster som kan skönjas här är att *hård* föredras vid baser som betecknar reciprok konflikt, som t.ex. *kamp*, *konkurrens* och *strid*, även om *stark* också är möjlig vid dessa ord. Kollokatorn *kraftig* används däremot bara med baser som betecknar *ensidig* aktion. Förutom *attack*, *beskjutning* och *motstånd* förekommer huvudord som *känga* (i uttrycket *ge ngn en kraftig känga*), *övertag* och *bombardemang*. Ibland kan användningsmönster således framträda även inom en och samma semantiska kategori (vilket kan ses som ett tecken på att kategorierna skulle kunna indelas ytterligare).

Avslutningsvis ska jag nu säga något sammanfattande om vart och ett av de tre undersökta adjektiven:

**HÅRD**

De områden där kollokatorn *hård* används mest frekvent för att uttrycka 'hög grad' är ARBETE/MÅLINRIKTAD HANDLING (med substantiv som *granskning*, *mangling* och *arbete*) och tre underkategorier till SOCIAL KONTAKT, nämligen '(verbal) kommunikation' (med t.ex. *debatt* och *kritik*), 'konflikt/angrepp' (med t.ex. *kamp*, *angrepp* och *motstånd*) och 'pålagor/tvång/beroende' (med t.ex. *press* och *tryck*). Inom underavdelningen 'konflikt/angrepp' är *hård* den dominerande kollokatorn, mer frekvent än *kraftig* och *stark*.

Kollokatorn *hård* används ytterst sparsamt i det mentala/fysiologiska huvudfältet (A). Inga substantiv anträffas i kategorin FYSIOLOGISKT TILLSTÅND och bara enstaka substantiv i övriga kategorier. Här är alltså användningen av

kollokatorn *hård* marginell. I det huvudfält som avser verksamhet och yttre tillstånd (B) återfinns också några kategorier utan träffar för adjektivet *hård*, nämligen 'ökning/minskning', SOCIAL KONTAKT, 'lukt/smak' och KONKRET FÖREMÅL. Inom dessa semantiska områden väljer man alltså inte heller *hård* som kollokator.

De kollokationer med ordet *hård* som ger flest träffar i korpusen är *hård kritik* och *hårt arbete* med 109 resp. 89 träffar, följt av *hårt tempo* med 45 träffar.

## KRAFTIG

Ordet *kraftig* har det minsta antalet träffar av de tre undersökta adjektiven, men används ändå i viss utsträckning inom nästan alla semantiska områden i indelningsmodellen. Inom området FYSIOLOGISKT TILLSTÅND är *kraftig* det adjektiv som dominerar (med substantiv som *abstinens*, *chock* och *trötthet*) och även inom detta huvudfälts underavdelning 'sjukdom/skador' väljer man helst ordet *kraftig* för att uttrycka 'hög grad' (t.ex. vid substantiv som *blåtira*, *blödning* och *förkylning*). Vad gäller 'sjukdom/skador' är *kraftig* t.o.m. den enda anträffade kollokatorn. Även inom kategorin HÄNDELSE och dess undergrupper 'ökning/minskning' och 'rörelse/accelerering' är det *kraftig* som är det mest frekventa av de adjektiv som undersökts.

Endast två kategorier saknar helt träffar för kollokatorn *kraftig* i betydelsen 'hög grad', nämligen KOGNITION under (A) och SOCIAL KONTAKT under (B). *Kraftig* används alltså inte gärna i samband med substantiv som betecknar tänkande, kunskap, tro och minne, såsom *livsåskådning* och *övertygelse*, eller med substantiv som betecknar mänskliga relationer, såsom *gemenskap* och *sambhörighet*. Vidare anträffas *kraftig* bara vid enstaka tillfällen tillsammans med substantiv i kategorierna EMOTION, PERSON och KONKRET FÖREMÅL. I kategorin KONKRET FÖREMÅL är *kraftig* med sina 3 förekomster, *bilbomb*, *hemmagjord bomb* och *sprängladdning*, dock den enda anträffade kollokatorn. Som nämns tidigare har denna kategori ganska begränsad omfattning.

De kollokationer med *kraftig* som har störst frekvens i Parole-korpusen är *kraftig ökning* med 36 träffar och *kraftig förstärkning* med 18 träffar.

## STARK

Med sina 2 130 träffar är *stark* det mest frekventa av de undersökta adjektiven och också det som förekommer i flest sammanhang. I det mentala och fysiologiska huvudfältet (A) är *stark* den dominerande kollokatorn med flest antal träffar i fyra av de fem huvudfälten. Vad gäller EMOTION (med huvudord som *besvikelse*, *empati*, ångest), KOGNITION (med huvudord som *medvetenhet* och

*misstro*) och PERSON (med t.ex. *anhängare* och *motståndare*) är *stark* i stort sett den enda anträffade kollokatorn. *Hård* och *kraftig* förekommer endast undantagsvis.

Inom huvudfält (B) är *stark* den enda anträffade kollokatorn i fältet SOCIAL KONTAKT, med huvudord som *gemenskap*, *närhet*, *relation* och *sammanhållning*. *Stark* dominerar också inom områdena '(verbal) kommunikation' och 'politik'. Några exempel på huvudord inom dessa områden är *argument* och *medhåll* resp. *konservatism* och *opinion*.

Endast inom två semantiska fält förekommer ordet *stark* inte alls, nämligen 'sjukdom/skador' (under A) och KONKRET FÖREMÅL (under B). Inom fältet 'sjukdom/skador' är användningen av *stark* ändå inte helt otänkbar. Google ger en del träffar bl.a. på *stark blödning* och *stark förkylning*. *Kraftig* är dock tveklöst den kollokator som föredras i dessa sammanhang.

*Stark* är således en kollokator som ofta fungerar ganska bra då man vill uttrycka hög grad av något. Högst antal träffar i Parole-korpusen har kollokationen *starkt stöd* med 63 träffar. *Stark* är också frekvent förekommande tillsammans med ord innehållande ordet *känsla*. Det enkla ordet *känsla* anträffas 37 gånger och ordet påträffas också i många sammansättningar, som t.ex. *familjekänsla*, *medkänsla*, *närvarokänsla* och *självkänsla*.

## 6. Avslutning

Sammanfattningsvis kan sägas att en bild har framträtt över hur de olika undersökta adjektiven används i den aktuella funktionen. Det är tydligt att det går att urskilja mönster i kollokationsbildningen. Det är dock svårt att uttala sig om *hur* kollokationsbildningen går till. Ibland motiveras valet av kollokator till en substantivisk bas tydligt av en grundbetydelse hos adjektivet. Då är betydelsen av myckenhet hos kollokatorn oftast sekundär. Så t.ex. i exemplet *hård kram*. Som nämns tidigare är en sådan kram framför allt hård till karaktären. Betydelsen 'hög grad' tillkommer som en bibetydelse. Andra gånger är det svårt att spåra anledningen till att ett visst adjektiv används. Hur kommer det sig t.ex. att endast adjektivet *stark* förekommer i samband med substantiv i fältet SOCIAL KONTAKT, som t.ex. *sambhörighet* och *vänskap*. Varför används inte ordet *kraftig* inom detta semantiska fält? Det är alltså tydligt att adjektivens grundbetydelser och substantivens semantiska tillhörighet spelar in vid valet av kollokationspartner, men att den språkliga konventionen delvis lever sitt eget liv.

Som nämns i inledningen utgör denna undersökning en pilotundersökning inför en större studie som ska omfatta ett stort antal adjektiv och adverb som används i svenskan för att uttrycka 'hög grad'. Jag planerar att inkludera bl.a.

adjektiven *bitter*, *djup*, *förhårdad*, *förkrossande*, *hög*, *riklig*, *stor*, *sträng* och *ymnig* och adverbena *blankt*, *djupt*, *enträget*, *fast* och *svårt*. Mitt mål är att tillföra kunskap om hur funktionen 'hög grad' realiseras i svenskan och att tydliggöra hur närsynonymer används i förhållande till varandra. Min arbetsmetod kan knappast komma i fråga vid utformandet av artiklar i en ordbok. Metoden är allt för tidskrävande för praktiskt lexikografiskt arbete. Undersökningsresultaten skulle däremot kunna användas som grund för en kritisk granskning av kollokationsbeskrivningen i svenska ordböcker.

## LITTERATUR

- Ekberg, Lena, 1989: *Gå till anfall och falla i sömn. En strukturell och funktionell beskrivning av abstrakta övergångsfraser*. Lund.
- Malmgren, Sven-Göran, 2003: *Begå eller ta självmord? Om svenska kollokationer och deras förändringsbenägenhet*. I: *Språk och Stil* 13 (ny följd). S. 123–168.
- Partington, Alan, 2004: *Utterly content in each other's company*. I: *Corpus-Based Language Studies*, 2006. S. 148–152.
- Rosqvist, Bodil, 2011: *Fast besluten och strängt upptagen*. Beskrivningen av kollokationer i SAOB. (Meddelanden från Institutionen för svenska språket 64.) Göteborg.
- SAOB = Ordbok över svenska språket, utgiven av Svenska Akademien, 1893 ff. Lund.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm.
- Tingsell, Sofia, 2000: *Gränslös glädje av visst värde. Några substantivs gradbestämmande kollokationer i svenska*. Göteborgs universitet: Inst. för svenska språket. (Otryckt D-uppsats.)
- Viberg, Åke, 1980: *Tre semantiska fält i svenskan och några andra språk*. Stockholms universitet: Inst. för lingvistik.

*Använda webb-adresser*

<<http://spraakbanken.gu.se/parole/>>  
[www.google.se](http://www.google.se)

*Bodil Rosqvist*

Svenska Akademiens ordbok (SAOB) och Göteborgs universitet.  
[bodil.rosqvist@svenskaakademien.se](mailto:bodil.rosqvist@svenskaakademien.se)



# *Wienerbröd* och andra finesser. Den nya svensk-litauiska ordboken

*Erika Sausverde*

The paper presents a new Swedish-Lithuanian dictionary (about 28,500 words) based on Swedish *Lexin* series. It presents a short survey of the Swedish-Lithuanian lexicographical situation, the background to the project, peculiarities and difficulties that we were confronted with while preparing the dictionary, with special attention to semantics and the problem of equivalents. The problems are typical in lexicography when compiling bilingual dictionaries and are based on different historical, cultural, social, etc. conditions of the two language regions.

*Nyckelord:* tvåspråklig lexikografi, svensk-litauiska ordböcker, Lexin-ordböcker, ekvivalenter

## Den svensk-litauiska lexikografiska situationen

Området svensk-litauisk lexikografi skämmer inte bort oss med talrika ordböcker – vare sig i det förflutna eller i moderna tider. Sedan länge finns det behov av en svensk-litauisk ordbok. Det som hittills har givits ut omfattar några parlörer samt ett par ordlistor, som exempelvis:

- De enklaste fraserna på litauiska. Trumpi švediški posakiai. R. Dambrauskaitė, redigerad av P. Åkesson. Stockholm, Innovera, 1994.
- Svensk-litauisk parlör. Švediški-lietuviški pasikalbėjimai. Parengė Virginija Jurgaitytė. Vilnius, Pradai, 1995.
- L. Dahlerus: Litauisk-svensk ordlista. Bokförlaget Instructor, 1995 (ca 7 000 uppslagsord).
- A. Birbalas, N. Valutyte, B. Jonasson, I. Ovhed: Ordlista svensk-litauisk, litauisk-svensk. Žodynėlis švedų-lietuvių, lietuvių-švedų. Blekinge FoU-enhet för socialtjänst och primärvård, BIR&D.
- I. Schmidt-Lagerholm. Svensk-litauisk pedagogisk ordbok. Vilnius, Mūsų knyga, 2004 (över 12 000 ord).
- Švedų kalbos pokalbiai ir žodynas. [Svensk-litauisk parlör.] Berlitz Publishing/Apa Publications GmbH & Co. Verlag KG, Euroknyga, 2004.

Dessutom kan man nämna den flerspråkiga *Norstedts baltiska (svenska-estniska-lettiska-litauiska) ordbok* med ”18 000 ord och fraser”, utgiven på Norstedts Akademiska Förlag, 2006.<sup>1</sup>

I de baltiska grannländerna Estland och Lettland skiljer sig den lexikografiska situationen ganska kraftigt. Även om våra grannar letterna påstår motsatsen befinner de sig i en bättre situation. Under de senaste decennierna har man givit ut en rad medelstora svensk-lettiska ordböcker (vissa ingår i en serie ordböcker som omfattar också andra nordiska språk samt finska) samt flera parlörer:

- M. Krastiņa: Svensk-Lettisk Ordbok: Zviedru-latviešu vārdnīca. Stockholm, Memento, 1992.
- M. Krastiņa: Lettisk-svensk ordbok. Stockholm, 1996 (ca 22 000 ord).
- R. Ledina, S. Elsberga: Norstedts svensk-lettiska ordbok. 1995 (ca 20 000 ord).
- E. Sanika: Lettisk-Svensk ordbok. Sanika och Harjula, 1999 (ca 30 000 ord).
- Ilze Kačevska: Latviešu-dāņu-norvēģu-zviedru-somu-krievu sarunvārdnīca. [Lettisk-dansk-norsk-svensk-finsk-rysk parlör]. Rīga, Norden AB. 1999.
- I. Kačevska, I. Balode, M. Grīnberga: Latviešu-zviedru, zviedru-latviešu vārdnīca. Lettisk-svensk, svensk-lettisk ordbok. Rīga, Norden, 1996. Pārstrādāts un papildināts izdevums. Rīga, Norden, 2004. Rīga, Zvaigzne ABC, 2009 (ca 20 000 ord).
- Ineta Balode un Ilze Kačevska: Latviešu-zviedru sarunvārdnīca. Lettisk-svensk parlör. Rīga, Norden, 2006. Rīga, Zvaigzne ABC, 2009.
- M. Krūma: Zviedru-latviešu vārdnīca/Svensk-lettisk ordbok. Rīga, Avots, 2008 (ca 10 000 ord).
- M. Krūma: Latviešu-zviedru sarunvārdnīca. Rīga, Avots, 2008.

I Estland har man länge använt *Rootsi-eesti sõnaraamat/Svensk-estnisk ordbok* från 1939 (Tartu, utarbetad av Per Wieselgren, Gustav Suits, Paul Ariste) och dess senare upplagor (1945, Stockholm; 1971 och 1976, Uppsala, bearbetad av Herbert Lagman). På 1970-talet kom en ny ordbok ut: *Rootsi-eesti sõnaraamat*, 1976, Tallinn (senaste oförändrade tryck 1996), utarbetad av Ü. Aaloe, I. Mullamaa, T. Mullamaa, A. Saluäär, J. Tuldava och E. Turmenefer. Till sist, efter ungefär 10 års arbete, gav man ut en stor svensk-estnisk ordbok omfattande ca 100 000 uppslagsord, i ett projekt lett av professor Stig Örjan Ohlsson vid Tartu universitet:

<sup>1</sup> Av andra nordisk-litauiska lexikografiska projekt i Litauen de senaste åren bör man nämna: Evalda Jakaitienė, Sturla Berg-Olsen, *Lietuvių-norvegų kalbų žodynas/Litauisk-norsk ordbok*. Vilnius, Baltos lankos, 2001 (cirka 25 000 uppslagsord med rikliga kommentarer och flertaliga exempel).

M. Aidla, S. Cousins, M. Elken, M. Kanarbik, K. Mullamaa, M. Pertšjonok, M. Praizner, R. Raag, V. Raag, J. Tuldava. *Rootsi-estli sõnaraamat. Svensk-estnisk ordbok*. Tallinn, 2004.

Den lexikografiska situationen i Estland, Lettland och Litauen återspeglar graden av historiska kontakter mellan de tre länderna och Sverige. I Litauens fall har dessa kontakter snarast varit fragmentariska. I motsats till Estland och Lettland som har haft täta historiska förbindelser med Sverige har Litauen gått andra historiska vägar. Å andra sidan har litauiskt språkmateriel alltid intagit en speciell plats inom lexikografien – vi syftar förstås då på relevanta etymologiska och historiska ordböcker – tack vare litauiskans arkaiska språkdrag. Sedan länge har litauiska intresserat språkvetare som först och främst som en ovärderlig källa inom indoeuropeisk forskning – därför hittar man talrika litauiska (och baltiska) exempel i alla relevanta etymologiska och historiska ordböcker.<sup>2</sup>

## Bakgrund

Kring 1995–96, efter flera långa samtal med Martin Gellerstam och andra lexikografer på Göteborgs Universitet, blev det bestämt att som underlag till den svensk-litauiska ordboken använda *Lexins* svensk-svenska ordbok som redan utgjorde bas för en rad tvåspråkiga större eller mindre lexikon. Trots att Lexinserien först och främst var avsedd som en ordbok för invandrare (Gellerstam 1999) och litauiska eventuella användare inte tillhörde denna målgrupp, och trots att man insåg både för- och nackdelar med att anpassa Lexin till Litauen blev de svenska kollegornas argument för att använda Lexins svenska del även för den litauiska ordboken övertygande. I början planerades en medelstor ordbok på omkring 17 000 ord motsvarande Lexins *stora* lexikon (som t.ex. det svensk-rumänska) – man önskade skriva en inte särskilt stor ordbok som man kunde ge ut på ett par år. Men när arbetet hade påbörjats blev det bestämt att försöka skriva en större ordbok – kring 28 000 ord – motsvarande Lexins så kallade lexikon i maxiserien, som Lexins svensk-engelska, svensk-ryska m. fl.

Översättningsarbetet gick i början mycket långsamt och svårt, det var ett riktigt pionjärarbete – först och främst därför att nordistiken i Litauen bara hade tagit sina första steg då och det inte fanns några erfarna specialister inom de båda språken, svenska och litauiska – för att inte tala om kvalificerade lexiko-

<sup>2</sup> Inom svensk lexikografi är det först och främst *SAOB* – i dess databas kan man nu hitta ungefär 300 litauiska, lettiska samt preussiska språkexempel inom den etymologiska delen. Detta baltiska material är ett intressant ämne *per se* (Sausverde 2010).

grafer. Arbetet har pågått med många pauser och långa avbrott.<sup>3</sup> Under tiden har behovet av svensk-litauiska och andra nordisk-litauiska ordböcker växt enormt. Vi återvände till projektet för ett par år sedan. Då konstaterade vi att den tidigare gjorda översättningen var långt ifrån perfekt och returnerade den till översättarna. Själva Lexinmaterialet hade också förändrat sig – så vi fick uppdatera det svenska basmaterialet också. Nu har Lexikaliska institutet vid Göteborgs universitet utarbetat en fjärde upplaga av Svenska ord där man betydligt har förändrat strukturen i artiklarna m.m. (Malmgren 2012). Tyvärr skulle det vara alltför stort arbete att göra om det som redan har gjorts – så i den svensk-litauiska ordboken har man hittills bevarat den version av svenska underlaget som man har haft tidigare.

## Ordbokens underlag

Eftersom vi har bevarat den struktur som redan fanns i Lexinordböckerna behövde vi inte bry oss om ordval, disposition av ordartiklar etc. Lexikonet innehåller alltså information om uppslagsordets uttal, ordklass, böjning, betydelse och litauiska ekvivalent(er). Den svenska delen inkluderar också eventuella grammatik- eller stilkommentarer, sakupplysningar om särskilda svenska realia (t.ex. *Lucia* o.d.), uppgifter om konstruktioner samt talrika språkexempel. Man har bara tagit bort de moment som inom Lexinprojektet speciellt var avsedda för invandrare, dvs. specifika kommentarer och dylikt som skulle se lite konstiga ut för en litauisk användare. I vissa fall har man lagt till eller ersatt ett eller annat språkexempel. De översatta orden i den litauiska delen får inte någon grammatisk beskrivning. Liksom andra Lexinordböcker är den blivande litauiska ordboken i första hand en inlärningsordbok, avsedd för en litauer som lär sig svenska. Om man utgår från den tvåspråkiga ordbokens funktioner (jfr Svensén 2004: 17–23) samt omformulerar Svensén är litauiskan för svenskar ett *litet* språk medan svenskan för litauerne är om inte ett *stort* så åtminstone ett *större* språk, alltså antalet litauer som behöver ordböcker mellan svenska och litauiska är större. Ur en sådan synpunkt är den blivande svensk-litauiska ordboken en L2→L1-ordbok. Den blir alltså först och främst en ordbok för reception av svensk text och för översättning av svensk text till litauiska (jfr Svensén

<sup>3</sup> Arbetsgruppen utgjordes av de första färdiga studenterna i svenska under handledning av skribenten vars modersmål är lettiska. Dessutom uppstod det tekniska problem – man fick rätt att använda själva Lexinmaterialet, dock inte det tekniska programmet för att arbeta med ordboken. Det utarbetades av en välvillig litauisk kollega. Arbetet pågick under fritiden och var ibland helt avbrutet på grund av mammaledighet.

2004: 21). Det hindrar dock inte att även de fåtaliga svensktalande som är intresserade av litauiskan kan dra en viss nytta av ordboken.

## Det tvåspråkiga perspektivet på ekvivalentfrågan

När man i vår ordboks fall använde färdigt svenskt material och helt enkelt ”bara fick översätta” svenska ord och uttryck till litauiska blev huvudmålet att överföra de svenska exemplen till litauiskan på det bästa möjliga sättet – ”om en tvåspråkig ordbok ska vara ett hjälpmedel för en översättare krävs att översättningen i ordboken är gjord med omsorg och presenterad på ett tillgängligt sätt” (Gellerstam 2004: 109). Under arbetets gång har man ständigt konfronterats med alla vanliga problem inom den tvåspråkiga lexikografin (jfr Svensén 2004, särskilt 310–340, Berkov 1996). Varje lexikograf måste lösa ett flertal problem som uppstår vid konfrontationen mellan två språkssystem eller – i vissa fall – konfrontationen mellan två skilda världsåskådningar.

Här följer nu ett urval exempel som visar vilken typ av svårigheter som har uppstått under arbetets gång. Oavsett kommentarernas preliminära karaktär (arbetet är inte färdigt än) kan man konstatera att de flesta problem gäller först och främst ekvivalenter till lemmat. Svenséns träffande anmärkning att ”skillnader när det gäller bl.a. historiska, geografiska, sociala, kulturella och ekonomiska förhållanden mellan de båda språkområdena orsakar ofta problem när det gäller kulturspecifika uttryck, liksom även språkspecifika företeelser som konnotationer, metaforer, rutinformler etc.” (Svensén 2004: 311) gäller egentligen alla nedanstående moment och exempel.

Ett problem blir ofta att översätta specifik svensk realia – att hitta en precis och elegant motsvarande formulering i litauiskan blir ibland en riktig utmaning. Vi försökte alltid komma ihåg adressaten – användaren av ordboken som skulle använda den inte bara för att förstå en svensk text utan också använda översättningen i en muntlig eller skriven form. Vår strävan var därför att hitta passande ekvivalenter och undvika beskrivningar där så är möjligt. Några exempel:<sup>4</sup>

**komvux** – *suaugusiųjų mokymas (pagal mokyklos/gimnazijos programas)* [de vuxnas undervisning (enligt grundskolans/gymnasiets utbildningsprogram)]<sup>5</sup>

**konsumentombudsman** eller **KO** – *vartotojų teisių ombudsmenas* [ombudsman för konsumentens rättigheter]<sup>6</sup>

<sup>4</sup> I exemplen utelämnas den grammatiska beskrivningen.

<sup>5</sup> I litauiskan som är ett utpräglat syntetiskt språk ska svenska sammansatta ord översättas med hjälp av genitivform(er).

<sup>6</sup> Som synes har man *ombudsman* som ett lånord i litauiskan.

Sådana svårigheter uppstår särskilt ofta vid översättningen av olika termer inom socialtjänst, socialvård, lokalförvaltning och liknande, men det kan hända även med ganska vanliga svenska ord, om de uttrycker realia som fattas i Litauen, till exempel:

**kollektiv** om grupp som bildar gemensamt hushåll (vidgad anv. sedan 1970-talet): bo i kollektiv *kolektyvas* (t.p. *komuna – viename name gyvenančių žmonių, turinčių bendrą ūkį, kolektyvas*) [kollektiv (som bor i samma hus och delar gemensamt hushåll)]

Man kunde översätta det svenska ordet till litauiskan med samma ord men fick lägga till en förklaring, för det fattas motsvarande betydelse i litauiskan.<sup>7</sup> Sannerligen, som den litauisk-norska ordbokens författare påpekar: ”ethvert språk bygger på en unik konseptualisering av verden; selv om to ord i to forskjellige språk i snever forstand betegner samme ting, er det ikke uvanlig at de to ordene uttrykker forskjellige konseptualiseringer av denne tingen” (Jakaitienė og Berg-Olsen 2004: 120 f.).

Mycket vanliga svenska ord kan bli riktiga utmaningar som kräver mycket tid och undersökningar för att hitta rätt passande motsvarighet i litauiskan, som till exempel:

**kanalbåt** – *laivas, skirtas plaukioti kanalais* [båt för färd på kanaler]

Man har i Litauen många vattendrag, men inga kanaler och som konsekvens ingen tradition av kanalbåtar (däremot paddlar man mycket i åar och sjöar), så man fick lägga till en förklaring.

**knäckebröd** – *traški duona* [frasigt bröd]<sup>8</sup>

Mat är ett ämne *per se*, för varje folk har ju sina matvanor, och även om man har likadana företeelser är det svårt att hitta motsvarande ekvivalenter som helt sammanfaller. Ett typiskt exempel är

**wienerbröd** – *sluoksniuotos teslos bandelė su įdaru* [bakverk av wienerdeg med fyllning]

<sup>7</sup> Översättarens arbete är välsignat och spännande – många ord i det tvåspråkiga perspektivet är värda större beskrivningar eller separata artiklar som i detta fall – ironiskt nog var att bo i kollektiv socialismens grundläggares dröm och ordet borde finnas i denna betydelse i de språk man talade i före detta Sovjetunionen, dock har ödet velat annorlunda.

<sup>8</sup> På nätet kallar man det ofta *traški švediška duona* [frasigt svenskt bröd], fast det finns också andra förslag till likadan realia (man kan t.ex. köpa tyskt *Knäckebröt* i litauiska affärer), så det pågår fortfarande diskussion om en passande motsvarighet.

I Sverige är det ett så vanligt bakverk. I Litauen är däremot benämningar på sådana bakverk varierande. Ur denna synpunkt är det intressant att undersöka vad man kallar en och samma sak på olika språk.<sup>9</sup>

Naturligtvis använder man under arbetets gång talrika varierande källor – från korpusar och juridiska handböcker till restaurangmenyer och Cosmopolitanskrifter samt alla möjliga ordböcker på olika språk. Ur översättarens perspektiv blir det ofta särskilt problematiskt när man hittar exakta ekvivalenter på ett annat närliggande språk medan de fattas i litauiskan; man är då tvungen att komplettera lemmat med olika slags kommentarer, som t.ex. *kompledig* [ledig som kompensation för övertidsarbete]. Det finns en exakt term i ryska, *находящийся в отгуле*, i litauiska däremot kan man inte undvara kommentarer: *turintis laisvų dienų (už išdirbtus viršvalandžius) (ypač už būdėjimo darbą)* [den som har lediga dagar (för övertidsarbete) (särskilt för jourarbete)].

En naturlig källa att jämföra litauiska exempel med är det närbesläktade lettiska språket. Ur den svensk-baltiska tvåspråkiga lexikografins synpunkt befinner sig lettiskan i en bättre situation – på grund av parallella svensk-lettiska historiska erfarenheter finns det en del gemensamma traditioner som återspeglas även i ordförrådet, t.ex. talrika gemensamma tyska lånord i svenskan och lettiskan som underlättar lexikografernas arbete. I litauiskan har man t.ex. inte någon exakt ekvivalent till

**pepparkaka** – (*kalėdinis imbirinis sausainis*) [(jul)ingefärskex]

I Litauen har man inte den traditionen<sup>10</sup>, så man måste ge en motsvarighet med förklaring, medan man i lettiskan har den exakta ekvivalenten *piparkūka*.

**kanna** – *ąsotis, indas (skysčiui)* [kanna, kärl (för vätska)]

**tekanna** – *arbatinukas*, vattenkanna – *laistytuvas*

Man måste lägga till en förklaring, för lit. *ąsotis* har inte en så bred betydelse som sv. *kanna*. Besläktade sammansatta ord får olika behandlingar i litauiska,

<sup>9</sup> Förutom i Norden och på franska har bakverket inte något namn som är relaterat till Wien – på engelska heter det *Danish pastry* eller bara *Danish* och i norra Tyskland är namnet *Kopenhagener* (i övriga Tyskland heter den *plunder*). Värt att notera är att bakverket i Köpenhamn heter *wienerbröd*, medan det i Wien heter *Kopenhagener Gebäck*. Efter skandalen med Muhammedbilderna i Jyllands-Posten gick (enligt svenska Wikipedia) iranska konditorförbundet ut med uppmaningar till sina medlemmar att de *Danish pastries* som säljs i Iran skulle döpas om till *Muhammeds rosor*.

<sup>10</sup> I stället gräddar man småkakor av mjöldeg – så kallade *kūčiukai* eller *šližikai* som man äter på julafton med så kallad *vallmomjolk* (av vatten och vallmo).

medan man i lettiska hittar exakta motsvarigheter även till dem, respektive *kanna*, *tėj(as)kanna*, *lejkanna*.

**krasse** – lit. *nasturtė* (blomma), *pipirė* (sallad); lett. *krese*

**kram, krams** – lit. *niekučiai*, *mažmožiai* [småsaker], *menkaverčiai papuošalai* [pyrdnader av ringa värde], lett. *krāms*

En särskild omständighet som gör lexikografiskt arbete i Litauen svårare än i Sverige är att litauisk språkpolitik till skillnad från den svenska är starkt normativ. Under arbetets gång har det uppstått många diskussioner om vilka motsvarande ekvivalenter man ska välja om de inte är representerade i DLKŽ – den mest auktoritativa ordboken över det moderna litauiska språket. Det kan gälla både vardagliga ord (ett typiskt exempel är *knullar* – man har motsvarande ekvivalenter i litauiska *kruštis*, *dulkintis* som inte finns med i DLKŽ, inte ens med kommentaren ”starkt vardagligt”), men även helt vanliga ord med lång användningstradition som de nämnda *kūčiukai* eller *šližikai*.

## Slutord

Frånvaron av en tvåspråkig lexikografisk tradition gör skrivandet av en svensk-litauisk ordbok svårt men samtidigt också spännande och roligt. En av de största lexikograferna, Valerij Berkov, har skrivit att ”leksikografi är både vetenskap och konst” (Berkov 1999: 4), och det har vi kunnat konstatera gång på gång. Oavsett alla svårigheter hoppas vi att kunna ge ut vår ordbok som ska omfatta ca 28 500 ord inom kort. Folk i Litauen behöver helt enkelt en svensk-litauisk ordbok!

## KÄLLOR OCH LITTERATUR

- Berkov, Valerij Pavlovič, 1996: Dvujazyčnaja leksikografija. St.Peterburg: Izdatel'stvo S.-Peterburgskogo universiteta.
- Jakaitienė, Evalda og Berg-Olsen, Sturla, 2004: Semantisk informasjon i tospråklige ordbøker (med utgangspunkt i *Litauisk-norsk ordbok*). I: LexicoNordica 11. S. 117–132.
- DLKŽ = Stasys Keinys et al. (eds.), 2000: Dabartinės lietuvių kalbos žodynas [Ordbok över det moderna litauiska språket]/Lietuvių kalbos institutas; red. kol.: St. Keinys (vyr. red.) ir kt. 4-asis leid. Vilnius: Mokslo ir enciklopedijų leidybos inst. Internet-adress: <http://www.lki.lt/dlkz/>
- Gellerstam, Martin, 1999: LEXIN – lexikon för invandrare. I: LexicoNordica 6. S. 3–18.



- Gellerstam, Martin, 2004: Översättning i och med tvåspråkiga ordböcker. I: *Lexico-Nordica* 11. S. 109–115.  
Lexin: <http://lexin2.nada.kth.se> (september 2011)
- Malmgren, Sven-Göran, 2012: Från svenska ord (Lexin) 3 till Svenska ord 4. I: *Nordiska studier i Lexikografi* 11. (Skrifter utgivna av Nordiska föreningen för lexikografi 12.) Lund.
- Sausverde, Erika, 2010: Baltų kalbų medžiaga žodyne *Svenska akademiens ordbok*. [Baltic Material in SAOB Dictionary]. I: *Baltu filologija*. XIX (1/2). S. 71–78. Rīga: LU Akadēmiskais apgāds.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Wikipedia: <http://sv.wikipedia.org/wiki/Wienerbr%C3%B6d> (september 2011)

*Erika Sausverde*

Skandinavistikos centras, Vilniaus universitetas  
erika.sausverde@flf.vu.lt

# Orientering om arbeidet med Norsk-russisk juridisk ordbok

*Aud Anna Senje*

Norway and Russia cooperate in many fields, and bilateral contact at a professional level largely depends on translators and interpreters. There is thus a great need for specialized dictionaries. This particular dictionary is a result of my twelve years of work within the Norwegian police. It provides a basic vocabulary of terms used in the fields of legal and police work, but it is of course not comprehensive. Its profile is general, and the main emphasis is on criminal law. Not all entries are strictly 'legal'. Terms that in a given context may have a specific meaning, but that are used so often in everyday language that they are perceived as 'common', are included. I have also included relevant terms from other special fields, such as finance, administration, medicine, environment, fisheries and IT.

The dictionary is mainly intended for translators, interpreters, the police, the prosecuting authority, and government officials, e.g. at Norwegian diplomatic representations in Russia and the former Soviet republics. My purpose has been to create a practical tool, rather than a work of reference with detailed information about the differences between the Norwegian and the Russian legal systems.

The supplement contains titles of selected Norwegian acts with suggested translations into Russian, and titles of selected conventions in Norwegian, Russian, and English.

*Norsk-russisk juridisk ordbok* ble utgitt på forlaget Unipub i august 2010. Dette er en av ytterst få fagordbøker mellom norsk og russisk.

Her vil jeg orientere om bakgrunnen for ordboken og litteratur og kilder jeg har brukt. Jeg vil begrunne utvalget av oppslagsord og vise konkrete eksempler på særlig problematiske termer. I tillegg vil jeg orientere om samarbeidet med fagkonsulentene og om kontakten med forlaget underveis i prosessen.

## Målgruppe og formål

Målgruppen for denne ordboken er først og fremst oversettere og tolker, ansatte og studenter ved universiteter og høyskoler, men også andre som på ett eller flere fagfelter samarbeider med russere, kan ha nytte av boken. Det kan være for eksempel forskere og fagmiljøer, departementer, norske utenriksstasjoner i Russland og tidligere sovjetrepublikker, sentrale og regionale organer innenfor offentlig forvaltning (kanskje særlig innenfor Barentssamarbeidet), politiet og påtalemyndigheten, tollvesenet, næringslivet og frivillige organisasjoner.

Mitt formål har vært å lage et praktisk verktøy, ikke et juridisk oppslagsverk med utdypende informasjon om forskjellene mellom norsk og russisk rettsystem og lovverk. Brukere som trenger slik bakgrunnsinformasjon, må derfor søke andre steder.

## Bakgrunnen for ordboken

Norge har et omfattende samarbeid med Russland på mange fagområder. Selv om engelsk stadig oftere brukes som forhandlingsspråk (mest på statsrådsnivå), er det norsk-russiske samarbeidet fortsatt i stor grad avhengig av tolker og oversettere. De fleste møter på forskjellige fagområder, enten det er på departementsnivå eller mellom underliggende etater og tjenester, foregår fremdeles på russisk og norsk med tolk.

Med andre ord er det et stort behov for fagordbøker og for en enhetlig terminologi innenfor de enkelte fagområdene. De som føler dette behovet sterkest, er oversettere og tolker. Men også fagmiljøene, undervisningsinstitusjonene og brukerne av oversetter- og tolketjenester vil være tjent med flere fagordbøker.

Av fagordbøker mellom norsk og russisk finnes fra før bare en håndfull:

Russisk-engelsk-norsk og norsk-engelsk-russisk skogordbok (2003)

Norsk-russisk militær ordbok (1993)

Russisk-norsk militær ordbok (1993)

Norsk-russisk og russisk-norsk fiskeriordbok (1986)

Norsk-russisk polyteknisk ordbok (1970)

Selv har jeg i en årrekke arbeidet med oversettelse av ulike fagtekster, både som fulltidsansatt oversetter og tolk i staten og på frilansbasis. Oversettere er generalister. Vi kan velge å spesialisere oss på ett eller flere områder. Da trenger vi kunnskap som vi ikke får gjennom et ordinært språkstudium. Ofte må vi famle og snuble oss inn i denne kunnskapen. Det er nettopp slik famling og snubling som er opphavet til denne ordboken.

## Hvorfor nettopp juridisk ordbok?

For det første kan en juridisk ordbok være nyttig også i ikke-juridisk sammenheng, siden jus overlapper mange andre fagområder. For det andre har Norge og Russland i flere år hatt et retts- og politisamarbeid. I 1998 inngikk de to landene en formell avtale om kriminalitetsbekjempelse, og Norge har utplassert to politisambandsmenn i Russland. De er kontaktpersoner for norske myndigheter, og de har behov for oversettelse av juridiske tekster både til og fra russisk. For det tredje etablerer stadig flere russere seg i Norge, og de har behov for å få oversatt blant annet dokumenter som gjelder statsborgerskap, skatt, eiendom, arv og skifte. I tillegg trenger de et begrepsapparat innenfor «hverdagsjus». For det fjerde hadde jeg som oversetter behov for et verktøy som ikke fantes, og derfor måtte jeg lage det selv. Som de fleste andre oversettere utarbeidet jeg mine egne termlister. Gjennom arbeidet mitt fikk jeg etter hvert en så stor samling av juridiske termer at jeg mente materialet burde gjøres tilgjengelig for andre oversettere. Alle er tjent med en enhetlig terminologi.

## Litteratur og kilder

I arbeidet med ordboken har jeg i stor grad basert meg på *Norsk-engelsk juridisk ordbok* av Åge Lind (2003) og valgt ut det jeg antok ville være relevant i norsk-russisk sammenheng. Videre har jeg benyttet andre tospråklige ordbøker, fagtekster og åpne kilder på Internett. I tillegg har jeg brukt avisartikler, norske og russiske dommer, tiltaler, rettsanmodninger og internasjonale konvensjoner. Dessuten har jeg brukt russiske kriminalromaner, der man kan finne nyttig informasjon som ikke finnes andre steder, blant annet intern politisjargong.

Blant de viktigste kildene mine er en dansk og en svensk ordbok. *Russisk-dansk juridisk ordbog* (Lykke Holst, Alice m.fl. 1999) inneholder både fagtermer og bakgrunnsinformasjon om det russiske rettsystemet. Utgangspunktet er russisk virkelighet, så enhver oversetter eller ordbokforfatter må vokte seg vel for å tro at den kan «speilvendes». *Svensk-rysk uppslagsordbok: juridik, förvaltning, ekonomi* (Rivelis, Eugene 1998) inneholder både termer og bakgrunnsinformasjon. Utgangspunktet er svensk virkelighet. Dette er en meget solid bok, men dessverre er den i begrenset grad anvendelig for norske oversettere, i og med at forskjellene mellom norsk og svensk forvaltningspråk er temmelig store.

## Utvalg av oppslagsord

*Norsk-russisk juridisk ordbok* har ca. 3800 oppslagsord. Jus er et omfattende fagområde, og denne ordboken er selvsagt ikke uttømmende. Det var helt nødvendig å avgrense stoffet for å gjøre utgivelsen overkommelig. Ordbokens profil er generell, med hovedvekt på terminologi fra strafferettens område og fra praktisk politiarbeid.

Gjennom tolv års arbeid i politiet har jeg tolket under avhør og i møter, og jeg har oversatt dokumenter i straffesaker og identifiseringssaker. Det dreide seg ikke bare om juridiske tekster, men også bank- og regnskapsdokumenter, obduksjonsrapporter, bryggesedler og tolldeklarasjoner, for å nevne noe. Innenfor politisamarbeidet kommer man altså inn på en rekke andre fagområder – noe som avspeiles i utvalget av oppslagsord i denne boken.

Eksempler på termer fra andre fagområder som er tatt med i boken:

- økonomi og forvaltning
  - o *anbud, bokføring, resultatregnskap, skattefradrag, tollklarering*
- medisin og psykologi
  - o *blodprøve, DNA-test, overdose, personlighetsforstyrrelse*
- miljøvern
  - o *forurensningskilde, miljøavgift, spesialavfall, utslipp, verneområde*
- fiskeri
  - o *bryggeseddel, fiskerioppsyn, fiskevernsoner, overfiske*
- IT
  - o *beskyttelsesbrudd, datavirus, hacker, phishing, sikkerhetskopier*

Termer som i en gitt kontekst kan ha et faglig innhold, men som opptrer så ofte i dagligspråket at de gjerne oppfattes som allmenne, er tatt med, f.eks. *alkoholmisbruk, bilbelte, dirk, håndvåpen, promillekjøring, skattekort, tips*. Jeg har også tatt med termer som kan være aktuelle i etterforskning av saker, f.eks. *etterforskningsskritt, fingeravtrykk, gripeenhet, konfrontasjon, overflatespor, spaning, tjenestehund*.

Utgangspunktet for ordboken er norsk virkelighet, og forklaringer og kommentarer gjelder norske forhold, også der det ikke er spesifisert. Ved hjelp av andre ordbøker og åpne nettkilder har konsulentene og jeg, så langt det er mulig, forsøkt å finne ekvivalente uttrykk på russisk. Dersom en norsk term ikke har noen tilfredsstillende ekvivalent i russisk, mener vi det er viktigere å forklare hva som ligger i den, enn å hente en kjent term fra russisk virkelighet som dekker innholdet i den norske termen kun delvis eller tilnærmet, og som følgelig kan føre til misforståelser.

## Grammatisk informasjon og trykk

Informasjon som forutsettes kjent eller som kan finnes i allmennordbøker, er ikke tatt med (f.eks. grammatisk kjønn og bøyningsmønster). Ordklasse er bare angitt ved grammatiske homonymer, eller der det er fare for forveksling. I de aller fleste tilfeller er begge aspektformer av verbet tatt med i den russiske oversettelsen.

Trykk er bare angitt der det er helt nødvendig, f.eks. hvis imperfektivt og perfektivt aspekt av verbet har samme stavemåte, og bare uttalen skiller. Eksempel: **avkorte** уреза́ть/урéзать.

## Konkrete eksempler på utfordringer i arbeidet med ordboken

### 1. Norsk term uten ekvivalent i russisk

Det hender at norske myndigheter ber om bevisopptak i Russland. Definisjonen på *bevisopptak* er «det å oppta forklaring avgitt under ed for å sikre bevis», altså rettslig avhør av et vitne før selve rettssaken. Formålet er å protokollere forklaringen og sikre den som bevis i tilfelle vitnet selv ikke kan møte under hovedforhandlingen. Noe tilsvarende praktiseres ikke i Russland, så termen må oversettes med en forklaring. Dersom den russiske termen for bevis – доказательство /dokazatel'stvo/ – blir brukt i en oversettelse av en rettsanmodning om bevisopptak, kan det føre til at norske myndigheter som ber om rettslig bistand, ikke får det de ber om.

### 2. Norsk term med delvis ekvivalent i russisk

I russisk rett skiller man ikke mellom *siktelse* og tiltale, slik man gjør i Norge. En siktelse er et foreløpig anklagedokument mot en person som påtalemyndigheten antar har begått en straffbar handling. Siktelsen utarbeides i innledningsfasen av etterforskningen. Når etterforskningen er ferdig, tar påtalemyndigheten ut *tiltale* dersom den mener at den siktede skal dømmes for den handlingen han er siktet for. (Tiltalen er et offisielt dokument; siktelsen er det ikke.) På russisk brukes bare én term for begge: обвинение /obvinenie/. *Siktelse* må dermed oversettes med samme ord som *tiltale*, med en forklaring i parentes.

### 3. Omfattende norsk term

Noen termer dekker et vidt spekter av begreper. Et eksempel på det er *overgrep*, som kan bety alt fra inngrep i rettigheter til fysisk vold. Under tvil kom konsu-

lentene og jeg fram til en inndeling i fire hovedbetydninger: 1 (*inngrep i rettigheter*) 2 (*vold*) 3 (*seksuelt misbruk*) 4 (*urettferdighet*).

#### 4. Misvisende eller ikke selvforklarende norsk term

En del norske termer er langt fra selvforklarende. Noen er til og med direkte misvisende, som *ekteskapsattest*. Denne termen kan lett bli oppfattet som *vigselsattest*, og dermed blir den oversatt feil. En ekteskapsattest er en bekreftelse på at norske myndigheter ikke har funnet noe som er til hinder for at en norsk borger gifter seg med en utenlandsk borger – altså nesten det motsatte av en vigselsattest.

En *navnemelding* er ikke en kunngjøring, slik man skulle tro. *Melding* er i denne sammenhengen en søknad (jf. *byggemelding*, som er en skriftlig søknad). En navnemelding er altså en søknad til myndighetene om å få endre etternavn eller fornavn.

#### 5. Termer som lett kan forveksles

Noen vanlige termer som ofte blir forvekslet, og dermed oversatt feil, er *asylsøker* og *flyktning*. Her er det svært viktig å få på plass standardiserte oversettelser. Utlendingsdirektoratets definisjoner:

«En *asylsøker* er en person som på egen hånd og uanmeldt ber myndighetene om beskyttelse og anerkjennelse som flyktning. Personen kalles asylsøker inntil søknaden er avgjort. *Flyktning* er et vanlig samlebegrep for alle som har flyktet fra hjemlandet, uavhengig av fluktårsak. I juridisk forstand en person som har fått innvilget politisk asyl.»

Andre eksempler på termer som lett kan forveksles, er *bortvisning* og *utvisning* (utvisning er en strengere reaksjon enn bortvisning), *dødsattest* (legeerklæring om dødsfall) og *dødsmelding* (melding til skifteretten fra de pårørende), *åstedsbefaring* (rettslig undersøkelse av et åsted, skjer under rettsforhandlingene) og *åstedsundersøkelse* (politiets tekniske undersøkelser på et åsted, skjer rett etter at en hendelse er avdekket).

#### 6. Betydningsforskyvning

Termene *gårdsnummer* og *bruksnummer* har fjernet seg fra grunnbetydningen, og de er ikke lenger selvforklarende. *Gårdsnummer* er registreringsnummeret til en landeiendom, som kan være delt opp i flere bruk. Opprinnelig var det gårdsbrukene som ble tildelt gårdsnummer. Etter hvert ble gårdsnummer innført også i byene. *Bruksnummer* er registreringsnummeret til en selvstendig fast eien-

dom. I våre dager har ikke disse numrene nødvendigvis med gårdsbruk å gjøre. Også en bygård eller en leilighet har et gårds- og bruksnummer. *Gårdsnummer* blir tidvis oversatt feil fordi oversetteren oppfatter *gård* som «bondegård».

## 7. Faste vendinger

Til jusspråket hører ikke bare fagtermer, men også en del faste vendinger som ikke så lett lar seg forklare. Et eksempel er uttrykket «gjøre gjeldende». Ordet *gjeldende* i seg selv betyr «gyldig». Men «gjøre gjeldende» betyr rett og slett å vise til, hevde, påstå, bruke som argument. «Å gjøre en mangel gjeldende» betyr nærmest å vise til eller hevde at det er en mangel ved noe, f.eks. en vare man har kjøpt, for deretter å kreve kompensasjon. Den russiske termen for *gjeldende/gyldig* – действующий, действительный /dejstuvjusjtsjij, dejstvitel'nyj/ – kan dermed ikke brukes i oversettelsen.

## Samarbeidet med konsulentene

For å kvalitetssikre manuset mitt bad jeg forlaget om å engasjere to konsulenter. Språkkonsulenten har russisk som morsmål og lang erfaring som oversetter. Den juridiske konsulenten har norsk som morsmål og behersker russisk.

Filologene (forfatteren og språkkonsulenten) og juristene (den juridiske konsulenten og hans medhjelpere) hadde nokså ulik tilnærming til stoffet. Juristene så nok for seg et oppslagsverk for jurister og ønsket derfor å legge inn omfattende og presiserende forklaringer til (eller i stedet for) flere av termene. Forklaringer kan selvsagt forsvares i tilfeller der det ikke finnes noen ekvivalent term i målspråket, men de bør ikke være for lange. Oversettere som skal bruke ordboken i arbeidet sitt, må kunne finne tilfredsstillende løsninger som fungerer i løpende tekst. For at ordboken skulle bli mest mulig funksjonell og brukervennlig, valgte jeg å sløyfe eller forkorte en del av forklaringene.

Filologene og juristene var av og til uenige om hvilke termer som var relevante å ta med. Det hendte at vi var uenige om løsninger og oversettelser. Selvsagt oppstod det også misforståelser underveis, og vi måtte stadig korrigere hverandre. Noen ganger måtte jeg rette meg etter juristenes forslag, andre ganger var det omvendt. Men vi hadde gode og nyttige diskusjoner og har forhåpentlig funnet brukbare løsninger. For meg har det vært en svært nyttig erfaring å samarbeide med fagpersoner.



## Samarbeidet med forlaget

Generelt er det svært vanskelig å skape interesse for en slik fagordbok i norske forlag. Det tok flere år å finne et forlag som våget å satse på prosjektet mitt, og dette forlaget hadde ingen erfaring med ordbokproduksjon. Jeg måtte kontrollere manuset svært nøye i alle faser og vise hvordan oppsettet skulle være. Det var en lærerik prosess både for meg og forlaget.

## Rettelser og tillegg til ordboken

Forlaget har opprettet en egen nettside der mine rettelser og tillegg er lagt ut: [www.norussjus.unipub.no](http://www.norussjus.unipub.no). Jeg tar gjerne imot tilbakemeldinger fra brukerne.

*Aud Anna Senje*  
Språkrådet, Oslo  
[aud.anna.senje@sprakradet.no](mailto:aud.anna.senje@sprakradet.no)

# Et informationsvidenskabeligt serviceeftersyn af Medicin.dk

*Henrik Køhler Simonsen*

This article discusses information science and modern metalexigraphy, arguing that both disciplines share many characteristics. On the basis of this theoretical discussion a new information-science model is introduced, and it is argued that this model encompasses the six most important dimensions in modern information science and metalexigraphy. The analysis model is then applied to Medicin.dk, which is one of Denmark's leading websites supplying health and medicinal information. The analysis of Medicin.dk is based on empirical data from a questionnaire survey involving 3,813 respondents, an interview survey comprising 11 doctors and other health care professionals, and two workshop surveys comprising 17 health care professionals. The empirical data also include data from a usability survey involving 8 citizens with different backgrounds. The analysis of the empirical data and the design and content analysis of Medicin.dk showed that the social relational needs of patients and relatives were not sufficiently met by the non-professional section of Medicin.dk called min.medicin.dk. Finally, a conceptual framework for the application of social media in a reference work such as Medicin.dk is proposed, and it is argued that social media facilitate the satisfaction of the social-relational needs of patients and relatives.

*Nøgleord:* Informationsvidenskab, metaleksikografi, stjernemodel, socialrelationelt behov, sociale medier

## 1. Indledning og problem

Opdateret og videnskabeligt baseret lægemiddelinformation er et vigtigt samfundsrelevante opgave af Dansk Lægemiddel Information A/S, som udgiver opslagsværket Medicin.dk. Denne artikel vil for det første diskutere informationsvidenskab og metaleksikografi og påvise på hvilke områder de to områder adresserer den samme type af problemer og løsninger. Jeg vil i artiklen desuden introducere den informationsvidenskabelige stjernemodel, som vil blive anvendt som analysemodel i diskussionen af Medicin.dk. Artiklen vil for det

tredje diskutere Medicin.dk og påpege især et område, hvor Medicin.dk med fordel kan forbedres. Endelig vil artiklen kort diskutere et koncept til anvendelse af sociale medier i opslagsværker som Medicin.dk med henblik på i højere grad at understøtte patienters og pårørendes socialrelationelle behov.

## 2. Leksikografi og informationsvidenskab

Informationsvidenskab og moderne leksikografi kan sammenlignes på flere områder. Popper (1989: 67) taler ikke om videnskaber, men mere om hvilke problemer og løsninger en videnskab beskæftiger sig med og anfører ”Subject matter or kind of things do not, I hold, constitute a basis for distinguishing disciplines...We are not students of some subject matter, but students of problems” (min understregning). Og man kan roligt sige, at både den moderne leksikografi og informationsvidenskab beskæftiger sig med mange af de samme problemer. Begge discipliner er interdisciplinære og er desuden i høj grad forbundet med informationsteknologi, og endelig har begge discipliner en stor snitflade til samfundet, det enkelte individ og indgår ofte i en social og organisatorisk kontekst.

Lad os se på hvorledes informationsvidenskab defineres. Merriam-Webster Online definerer informationsvidenskab således: ”Information science is an interdisciplinary science primarily concerned with the analysis, collection, classification, manipulation, storage, retrieval and dissemination of information (min understregning)”, hvilket synes at være en dækkende definition. En langt mere interessant diskussion tilbydes dog af Saracevic (1999: 1052), som anfører i sin diskussion af informationsvidenskab, at den som videnskab har en række karakteristika. Informationsvidenskab er for Saracevic karakteriseret af at være ”...interdisciplinary in nature...”, ”...inexorably connected to information technology...”, og at ”...Information science has a strong social and human dimension...”.

Lad os nu se på hvorledes leksikografi defineres. Merriam-Webster Online definerer leksikografi som ”the editing or making of a dictionary”, hvilket må siges at være en gammeldags og ikke særlig dækkende definition, idet moderne leksikografi i stigende grad har fokus på behovstilpasset informationstilgang og mere opfatter sig selv som et delområde inden for informationsvidenskaben. Leksikografi defineres mere dækkende af f.eks. Bergenholtz & Tarp (1995: 31, som henholdsvis ”metalexigraphy (theoretical lexicography or dictionary research)” og ”practical lexicography”. Den moderne leksikografi har udviklet sig eksplosivt de senere år, men sammenhængen mellem leksikografi og informationsvidenskab blev allerede påvist i 1995, hvor Reinhard Hartmann i en

uformal samtale med Tom McArthur foreslog termen *reference science* som en kombination af *library science* og *information science* (McArthur 1998: 220).

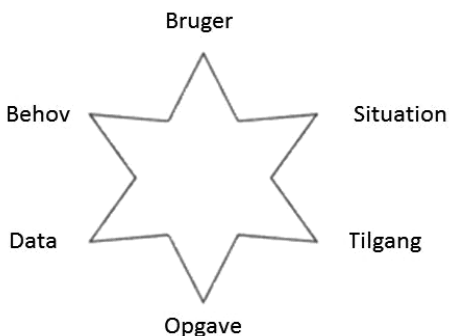
Sammenfattende argumenteres det, at begge discipliner netop udnytter det interdisciplinære samarbejde, samt i varierende grad har fokus på især opsamling, klassifikation, tilpasning, tilgang og spredning af information.

### 3. Den informationsvidenskabelige stjernemodel

”Relevant data should be retrieved and processed according to the external situation that motivated consultation in the first place, and the information needed to change a state of affairs in the outside world should be operationalized” (Verlinde et al. 2010: 5) er et meget rammende citat, og det er meget vanskeligt at være uenig med de tre forfattere i, at det netop er det hele handler om. Og her skal såkaldte forskellige teoretiske skoler ikke stå i vejen for det i grunden åbenlyse, men samtidig meget vanskelige: at tage udgangspunkt i brugerens kompetencer, behov, situation, og den opgave brugeren står overfor, og derefter give brugeren behovs- og opgavetilpasset tilgang til de rette informationstyper. De tre forfattere argumenterer, at ethvert opslagsværk skal kalibreres i forhold til tre dimensioner som er: bruger, data og tilgang. Disse overvejelser benævnes ”the lexicographic triangle” (Verlinde et al. 2010: 5), og hovedargumentet er, at hvis ikke de tre elementer er kalibreret i forhold til hinanden, så vil opslagsværket så at sige få slagside.

Der er dog langt flere dimensioner på spil end de tre, og Figur 1 nedenfor viser den informationsvidenskabelige stjernemodel, som til forskel fra den lexicografiske trekant er baseret på hele seks dimensioner nemlig bruger, situation, tilgang, opgave, data og behov.

Figur 1. Den informationsvidenskabelige stjernemodel



*Brugeren* kan både være en person og en organisation, se f.eks. Simonsen (2002: 46), som taler om et "dual-focus approach", hvor både organisationen som et strategisk miljø og den enkelte bruger opfattes som bruger. Tilsvarende kan man tale om et helt samfund som bruger af et opslagsværk, hvilket der er tale om i denne situation, hvor medicin.dk. tilfredsstillende et samfundsmæssigt relevant informationsbehov efter opdateret og videnskabeligt baseret lægemiddelinformation.

*Situation* er den situation, hvor information tilgås. Tarp (2006) omtaler to typer af brugssituationer, nemlig "punktuelle" og "systematiske" og Simonsen (2008) diskuterer "the lexicographic workflow cycle", og argumenterer for, at en brugssituation ofte er fluktuerende og består af indbyrdes forbundne kognitive, kommunikative, punktuelle og systematiske brugssituationer. Situationen får i øvrigt stigende betydning, idet brugersituationen ikke længere er fysisk bundet til en lokation via et opslagsværk eller en computer, men nu i stigende grad er bevægelig og forekommer på smart phones eller andre mobile devices, og dermed en mobil informationstilgang med den støj, som det afstedkommer.

*Tilgang* er den måde brugeren tilgår data på og omfatter den eller de søgemaskiner og -metoder, hvorpå data kan fremfindes. Datatilgangen bør være variabel og afhængig af hvilken type af data brugeren har brug for til løsning af en specifik opgave. Datatilgang indeholder også datapræsentation, som er den måde data præsenteres på i en given situation.

*Opgave* er den opgave, som brugeren er ved at løse. Brugeren kan have behov for information i forbindelse med tekstreception, tekstproduktion, oversættelse, tilegnelse af viden, dvs. understøttelse af både kommunikative eller kognitive funktioner, eller fordi han skal bruge bestemte informationer med henblik på at udføre en handling eller opgave. Tarp (2006) giver en omfattende oversigt over de mange forskellige funktioner et opslagsværk kan tilfredsstillende, men funktionslæren er i sig selv blevet begrænsende, idet information efterspørges i langt flere situationer end blot i kommunikative og kognitive brugssituationer. Tænk blot på information, som tilgås i forbindelse med underholdning eller ved dannelse af sociale relationer etc.

*Data* er de semiotiske tegn, som brugeren søger frem. Mængden og typen af data er helt naturligt afhængig af behovet, opgaven og situationen, og tilgangen skal helt naturligt være tilpasset de samme faktorer. Data kan således sagtens være alt lige fra traditionelle tegn til lyd- og videofiler. En patient med en kronisk sygdom og en speciallæge har helt naturligt ikke samme informationsbehov, og de har derfor forskellige ønsker og krav til datatilgangen.

*Behov* er det informationsbehov, som brugeren har. Ifølge [www.informationsordbogen.dk](http://www.informationsordbogen.dk) kan informationsbehov både være "erkendte" og "ikke-erkendte" samt "subjektive" og "objektive". Subjektive er de informationsbe-

hov, som en person mener at have og objektive informationsbehov omfatter de informationer, som er nødvendige for at løse et bestemt problem. Det er helt afgørende, at huske på, at et informationsbehov jo er uløseligt forbundet med de øvrige dimensioner, men især tilgang, opgave og data, hvilket jo i den grad skal afstemmes i forhold til hinanden for at tilfredsstille brugerens informationsbehov.

Sammenfattende argumenteres det således, at den informationsvidenskabelige stjernemodel er velegnet til både at beskrive samt analysere de forskellige faktorer i informationsvidenskab og metaleksikografi, og at modellen er velegnet til at beskrive og anvende ved udviklingen af moderne informationsvidenskabelige opslagsværker. Der er således ikke kun tale om brugertilpassede adgang til data, men også i høj grad situations- og opgaveafhængige data, som er dikteret af en lang række faktorer, herunder brugerens kompetencer, opgaveløsning samt opgavesituation og i realiteten også det formål, som brugeren har med det søgeresultat, som informationssøgningen fører til.

### 3. Medicin.dk

Medicin.dk indeholder videnskabeligt baseret information om lægemidler, sygdomme og behandling. Medicin.dk anvendes af sundhedsprofessionelle på f.eks. hospitaler, i lægehuse og på apoteker, ligesom lægemiddelinformationen fra Medicin.dk kan tilgås fra websites som f.eks. apoteket.dk, netdoktor.dk og på mange patientforeningers websites. Logfilldata fra Google Analytics fra perioden 18. april – 24. april 2011 viser f.eks., at Medicin.dk på en uge har 553.318 sidevisninger. Medicin.dk er opdelt i to indgange:

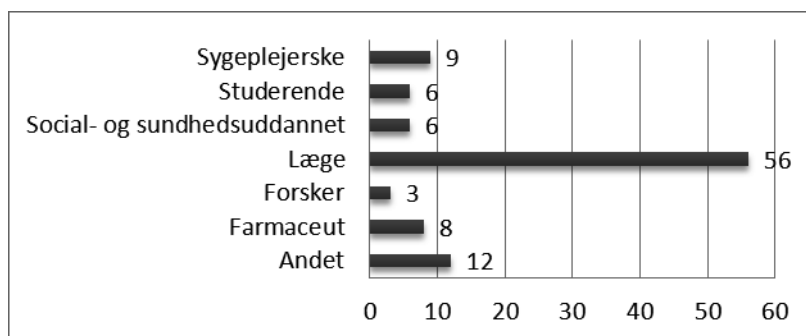
- **pro.medicin.dk**, hvis målgruppe er fagfolk, dvs. læger, farmaceuter, sygeplejersker, lægesekretærer, etc., der har en sundhedsfaglig baggrund
- **min.medicin.dk**, hvis målgruppe lægfolk, dvs. patienter, pårørende og andre interesserede borgere, der ikke nødvendigvis har nogen sundhedsfaglig baggrund.

Medicin.dk udgives af Dansk Lægemiddel Information A/S (DLI). Medicin.dk har en samlet redaktion, men to forskellige ansvarshavende redaktører for hhv. min.medicin.dk og pro.medicin.dk. Redaktionerne gør brug af mere end 400 lægefaglige eksperter, der laver og opdaterer lægemiddelinformation inden for hver deres speciale. Redaktionen varetages af sekretariatet, som består af informationsfarmaceuter, der koordinerer og opdaterer alle tekster.

#### 4. Brugerne og brugssituationerne

For at kunne udføre det informationsvidenskabelige serviceeftersyn, skal vi først se lidt nærmere på brugerne af Medicin.dk, de typiske brugssituationer samt på de relevante informationskategorier. Første del af dette kapitel ser på brugerne af pro.medicin.dk, og tager udgangspunkt i data fra en spørgeskemaundersøgelse med 3813 respondenter, en interview-undersøgelse og en workshop-undersøgelse, jf. Devoteam (2009). Interviewundersøgelsen omfattede 11 interviews af læger, farmaceuter og sygeplejersker, og de to workshops omfattede 17 læger, farmaceuter og sygeplejersker. Som det fremgår af Figur 2 nedenfor er over halvdelen af brugerne læger (56%), men også sygeplejersker, farmaceuter og social- og sundhedsuddannede bruger Medicin.dk.

Figur 2. Profil af Medicin.dk's brugere (pro.medicin.dk)



Som det fremgår af Figur 3 nedenfor afspejles dette også i de fem hyppigste brugssituationer nemlig *Valg af behandling*, *Ordination*, *Håndtering af lægemidler*, *Information til patienter*, og *Opdatering af viden*, som er typiske brugssituationer for sundhedsprofessionelle, når de ønsker at tilegne sig ny viden og træffe beslutninger, jf. Devoteam (2009). De fem hyppigste brugssituationer kan karakteriseres som primært kognitive brugssituationer, jf. Tarp (2006:58), som inddeler brugssituationer i hhv. ”kommunikative situationer (dvs. situationer, hvor der opstår et problem i forbindelse med en igangværende eller planlagt kommunikation) og kognitive situationer (dvs. situationer, hvor der eksisterer et ønske om at få viden om et eller andet emne)”.

Figur 3. Informationstyper i de fem hyppigste brugssituationer i pro.medicin.dk

Valg af behandling	Ordination	Håndtering af lægemidler	Information til patienter	Opdatering af viden
Bivirkninger (80%)	Dosis (74%)	Dosis (49%)	Bivirkninger (86%)	Bivirkninger (75%)
Dosis (72%)	Bivirkninger (73%)	Bivirkninger (46%)	Interaktioner (67%)	Interaktioner (71%)
Interaktioner (71%)	Interaktioner (65%)	Interaktioner (42%)	Dosis (57%)	Indholdstof (61%)
Særlige forsigtighedsregler (55%)	Særlige forsigtighedsregler (50%)	Særlige forsigtighedsregler (32%)	Pris- og pakningsinformation (52%)	Dosis (60%)
Indholdsstoffer (50%)	Pris- og pakningsinformation (49%)	Indholdsstoffer (30%)	Særlige forsigtighedsregler (50%)	Særlige forsigtighedsregler (59%)

Figur 3 viser således sammenfattende, hvordan 3813 respondenter har rangordnet de fem mest foretrukne informationstyper i de fem hyppigste brugssituationer. Lægen bruger således pro.medicin.dk i sit daglige arbejde delvist til beslutningsstøtte og delvist i en kognitiv brugssituation til at træffe beslutninger og til at besvare spørgsmål som *Hvilket lægemiddel skal jeg vælge til patienten?*, *Hvilken dosis skal patienten have?* og *Hvilken information skal jeg huske at anføre på recepten?*, etc.

Anden del af dette kapitel diskuterer brugerne af min.medicin.dk, og der tages her først og fremmest udgangspunkt i data fra en brugertest fra 2010 med otte borgere med forskellig baggrund, jf. SnitkerGroup (2010). Undersøgelsen omfattede otte brugervenlighedstests, som blev gennemført efter tænke-højtmetoden og optaget på video. Testpersonerne omfattede yngre og ældre borgere, internervante og mindre internervante borgere samt syge og raske borgere. Undersøgelsen omfattede mere specifikt en kvinde på 41 år (ufaglært) mor til to, en mand på 52 år (cand.merc.) far til to, kronisk astma, en kvinde på 78 år (faglært), en kvinde på 46 år (korrespondent), pårørende medicineret for depression, en kvinde på 41 år (cand.mag), en mand på 24 år (elektriker), kronisk tarmsygdom, en kvinde på 55 år, datter med Aspergers syndrom og en mand 78 år (pensionist).

En decideret brugerprofil i lighed med den i Figur 1 er således ikke tilgængelig, hvorfor følgende brugerprofil opstilles på basis af data fra de otte brugervenlighedstests samt den formulerede målgruppe for min.medicin.dk. Den intenderede målgruppe for min.medicin.dk er ”patienter, pårørende og andre



interesserede brogere, der ikke nødvendigvis har nogen sundhedsfaglig baggrund” (www.medicin.dk), og min.medicin.dk bør derfor optimalt set være et opslagsværk, som primært henvender sig til en meget bred målgruppe bestående af lægfolk, som samtidig ofte er påvirket af bekymringer, angst og sorg over egen eller pårørendes sygdom.

Patienten og den pårørende derimod har en speciel brugssituation. Patienten og den pårørende bruger min.medicin.dk i kognitive brugssituationer, idet de ønsker mere information om den sygdom, de eventuelt er blevet diagnosticeret med og/eller, når de ønsker at få mere viden om den medicin, de er blevet ordineret af lægen. Patienten og den pårørendes brugssituation er karakteriseret ved, at de ofte er bekymrede, kede af det, bange, irrationelle og følelsesmæssigt berørte, og nogle af de spørgsmål, som de ønsker svar på er *Hvad kendetegner min sygdom?*, *Hvad er udsigterne til at blive helbredt?*, *Hvordan kan den medicin lægen har ordineret hjælpe mig?*, *Hvilke bivirkninger har medicinen?* og *Hvordan skal jeg tage medicinen?* etc.

Alle disse spørgsmål kan patienten og den pårørende sagtens få svar på i den nuværende udgave af min.medicin.dk, men som det informationsvidenskabelige serviceeftersyn påviser, så har patienter og pårørende også andre, vigtige og ofte mere socialt orienterede spørgsmål i forbindelse med deres sygdom og dermed ikke udelukkende kognitive behov. I forbindelse med sygdom eller som pårørende til en syg, vil brugere af min.medicin.dk ofte have spørgsmål som *Hvordan klarer andre sig med denne sygdom?*, *Hvilke erfaringer og gode råd kan andre patienter med samme sygdom give mig?* og *Hvordan håndterer andre patienter denne sygdom?* etc. Det er væsentlige spørgsmål, som går langt ud over det blot at tilegne sig faktuel viden om en sygdom og en udskrevet medicin. Der er med andre ord tale om en mere social brugssituation, som er kendetegnet ved, at der indgår sociale relationer mellem mennesker, i første omgang ofte faciliteret vha. sociale medier og eventuelt senere vha. personlige relationer, hvor patienter med samme type sygdom mødes i patientgrupper etc. Denne type brugssituation vil i det efterfølgende blive omtalt som en *socialrelationel brugssituation* (dvs. en situation, hvor der eksisterer et ønske om at skabefastholde sociale relationer mhp. at dele bekymringer og viden om ens sygdom).

Men dette behov for at komme i kontakt med andre patienter og dele bekymringer med andre med samme sygdom tilfredsstilles kun i ringe grad med den nuværende udgave af min.medicin.dk. Denne mangel fremgår især tydeligt, når man ser på de mest sandsynlige brugssituationer og de tilhørende informationstyper, hvoraf meget få af disse er til stede i eller kan tilgås fra min.medicin.dk, se Figur 4 herunder.

Figur 4. Informationstyper i de fem mest sandsynlige brugssituationer i min.medicin.dk

Information om min sygdom	Information om min medicin	Selvhjælp	Hjælp og rådgivning	Temamøder
Sygdomsbeskrivelse	Brug af medicin	Forum	Netrådgivning	Emne for møder
Symptomer	Bivirkninger	Blog	Psykologhjælp	Mødedato
Sygdomsforløb	Priser og pakning	Patientdagbøger	Socialrådgiver	Mødested
Behandling	Medicinkombination		Behandlingsplan	
Overlevelseschance				

At patienter og pårørende har andre behov end blot information, underbygges af følgende citat fra kvinde 55 år, som var en af forsøgspersonerne i testen: ”Jeg søger information på blogs og i chat-grupper hos patientforeninger. Her finder jeg erfaringer fra bruger til bruger” Devoteam (2009). Også information om overlevelseschancer etc. er efterspurgt, hvilket kommer til udtryk af mand 52, som stillet overfor en opgave i testen om at finde relevant information om brystkræft, udtaler, at ”Måske har man også brug for fakta om, hvis man får brystkræft, hvad chancerne er for helbredelse osv., hvor mange der kommer godt igennem sygdommen” Devoteam (2009). Der synes derfor også at være empirisk basis for at argumentere, at min.medicin.dk i højere grad bør tilfredsstille socialrelationelle behov. Det informationsvidenskabelige serviceeftersyn og appliceringen af stjernemodellen på hhv. brugerne, brugersituationerne, datatilgængene, brugeropgaverne, de nødvendige data og brugerbehovene viste, at der stadig er en del at gøre, før min.medicin.dk tilfredsstiller de socialrelationelle behov.

## 5. Adgang til sociale medier fra min.medicin.dk?

En dækkende definition af sociale medier gives af Kaplan & Haenlein (2010: 61), som definerer sociale medier som ”Social media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of User Generated Content”. Grunden til at netop sociale medier anbefales til at understøtte de socialrelationelle behov i min.medicin.dk er jo netop, at sociale medier understøtter brugernes ønske om social tilstedeværelse, selvpræsentation og ønsket om at afsløre informationer om sig selv. Netop det at afsløre noget om sig selv er helt

afgørende, når mennesker vil udvikle tætte sociale relationer, og netop i en situation, hvor en pårørende eller kronisk syg ønsker at dele sine bekymringer med andre i samme situation, synes sociale medier at være oplagt at bruge, se i øvrigt Kaplan & Haenlein (2010: 62), som bruger de to dimensioner *Social Presence* og *Self Presentation* til at klassificere seks forskellige typer sociale medier. Brogan (2010:5 ff.) giver også en række gode grunde til at anvende sociale medier i forretningsøjemed herunder at "Social networks can amass like-minded people around shared interests with little external force, no organizational center, and a group sense of what is important and what comes next" og endelig "People feel heard" (min understregning).

Derfor er der alt mulig god grund til at give brugerne med socialrelationelle behov adgang til sociale medier fra min.medicin.dk. Tanken i konceptet til anvendelse af sociale medier i min.medicin.dk er, at min.medicin.dk udbygges med et ekstra felt, der f.eks. hedder *Sygdom*, som vises i lighed med de andre informationstyper. Det ekstra felt skal naturligvis vises både i navigationsmenuen øverst i skærmbilledet og i navigationsfeltet i venstre side af skærmbilledet på min.medicin.dk.

Hvis en patient f.eks. har fået ordineret lægemidlet Tamoxifen til efterbehandling af østrogenfølsomme brystkræftsygdomme, så vises *Brystkræft* i sygdomsfeltet. Præparater og sygdomme forbindes således, at de relevante præparater vises i forbindelse med de relaterede sygdomme og omvendt. Så udover at få de kognitive behov dækket, kan patienten eller den pårørende således klikke på linket *Brystkræft*, forlade min.medicin.dk og gå til det sygdomsspecifikke subsite, som tilbyder yderligere muligheder for patienter, der er i behandling for brystkræft. Dette og øvrige subsites stilles til rådighed af min.medicin.dk for patienter, pårørende og patientforeninger, som så vil kunne anvende sitet efter eget ønske. Når brugeren forlader min.medicin.dk forlader vedkommende også den kontrollerede del og begiver sig ind på et subsite, som Medicin.dk ikke har ansvaret for og som er underlagt de generelle regler for anvendelse af sociale medier. Sociale medier lader sig vanskeligt styre, og derfor skal Medicin.dk fraskrive sig ethvert ansvar for det, som måtte foregå på subsitet. Men samtidig er sociale medier særdeles velegnede til netop understøttelse af socialrelationelle behov og er en stigende magtfaktor for patienter og patientforeninger, fordi de via disse medier endelig har fået mulighed for at påvirke læger, politikere og offentligheden.

Når brugeren så er kommet til det pågældende subsite – her *Brystkræft* – så vil han kunne vælge mellem forskellige typer af sociale medier, som alle hver for sig på forskellig vis understøtter de socialrelationelle behov og tilfredsstiller behovet for at have social tilstedeværelse og selvpræsentation, se Kaplan & Haenlein (2010: 62).

Subsitet *Brystkræft* kan med fordel indeholde forskellige typer af sociale medier og konceptet her indeholder funktionaliteterne *Forum*, *Behandlingslan*, *Dagbog*, *Lokale foreninger*, *Møder*, *Psykolog* og *Socialrådgiver*. Patienten eller den pårørende vil således kunne udveksle viden og bekymring om egen eller pårørendes brystkræft i *Forum* og få adgang til personligt og NEM-ID-beskyttet *Behandlingsplan*, som læge og hospital har udarbejdet. Patienten vil også kunne skrive om sit sygdomsforløb under *Dagbog* på en blog, hvor patienten kan blogge om sin sygdom, se Natt och Dag (2012), kunne tilgå information om *Lokale foreninger* og *Møder* om brystkræft og endelig kunne tilgå netrådgivning eller personlig rådgivning fra psykologer og socialrådgivere under *Psykolog* og *Socialrådgiver*.

## 6. Konklusion

Denne artikel diskuterede informationsvidenskab og leksikografi og introducerede den informationsvidenskabelige stjernemodell. Stjernemodellen er velegnet som analysemodel og til at anvende i udviklingen af moderne informationsvidenskabelige opslagsværker, som skal tage højde for dimensioner som bruger, situation, tilgang, opgave, data og behov. Stjernemodellen blev brugt som analysemodel på Medicin.dk, og analysen viste, at især patientens og den pårørendes socialrelationelle behov ikke i tilstrækkelig grad blev tilfredsstillet i den nuværende udgave af sitet. På basis af denne analyse blev der derfor til sidst i artiklen præsenteret et enkelt koncept til anvendelse af sociale medier i opslagsværker som f.eks. Medicin.dk, og det argumenteres, at anvendelsen af sociale medier i særlig grad er velegnet til ikke kun at give patienten og den pårørende faktuelle oplysninger om egen sygdom og behandling, men også til at få tilfredsstillet eller faciliteret sine socialrelationelle behov.

## LITTERATUR

- Bergenholtz, Henning & Tarp, Sven, 1995 (ed.): *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. With contributions by: Grete Duvå, Anna-Lise Laursen, Hans Kristian Mikkelsen, Sandro Nielsen, Ole Norling-Christensen and Jette Pedersen. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Brogan, Chris, 2010: *Social Media 101: Tactics and Tips to Develop Your Business Online*. John Wiley & Sons, Inc. New Jersey.
- Devoteam Consulting 2009: *Brugerundersøgelse af Medicin.dk. Brugerundersøgelsesrapport*.

- Kaplan Andreas M., Haenlein Michael, 2010: Users of the world, unite! The challenges and opportunities of social media, *Business Horizons*, Vol. 53, Issue 1. S. 59–68.
- McArthur, Tom, 1998: *Living Words, language, lexicography and the knowledge revolution*. Exeter: University of Exeter Press. S. 215–222.
- Natt och Dag, Sara, 2012: *Den Bloggande Patienten: sociale medier i vården*. Stockholm. Gothia
- Popper, K.R., 1989: *Conjectures and refutations* (5th ed.). London: Routledge.
- Saracevic, Tefko, 1999: Information Science. I: *Journal of the American Society for Information Science* 50(12). S. 1051–1063.
- Simonsen, Henrik Køhler, 2002: *TeleLex – Theoretical Considerations on Corporate LSP Intranet Lexicography: Design and Development of TeleLex – an Intranet-based Lexicographic Knowledge and Communications Management System*. Ph.d.-afhandling, ASB.
- Simonsen, Henrik Køhler 2008: When Two Become One in Functional Bliss: Integration of a Corporate Language Policy and a Corporate Dictionary? I: *Lexikographica, Internationales Jahrbuch für Lexikographie*, Band 24, 2008, Max Niemeyer Verlag, Tübingen, Germany. S. 317–346.
- SnitkerGroup 2010: Rapport – Brugertest af prototype af min.medicin.dk. Brugertesrapport.
- Tarp, Sven, 2006: *Leksikografien i grænselandet mellem viden og ikke-viden: Generel leksikografisk teori med særlig henblik på lærerleksikografi*. Doktorafhandling. ASB.
- Verlinde, Serge, Leroyer, Patrick, Binon, Jean, 2010: Search and You Will Find. From Stand-Alone Lexicographic Tools to User Driven Task and Problem-Oriented Multifunctional Leximats. I: *International Journal of Lexicography*, Vol. 23, Issue (1) 2010. S. 1–17.

*Henrik Køhler Simonsen*

Copenhagen Business School, Institut for Internationale Sprogstudier og Vidensteknologi  
hks.isv@cbs.dk

# A och O om akademiska ord

## Om framtagning av en svensk akademisk ordlista

*Emma Sköldbberg & Sofie Johansson Kokkinakis*

This paper reports on the development of a new electronic Swedish academic word list, which will be based on a corpus consisting of texts from a wide range of university subject areas. (Cf. the English Academic Word List, AWL, presented in Coxhead 1998, 2002.) Once finalized, the dictionary and corpus will be publicly available. The aim of the word list is to provide an indication of important words for university students, in particular non-native speakers of Swedish. In order to succeed in higher education, the mastery of an academic vocabulary may be crucial. A closely related study of extraction of a school-related vocabulary of secondary school textbooks is the OrdiL project (Lindberg & Johansson Kokkinakis 2008). The Swedish academic word list will be developed in cooperation with other Nordic countries in a multi-disciplinary project focusing on language use in Nordic academic settings.

*Nyckelord:* Academic Word List, akademiskt ordförråd, svenska, språkinläring

### 1. Inledning

Det svenska akademiska ordförrådet, dvs. de lexikala enheter som återkommer i universitetstexter oavsett vilket ämne som behandlas, vållar problem för många akademiska nybörjare – oberoende av vilket modersmål de har. Exempel på verb, som är problematiska för svenska studenter, är *analysera*, *argumentera*, *beskriva*, *diskutera*, *referera*, *reflektera*, *karaktärisera*, *motivera*, *problematisera*, *redogöra* och *värdera* (se vidare om den hjälp som erbjuds alla studenter inom Göteborgs universitets Språkhandledning; Carlund & Andersson 2011). För att möta studenters behov av kunskap om den här typen av vokabulär har det utarbetats en rad engelska akademiska ordlistor. Nedan presenteras i korthet några sådana. Men tyngdpunkten i artikeln ligger på en beskrivning av vårt påbörjade arbete med att ta fram en webbaserad, svensk, fritt tillgänglig akademisk ordlista – en typ av resurs som hittills saknas för vår del. Bland annat kommer vår metod för framtagning av en ordlista och dess tänkta innehåll och utformning

att diskuteras. Avslutningsvis presenteras också kortfattat ett nordiskt nätverk kring språkbruket i nordiska akademiska miljöer som nyligen har bildats.

## 2. Tidigare akademiska ordlistor

För att stödja i synnerhet mer avancerade engelskinlärare vid högre utbildningar, men även t.ex. universitetslärare med annat modersmål än engelska, har det som sagt tagits fram ett antal mer renodlade ordlistor över ett akademiskt ordförråd (se vidare Coxhead 2000: 213–214 för en genomgång av sådana). I dessa ordlistor upptas sålunda engelska ord (som *substitute*, *underlie* och *establish*) vilka, vid sidan av ett mer allmänt ordförråd och specifika facktermer, typiskt förekommer i texter som skrivs av och för personer med direkt anknytning till högre studier. Det rör sig bl.a. om vetenskapliga artiklar och studenters kursböcker. En viktig ordlista är Xue & Nations *The University Word List* (1984; hädanefter UWL). Denna lista utgörs i huvudsak av en sammanställning av redan existerande listor från 1970-talet och har, enligt Coxhead (2002: 73), använts flitigt av inlärare och lärare men också som underlag för att ta fram ordtester och undervisningsmaterial. En senare ordlista, som kommit att ersätta UWL, är Coxheads *The Academic Word List* (1998; hädanefter AWL). Denna lista presenteras mer ingående i nästa avsnitt (se även Coxhead 2000, 2002). Men utöver dessa specialordböcker lyfter även vissa allmänspråkliga engelska ordböcker fram ett akademiskt ordförråd. Som exempel kan *Longman Dictionary of Contemporary English* (2009) anföras. Typiskt akademiska ord markeras bl.a. i lemmalistan, och det finns mycket övningar på dvd-versionen och online för främst avancerade inlärare. Däremot har det alltså inte utarbetats någon svensk akademisk ordlista av det slag som vi åsyftar. Men det finns en annan typ av ordlista som är kopplad till högre studier, t.ex. den på Högskoleverkets portal vilken innehåller förklaringar till bl.a. mer administrativt relaterade ord som *antagningsbesked*, *behörighet* och *doktorsexamen* (se vidare studera.nu). Ett svenskt projekt som är relevant i detta sammanhang är också OrdiL vilket beskrivs närmare i Lindberg & Johansson Kokkinakis (2007). Inom detta projekt ligger fokus på identifiering av ett skolrelaterat ordförråd i läromedel för grundskolans senare år.

Det finns flera skäl till att det akademiska ordförrådet bör lyftas fram och till att det är ett viktigt inlärningsmål för inlärare som genomgår eller siktar mot högre studier. Det akademiska ordförrådet uppträder i ett antal typer av akademiska texter, men däremot inte lika ofta i t.ex. skönlitterära sådana. Vidare utgör det akademiska ordförrådet en betydande del av orden i akademiska texter. Dessutom är det akademiska ordförrådet inte lika känt – och iögonenfal-

lande – som t.ex. ett mer tekniskt ordförråd vilket leder till att den här typen av ord ofta glöms bort (Nation 2001: 189–191).

### 3. Kort om Coxheads AWL

Enligt Coxhead (2000: 214) har akademiska ordlistor satts samman på olika sätt, bl.a. utifrån ord som studenterna markerat som svåra i kurslitteraturen. Författaren menar dock att det är viktigt att innehållet i en akademisk ordlista baseras dels på relevanta principer inom korpuslingvistik, dels på kunskaper om ordinläring och ordundervisning (Coxhead 2002: 74). När Coxhead skulle utarbeta sin ordlista satte hon därför först samman en korpus av texter med akademiskt språkbruk. Utifrån denna korpus togs sedan själva ordlistan fram.

Coxheads korpus av akademiska texter (hädanefter AWL-korpusen) består av texter hämtade från bl.a. vetenskapliga tidskrifter och kursböcker. Hon har strävat efter att inkludera kortare texter som författats av många olika personer och det har resulterat i att korpusen sammanlagt rymmer 414 texter skrivna av mer än 400 olika författare. AWL-korpusen är begränsad till sin storlek; sammanlagt innehåller den ca 3,5 miljoner löpande ord.

Coxheads AWL-korpus utgörs av fyra delkorpusar (på ca 875 000 löpande ord vardera) med texter från de fyra ämnesområdena *arts*, *commerce*, *law* och *science*. Varje delkorpus består i sin tur av texter från sju mer specifika ämnesområden (se tabell 1).

<b>Arts</b>	education, history, linguistics, philosophy, politics, psychology, sociology
<b>Commerce</b>	accounting, economics, finance, industrial relations, management, marketing, public policy
<b>Law</b>	constitutional, criminal, family and medicolegal, international, pure commercial, quasi-commercial, rights and remedies
<b>Science</b>	biology, chemistry, computer science, geography, geology, mathematics, physics

Tabell 1. Ämnesområdena företrädna i de fyra AWL-delkorpusarna (Coxhead 1998).

Coxhead tog sedan fram ett antal lexikala enheter med hög frekvens och stor spridning i de fyra delkorpusarna (se vidare Coxhead 2002: 75). Därefter utslöt hon från dessa ord de 2000 vanligaste orden som ingår i *A General service list* (GSL, West 1953), en lista över frekventa ord i allmän engelska som West ansett särskilt användbara för inlärare.



Denna metod resulterade i, som Coxhead säger, 570 word families (sv. ordfamiljer). Enligt Coxhead (2000: 218) är en ordfamilj detsamma som ”a stem plus all closely related affixed forms”. Som exempel kan verbet *conclude* nämnas, vilket enligt AWL ingår i följande ordfamilj: *conclude, concluded, concludes, concluding, conclusion, conclusions, conclusive, conclusively, inconclusive, inconclusively*. I och med angivelserna av ordfamiljer, ansluter sig Coxhead (1998) till en tydlig tradition bland dem som utarbetat ordlistor för engelskinlärare (se t.ex. West 1953, Xue & Nation 1984). Hon motiverar också detta val med att bruket av ordfamiljer ”is supported by evidence suggesting that word families are an important unit in the mental lexicon” (Coxhead 2000: 217 f.).

De 570 ordfamiljer som utgör AWL är fördelade på 10 listor (eng. *sublists*), vars innehåll baseras på de ingående ordfamiljernas frekvens och spridning i delkorpusarna. Den första av dessa listor, sublist 1, innehåller de mest frekventa ordfamiljerna, medan sublist 2 utgörs av de näst mest frekventa ordfamiljerna etc. I figur 1 återges de första av ordfamiljerna i AWL:s sublist 1.

analyse	reassessed	conceptually
analysed	reassessing	consist
analyser	reassessment	consisted
analyser's	unassessed	consistency
analyses	<i>assume</i>	<i>consistent</i>
analysing	assumed	consistently
<i>analysis</i>	assumes	consisting
analyst	assuming	consists
analysts	assumption	inconsistencies
analytic	assumptions	inconsistency
analytical	<i>authority</i>	inconsistent
analytically	authoritative	constitute
analyze	authorities	constituencies
analyzed	<i>available</i>	constituency
analyzes	availability	constituent
analyzing	unavailable	constituents
<i>approach</i>	<i>benefit</i>	constituted
approachable	beneficial	constitutes
approached	beneficiary	constituting
approaches	beneficiaries	constitution
approaching	benefited	constitutions
unapproachable	benefiting	<i>constitutional</i>
<i>area</i>	benefits	constitutionally
areas	<i>concept</i>	constitutive
assess	conception	unconstitutional
assessable	concepts	<i>context</i>
assessed	conceptual	contexts
assesses	conceptualisation	contextual
assessing	conceptualise	contextualise
<i>assessment</i>	conceptualised	contextualised
assessments	conceptualises	contextualising
reassess	conceptualising	uncontextualised

Figur 1. Utdrag ur sublist 1 i AWL (Coxhead 1998).

Ytterligare ord (med ordfamiljer) som ingår i sublist 1 är bl.a.: *establish, estimate, evident, factor, function, indicate, individual, involve, issue, legislate, major, method, occur, percent, period, principle, process, require, research, respond, section, significant, similar, source, specific, structure* och *theory*.

Av figur 1 framgår tydligt att AWL, som namnet indikerar, är en renodlad ordlista. I den ingår ordfamiljer, grafiskt markerade med ett inledande huvudord som följs av familjemedlemmar – i den mån det finns några sådana. Däremot finns det inga uppgifter om huvudordens eller familjemedlemmarnas uttal, böjning, betydelse, kombinatoriska egenskaper etc. Det enda användarna får veta är vilken familjemedlem som är den mest frekventa i korpusen; denna står nämligen i kursiv stil. Ibland är det huvudordet som har högst frekvens (som i fallet *approach*), i andra fall är det någon av familjemedlemmarna (som i fallet *consist* där adjektivet *consistent* är mest frekvent).

Det faktum att det finns så lite information om orden i listan begränsar givetvis ordlistans användningsområden. Över huvud taget resonerar Coxhead i sina tidigare arbeten mycket litet kring bruket av listan och vilka lexikografiska funktioner den är tänkt att fylla. Författaren konstaterar, i mer allmänna ordalag, att:

An academic word list should play a crucial role in setting vocabulary goals for language courses, guiding learners in their independent study, and informing course and material designers in selecting texts and developing learning activities. (Coxhead 2000: 214)

På AWL:s webbsida finns det numera många handfasta råd till (in)lärare hur de kan arbeta med ordlistans innehåll (se vidare AWL: How to use the list). Precis som UWL har också AWL legat till grund för flera ordkunskapstest och t.ex. många praktiska kursböcker med övningsuppgifter (se vidare AWL Useful Links).

#### 4. Framtagning av en svensk akademisk ordlista: val av metod

Vi har nu påbörjat arbetet med att ta fram en fritt tillgänglig elektronisk svensk *Akademisk ordlista* (hädanefter AO). Ordlistan vänder sig främst till studenter och lärare med annan språklig bakgrund än svenska, men också till modersmålsstudenter med mindre erfarenhet av akademiska studier. Vi hyser också en förhoppning om att AO även kan användas vid gymnasiala studier.

En viktig fråga är givetvis vilken metod som är lämpligast för att ta fram denna ordlista. Ett möjligt tillvägagångssätt är att översätta AWL till svenska. AWL innehåller många engelska lexikala enheter vars svenska motsvarigheter – enligt vår språkliga intuition – är vanligt förekommande i svensk akademisk

prosa och därför kan försvara sin plats i en svensk lista (t.ex. *analys*, *indikera* och *struktur*). Översättning har använts för att ta fram olika typer av lexikala resurser, t.ex. inom det norska Lexin-projektet och vid utarbetandet av ett finskt ordnät, FinnNet (se vidare Bjørneset 2001 och Lindén & Carlson 2010). Metoden har också visat sig vara snabb och kostnadseffektiv, men samtidigt är den problematisk (se t.ex. Martola 2011 om brister hos FinnNet). En pilotundersökning som vi genomfört om möjligheterna att översätta AWL till svenska visar också på tydliga svårigheter. Närmare bestämt har vi granskat vad de 60 huvudorden i AWL:s sublista 1 har för svenska ekvivalenter i *Norstedts stora engelsk-svenska ordbok* (2000). Ett fåtal ord i listan, t.ex. *percent*, är enkla att översätta. Men mer än en tredjedel av huvudorden i listan, t.ex. *contact* och *issue*, är homografa. Dessutom är nästan samtliga huvudord i listan polysema. Som exempel kan substantivet *approach* nämnas. I *Norstedts stora engelsk-svenska ordbok* är de svenska ekvivalenterna till substantivet fördelade på sju olika betydelsemoment. Ordet *approach* kan därmed översättas med *närmande*, *annalkande*; *tillträde*; *tillvägagångssätt*; *inställning*; *första steg* etc. Ett annat exempel är adjektivet *significant* som har ett flertal svenska ekvivalenter, t.ex. *betydelsefull*, *viktig*, *meningsfull*, *betecknande för*; *markant* och *signifikant*. I praktiken vore det svårt att avgöra vilken av ekvivalenterna som skulle inkluderas i den svenska listan. Men även om det skulle gå att översätta huvudorden i AWL kan man givetvis också ifrågasätta om just dessa ekvivalenter utgör ett representativt svenskt urval.

Pilotstudien tydliggör ännu mer de begränsningar som AWL har. Listan innehåller många abstrakta ord som dessutom alltså är polysema, vilket torde leda till svårigheter för användarna. Det är t.ex. rimligt att de undrar om det är någon särskild betydelse hos orden i listan som är typiskt förekommande i akademiska sammanhang och vilken den i sådana fall är.

Med tanke på de problem som en översättning av AWL skulle medföra har vi valt att använda en liknande metod som Coxhead (1998), nämligen att ta fram en svensk ordlista utifrån en akademisk korpus. En sådan svensk korpus finns ännu inte, men en framtida sådan ska givetvis innehålla många olika, moderna svenska texter hämtade från vetenskapliga tidskrifter, kursböcker etc. För att resultatet ska bli en så allmän svensk akademisk ordlista som möjligt är det viktigt att dessa svenska texter behandlar olika universitetsämnen. Coxhead (2000: 219 f.) hämtade engelska akademiska texter från olika delar av världen och hon använde texter som redan ingick i existerande samlingar. För vår del är det annorlunda eftersom det svenska språket är så mycket mindre än det engelska. Engelskan dominerar också som undervisnings- och forskningspråk inom en rad ämnen på Sveriges universitet, i synnerhet inom naturvetenskapliga sådana. Det är alltså inte lika lätt att finna svenska tidskriftsartiklar m.m. inom

ämnen som medicin, fysik och kemi. Inom humaniora har dock svenska språket en starkare ställning. Med andra ord kan det vara svårt att uppnå lika goda resultat som Coxhead (1998) när det gäller de svenska texternas spridning över ämnesområden.

## 5. Två akademiska testkorpusar

I skrivande stund (juni 2011) har vi i testsyfte tagit fram två korpusar. Den första av dessa kallas hädanefter *Avhandlingskorpusen*. Den består i nuläget av 9 språkvetenskapliga doktorsavhandlingar som utarbetats vid Institutionen för svenska språket vid Göteborgs universitet och som är publicerade under 2000-talet. Avhandlingarna är främst skrivna inom forskarutbildningsämnet nordiska språk, men också inom svenska som andraspråk. Texterna som ingår i denna korpus har bearbetats en hel del; precis som Coxhead (2000: 219) har vi tagit bort litteraturlistorna, men vi har också rensat bort längre (utbrutna) citat och språkexempel. Avhandlingarna innehåller nämligen inte enbart de texttyper som vi är intresserade av; de inkluderar också längre engelska citat och t.ex. transkriberat ungdomsspråk som är mindre relevant för oss. Vidare har texterna i avhandlingskorpusen bearbetats språkteknologiskt; närmare bestämt har de tokeniserats, lemmatiserats och ordklasstaggats och de finns nu i SketchEngine (se vidare om detta verktyg i Kilgarriff et al. 2004). Avhandlingskorpusen är i nuläget på ca 800 000 löpande ord (tokens). Ett plus med denna korpus är att vi känner samtliga textförfattare väl och har fått rättigheter till hela texterna, vilket är en fördel eftersom vi också vill att vår akademiska korpus i framtiden ska bli fritt tillgänglig. Det har dock varit mycket tidskrävande att manuellt ta bort citat, exempel m.m. från de ingående texterna. Vidare är användningen av korpusen i nuläget begränsad eftersom den bara innehåller språkvetenskapliga texter. Den blir fullt användbar för vårt specifika ändamål först när andra delkorpusar med texter från andra universitetsämnen har byggts upp.

Den andra testkorpusen, som hädanefter kallas *SweAc*, har tagits fram av Håkan Jansson vid Institutionen för svenska språket, Göteborgs universitet. Den har utarbetats med hjälp av verktyget WebBootCat som ingår i Sketch Engine och med vars hjälp man kan skapa en korpus utifrån Internetdokument. (Verktyget beskrivs närmare av Baroni et al. 2006.) SweAc-korpusen innehåller olika typer av i huvudsak moderna svenska akademiska texter. De texter som ingår i denna korpus är i skrivande stund kompletta, dvs. litteraturförteckningar m.m. har inte tagits bort. De är heller inte språkteknologiskt bearbetade. I nuläget vet vi också mycket litet om exakt vilka ämnen de hämtade texterna behandlar, men detta går givetvis att kontrollera om man granskar varje enskilt

dokument. Däremot är det inte lika lätt att ta fram bakgrundsdata om skribenterna m.m. Vi har med andra ord i nuläget inte alls samma kontroll över innehållet i SweAc-korpusen, men den är å andra sidan betydligt större än avhandlingskorpusen; i nuläget utgörs den av ca 20 miljoner tokens. En fördel med verktyg som WebBootCat är att man kan sätta samman förhållandevis stora korpusar på en begränsad tid; det tog nämligen betydligt kortare tid att ta fram denna korpus i det skick som den är nu.

Vi är alltså intresserade av lexikala enheter som dels har hög frekvens, dels god spridning i de texter som kommer att ingå i vår akademiska korpus, men som *inte* är högfrekventa i allmänspråket. För att särskilja ett mer akademiskt engelskt ordförråd uteslöt Coxhead (1998), som redan nämnts, de ord som ingår i GSL (West 1953) ur sin akademiska ordlista. Eftersom det inte finns en sådan lista för svenskans del (jfr dock Allén 1972), kan man jämföra ordförrådet i de akademiska texterna med ordförrådet i mer allmänspråkliga korpusar. En sådan jämförelse kan göras med hjälp av Keywords-funktionen i Sketch Engine. I vårt fall kan man tänka sig att man jämför ordförrådet i våra akademiska delkorpusar med ordförrådet i befintliga svenska korpusar eller med en ny korpus som vi själva håller på att utarbeta och som innehåller moderna Norstedtsromaner.

## 6. Några första resultat

En inledande granskning av de två testkorpusarna tyder på intressanta resultat. Som exempel kan några av de mest frekventa lexikala enheterna i SweAc-korpusen anföras. Dessa tydliggör några väsentliga skillnader mellan de ord som ingår i AWL och de ord som kan tänkas ingå i en svensk akademisk ordlista. Bland de aktuella lexikala enheterna återfinns:

*angreppssätt, avgränsning, intervju, reliabilitet, metod, studie, tillvägagångssätt, datainsamling, undersökning, förförståelse, tillförlitlighet, implementering, generaliserbarhet, referensram, strukturering, frågeställning, poängtera, problemlösning*

Många av orden, t.ex. *avgränsning* och *tillförlitlighet*, är av germanskt ursprung. I AWL är däremot ca 80 % av orden av grekisk eller latinsk härkomst (Coxhead 2000: 228 f.). Vidare är som synes många av orden sammansättningar eller sammansatta avledningar.

Man kan också fundera kring möjligheterna att fastställa vilka ordfamiljer dessa ord ingår i. Som exempel väljer vi substantivet *avgränsning*. Några möjliga familjemedlemmar skulle kunna vara *avgränsa*, *avgränsa sig*, *avgränsad*, *avgränsande*, *avgränsning*, *avgränsbar*, *avgränsbarhet*, *oavgränsbar*, *oavgränsbarhet* etc.

Det är uppenbart att listan kan göras lång bara genom att man tillför olika produktiva affix. Man kan dock ifrågasätta användbarheten av dessa ordfamiljer. Även om familjemedlemmarna kan kopplas till huvudordet genom produktiva regler och även om familjemedlemmarna skulle vara relaterade till varandra i det mentala lexikonet (se avsnitt 3 ovan), kan man diskutera vissa medlemmars värde för inlärarna. Många potentiella svenska ordfamiljsmedlemmar, som t.ex. *oavgränsbar* och *oavgränsbarhet*, torde ha synnerligen låg frekvens oavsett vilken korpus man söker i.

## 6. Sammanfattande diskussion

I vårt föredrag presenteras ett nytt projekt som har som mål att ta fram en allmänt tillgänglig nätbaserad svensk *Akademisk ordlista* (AO). Ordlistan ska ge en fingervisning om vilka ord som kan vara värdefulla att lära sig för universitetsstudier med olika språklig bakgrund. På så sätt kan ordlistan främja akademiska studier. I artikeln har vi lagt stor vikt vid att presentera en viktig föregångare, nämligen Coxheads AWL (1998). Vi tror att det sätt som den utarbetades på, dvs. utifrån en speciellt framtagna korpus, är lämpligt. Men det återstår dock en hel del frågor kring en framtida svensk akademisk korpus, bl.a. hur stor den måste vara, hur stor spridning vi kan uppnå vad gäller ämnesområden etc. (se vidare Hyland & Tse 2007). Möjligen kan olika typer av delkorpusar, som tagits fram på olika sätt och vars innehåll är lite olika, komplettera varandra.

Innehållet i den kommande ordlistan måste givetvis anpassas till svenska språkets strukturella egenskaper. Vidare måste bruket av ordfamiljer tänkas igenom noggrant. Om listan ska kunna användas i många olika användningssituationer krävs också mer information om de ingående orden. Samtidigt tar det givetvis längre tid att ta fram en ordlista om t.ex. betydelser ska infogas.

Det är också önskvärt att det finns tydliga kopplingar mellan den akademiska korpusen och själva ordlistan, så att användarna bl.a. kan se exempel på hur de ingående orden används i konkreta kontexter. Det kan även bli aktuellt med frekvensangivelser. Det är dock viktigt att listans innehåll och funktioner testas fram genom olika användarundersökningar, i undervisningssituationer (med ordtester etc.) och även genom metodiskt arbete inom verksamheter som Språkhandledningen inom Göteborgs universitet.

Det återstår alltså många frågor – men intressanta sådana. Arbetet med en svensk akademisk korpus och AO har tvärvetenskaplig karaktär. Det inkluderar många forskningsområden, t.ex. andraspråksinlärning, språkteknologi, lexikologi, lexikografi, textforskning och språkpolitik. Resultatet av vårt arbete kan

också få nordiska förgreningar. Det har bildats ett nätverk för att det i framtiden ska utarbetas akademiska ordlistor för flera nordiska språk. I maj 2011 arrangerades en första workshop i Göteborg med inbjudna forskare och doktorander från Sverige, Norge och Danmark. Planerna är att de metoder som används för framtagning av en svensk ordlista även ska kunna tillämpas för en norsk och en dansk motsvarighet.

## KÄLLOR OCH LITTERATUR

- Allén, Sture, 1972: *Tiotusen i topp*. Stockholm.
- AWL – How to use the list: <http://www.victoria.ac.nz/lals/resources/academicwordlist/howto.aspx> (september 2011)
- AWL – Useful Links: <http://www.victoria.ac.nz/lals/resources/academicwordlist/links.aspx> (september 2011)
- Baroni, Marco, Adam Kilgarriff, Jan Pomikálek & Pavel Rychlý, 2006: *WebBootCaT: a web tool for instant corpora*. I: E. Corino, C. Marelló & C. Onesti (eds.), *Proceedings XII Euralex International Congress*. Torino, Italia, September 6–9, 2006. Alessandria. S. 123–131.
- Bjørneset, Tove, 2001: Introduksjon til ordboksprosjektet *NORDLEXIN-N*. I: M. Gellerstam, K. Jóhannesson, B. Ralph & L. Rogström (red.), *Rapport från Konferens om lexicografi i Norden Göteborg 26–29 maj 1999*. (Nordiska studier i lexicografi 5.) Göteborg. S. 44–53.
- Carlund, Carina & Sören Andersson: [http://www.utbildning.gu.se/digitalAssets/1344/1344562\\_minioralista.pdf](http://www.utbildning.gu.se/digitalAssets/1344/1344562_minioralista.pdf) (september 2011)
- Coxhead, Averil, 1998: *An academic word list*. (English Language Institute Occasional Publication no 18.) Wellington.
- Coxhead, Averil, 2000: *A new academic word list*. I: *TESOL Quarterly*, 34:2. S. 213–238.
- Coxhead, Averil, 2002: *The Academic Word List: A corpus-based word list for academic Purposes*. I: B. Ketteman & G. Marks (eds.), *Teaching and Language Corpora (TALC) 2000 Conference Proceedings*. Atlanta: Rodopi. S. 73–89.
- Hyland, Ken & Polly Tse, 2007: *Is There an “Academic Vocabulary”?* I: *TESOL Quarterly*, 41:2. S. 225–253.
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrz & David Tugwell, 2004: *The Sketch Engine*. I: G. Williams & S. Vessier (eds.), *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France July 6.–10. 2004*. Lorient. S. 105–116.
- Lindberg, Inger & Sofie Johansson Kokkinakis, 2007: *OrdiL. En korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år. (Rapporter om svenska som andraspråk 8.)* Red. I. Lindberg & S. Johansson Kokkinakis. Göteborg.

- Lindén, Krister & Lauri Carlson, 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica* 17. S. 119–140.
- Longman dictionary of contemporary English 2009: 5 uppl. Harlow.
- Martola, Nina, 2011: FinnWordNet och kulturbundna ord. I: *LexicoNordica* 18. S. 111–133.
- Nation, I. S. Paul, 2001: *Learning vocabulary in another language*. Cambridge.
- Norstedts stora engelsk-svenska ordbok 2000: 3 utökade uppl. Stockholm.
- Språkhandledning: <http://www.utbildning.gu.se/student/sprakhandledning> (september 2011)
- Studera.nu: <https://www.studera.nu/studera/703.html> (september 2011)
- The Academic Word List: <http://www.victoria.ac.nz/lals/resources/academicwordlist/> (september 2011)
- West, Michael, 1953: *A General Service List of English Words*. London.
- Xue, Guoyi & Nation, I. S. Paul, 1984: A university word list. *Language Learning and Communication* 3, 2. S. 215–229.

Anmärkning: Emma Sköldberg har skrivit sin del av bidraget inom ramen för sin forskartjänst som finansieras av Knut och Alice Wallenbergs Stiftelse.

*Emma Sköldberg*

Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet  
emma.skoeldberg@svenska.gu.se

*Sofie Johansson Kokkinakis*

Institutet för svenska som andraspråk/Språkbanken, Institutionen för svenska språket,  
Göteborgs universitet  
sofie.johansson.kokkinakis@svenska.gu.se



# Svordomar i ett lexikografiskt perspektiv

*Ulla Stroh-Wollin*

During the work on a list of swear words found in Swedish dramas from the 18th, 19th and 20th centuries, it became clear that the art of swearing has hitherto not received much attention from Swedish lexicographers. An especially intriguing challenge is how to describe the more or less conventionalized phrases and constructions of which swear words are a part. The specific contribution of a swear word phrase to the meaning of an utterance is not always obvious and requires a very careful analysis. In some cases, phrases that look quite different are pragmatically and syntactically interchangeable – sometimes fully so, sometimes partly. On the other hand, even minor differences between two phrases may have a considerable impact on their meaning. Thus, the lexicographer's challenge, besides explaining every single expression, may include accounting for the relations between different words and phrases as well. Further, information on the syntactic functions that swear words and swear word phrases can have is normally not included in dictionaries, but may in fact be very valuable to their users. However, several syntactic functions are more or less fulfilled exclusively by swear words. In these cases, grammatical labels are missing, swear words being neglected even by traditional grammar.

*Nyckelord:* svordomar, svordomskonstruktioner

## Bakgrund

Under ett par år av detta sekels första decennium ägnade jag en stor del av min uppmärksamhet åt svordomarna i en korpus med dramatexter från 1700-, 1800- och 1900-tal som sammanställts vid Institutionen för Nordiska språk i Uppsala.<sup>1</sup> Primärt var jag ute efter att kunna studera de svenska svordomarnas diakrona utveckling med hänsyn till deras plats i grammatik och lexikon. I ett par skeden av arbetet tänkte jag att det var lämpligt att först ta del av svenska

---

<sup>1</sup> Projektet Svenska dramadialog under tre sekler och dess korpus presenteras i Melander Marttala & Östman 2000 resp. Melander Marttala & Strömquist 2001. En del resultat från projektet kan man ta del av i Strömquist (red.) 2003 och Thelander 2007.

ordböckers behandling av svordomarna. Det gällde dels i jakten på en definition av begreppet svordom och en uppfattning om vilka enskilda ord som allmänt kan anses ha svordomspotential, dels, senare, i sökandet efter en modell att sortera och beskriva de svordomar jag funnit. Vid båda tillfällena fann jag emellertid att ordböckerna egentligen inte var till så stor hjälp för mina behov.

Det blev t.ex. uppenbart för mig att ordboksredaktionerna knappast använde beteckningen svordom på något särskilt precist sätt. När jag med ordböckernas hjälp försökte ringa in ett antal enskilda potentiella svärords status möttes jag omväxlande av beteckningarna *svordom* och *kraftuttryck*, ibland med modifikationer som *mild*, *mildare* eller *lindrigare*, utan någon alldeles uppenbar princip bakom variationen. Ordet *tusan* t.ex. betecknas som ”tämligen milt kraftuttryck” i Nationalencyklopedins ordbok (NEO, 1995), men förses med beskrivningen ”i svordomar” i 13:e upplagan av Svenska Akademiens ordlista (SAOL, 2006). En del av de ord jag testade saknade helt karaktär av det nämnda slaget. Ordet *skit* (använt som interjektion) t.ex. markerades inte som vare sig svordom eller kraftuttryck i någon av de konsulterade ordböckerna. (Se vidare Stroh-Wollin 2008: 22 ff.)

När jag så småningom skulle sortera mitt material uppstod tanken att samtidigt åstadkomma en regelrätt ordlista med de funna svordomarna. Med tanke på att svordomar ofta utgörs av idiomatiska uttryck om flera ord kändes det i det sammanhanget naturligt att först studera hur man tagit sig an sådana förbindelser i Svenskt språkbruk (Ordbok över konstruktioner och fraser, 2003). Men det visade sig att sorteringen av svordomsuttryck i Svenskt språkbruk är tämligen godtycklig. I t.ex. artikeln *fan* sammanförs olika fraser som innehåller ordet *fan* under beskrivningar som ”allmänt förstärkande”, ”i uttr. när man säger emot ngn el. vägrar ställa upp på ngt” eller ”i uttr. för förvåning eller överraskning” etc., utan att det är uppenbart vilka fraser som bör hamna var. Uttrycket *ta mig fan* t.ex. dyker dessutom upp på flera ställen, dock inte under ”allmänt förstärkande”. Där hittar man i stället bl.a. frasen *fan i mig*, som enligt mina rön är både grammatiskt och pragmatiskt i stort sett likvärdig med *ta mig fan*.

Mina studier i svordomarnas plats i lexikon och grammatik fick helt enkelt bedrivas utan stöd i befintliga ordböcker.<sup>2</sup> Nu när projektet väl är avslutat

<sup>2</sup> Jag bortser här ifrån att jag självfallet har funnit förklaringar till äldre svordomar i Svenska Akademiens ordbok (SAOB). Jag har också hämtat etymologiska uppgifter både i SAOB och i Hellquists Svensk etymologisk ordbok (1948), även om jag upptäckte att sådana uppgifter ibland måste tas med en nypa salt, vilket föranledde en del kompletterande studier i diskussioner rörande olika svordomars ursprung. (Se vidare Stroh-Wollin 2008, kap. 5.)

tänkte jag att man kan vända på steken och fundera över om resultatet i stället kan bidra med något till lexikografin. Det är temat för denna artikel.

## Begreppet svordom

För att kunna exciperera min dramakorpus var jag tvungen att bestämma mig för en operationell definition av begreppet svordom. Det är inte av nöden att jag i detta sammanhang går in på detaljer i den problematiken. (Den intresserade hänvisas till Stroh-Wollin 2008, kap. 2–3.) Dock vill jag gärna göra klart att jag inte alls beaktat graden av anstötlighet när jag identifierat mina svordomar. Det innebär att rätt många av de uttryck jag excerperat, och som jag faktiskt betraktar som svordomar, kan te sig väldigt oskyldiga, särskilt för svenskar av idag.

Ett kriterium som tar fasta på hur ”fult” ett visst ord eller en viss fras är fungerar naturligtvis inte när man arbetar med historiskt material, eftersom uttryck av de här slaget gärna omvärderas från tid till annan. Men det finns ännu mer avgörande skäl att bortse ifrån den moraliska värderingen. Det går nämligen inte att skilja ut fulare svordomar från mindre anstötliga på grammatiska, lexikala eller pragmatiska grunder – frasen *fy sjutton* t.ex. signalerar lika bra som *fy fan* att man menar ’usch’ snarare än ’aja baja’ som normalt är innebörden av ett enkelt *fy* – men däremot går det någorlunda bra att avgränsa de mer eller mindre fula uttryck som jag kallar svordomar från andra ord och uttryck i lexikonet.

En svordom, enligt mitt sätt att se det, består av eller innehåller ett ord som (genom att det hämtats från en viss betydelsedomän, som religion, avföring, sexualitet m.fl.) kan ge det en viss laddning, t.ex. *gud*, *kors*, *djävul*, *helvete* eller *skit*. Ordet eller ordförbindelsen tolkas inte bokstavligt, men normalt heller inte metaforiskt. Använder man t.ex. *fan* eller *skit* som interjektioner fungerar ordens referentiella innebörd endast som klangbotten för den frustration man vill uttrycka; frasen *gud ske lov* är inte längre en religiös formel utan ett allmänt uttryck för t.ex. lättnad; osv. Till svordomarna räknar jag även de ord och uttryck som uppstår då man byter ut ett primärt kraftord mot ett eufemistiskt, dvs. ett ord som inte har någon användning utanför svordomarna, t.ex. *hundan*, *attan* eller *tusan* i stället för *fan* eller *jösses* i stället för *Jesus*.

## Ordlista över dramernas svordomar

I arbetet med dramakorpusens svordomar bestämde jag mig för att samla materialet, de exciperade svordomarna, i en ordlista. (Ordlistan utgör en separat del

av rapport nr 10 i rapportserien Svensk dramadialog, Stroh-Wollin 2008.) Eftersom det rör sig om en materialredovisning skiljer sig ordlistan på vissa punkter från vanliga ordlistor, t.ex. genom att endast ord och uttryck som verkligen är belagda i materialet kommit med och genom att kvantitativa uppgifter av antalet förekomster av de olika uttrycken, drama för drama, redovisas. Jag ville emellertid också att ordlistan skulle kunna användas som uppslagsverk och var därför angelägen om att utforma den efter genomtänkta lexikografiska principer.

En konsekvens av ordlistans dubbla roller är att artiklarna innehåller mycket information. Här ges bl.a. både grammatiska kategoriseringar av olika uttryck och deras faktiska användning i materialet, preciserat som syntaktiska funktioner. Om man t.ex. vill veta vilka adjektiv som förekommer i dramerna kan man göra en sökning på adjektiv. Ett typiskt svordomsadjektiv är *djävla*. Det används dels attributivt (*den där jävla bönan*) med allmänt pejorativ innebörd, dels som gradadverbial (*så typiskt jävla svenskt*) med allmänt förstärkande funktion. Av artikeln framgår dessutom att adjektivet *djävla* är oböjligt, vilket förklaras av den etymologiska uppgiften ”av (fsv.) gen.pl. av *djävul*”. Ett annat adjektiv som faller ut vid sökningen är *förbannad* som används i samma funktioner som *djävla* men dessutom med två mera pregnanta betydelser, där predikativ användning dominerar: dels ’besvärlig’ (*Det är rent förbannat!*), dels ’arg’ (*Du är förbannad för att ...*).

Ett annat exempel på vilken typ av information man kan hämta i ordlistan är hur följande satsformade förbindelser, vilka ytligt sett påminner om varandra, ändå skiljer sig grammatiskt och pragmatiskt: *det var fan!*; *det var fan vad/att ...*; *det var väl fan (att/vad ...)*. Den första, som i regel uttrycker att talaren är förvånad eller imponerad, fungerar, får man veta, endast med verbet i preteritum; de båda övriga kan även ha tempus presens. Beträffande den andra förbindelsen framgår att den förutsätter en expressiv komplementsats vars innehåll talaren både kan gilla och ogilla (*det var fan vad hon kunde kuta/vad hon tjarar*). Om den tredje förbindelsen får man veta att den innehåller ett adverbialt led som signalerar någon form av frustration och kan kompletteras med en expressiv sats som förklaring till ogillandet (*det var väl då!(som) själva fan (också) att ...*). Dessa upplysningar ges i artikeln *fan*, men ordlistan innehåller även en s.k. översiktsartikel som tar upp de nämnda tre och två andra konstruktionstyper som inleds med *det var (är)*. Översiktsartikeln förklarar skillnaderna och går också igenom vilka kraftord som kan användas, och de facto används i korpussens belägg, i de olika konstruktionstyperna.

Många av de traditionella svenska svordomarna är konventionaliserade användningar av olika språkhandlingsformler, t.ex. *gud ske lov* som nämndes ovan. När så är fallet anges vid redogörelsen för förbindelsen i fråga vilken typ

av bakomliggande språkhandling det rör sig om. Följande språkhandlingsetiketter används i ordlistan: edsfras (t.ex. *vid gud, min själ*), böneformel (t.ex. *gud bevara mig*), lovprisning (t.ex. *gud ske lov*), vädjan till ett kristet sinnelag (t.ex. *för guds skull*), vokativ fras (t.ex. *gode gud, herre jösses*), diabolisk ed (t.ex. *ta mig fan, djävlar i mig, förbanne mig*), förbannelse (t.ex. *fan ta dig, förbannelse över ...*), illasinnad uppmaning (t.ex. *dra åt helvete*), illasinnad önskan (t.ex. *för bödeln i våld med ...*), diabolisk lovprisning (t.ex. *tacka fan för ...*).

Naturligtvis ska språkhandlingsformeln inte uppfattas bokstavligt, men ursprunget till svordomen påverkar ändå dess innebörd eller pragmatiska funktion. Och inte nog med det; det påverkar även svordomens syntaktiska egenskaper. Några exempel på detta tas upp i nästa avsnitt.

## Några svordomar – en jämförelse

Eftersom svordomar till sin karaktär är emfatiska och emotionella och oftast saknar en mera pregnant referentiell betydelse, kan det ibland vara svårt att fånga vad ett uttryck egentligen tillför det yttrande där det dyker upp eller om olika uttryck är pragmatiskt likvärdiga eller skiljer sig åt. Detta gäller förstås i all synnerhet uttryck som dyker upp i samma syntaktiska funktion. Många svordomar kan t.ex. fungera som modala adverbial. I sådana fall måste man se upp med om en viss innebörd verkligen tillförs yttrandet genom svordomen eller om den faktiskt föreligger redan i den grundläggande propositionen.

Ett misstag på den här punkten kan föreligga när Svenskt språkbruk anför satsen ”ni är ta mig fan inte kloka!” som ett exempel på att *ta mig fan* används ”i uttr. när man säger emot ngn eller vägrar ställa upp på ngt” (a.a. artikeln *fan*, moment 5). Här framgår det ju att talaren opponerar sig även om man tar bort svordomen. Fastän uttrycket *ta mig fan* här i och för sig används i oppositionell kontext, så är exemplet alltså föga upplysande beträffande svordomens roll i sammanhanget.

Det hade för övrigt varit möjligt att byta *ta mig fan* i ”ni är ta mig fan inte kloka!” mot t.ex. *fan i mig* eller *för fan*. Inte heller dessa uttryck påverkar det grundläggande budskapet (’ni är inte kloka’). Men en intressant fråga är onekligen vad de olika svordomarna faktiskt skulle tillföra yttrandet och om bidraget skulle vara detsamma eller ej. Låt oss därför göra en jämförande analys av de nämnda uttrycken. Vi kan starta med meningarna i (1a–c), där de tre svordomarna uppträder som modalt adverbial (liksom *ta mig fan* i det just diskuterade exemplet). Meningarna är identiska sånär som på svordomarna.

- (1) a. De har **ta mig fan** aldrig varit gifta.
- b. De har **fan i mig** aldrig varit gifta.

c. De har **för fan** aldrig varit gifta.

Jag vill nu hävda att svordomarna i (1a–b) är pragmatiskt tämligen lika, medan den i (1c) har en lite annan pragmatisk funktion. Det är inte säkert att skillnaden är uppenbar för alla språkbrukare i den givna kontexten. Om man däremot försöker använda de respektive svordomarna tillsammans med imperativer blir skillnaden tydligare. Uttrycken *ta mig fan* och *fan i mig* går inte ihop med imperativer, se (2a–b). Frasen *för fan*, liksom även *för guds skull*, passar däremot utmärkt i imperativiska satser, se (3a–b).

- (2) a. \*Kom **ta mig fan** i tid nästa gång.  
 b. \*Kom **fan i mig** i tid nästa gång.
- (3) a. Kom **för fan** i tid nästa gång.  
 b. Kom **för guds skull** i tid nästa gång.

Skillnaden mellan de olika svordomarna märks också därigenom att *ta mig fan* och *fan i mig* kan användas före en *om*-sats för att uttrycka motsatsen till innehållet i bisatsen (även om denna konstruktion inte är så vanlig längre), medan något motsvarande är helt omöjligt med *för fan* och *för guds skull*, jfr (4a–b) och (5a–b).

- (4) a. **Ta mig fan** om jag tänker stanna här.      - 'Jag tänker inte stanna här.'  
 b. **Fan i mig** om jag tänker stanna här.
- (5) a. \***För fan** om jag tänker stanna här.  
 b. \***För guds skull** om jag tänker stanna här.

De här grammatiska skillnaderna hänger samman med uttryckens pragmatiska funktioner, som utgår ifrån att *ta mig fan* och *fan i mig* är (diaboliska) edsformler, medan *för fan* är en blasfemisk variant på *för guds skull*, en fras som uttrycker en vädjan till en kristen medmänniska.

Alla edsformler, vid sidan om *ta mig fan* och *fan i mig* även *förbanne mig*, *min själ* och *vid gud*, fungerar på samma sätt. De kan användas för att bedyra sanningshalten i ett påstående eller talarens allvar med en utfästelse, utifrån logiken "må fan ta mig om jag inte talar sanning/gör som jag lovar etc.", jfr exemplen i (4a–b). De kan även understryka ett värdeomdöme. Men de fungerar inte när det inte finns någon proposition som kan hävdas, t.ex. i imperativiska meningar. De är starkt emfatiska till sin karaktär med emfasen primärt riktad mot innehållet i propositionen. Sekundärt kan de signalera olika känslor, t.ex. att innehållet i propositionen är förvånande eller något man ogillar.

Det vädjande ursprunget i *för guds skull* innebär att frasen är tydligt emotiv från början och företrädesvis emfatisk visavi den tilltalade. Det gäller även de

blasfemiska varianterna *för fan* och *för helvete*. Detta drag låter sig lätt förenas med imperativer. Uttrycken *för fan* och *för helvete* signalerar normalt starkare irritation än *för guds skull*, men för alla dessa fraser gäller att de kan uttrycka både irritation och vädjan. Irritationen kan även spilla över på det sakförhållande som en sats uttrycker.

## Lexikografiska svårigheter och möjligheter

Man kan nu fråga sig vilka lärdomar av projektet om dramernas svordomar som kan vara av nytta för en lexikograf. Till en början hoppas jag då att det blivit alldeles uppenbart att man inte kan utgå ifrån ett enskilt svärord, t.ex. *fan*, och placera in olika exempelmeningar som innehåller flerordsförbindelser som i sin tur innehåller detta ord under mer eller mindre godtyckligt valda rubriker. Man måste nog analysera varje specifik förbindelse för sig. I vissa fall är det så att ganska små detaljer har stor betydelse för hur vi använder och tolkar en viss förbindelse. Ganska ofta finns en bakomliggande logik, t.ex. när svordomen utgörs av en konventionaliserad språkhandlingsformel, som kan förklara dess pragmatiska och syntaktiska funktioner. Och även om svordomarnas emotiva uttryck kan uppfattas som glidande, förefaller det som om ursprunget ändå sätter tydliga gränser för hur de används. Detta innebär att vissa drag kan identifieras som primära och inherent, att skilja från mera tillfälliga innebörder som kan tolkas in i specifika kontexter.

Den analys jag förespråkar visar dessutom att olika svordomars funktionella utbytbarhet läggs i dagen; olika edsformler t.ex. tycks i stort sett vara helt utbytbara mot varandra. Å andra sidan bygger dessa formler sällan på samma kraftord, vilket ur ordboksperspektivet innebär att de normalt inte behandlas inom en och samma artikel. Här föreligger ändå möjlighet till korsreferenser om man föredrar att samla en utförligare redovisning till ett ställe.

Korsreferenser kan naturligtvis användas också när det gäller grupper av enskilda kraftord som kan användas i samma förbindelsetyper. Ordet *fan* kan i de flesta fall bytas mot t.ex. *attan*, *hundan*, *katten*, *sjutton* och *tusan* och naturligtvis mot alla fonetiska varianter som *fanken*, *fasen*, *faderullan* etc. På samma sätt fungerar *helsike*, *helskotta* etc. i stället för *helvete*. Osv.

En komplikation, som jag inte varit inne på, är de fall när det råder en partiell utbytbarhet. Man kan t.ex. säga såväl *var fan* som *var i helvete* som *var i herrans namn*, dvs. valfritt förstärka ett frågeord med *fan*, *i helvete* eller *i herrans namn*. Men det är företrädesvis *fan* (och dess ekvivalenter) som fungerar som förstärkning till adverbena *inte*, *nog* och *visst* (*inte fan*, *nog fan*, *visst fan*); alla accepterar inte *i helvete* eller *i herrans namn* efter de här adverbena. För att kom-

plicera saken ytterligare kan man konstatera att både *fan heller* och *i helvete heller* fungerar som någon sorts negation, medan *i herrans namn* inte används i negationer. Om man vill ange pragmatisk synonymi måste man alltså i dessa fall vara noga med i vilka kontexter uttrycken är utbytbara.

Eftersom den ordlista jag upprättade i mitt svordomsprojekt också skulle tjäna som materialsortering för vidare bearbetningar var det ytterst väsentligt att även de olika svordomarnas syntaktiska funktioner redovisades. Sådana uppgifter förekommer endast sparsamt i vanliga ordböcker. Jag menar emellertid att denna information i vissa fall borde vara mycket givande även för en vanlig ordboksanvändare. Jag har t.ex. ovan varit inne på att adjektiviska svordomar gärna används som attribut och gradadverbial, medan en predikativ användning förutsätter en mer pregnant betydelse, som t.ex. då *förbannad* betyder 'arg'.

Det har också framkommit att edsformler både kan fungera som satsintegrerade modala adverbial (ex. 1a och 1c) och föregå *om*-satser (ex. 4a–b), vilket är två helt olika användningar. Av både teoretiska och empiriska skäl kan man förstå den satsintegrerade användningen som en produkt av en successiv grammatikalisering där konstruktionen med *om*-sats utgör startpunkten, utifrån logiken att man sätter sin själ i pant – *om* man inte står vid sitt ord. (Se vidare Stroh-Wollin 2008: 92 ff.) Samma logik förklarar också varför konstruktionen med *om*-sats är unik för just edsformler.<sup>3</sup> En besläktad – också specifik – användning av edsfraser illustreras i (6a). Denna användning och den i (6b) kan ses som steg på vägen i grammatikaliseringsprocessen.

- (6) a. **Förbanne mig**, har jag inte sett den mannen förut.  
 b. **Förbanne mig**, jag har sett den mannen förut.

I (6a) klargör edsfrasen att den följande satsen, som alltid är negerad, ska uppfattas som en retorisk fråga, vars svar är *jo*. Helheten i (6a) betyder alltså 'jag har sett den mannen förut'. I (6b) är budskapet detsamma, men här framgår detta entydigt av den påståendesats som följer edsformeln. Svordomen utgör här alltså endast ett emfatiskt, fritt annex i meningen (varifrån vägen till den satsintegrerade användningen inte är så lång).

Ett problem i arbetet med min ordlista över svordomar var att den traditionella grammatiska terminologin inte räckte till, och framför allt gällde det just den syntaktiska analysen. Vilken satsdelsbeteckning ska man t.ex. ge svordomar

<sup>3</sup> Konstruktionen edsformel + *om*-sats förutsätter att helheten uppfattas som ett hävdande (av motsatsen till bisatsens proposition, jfr. (4a–b) ovan). Det är den konstruktionen jag avser här. Det förekommer även att interjektionella svordomar föregår *om*-satser. I dessa fall kan svordomen bytas mot *tänk* och bisatsen uppfattas som hypotetisk: jfr *Herregud om någon kommer in när jag sover. – Tänk om någon kommer in när jag sover.*



som förstärker frågeord eller edsfraser som föregår *om*-sats och retoriska frågor? Jag har löst sådana problem helt enkelt med beskrivande etiketter som "efter wh-ord" resp. "före 'falsk' *om*-sats" och "före 'falsk' neg. ja/nej-fråga" (där "falsk" är ett sätt att markera att den proposition man "hävdar" är icke-negerad om satsen efter edsformeln är negerad och vice versa; den retoriska ja/nej-frågan kan endast vara negerad, men *om*-satsen kan vara både och).

## Slutord

Till sist kan man naturligtvis fråga sig vilken betydelse man ska tillmäta svordomarna inom lexikografin, liksom förstås i språkforskningen överlag. Jag tror emellertid inte att man kan avfärda svordomarna bara som en marginell del av lexikonet. Däremot har de sin plats i talat språk, mycket mer än i skrift, vilket kanske delvis förklarar att de etablerade ordböckernas beskrivningar av svordomar kan te sig en aning förströdd. Som många har påpekat har studiet av tal-språket länge fått stå tillbaka för studiet av det skrivna språket, t.ex. i den grammatiska forskningen. Att svordomarnas egenart då inte uppmärksammats av lexikograferna är egentligen bara följdriktigt.

Samtidigt vill jag gärna framhålla att ordböckerna inte kommer att bli bättre på att beskriva svordomar endast därigenom att man blir mer uppmärksam på talspråk i största allmänhet. Svordomarna utgör verkligen en egen, någorlunda välavgränsad, del av vårt lexikon. Att förstå deras egenart kräver att man går i närkamp med dem. I den kampen kan man möjligen också finna inspiration i en del konstruktionsgrammatisk forskning. Att finna nya vägar att beskriva "orden i kombination" (jfr Svenskt språkbruk, s. 10) är fortsatt en utmaning för lexikografin.

## LITTERATUR

- Hellquist, Elof, 1948: Svensk etymologisk ordbok, 3:e uppl. Lund.  
 Melander Marttala, Ulla & Strömquist, Siv, 2001: Korpusen svensk dramadialog. Användarhandbok. Svensk dramadialog nr 2 (= FUMS Rapport nr 202). Uppsala.  
 Melander Marttala, Ulla & Östman, Carin, 2000: Svensk dramadialog under tre sekler – en projektbeskrivning. Svensk dramadialog nr 1 (= FUMS Rapport nr 200). Uppsala.  
 NEO = Nationalencyklopedins ordbok 1–3. 1995–96. Höganäs.  
 SAOB (Svenska Akademiens ordbok) = Ordbok över svenska språket utg. av Svenska Akademien. 1893 ff. Lund.  
 SAOL = Svenska Akademiens ordlista, 13:e uppl., 2006. Stockholm.

- Stroh-Wollin, Ulla, 2008: Dramernas svordomar. Svensk dramadialog nr 10. (= FUMS Rapport nr 224). Uppsala.
- Strömquist, Siv (red.), 2003: Svenska på scen. Språk och språkanvändning i svensk dramatekt från tre sekler. Ord och stil (Språkvårdssamfundets skrifter) 34.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser, 2003. Utarb. av Svenska språknämnden. Stockholm.
- Thelander, Mats, 2007: "De där befängda moderna idéerna ...". Svenskt dramapåk och det moderna genombrottet. I: Wollin, Lars, Saarukka, Anna & Stroh-Wollin, Ulla (red.): Studier i svensk språkhistoria 9. Det moderna genombrottet – också en språkfråga? Skrifter från Svenska institutionen vid Åbo Akademi. 5/2007. S. 35–63.

*Ulla Stroh-Wollin*

Institutionen för nordiska språk vid Uppsala universitet  
ulla.stroh-wollin@nordiska.uu.se

# Folklig dialektlexikografi 1991–2011

*Eva Thelin*

This article investigates Swedish dialect dictionaries published from 1991 to 2011. Generally, these dictionaries have been compiled by lay lexicographers and show the vocabulary of a dialect spoken in one or a few adjacent parishes.

Right from the start in the 18th century, the main purpose of publishing dialect dictionaries has been to document dialects and preserve them for posterity. As a consequence, the dialects were always presented as the source language. In 1991, however, a dictionary emerged using standard Swedish as the source language, thereby focusing on the needs of users who wanted to learn the dialect. Twenty years afterwards it can be determined whether this dictionary was an isolated phenomenon or if other similar publications have followed. Out of 115 dictionaries, I have identified 19 which in my opinion tend to learners' needs in various ways. Quite a few of them treat dialects that are incomprehensible from a standard Swedish perspective.

Another apparent change taking place during this period is the development of web-based dictionaries. I have been able to trace 59 presenting Swedish dialects. They differ from print publications in several ways. Two of the most obvious differences, concerning the number of words and the use of illustrations, are addressed and discussed. Somewhat surprising is the discovery that web publishers rarely invite interactivity. Nor do they provide for performing various types of searches or listening to sound samples.

*Nyckelord:* dialektlexikografi, svenska dialekter, webbpublicering

## 1. Inledning

Den här artikeln behandlar de dialektordböcker som publicerats i Sverige under de senaste tjugo åren. Ordböckerna beskriver vanligen ordförrådet hos en dialekt som talas i en eller några få intilliggande socknar. I vissa fall omfattas ordförrådet i ett större område, ibland ett helt landskap. De som publicerat ordböckerna är oftast helt oskolade lexikografer, därav rubriceringen *folklig* dialektlexikografi.

Artikeln inleds med en presentation av vad som publicerats om svenska dialektordböcker fram till idag, vilken följs av en kortfattad redogörelse för hur utgivningen beskrivits. Därefter övergår jag till att närmare undersöka de tryckta ordböcker som publicerats fr.o.m. 1991. Det året kom det nämligen ut en ordbok som skulle kunna tyda på att syftet med att ge ut dialektordböcker höll på att förändras.

En tydlig förändring som inträffat under de senaste tjugo åren är att dialektordböcker även börjat publiceras på Internet. De webbpublicerade och tryckta dialektordböckerna skiljer sig åt i flera avseenden och två av de mest påtagliga skillnaderna presenteras. För webbordböckernas del undersöker jag dessutom i vilken mån de specifika möjligheter som en webbpublicering medger har tagits tillvara.

Avslutningsvis formulerar jag några förslag till fortsatt forskning om dialektordböcker.

## 2. Bakgrund

Vid några få tillfällen har den samlade utgivningen av svenska dialektordböcker behandlats. Första gången, så vitt jag vet, var av Adolf Noreen i *Vårt språk* (1903–07: 268–286). Noreens genomgång har huvudsakligen karaktären av en kommenterad uppräkningslista, där han kortfattat beskriver de enskilda ordböckerna och avger summariska omdömen om deras vetenskapliga värde.

1939 publicerade Karl Gustav Ljunggren en översikt över svenska dialektordböcker, där han, till skillnad från Noreen, sätter in ordböckerna i ett sammanhang efter hur de präglats av rådande tidsanda och vetenskapssyn. Som en bakgrund till dagens utgivning refererar jag hans översikt kortfattat i det följande.

Ljunggren inleder med det första trevande insamlandet av dialektord under 1600-talets göticism, vilket inte ledde till några publikationer. Under början av 1700-talet vidtog ett mer systematiskt insamlande under ledning av professorn, sedermera ärkebiskopen Erik Benzelius (1675–1743). Hans hopbragta samlingar kom senare att användas av Johan Ihre för *Swenskt Dialect Lexicon* (1766). Under senare delen av 1700-talet avtog intresset för att studera dialekter och tog inte fart igen förrän under 1800-talets romantik och återuppväckta göticism. Dialekterna kom då att betraktas som levande fornminnen, vilket gjorde att det största intresset riktades mot deras ålderdomliga drag. Detta synsätt avspeglades även i urvalet av ord i den tidens dialektordböcker, så också i Johan Ernst Rietz' *Svenskt dialekt-lexikon* (1862–67).

När den junggrammatiska skolan slog igenom på 1870-talet förändrades synen på dialekter radikalt. Nu sågs de som viktiga, inte för att de var ålderdomliga, utan för att de utgjorde levande exempel på vad som händer när språket får utvecklas naturligt. Och i och med att junggrammatikerna huvudsakligen intresserade sig för ljudutvecklingar, blev i princip alla dialektord lika intressanta att studera. I ordböckerna lades stor vikt vid uttalsuppgifterna. Betydelseangivelser och fraser ägnade man däremot ingen större uppmärksamhet.

När det gäller utgivningen under 1900-talet konstaterar Vidar Reinhammar (1984) att den tog ordentlig fart på 1960-talet. Han menar att det har sin grund i att landsbygdens kultur då började uppfattas mer positivt och att många därmed såg ett värde i att dokumentera och bevara dialektord åt eftervärlden (Reinhammar 1984: 4). I och med detta uppsving träder också en ny typ av ordboksförfattare in. Från att dialektlexikografi varit en uppgift i huvudsak för lärda män, har utgivningen under senare decennier kommit att domineras av entusiaster på gräsrotsnivå, som ofta arbetar tillsammans i studiecirklar och liknande.

Synen på dialekter har alltså varierat under århundradena och därmed också synen på vad som är syftet med att ge ut en dialektordbok. *Vem* det är som ger ut dialektordböcker har också förändrats. Något som dock förenat dialektordböckerna genom tiderna är att de varit enriktningsordböcker med dialekten som källspråk och standardsvenskan som målspråk. De dialektala uppslagsorden har återgetts med något slags ljudskrift eller i normaliserad form.

### 3. En ny inriktning?

För tjugo år sedan kom det ut en ordbok som bröt mot mönstret med dialekten som källspråk. Det uppmärksammades i en artikel av Birgit Eaker (1999). Ordboken det handlar om är *Pitemålet -ållt mila àagg å ö`öx-* (Lidström m.fl. 1991). Denna ordbok över det norrbottniska pitemålet har alltså, till skillnad från tidigare ordböcker, standardsvenska uppslagsord med ekvivalenter på dialekt, återgivna med grov ljudskrift. Den som i ordboken vill söka på ett dialektord har emellertid inte lämnats i sticket. Boken avslutas nämligen med ett 80-sidigt register över dialektformer med hänvisningar till vilka artiklar de återfinns i (figur 1).

SOMNA v. 1. såmen, sãmne, =. "I var et så såmen." (Jag kunde inte somna) 2. da:n å:, dane å:, =, somna tvärt så snart man lagt sig. 3. d<sup>ur</sup> å:/du<sup>r</sup> å:, du<sup>r</sup>e å:, =, slumra till, halvsova. 4. fåra opa Fjæderhøimen å nååpp snåttren, uttr. (fara till Fjäderholmen för att plocka hjortron) "Fjäderholmen" var antagl. en omskrivning av "dunkudde"; sades till/om barn när man ville narra dem att sova. 5. I<sup>uk</sup> å:t/lu<sup>w</sup>k å:t, lu<sup>w</sup>ke å:t, =, ofta om barn. "Häv- en lu<sup>w</sup>ke å:t nö?" (Har han somnat nu) 6. shlågk/en, -ne, =, "slockna", somna när man är uttröttad, t ex efter hårt arbete. 7. småll å:, small å:, smolle å:, somna snabbt och hårt; oftast om berusad.

lu<sup>w</sup>k å:t v 1. SOMNA. 2. STÅNGA.

Figur 1. Artikeln *somna* i *Pitemålet -ållt mila åagg å ö`öx-*. Från *lu<sup>w</sup>k å:t* i registret hänvisas till *somna* och även till *stånga*.

Eaker frågade sig i sin artikel, om vi kanske kommer att få se fler ordböcker av pitemålsordbokens typ i framtiden, alltså dialektordböcker som ”i första hand är uppställda för att passa användare som inte längre behärskar sin hembygds dialekt men som vill lära sig den” (Eaker 1999: 79). Nu har det gått tjugo år sedan den kom ut. Vad har hänt sedan dess? Har den fått några efterföljare och i så fall hur många? För att kunna besvara dessa frågor, har jag tagit del av de dialektordböcker som, enligt den nationella bibliotekskatalogen Libris (i april 2011), kommit ut fr.o.m. 1991. Totalt rör det sig om 129 stycken<sup>1</sup>. Av dem har 14 inte varit tillgängliga för mig. Undersökningen omfattar således 115 ordböcker.

### 3.1 Resultat av undersökningen

Efter en närmare granskning av de 115 ordböcker som ingår i undersökningen, visar det sig att det endast är *en* ordbok som, liksom pitemålsordboken, har en lemmalista med uppslagsord på standardsvenska samt ett register över dialektformer. Det rör sig om *Orsaord. En orsamålsordbok* (Ohlsén & Olander 2010). *Orsaord* är en 496-sidig ordbok, där s. 17–428 utgörs av lemmalistan och de följande 56 sidorna av en ordlista ”som ska fungera som ett register till ordboken”

<sup>1</sup> I siffran innefattas endast böcker som utkommit i förstautgåva 1991 eller senare. Häften av ordböcker där utgivningen påbörjats före 1991, äldre ordböcker i nytryck, utgåvor av tidigare opublicerade handskrivna ordböcker samt supplementutgåvor är inte medräknade. Detsamma gäller den vetenskapliga Ordbok över Sveriges dialekter, vars första häfte utkom 1991. Inte heller ingår ordlistor publicerade i tidskrifter, årsböcker o.d., eftersom det inte varit möjligt att få fram någorlunda fullständiga uppgifter om vilka publikationer de återfinns i. Ordlistor som endast har till funktion att förklara ord i texter på dialekt har också de förbigåtts.

(Ohlsén et al. 2010: 429). Hänvisningar ges där från dialektformerna till uppslagsorden i lemmalistan (figur 2).

<p>somna <b>sömn ä, -öd, -äd</b>          Sömnör o snärt? <i>Somnar hon snart? Du fo'nt sömmön</i> um. <i>Du får inte somna om.</i>          I ska bara <b>lur til/ta mi jänn lur.</b> <i>Jag ska somna till. – ta mig en tupplur</i>          ”Nu a krippär <b>feri börti stiftälds-budär.</b>”  <i>Nu har barnen farit till ”skinnfällsbodarna”. – somnat under fällen</i></p>	<p><b>lur til</b> → somna</p>
--	-------------------------------

Figur 2. Artikeln *somna* i *Orsaord*. Från *lur til* i registret hänvisas till *somna*.

Det finns emellertid ytterligare några ordböcker som till sin uppställning påminner om ordböckerna över pitemål och orsamål. De tre publikationerna *Våmhus. Ordlista. Svenska–våmhusmål jämte kortfattad grammatik samt våmhusmål–svenska. Ett urval grundformer* (Ask 2000), *Limamålet* [2006] samt *Åolleist opa köllismåle* 'Ordlista över kalixmålet' [2006] innehåller också de lemmalistor med standardsvenska uppslagsord. Alla tre är dessutom försedda med lemmalistor i motsatt riktning, men i dessa ges endast kortfattade upplysningar om de enskilda orden och urvalet av dialektformer är ofta (mycket) begränsat. Även *Rånemålet i Böle* (2004) innehåller en lemmalista med standardsvenska uppslagsord, men till skillnad från de ovannämnda saknas lemmalista eller register i motsatt riktning.

Det är värt att notera att samtliga hittills nämnda ordböcker, inklusive pitemålsordboken, beskriver dialekter i norra eller västra Dalarna eller i Norrbotten. Det är sannolikt ingen tillfällighet. Dialekterna i de här områdena skiljer sig så kraftigt från standardsvenskan att talarna upplever sig som tvåspråkiga. Behärskar man inte dialekten krävs det, liksom vid inläring av ett främmande språk, en ordbok med standardsvenskan som källspråk.

I tillägg till ovannämnda ordböcker kan ytterligare två nya typer av ordböcker urskiljas, vilka har en uppställning som ger anledning att uppmärksamma dem här.

Den ena gruppen utgörs av tio ordböcker, som på traditionellt sätt har en lemmalista med dialektala lemman, men som också innehåller någon typ av register eller enklare ordlista i motsatt riktning. Utgivarna benämner ibland registret/ordlistan som index, referensordlista eller enkel ordlista. De ordböcker som ingår i denna grupp är följande:

<p>Färdär frå Soldn. Spår från Sollerön. En ordbok på soldmål.          (Andersson et al. 1999)</p>	<p>Dalarna</p>
<p>Go' sup ... ord men inga visor på idremål. (Erikson 2007)</p>	<p>Dalarna</p>

Orda-längd över Järnamålet. (1999)	Dalarna
Ordbok över Rättviksmålet. (1998)	Dalarna
Wenjad. En ordbok. (Wennberg 2007)	Dalarna (Venjan)
Klövsjöord. (Edlund et al. 2001)	Jämtland
Åremålet. Uppptecknat 1986–1996. Jämtska–svenska.	Jämtland
Svenska–jämtska. (Forsåker 1999)	
Skånsk–svensk–dansk ordbok från A! Till Örrk! (Lång et al. 1995)	Skåne
Västra Torsås' dialekt under tidiga 1900-talet. [2006]	Småland
Vitvattnsmåle. En Bjurholmsdialekt. (Lindberg et al. [1991])	Ångermanland

Som synes är den geografiska spridningen i utgivningen här större, men tyngdpunkten ligger i norra och västra Dalarna.

Den andra gruppen består av fyra tvåriktningsordböcker där de ingående lemmalistorna är (någotsånär) jämbördiga. I det här fallet är utgivningen ännu mer spridd över landet. Gruppen utgörs av:

Oa frau Listalannet. Parlör för den listerländska dialekten. (2007)	Blekinge
Flomål–svenska. Svenska–flomål. (Eivinson 2010)	Dalarna (Floda)
Ordlista över Älvros-målet. (2004)	Härjedalen
ÖverköLisma:Le. Överkalixmålet. (2004)	Norrbottnen

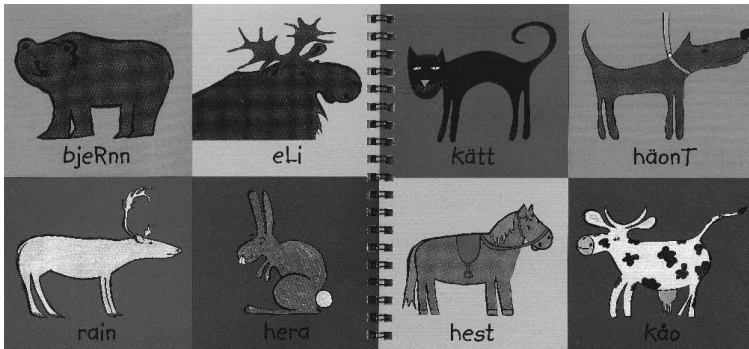
### 3.2 Starkare betoning på inläring

Det kan konstateras att huvuddelen av de dialektordböcker som utkommit fr.o.m. 1991 är uppställda på samma sätt som tidigare i så måtto att de innehåller en enda lemmalista med lemman på dialekt och ekvivalenter på standardsvenska. Trots det vill jag hävda att de sammanlagt 19 ordböcker (pitemålsordboken ej inräknad) som enligt ovan avviker från detta mönster genom att innehålla en lemmalista med standardsvenska lemman alternativt ett register med hänvisningar från standardsvenska till dialekt, måste ses som ett tecken på att något ändå har förändrats i synen på vad en dialektordbok är till för och vilket syfte utgivarna har med dem. Dessa ordböcker är alla på något sätt utformade för att ta hänsyn (också) till användare som inte längre helt behärskar sin dialekt. Genom valet av utformning har utgivarna i högre grad än tidigare riktat fokus mot inläring av (fler) dialektord samtidigt som viljan att dokumentera och bevara dialekten inte har försvunnit – den framhålls ofta i förorden. Man kan kanske säga att utgivarna med dessa ordböcker vill bevara *genom* inläring, genom att underlätta och uppmana till praktisk användning av dialekten. En dialekt lever ju vidare genom att överförs muntligt från generation till generation.



Att man på senare år kommit att lägga större vikt vid inläring kan också två dialektböcker som vänder sig till barn vara exempel på. Också de beskriver dialekter i Norrbotten resp. västra Dalarna. Det handlar om *Leir di ÖverköLisma:Le*, 'Lär dig överkalixmålet' (Jonsson 2002) och *Lär däg talö malungsmål* (Thors & Lissmyr 2010). De benämns inte av utgivarna som ordböcker, men kan i praktiken ses som enkla sådana.<sup>2</sup>

*Leir di ÖverköLisma:Le* är en pekbok som innehåller sammanlagt 48 bilder. Till varje bild anges med grov ljudskrift vad det avbildade heter på överkalixmål (figur 3).



Figur 3. Uppslag ur *Leir di ÖverköLisma:Le*

I baksidestexten skriver utgivaren Ella Jonsson så här:

Den unika dialekten håller på att dö ut. Vi måste skynda oss att rädda den. Och det är bråttom. I din hand håller du mitt bidrag till räddningsaktionen. En pekbok med bilder, där orden skrivs på ÖverkåLisma:Le. För det är ju med våra barn vi måste börja. Tanken är att boken leder till ett dubbelt lärande; den vuxne (som ofta glömt), lär sitt barn eller barnbarn och samtidigt sig själv.

*Lär däg talö malungsmål* är utgiven av nioåriga Gabriella Thors och hennes mamma Ulrika Lissmyr. Den innehåller tematiska ordlistor, parlörmeningar och andra meningar, allt från standardsvenska till dialekt (figur 4).

<sup>2</sup> De ingår inte bland de 115 böcker som undersökningen omfattar.



Figur 4. Två uppslag ur *Lär dig talö malungsmål*

## 4. Elektronisk publicering

Under de senaste tjugo åren har som bekant möjligheten till elektronisk publicering tillkommit i form av off-line-ordböcker på cd-rom, dvd m.m. och on-line-ordböcker på Internet. Också dialektordböcker finns representerade bland dessa typer.

### 4.1 Publicering på cd-rom och dvd

När det gäller dialektordböcker på cd-rom eller dvd finns endast en ordbok, *Dalbyordboken* (Bergkvist & Jacobsson 2001), utgiven. I det fallet föreligger också en tryckt 654-sidig parallellutgåva, som kom ut året dessförinnan.

Att cd-rom och dvd inte fått större användning är det sannolikt flera skäl till. Ett kan vara ekonomiskt. Dalbyordboken fick stöd från flera håll för sin utgivning. Ett annat skäl är nog att många dialektordböcker helt enkelt innehåller alltför få ord för att utgivarna ska anse att en cd-rom eller dvd är rätt avpassat medium.

### 4.2 Publicering på Internet

På Internet är publiceringsverksamheten desto mer livaktig. Hittills har jag kunnat räkna in 59 dialektordböcker (april 2011).<sup>3</sup>

Cirka två tredjedelar av ordböckerna återfinns antingen på olika slags hembygdsinriktade sidor eller på enskilda personers webbplatser eller bloggar. Bland

<sup>3</sup> Här har Lars Törnqvists webbplats *Thesaurus Lex*, där han samlat länkar till bl.a. dialektordböcker, varit till stor hjälp.

övriga webbutgivare kan nämnas tre kommuner, en lokalradiostation, en dagstidning, ett förlag, ett musikband och ett bostadsbolag. Ofta får ordböckerna dela utrymme med de mest skilda ämnen som utgivaren valt att presentera. Endast ett fåtal ordböcker återfinns på webbplatser som skapats enkom för ändamålet eller som handlar om dialekten i vidare mening.

Vanligtvis är orden inskrivna direkt på webbplatsen, men i åtta av fallen är ordböckerna utlagda som pdf-filer. Det rör sig då vanligen om tidigare tryckta ordböcker eller datorutskrift, vilka ursprungligen inte sammanställts för publicering på webben. Tre publikationer är från 1800-talet. Två av dem är utlagda inom *Projekt Runeberg* resp. *Google böcker*. Det gäller Rietz' *Svenskt dialekt-lexikon* (1862–67) och Peter von Möllers *Ordbok öfver halländska landskaps-målet* (1858). Den tredje, en ordlista ur N. Aug. Lundgrens *Beskrifning öfver Vestra Vingåkers Socken* (1873), föreligger i form av en avskrift.

### 4.3 Skillnader mellan tryckta och webbpublicerade ordböcker

Vad gäller skillnader mellan tryckta och webbpublicerade ordböcker kan säkerligen en rad sådana påvisas. Här vill jag lyfta fram två av de mest iögonenfallande.

En väsentlig skillnad mellan de tryckta och webbpublicerade ordböckerna är att de senare innehåller långt färre uppslagsord. Av de 59 webbpublikationerna är det 14 som innehåller färre än 100 uppslagsord och höjer man till max 200 uppslagsord stiger antalet till 22. Inte någon av de 115 tryckta ordböckerna innehåller så få ord.<sup>4</sup> Dessutom förhåller det sig så att de ordböcker på webben som innehåller det högsta antalet uppslagsord i många fall är pdf-publikationer.

Den andra tydligt framträdande skillnaden gäller förekomsten av illustrationer till artiklarna. I tryckta dialektordböcker är det ett förhållandevis vanligt inslag. Av de 115 ordböckerna är 28 försedda med sådana i större eller mindre utsträckning. Vad gäller de webbpublicerade ordböckerna är det endast två, de bohuslänska *Dialekt i Tanum* och *Ordbok över Skeemålet*, som har illustrationer i nämnvärd omfattning, men båda dessa är ursprungligen tryckta publikationer.

### 4.4 Utnyttjas möjligheterna med webbpublicering?

Den mest uppenbara skillnaden med att publicera på nätet jämfört med i tryck får väl sägas vara tillgången på utrymme. En nätutgivare kan dessutom göra det

---

<sup>4</sup> Som tidigare nämnts är ordlistor publicerade i tidskrifter, årsböcker o.d. inte medtagna bland de 115 ordböckerna. Sannolikt innehåller några av dem färre än 100 eller 200 ord.

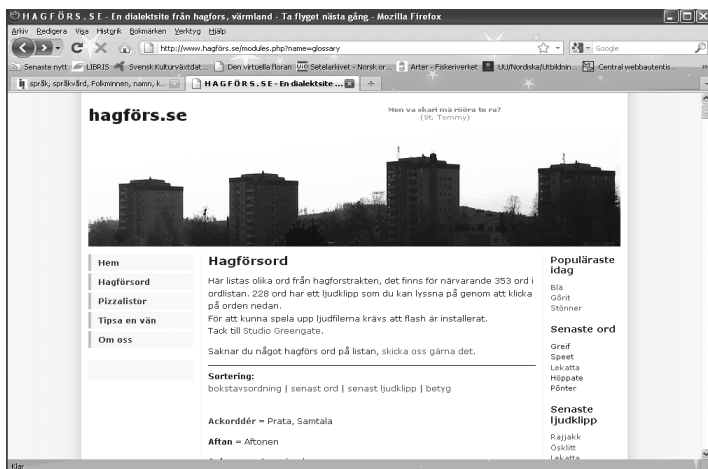
möjligt för brukaren att t.ex. utföra flera typer av sökningar, lyssna till ljudexempel och via länkar nå ytterligare information internt eller externt. Ordboksanvändaren kan dessutom uppmanas att själv bidra på olika sätt. Frågan är i vilken utsträckning dessa möjligheter har utnyttjats för dialektordböckernas del.

Som redan torde ha framgått är det obegränsade utrymmet inget som tagits i anspråk. Ett högre antal ord och förekomsten av illustrationer tycks snarare vara ett kännemärke för tryckta publikationer. Vad gäller sökmöjligheter och länkningar är förhållandet inte mycket bättre. Vanligtvis kan man nämligen inte söka alls i ordböckerna, utan alla ord presenteras på en och samma sida, i synnerhet när de är få till antalet. Är uppslagsorden något fler, kan det i ordbokens topp eller på en förstasida finnas klickbara begynnelsebokstäver (figur 5).



Figur 5. Parlör från Kålaboda by i Västerbotten med klickbara begynnelsebokstäver

Att inkludera ljud i webbordböcker kan tyckas särskilt naturligt när det gäller just dialekter, i och med att de huvudsakligen är talat språk. Ljudexempel visar sig emellertid förekomma i endast två fall i nämnvärd utsträckning, dels i *Öschöte guschelöv* med dialektord från Östergötland, dels i *Hagförsord* som ägnas den värmländska hagförsdialekten. Den senare skiljer sig påtagligt från övriga webbpublikationer genom att erbjuda brukarna olika funktioner som: Populäraste idag, Senaste ord, Senaste ljudklipp och Senaste kommentarer (figur 6).



Figur 6. Webbplatsen hagförs.se med *Hagförsord*

Till Hagförsord levereras dialektorden in av användarna. Det förekommer för övrigt i endast tre fall. Trelleborgs kommun i Skåne har på sin webbplats publicerat en ordbok som är helt baserad på ord som skickats in. Den har senare också kommit ut i tryck. Radio Skaraborg i Västergötland har utifrån ett radio-program, där dialekt ofta diskuteras med lyssnarna, skapat en ordbok som fylls på efterhand som nya ord tas upp i programmet. Slutligen innehåller nättidningen *Nära* i Ångermanland ett ”dialektikon”, som, liksom tidningen i sin helhet, byggs upp av insända bidrag. För övrigt nöjer sig webbutgivarna med att på sin höjd uppmana besökarna att via en e-postlänk bidra med fler ord eller gesynpunkter.

Mot bakgrund av hur tryckta dialektordböcker brukar tillkomma, ter det sig anmärkningsvärt att inte fler webbutgivare tar vara på möjligheten att interagera med användarna. Bakom en tryckt ordbok ligger inte sällan ett långt, ofta kollektivt förarbete, där man ord för ord arbetat sig fram till ett innehåll. När ordboken väl kommit ut, följs den inte sällan av ett eller flera supplement eller av nya utökade upplagor, vilka innehåller ord som utgivarna kommit på i efterhand eller som ordbokens användare rapporterat in.

#### 4.5 Sammanfattande intryck av webbutgivningen

Helhetsintrycket av webbutgivningen är att Internet i hög grad verkar vara ett forum för publicering av korta, osofistikerade ordlistor. Ordlistorna tycks föga in sig i det allmänna mönstret av stor diversifiering som råder på nätet. Som regel ingår de som en del av en webbplats, där också information av helt annat,

ofta lika kortfattat, slag finns. Ibland utgör de bara ett slags sidospår till den egentliga verksamheten som beskrivs på webbplatsen.

De specifika faciliteter som nätet erbjuder utnyttjas sällan. Det kan knappast tillskrivas bristande medvetenhet hos utgivarna om vad som skulle kunna göras. Att länkar samt ljud- och bildillustrationer är vanligt förekommande torde inte ha undgått någon nätanvändare. *En* förklaring till varför dessa funktioner inte tagits i bruk oftare, skulle helt enkelt kunna vara bristande datakunskaper. Många inom hembygdsrörelsen inklusive utgivare av dialektordböcker är ofta lite till åren komna och kanske inte alltid insatta i hur det faktiskt går till att inkludera ljudexempel m.m. Som jämförelse kan nämnas att hagförs.se administreras av fem personer som var och en uppges ha specifika uppgifter inom ljudteknik, databasdesign osv.

## 5. Fortsatt forskning

I den här artikeln har jag belyst några förändringar i utgivningen av dialektordböcker som ägt rum under de senaste tjugo åren. Det finns flera andra aspekter hos såväl äldre som yngre publikationer, vilka vore värda att undersöka. En sådan är de tryckta ordböckernas megastruktur, som ofta kan vara både rikhaltig och varierande. Ett annat område som skulle förtjäna uppmärksamhet är den redan nämnda relativt frekventa förekomsten av illustrationer i de tryckta ordböckerna. Själva lemmanvandet vore också intressant att studera närmare. Vad har utgivarna ansett vara dialektord och värt att inkludera? Jag har inte haft möjlighet att gå in på dessa frågor idag, men hoppas kunna återkomma till några av dem längre fram.

## KÄLLOR OCH LITTERATUR

- Andersson, Margit & Danielsson, Suzanne, 1999: Färdär frå Soldn. Spår från Sollerön. En ordbok på soldmål. Sollerön.
- Ask, Bertil, 2000: Våmhusmål. Ordlista. Svenska–våmhusmål jämte kortfattad grammatik samt våmhusmål–svenska. Ett urval grundformer. Våmhus.
- Bergkvist, Karl L.son & Jacobsson, Jacob, 2000: Dalbyordboken. Ordbok över det värmländska dalbymålet. Som cd-rom-utgåva 2001. Sysslebäck.
- Eaker, Birgit, 1999: Vem läser en dialektordbok? – Vem skriver vi för? I: Nordiska studier i lexikografi 4. S. 75–83.
- Edlund, Gösta, Johansson, Petter & Andersson, Einar, 2001: Klövsjöord. [Klövsjö.]
- Eivnsson, Karin, 2010: Flomål–svenska. Svenska–flomål. Nora.
- Erikson, Arne H., 2007: Go' sup ... ord men inga visor på idremål. Saltsjö-Boo.

- Forsåker, Anna-Lena, 1999: Åremålet. Upptecknat 1986–1996. Jämmtska–svenska. Svenska–jämmtska. Åre.
- Ihre, Johan, 1766: Swenskt Dialect Lexicon. Upsala.
- Jonsson, Ella, 2002: Leir di ÖverköLisma:Le. Överkalix.
- Lidström, Gun, m.fl., 1991: Pitemålet -ållt mila åagg å ö`öx-. Piteå.
- Limamålet. [2006]. Utan ort.
- Lindberg, Harald & Alf-Göran, [1991]: Vitvattsmåle. En Bjurholmsdialekt. Bjurholm.
- Ljunggren, Karl Gustav, 1939: Svenska dialektordböcker. En översikt. I: Årsbok 1938. Vetenskaps-societeten i Lund. S. 21–37.
- Lundgren, N. Aug., 1873: Beskrifning öfver Vestra Vingåkers Socken. Örebro.
- Lång, Helmer & Vide, Sten-Bertil, 1995: Skånsk–svensk–dansk ordbok från A! Till Örrk! Vejbystrand.
- von Möller, Peter, 1858: Ordbok öfver halländska landskaps-målet. Lund.
- Noreen, Adolf, 1903–07: Vårt språk. Nysvensk grammatik i utförlig framställning. 1. Lund.
- Oa frau Listalannet. Parlör för den listerländska dialekten. 2007. Karlskrona.
- Ohlsén, Karin & Olander, Eva, 2010: Orsaord. En orsamålsordbok. Mora.
- Orda-längd över Järnamålet. 1999. [Dala-Järna.]
- Ordbok över Rättviksmålet. 1998. Boda kyrkby.
- Ordbok över Sveriges dialekter. Häfte 1–3. 1991–2000. Uppsala.
- Ordlista över Älvros-målet. 2004. Älvros.
- Reinhammar, Vidar, 1984: Dialektordbok – nationella och lokala aspekter. I: Mot-skrift. Arbeidsskrift for språk og litteratur. Nr. 2. S. 1–15.
- Rietz, Johan Ernst, 1862–67: Svenskt dialekt-lexikon. Ordbok öfver svenska allmoge-språket. 1–2. Malmö, Köpenhamn, Leipzig & London.
- Rånemålet i Böle = RanmaLe opa BöLe. Red. Dagmar Wikberg. 2004. Råneå.
- Thors, Gabriella & Lissmyr, Ulrika, 2010: Lär dag talö malungsmål. Malung.
- Wennberg, Karin, 2007: Wenjad. En ordbok. Orsa.
- Västra Torsås' dialekt under tidiga 1900-talet. [2006.] Växjö.
- Åolleist opa kölismåle. [2006.] Kalix.
- ÖverköLisma:Le. Överkalixmålet. 2004. Överkalix.

*Webbplatser* (URL-adresserna är kontrollerade i september 2011)

Dialekt i Tanum: <http://www.tanum.se/vanstermenykommun/kommunen/faktaomta-num/dialektitanum.4.489a3ea711b721a577680001443.html>

Google böcker: <http://books.google.com/intl/sv/googlebooks/about.html>

Hagförsord: <http://www.hagfors.se/modules.php?name=glossary>

Libris: <http://libris.kb.se/>

Nära Dialektikon: <http://nara.allehand.se/dialektikon.asp?visa=A>

Ordbok över Skeemålet: <http://www.stromgen.se/dialekter/skeedialekt.pdf>

Parlör Kålabodamål: <http://www.kalaboda.se/index.php?id=Parlor>

Projekt Runeberg: <http://runeberg.org/>

Radio Skaraborg: <http://sverigesradio.se/sida/artikel.asp?programid=3670&artikel=3623273>

Thesaurus Lex: <http://thesauruslex.se/lexikon/typ/dialektordbok.htm#Svenska>

Trelleborgs kommun: <http://www.trelleborg.se/Lar-kanna-Trelleborg/Trelleborgska/>  
Öschöte guschelöv: <http://www.ulfandersson.se/dialekt.html>

*Eva Thelin*

Institutet för språk och folkminnen, Uppsala.  
[eva.thelin@sofi.se](mailto:eva.thelin@sofi.se)



# Udtale i netordbøger – hvad, hvordan, hvor meget?

*Lars Trap-Jensen*

It is a well-known fact that dictionary users have difficulties in reading and comprehending the numerous special characters and diacritics of a phonetic writing system. An indisputable advantage of e-dictionaries is their capability to convey pronunciation in the form of sound files. In the process of converting The Danish Dictionary from print to screen, recorded sound files were among the editorial priorities as an alternative or supplement to the phonetic notation of the print dictionary. But the change of medium does not simply mean transformation of phonetic symbols to sound on a one-to-one basis. First, a number of choices must be made, and the article deals with the central questions based on the model we have chosen for The Danish Dictionary: What constitutes a suitable speaker voice? Is it advisable to manipulate the speaker towards a preferred pronunciation if it is not part of her natural language? Is agreement between the phonetic transcription and the recordings important, perhaps necessary? How should variable pronunciation be handled?

*Nøgleord:* udtale, lydfiler, lydskrift, IPA, indtalerstemme, udtalenorm, udtalevariation.

To my mind, hearing the pronunciations is the only unarguable improvement of a CD over a book (Landau, 2001: 97)

The most obvious use of multimedia is to convey the pronunciations, instead of the awkward symbology print dictionaries use (Sean M. Burke, 2003: 249)

## 1. Baggrund

Selvom man ikke behøver være enig med Landau og Burke i at lydfiler er det eneste eller for den sags skyld det vigtigste træk hvorved den elektroniske ordbog er den trykte overlegen, kan de fleste nok tilslutte sig det synspunkt at lydskrift, sådan som den kendes fra trykte ordbøger, aldrig bliver populær ud over en snæver kreds af sprofolk og særligt interesserede – uanset om det gøres i den

amerikanske tradition med en specialudviklet notation baseret på det almindelige alfabets bogstaver eller i den europæiske tradition, der bygger på lydskriftsystemer udviklet inden for fonetikken, med IPA som den internationalt bedst kendte og mest benyttede notation.

Hvis man med en ordbog ønsker at henvende sig til en bredere kreds af brugere end lige specialisterne, er det derfor oplagt at oplyse om udtale på den måde som computeren er velegnet til: multimedialt, altså sådan at brugeren hører udtalen frem for at skulle læse den. Den Danske Ordbog har ca. 15.000 besøgende på en almindelig hverdag, og dermed rækker brugergruppen selvfølgelig langt ud over specialisternes snævre kreds. Udtale i form af lydfiler har da også været et højt prioriteret ønske siden konverteringen af Den Danske Ordbog til netordbog begyndte. Vi er derfor glade for at have fået mulighed for at gennemføre en indtaling af samtlige opslagsord i ordbogen. I det følgende demonstreres modellens enkeltdele og de valg der er truffet undervejs.

## 2. Hvor mange ord?

Udgangspunktet er naturligvis de eksisterende udtaleoplysninger. Den aktuelle netordbog adskiller sig i udgangssituationen ikke væsentligt fra den trykte DDO's angivelser; de vigtigste tal fremgår af tabel 1.

Opslagsord med udtale	40.636 (af 95.846)
Samlet antal udtaleangivelser	49.164
Andel af artikler med mere end én udtale	15-20 %
Lydskriftnotation, trykt DDO	Dania: [kɔnfə'raŋsə]
Lydskriftnotation, netordbog	IPA, halvgrov: [kʌnfə'βɑŋsə]


Tabel 1: Nøgletal for udtaleangivelser i DDO, maj 2011

Af tabel 1 ses det at under halvdelen (ca. 42 %) af opslagsordene i DDO har oplysning om udtale. I den trykte DDO var hovedprincippet at sammensætninger ikke fik udtaleangivelse hvis udtalen kunne forudsiges på baggrund af de indgående ord i sammensætningen. Kun i tilfælde hvor sammensætninger har uforudsigelig udtale, herunder tryk- eller stødfordeling, har disse altså fået en udtaleangivelse. Dette princip grunder sig hovedsagelig på pladsøkonomi og er derfor ikke relevant for netordbogen.

Endvidere er DDO's udtaleangivelser ved sammensætninger ofte ufuldstændige: Kun den del af ordet der udviser det uforudsigelige træk, lydskrives, jf. figur 1.

**bechamelsauce** eller **bechamelsovs** substantiv, fælleskøn

**BØJNING** bechamelsauce: -n || bechamelsovs: -en

**UDTALE** [**be**ca'mel-] eller [**be**ca-] 

Figur 1. Udtaleangivelse for *bechamelsauce* i DDO

Baggrunden for denne praksis er for det første pladsøkonomisk: Ved ikke at nævne udtalevariationen for sidsteleddet – som også er beskrevet under *sauce/sous* – reduceres udtaleangivelserne til to. For det andet er den pædagogisk: De fleste almindelige brugere har svært ved at læse lydskrift, og derfor er det vigtigt at øjet hurtigt finder det udtalemæssigt problematiske sted – i dette tilfælde variation i udtalen af ordets første vokal. Det er grunden til at sidsteleddet er udeladt ved den første angivelse, mens den anden, udtalevarianten, er yderligere reduceret for at fokusere på det kontrastive.

Problemstillingen er en anden når det gælder lydfiler. Her er der ikke noget der taler imod at bringe udtale for samtlige opslagsord, og hensynet til afkodning af notationen er selvsagt irrelevant. Som hovedregel forsynes alle opslagsord derfor med lydfiler. Undtaget fra reglen er et begrænset antal ord af en særlig art, nemlig de (fleste) forkortelser og symboler der er begrænset til det skriftlige medie, samt præfikser, suffikser og andre lemmatiserede orddele der ikke har en selvstændig udtale. På den måde vil lydfilmodellen føre til at antallet af ord med udtaleangivelser mere end fordobles.

Af tabel 1 fremgår det endvidere at lidt over 9.000 artikler har mere end én udtaleangivelse. Det kan være opslagsord med variation i udtalen (figur 1 kan også her tjene som eksempel), der kan være oplysning om udtalen af bestemte bøjningsformer, fx en uregelmæssig flertalsform eller en præteritumsform af et verbum, og endelig kan der være en oplysning om hvordan ordet udtales når det indgår i sammensætninger. Sidstnævnte angivelse skyldes forlægget fra den trykte DDO, der indeholder mange sammensætninger som ikke er hovedopslagsord med egne artikler. Af samme grund har disse ord ikke udtaleoplysninger, så hvis der er noget at sige om udtalen, må man søge oplysningen i en simpleksartikel, typisk førsteleddet. Det er naturligvis ikke nødvendigt i lydfilmodellen, hvor alle ord får egen udtaleangivelse. Derimod vil det fortsat være relevant at oplyse om særlige udtaler af uregelmæssige bøjningsformer.

I den trykte DDO er der valgt en lydskrift der har tradition i dansk, nemlig lydskriften *Dania*, udviklet af Otto Jespersen (i 1890) specielt til det danske lyd-

system. Dania's tegn er ikke en del af Unicode-standarden, der bruges af browserne på nettet, og derfor kan *Dania*-lydskriften ikke umiddelbart vises i netudgaven af DDO. Vi har derfor konverteret udtaleangivelserne til en halvgrov form af IPA. IPA er en international notation, den er en del af Unicode og kan vises uden problemer i alle browsere.

### 3. Hvem?

Ved valg af indtalerstemme er der flere faktorer at tage hensyn til. Der er for det første noget rent biologisk: Skal det være en mands- eller kvindestemme? De fleste er enige om at en lys stemme er tydeligere end en mørkere stemme. Det kunne tale for at vælge en kvindestemme, men noget tyder på at andet end det rent akustiske signal spiller en rolle for afkodningen, fx også brugerens holdning til stemmen. I en undersøgelse foretaget på universitetet i Sheffield (Sokhi m.fl. 2005) har et hold psykologer peget på at der er forskel på hvordan mænd og kvinder processerer stemmer fra personer af det modsatte køn. De opdagede at mænd processerer mands- og kvindestemmer i forskellige centre i hjernen, og for mændene var processing af en kvindestemme mere kompliceret end en mandsstemme. Samlet set betød den mere komplekse proces at mændene i forsøget blev hurtigere trætte af at lytte til kvindestemmen – uanset at det fysisk-akustiske input fra kvindestemmen har været klarere. På den baggrund er det ikke helt så entydigt hvilken stemme man skal vælge – og næppe overraskende at eksisterende ordbøger har gjort forskellige valg. Foruden de nævnte faktorer er der også andre kvaliteter ved stemmen at tage hensyn til: Den skal have den rigtige kombination af anatomiske og akustiske parametre der gør den behagelig at lytte til – foruden at signalere troværdighed, sympati, effektivitet eller andre kvaliteter som man ønsker forbundet med stemmen. Der findes en lang række opfindsomme og mere eller mindre lodige bud på forbindelsen mellem stemmekvalitet og psykologiske typer. Jeg vil ikke foregive at være nogen ekspert på området, og nedenstående skal derfor blot opfattes som et eksempel til illustration (efter Hultmann 2004):

**Ulven** med den hurtige, aggressive stemme.

**Ekkoet** der som en klang-kamæleon skifter stemme efter situationen.

**Vandfaldet** der taler uafbrudt i et højt toneleje og væk fra sine egne følelser.

**Bremsen** med den afbrudte og hakkende stemmeføring som om der hele tiden sker en indre opbremsning.

Den **kontrolgale** der med massiv tungerodsspænding får stemmen til at lyde klemt.

**Basunen** med den overdrevne kraftige stemme der overdøver alt og alle – og endda måske sig selv.

Hertil kommer **Optimisten**, **Pessimisten**, **Starteren**, **Cocktailstemmen**, **Lyseslukkeren**, **Forførelsen**, **Rømmeren** og flere andre [i alt 20 typer opregnes, LTJ]

#### 4. Én eller flere stemmer?

Et andet spørgsmål der skal besvares, er om man skal vælge mere end én stemme. Hvis ordbogens lemmabestand opdateres løbende, bliver samarbejdet med indtalerstemmen tilsvarende længerevarende, og det kan godt blive problematisk hvis brugeren vænner sig til altid at høre én og kun en stemme. Der kan være helt praktiske faktorer der nødvendiggør en afbrydelse af samarbejdet, fx at indtaleren flytter langt væk fra lydstudiet og teknikkerne, skifter til andet arbejde, får andre opgaver som er uforenelige med fortsat indtaling, ændrede økonomiske betingelser osv. Fordi samarbejde med en enkelt indtaler er sårbart, kan det være betænkeligt at vænne brugerne til at høre én og kun én stemme. Man kan med fordel operere med fx to forskellige stemmer så brugerne vænner sig til at høre forskellige stemmer. På den måde vil det blive nemmere at introducere en ny stemme engang i fremtiden hvis det bliver aktuelt. Omvendt er det praktisk og logistisk vanskeligere at håndtere flere indlæsere i det indledende forløb, og i sidste ende har det vejet tungere, men det kan godt komme på tale at supplere med en yderligere stemme senere, jf. afsnit 6.2.

Ud over den slags praktiske overvejelser er der også nogle sproglige hensyn. Ikke mindst er det vigtigt at tage stilling til den sprogvariant som er indtalerens naturlige. Sammenhængende dermed er dernæst spørgsmålet om man skal lade indtaleren bruge sin egen naturlige udtale eller hellere instruere vedkommende i at bruge en bestemt udtale, fx den der svarer til lydskriften i den trykte ordbog.

Med tanke på disse faktorer blev der gennemført prøveoptagelser med et antal kandidater, og til sidst blev én stemme valgt ud. Valget faldt på en professionel indtaler, lige under 40 år, med østdansk baggrund, dvs. med en naturlig udtale der ligger ret tæt på dansk standardudtale. Hendes sprog er let konservativt i forhold til hendes alder, hvilket nok skal ses i sammenhæng med hendes professionelle baggrund: Hun er bevidst om sit sprog og hvad omgivelserne forventer af hende. Desuden spiller det en rolle at det drejer sig om ord udtalt i isolation, hvor man kan gå ud fra at udtalen er mere distinkt og med større opmærksomhed på normen end det er tilfældet i spontan tale.

## 5. Indtalingsprocessen

Det er vigtigt at afklare præcis hvilke oplysninger indtaleren skal have som input under optagelserne. På den ene side er det en fordel at forlægget er så enkelt og overskueligt som muligt, dvs. indeholder færrest mulige oplysninger. Herved sikres det at der er et minimum af distraherende elementer til stede så hun kan koncentrere sig om indtalingen af hvert enkelt ord og optagelserne forløber i et konstant flow uden afbrydelser. På den anden side er supplerende oplysninger nødvendige for at identificere flertydige ord (pga. homografi) eller ord der er kilde til usikkerhed (fx fordi de er sjældne, faglige, nye eller uetable-rede). Det gælder altså om at finde en balance så der vises den minimale, men netop tilstrækkelige mængde information. Forud for selve optagelserne gennemførte vi en række prøveoptagelser i studiet, og også efter de indledende optagesessioner har vi evalueret og justeret inputoplysningerne.

De vigtigste erfaringer der er gjort, er især:

- input bør være randomiseret. Hvis ordene præsenteres for indtaleren i alfabetisk rækkefølge, er der forhøjet risiko for at hun – ubevidst – udtaler dem med modsætningstryk (adgangskon'trol, adgangs'kort osv.)
- for den samlede produktivitet er høj indtalingskadence ikke afgørende. Langt vigtigere er det at undgå at indtalingsflowet afbrydes fordi indtaleren læser forkert, er usikker på ordet eller mister koncentrationen

Det har ført til følgende justeringer af modellen: Det samlede antal opslagsord er inddelt i alfabetiske grupper, og inden for hver gruppe er rækkefølgen randomiseret. I skrivende stund er det ikke endeligt fastlagt hvilke oplysningselementer der skal være til stede. Redaktionens forslag er at opslagsordene grundlæggende opdeles i to grupper: “problematisk” og “uproblematisk” ord. Hvis et ord er uproblematisk, kan input begrænses til at bestå af opslagsord og ét citat. Denne gruppe omfatter ca. 80 % af alle opslagsord. De resterende 20 %, som altså regnes for “problematisk”, samles i selvstændige grupper som forsynes med yderligere oplysninger. Problematisk betyder i denne sammenhæng alene at ordet indeholder et eller flere træk der fremkalder fejl eller afbrydelser i indtalisstrømmen. De ord der efter de første optagelsesrunder synes at være problematiske, er: sjældne ord, faglige ord, ord med varierende udtale, homografer, nye engelske lån, ord der indeholder taltegn og store bogstaver, ord fra ordklasser ud over de fire store (substantiver, verber, adjektiver, adverbier). Derudover er der individuelle forhold som virker i begge retninger: Nogle “uproblematisk” ord viser sig alligevel problematiske, typisk fordi indtaleren ikke kender dem, mens andre, der er klassificeret som “problematisk”, ikke giver problemer.

Gruppen af problematiske ord kan yderligere opdeles i typer og forsynes med de ekstra oplysninger der er relevante for typen. For homografer kan ordklasse og betydningsforklaring være nyttige ekstraoplysninger, ligesom sammenhørende homografer ikke bør omfattes af randomiseringen. Ved sjældne og meget faglige ord kan lydskrift være nødvendig, evt. suppleret med en redaktionel bemærkning med en anvisning til indlæseren, en mulighed der er indbygget i indlæsningsprogrammet.

Rytmen i optagelserne er planlagt til to ugentlige runder a en halv dags varighed. Beregninger viser at med det estimerede tempo og den optagefrekvens der er aftalt, vil det tage i alt omkring et halvt år at optage udtaler til samtlige ord i DDO. Det er aftalt at der efter hver enkelt studieoptagelse sker en efterbehandling på Det Danske Sprog- og Litteraturselskab hvor redaktionen kontrollytter samtlige optagelser og noterer de udtaler der ikke er tilfredsstillende. Anmærkningerne kan vedrøre alt fra teknisk dårlig optagekvalitet, uvedkommende støj og fejl fra indtalerens side til utilfredshed fra vores side med den udtaleform hun har valgt. På den måde returneres en del af optagelserne og bliver indtalt igen i en ny optagerunde.

Efter kontrollytning og evt. en genindlæsning er udtalerne klar som wav-filer forsynet med id-numre til indsættelse i ordbogsbasen under det pågældende opslagsord. Softwaren der bruges til optagelserne, holder selv styr på forbindelsen mellem filerne og deres id-numre med henblik på den senere brug. Filerne kan selvfølgelig konverteres til andre filformater.

## 6. Problemområder

### 6.1 Abstrakt lydskrift vs. konkret stemme

Man kunne tro at lydskrift og lydfil indeholder den samme information i to forskellige medier, men sådan er det ikke. Den trykte DDO's lydskriftsangivelser er abstrakte i den forstand at de generaliserer over en række individuelle udtaler og gengiver en almindelig udbredt standardudtale. Det betyder at det er fuldt ud muligt, ja vel endda mest sandsynligt, at der ikke findes én sprogbruger i Danmark som i alle tilfælde har netop de udtaler der anføres i DDO, i sit eget sprog. De enkelte udtaler stammer ikke fra en konkret person, men er den mest udbredte udtale hos ikke-dialekttalende danskere. Dog er der af systematiske grunde taget stilling til fordel for en let konservativ norm i den trykte DDO, dels fordi vi ønskede en udtale der ikke virkede alt for provokerende (læs: ung), dels fordi det nu er over 10 år siden dette arbejde pågik, hvorved træk der dengang var almindelige, i nogle tilfælde forældes, mens andre træk, der dengang

blev opfattet som ungt sprog, i den mellemliggende periode er blevet mainstream.

Anderledes er det med indtalerstemmen. Den er anderledes konkret, og der er et rigtigt menneske bagved som har sin egen naturlige udtale af ordene. Når man bruger en bestemt stemme i en ordbog, kan det jo ikke undgås at den får status som en slags norm, og derfor skal den udvælges med omhu. Vi har som udgangspunkt valgt at lade indtaleren bruge sin egen spontane udtale frem for at instruere hende i at anvende en bestemt udtale. Den beslutning vil nogle være uenige med os i. Fx har Lars Brink (personlig kommunikation) hævdet at man forholdsvis nemt kan træne en person i en bestemt udtale og på den måde styre udtalen i en ønsket retning. På Den Danske Ordbogs redaktion tillader vi os at betvivle det fordi det dels går ud over naturligheden i indtalingen, dels vil påvirke optagetempoet hvis indtaleren skal indtale på grundlag af en lydskriftangivelse eller en nedskrevet instruktion frem for sin egen naturlige udtale.

Men heller ikke løsningen med spontan udtale er uden problemer. Den medfører at der vil opstå et – endnu ukendt – antal tilfælde hvor lydfilens udtale afviger fra det der står i lydskriften, og da må vi tage stilling til hvordan det skal håndteres. Logisk set kan man tænke sig fire muligheder.

Vi kan vælge at rette i lydskriften. Især hvis uoverensstemmelsen skyldes at udtalen i rigsmålet har ændret sig sådan at den trykte DDO's udtale er blevet for konservativ, er det en nyttig og tiltrængt revision der på den måde kan foretages. Et eksempel kunne være udtalen af ord som *prud*, *brud*, *rudel* over for *hud*, *klud*, *tud*, *nudel*, *puddel*, hvor lydskriften i DDO anfører en udtale med lang vokal og stød på vokalen for de tre første ord, og en udtale med kortere vokal og stød på konsonanten for de sidste. Her er der tale om en generel lydændring som er under udbredelse, og der har derfor været vaklen i sprogbrugen. Det er nok ikke tilfældigt at den ældre udtale med lang og stødt vokal er anført ved de lidt gammeldags ord *prud* ('smuk og ædel') og *rudel* ('flok af hjortedyr'), mens mere neutrale og dagligdags ord *hud*, *klud*, *nudel* har fået den nyere udtale hvor stødet er flyttet til den følgende konsonant. I skrivende stund har det ikke været muligt at sammenligne med de indtalte lydfiler (fordi indtalingen af ordene ikke er afsluttet), men hvis vi bekræftes i at de første ord nu også er omfattet af den nye udtale, er det et oplagt eksempel på at lydskriftens notation bør ændres. Det vægtigste argument mod at revidere lydskriften generelt er at det er en resursekrævende opgave.

En anden løsning er at lade indtalingen gå om. Det kan overvejes hvis der er nogle få udtaler som afviger fra et ellers konsistent lydssystem. Det er det mest oplagte alternativ når vi har fravalgt at styre indtalingen gennem oplæring: at rette udtaler der er utilfredsstillende, ikke af tekniske, optagemæssige grunde, men af sproglige fordi vi ønsker en anden udtale end den der er indtalerens



naturlige. Hvor ofte den situation vil opstå, er erfaringerne endnu for sparsomme til at bedømme, men den indebærer en risiko for belastning af projektet, dels fordi det kan forlænge indtalfingsfasen tidsmæssigt, dels fordi yderligere optagelser fordyrer studieomkostningerne.

En mere radikal løsning består i helt at undertrykke den lydskrevne udtale. Derved elimineres problemet med afvigelser mellem lydskrift og lydfil effektivt, og kontrolytningen skal alene sikre at udtalen er i orden i hvert enkelt tilfælde. En lidt mindre radikal løsning vil være at beholde en – måske afvigende – udtale, men ikke vise den medmindre brugeren aktivt vælger den til. På den måde kan specielt interesserede se og sammenligne med den trykte ordbogs angivelser, mens den almindelige bruger kan nøjes med at klikke på højtaler-symbolet for at høre udtalen.

Den fjerde og sidste mulighed består i at se stort på uoverensstemmelserne og vise lydskrift og lydfiler uafhængigt af hinanden, med de modsigelser som det giver. Det er ikke en løsning vi er umiddelbart tilhængere af, men først når vi har fået et mere fyldigt materiale i form af filer fra optagelserne, kan vi for alvor analysere hvor stor forskel der er mellem filer og lydskrift. I sidste ende bliver det et resursepørgsmål.

## 6.2 Valg af udtalenorm

Et andet problem består – ikke overraskende – i valg af udtalenorm. Denne ene stemme er jo kun én ud af en række mulige, mens sprog og udtale som bekendt varierer på mange parametre, og hvor DDO ellers gør en dyd ud af at være beskrivende, er vi det ikke på samme måde her. Der er tale om ét eksempel – måske nok et typisk eksempel – der trækkes frem og derved gøres til norm. Der findes andre og fuldt ud lige så gyldige varianter vi kunne have valgt: en ældre udtale, en yngre, en repræsentant for vstdansk standardsprog osv. Den oprindelige plan opererede faktisk med to udtaler: en østdansk og en vstdansk, repræsenteret ved en yngre og en lidt ældre stemme. På den måde ville det blive signaleret at der er flere måder at udtale ordene på, og den enkelte bruger var frit stillet til at vælge sin foretrukne stemme – eller sammenligne mulighederne.

Det er på mange måder en god model, men det er også en dyrere model fordi der skal laves dobbelt så mange optagelser. I den indledende fase af projektet var der flere interessenter involveret, og derfor var modellen mere realistisk, for det kræver nok en partner mere hvis økonomien skal holde. Sådant er situationen ikke i øjeblikket, men der er ikke noget der hindrer at vi på et senere tidspunkt kan supplere med flere stemmer for at vise en variation inden for rigsmålet.

### 6.3 Nye udtaleangivelser

Et andet område der kan give problemer, er de ord der nu forsynes med udtale, hvor der i den trykte DDO ikke var nogen. Det gælder som nævnt et ganske betydeligt antal, omkring 55.000 ord. Mange af dem er sammensætninger og afledninger, men også det er et område med variation og tvivlstilfælde. Lad os se på et par eksempler: Ordet *uddannelse* udtales på dansk med stød på anden stavelse, men i en sammensætning som *uddannelsessted* er der vaklen mellem stødt og ustødt udtale af samme stavelse. Og i ordet *egen* udtales den første vokal med åbent a – [α] – mens der i sammensætninger, fx *egeninteresse*, er vaklen mellem udtalen [α] og [e]. Her må redaktionen i hvert enkelt tilfælde tage stilling til om begge udtaler skal med eller kun én af dem, og i givet fald hvilken.

Hvis man tager princippet om at bruge en enkelt stemme alvorligt, bør der skelnes mellem den variation som kan henføres til forskelle mellem personer, baseret på sociologiske parametre som køn, alder, region og socialt tilhørsforhold, og den individuelle variation som vi alle har naturligt i mere eller mindre udstrakt grad i vores sprog, enten som fri eller situationsbetinget variation. Principielt bør kun den individuelle variation udløse flere indtalingen. I praksis er skellet dog ikke så nemt at opretholde, bl.a. fordi indtaleren typisk vil have et passivt kendskab til én eller flere andre varianter. Det er også velkendt at vi er heller ikke altid de mest pålidelige bedømmere af egen udtale.

Endelig er der situationer hvor indtaleren ikke er fortrolig med ordet og derfor ikke på samme måde kan siges at have en naturlig udtale. Det gælder mest oplagt for en række neologismer hvor man må regne med at ikke alle nye ord er kendt af indlæseren – og ofte udviser de også større variation i udtalen fordi den endnu ikke er etableret. Bedøm selv nogle eksempler på nye ord med usikker udtale: *ASAP*, *bake-off*, *eid*, *niqab*, *situp*, *streame*, *tabbouleh*, *wallah*, *wiki*, *zumba*. De fleste læsere kan sikkert også komme i tvivl om hvordan ordene udtales. Set fra indlæserens side adskiller neologismerne sig ikke fra andre ord hun – i denne sammenhæng tilfældigvis – ikke kender. Ud over de ord der tidligere blev nævnt som “problematiske”, er det uundgåeligt at der vil være ord som indtaleren ikke kender. Ikke to menneskers ordforråd er det samme, og med op mod 100.000 ord der skal indtales, er det helt naturligt at der vil dukke ukendte ord op, også ud over de ca. 20.000 der er skilt ud som potentielt problematiske.

I begge situationer finder redaktionen det rimeligt at fravige princippet om spontan udtale og i stedet vejlede indtaleren, enten i form af bemærkninger i input-softwaren eller ved møder ansigt til ansigt. Det er måske ikke den optimale løsning, men alternativet er værre: udlæsning af ukendte ord eller manglende lydfiler.

Beslægtet med dette problem er systematiske idiosynkrasier hos indtaleren. Prøveoptagelserne viste fx at hun valgte en mere engelsk udtale af engelske låneord end de fleste danskere har, enten fordi hendes eget sprog er påvirket af ophold og uddannelse i England, eller fordi hun ikke har været opmærksom på forskellen mellem en dansk- en og engelsksproget sammenhæng. Også her kan det være berettiget at give anvisninger så hun gøres opmærksom på forskellen og indretter sin udtale så den svarer til den almindelige udtale, på samme måde som når nyhedsoplæsere tillægger sig en bestemt udtale (fx [αf-] frem for normalt [αw'-] i *Afghanistan*), enten som følge af en eksplicit norm eller ved afsmitning fra en dominerende kollega.

Til sidst en kommentar til dem der undrer sig og spørger: Er tiden egentlig ikke vokset fra indtalte lydfiler? Er det ikke bedre at tilbyde al tekst i ordbogen som syntetisk tale og med ét slag løse de praktiske problemer der knytter sig til brug af menneskelige stemmer: opdatering, skiftende stemmer, økonomi osv. Hertil er der desværre ikke andet at sige end at kvaliteten af eksisterende talesynteser endnu ikke har nået det niveau der kræves i en ordbog som skal give vejledning om ordenes udtale. Vi vil gerne tilbyde syntetisk tale, men i et stykke tid endnu kan det kun blive som et supplement til og ikke en erstatning for indtalte lydfiler.

## LITTERATUR

- Burke, Sean Michael, 2003: The design of online lexicons. In: van Sterkenburg, Piet: A Practical Guide to Lexicography. Amsterdam/Philadelphia: John Benjamins Publishing Company. S. 240–249.
- DDO = Den Danske Ordbog: <http://ordnet.dk/ddo>.
- Hultmann, Lisbeth, 2004: Stemmens kraft. Aschehoug.
- Landau, Sidney I., 2001: Dictionaries – The Art and Craft of Lexicography. 2nd edition, Cambridge University Press.
- Sokhi, Dilraj S., Michael D. Hunter, Iain D. Wilkinson, Peter W.R. Woodruff, 2005: Male and female voices activate distinct regions in the male brain. In: *NeuroImage* 27. [www.elsevier.com/locate/ynimg](http://www.elsevier.com/locate/ynimg). S. 572–578.

*Lars Trap-Jensen*

Det Danske Sprog- og Litteraturselskab  
ltj@dsl.dk

# Ordfrekvens och ordboksutrymme

*Bo-A. Wendt*

In a linguistic description there is normally a more or less conspicuous lack of balance between the space a phenomenon requires to be fully described and its frequency in actual language usage, here called a lack of *iconicity*. The article investigates to what degree some Danish and Swedish dictionaries are iconic in this sense, i.e. whether the space of a word in the dictionary corresponds with its actual frequency. As expected, iconicity is rather meager, the most frequent words being treated much shorter than their actual frequency would suggest and the less frequent ones the other way around, though the more comprehensive the dictionary the more iconic it normally is regarding the first category. Further, frequent nouns and adjectives can be shown to exhibit a more iconic description than frequent verbs, and compounds and derivations treated together with their base words often exhibit a (shorter and thus) more iconic description than those treated as their own entries (often iconically much too extensive). Iconicity is of course not the overall ideal state of good dictionary entries: frequent forms are often best treated rather generalized (and thus rather shortly), some less frequent words can only be described accurately if described rather extensively. Yet, it is argued that iconicity should not be dismissed altogether but to be considered as one factor (among many other ones) when deciding upon how much space a word is worth in a dictionary.

**NYCKELORD:** ordböcker, ikonisk, beskrivningsutrymme, frekvens, högfrekventa ord, lågfrekventa ord, ordboksartiklars längd.

När jag arbetade som assistent inom projektet Svenska Akademiens grammatik (SAG), uttryckte någon gång redaktören Staffan Hellberg bryderi för att den grammatiska beskrivningen blev så oikonisk. Med detta avsåg han det förhållandet att de mest högfrekventa grammatiska strukturerna ofta kunde beskrivas med bara några få ord, till exempel finitverbets andraplacering i svenska huvudsatser, medan ovanliga strukturer, undantagen från huvudregeln, gärna krävde mycket utförliga beskrivningar för att man skulle förmå ringa in dessas förutsättningar och giltighet. Ja, också rena gränsfall med avseende på grammatikaliteten över huvud taget kunde ibland kräva mer utrymme än den regel som gäl-

ler 99 fall av 100. Detta gjorde alltså grammatikframställningen oikonisk, med en ojämvikt mellan det man ser eller hör mycket av i verkligheten och det man ser mycket av vid läsningen: hög frekvens i språket – litet utrymme i texten, låg frekvens i språket – stort utrymme i texten. Den bild, ἡ εἰκὼν, som textens omfång ger läsaren svarar helt enkelt inte mot det avbildades vikt. Närmast ansluter väl Hellbergs benämning till Charles S. Peirce' användning av *icon* som semiotisk term, avseende ett tecken med likhet mellan form och betydelse (jfr Liszka 1996: 37).

Denna principiella synpunkt på språkvetenskapliga beskrivningar (eller rentav vetenskapliga beskrivningar över huvud taget) har sedan dess gjort sig påmind i mina tankar allt emellanåt. I min egen forskning har jag mest sysslat med stilhistoria eller åtminstone språkbruk i historisk belysning. Där är det kanhända enklare att åstadkomma detta att det vanligaste bruket också får störst utbredning i beskrivningen, men det är ändå lätt hänt också där att de udda fallen, ibland av avgränsningsskäl, måste ägnas jämförelsevis stort utrymme. Så småningom har jag tillika hamnat i lexikografiens våld och därvid ibland kommit att fundera litet kring hur det är med ordböckers ikonicitet, och det är detta som den följande framställningen kommer att avhandla, dels resonemangsvis (i avsnitt 1), dels på grundval av en liten stickprovsundersökning jag har gjort i några ordböcker, varvid jag försökt sätta siffror på graden av ikonicitet (i nyssnämnda mening, avsnitt 2).

## 1. Ordböckers ikonicitet i teorin

När det gäller förhållandet mellan frekvens och beskrivningsutrymme har ordböcker – där man beskriver ett ord i taget – givetvis det försteget framom grammatiken att de ovanligaste orden svårligen kan komma att påtagligt breda ut sig så mycket mer än de vanligaste. Däremot är det ingalunda självklart att de nödvändigtvis skulle få mindre utrymme heller, åtminstone näppeligen så mycket mindre som deras förhållandevis lägre frekvens borde givit upphov till – om nu beskrivningen skall vara någorlunda ikonisk i Staffan Hellbergs mening. Ikoniciteten härvidlag är förstås i hög grad avhängig av typen av ordbok. I en ordbok som bara redovisar stavning och böjning tar de allra flesta orden ungefär lika stort utrymme, vilket givetvis är oikoniskt i förhållande till frekvensen men ändå inte alls på det negativt korrelerande sätt som så ofta i grammatiken. Ett litet försteg har dessutom vissa av de vanligare orden i det att hög frekvens gärna medför större variation att redovisa. Vad sedan gäller ordböcker med fylligare lexikografisk beskrivning, omfattande också semantik, konstruktionsmönster och idiomatiska uttryck, så finns det här rimligtvis än mer att säga om framför-

allt de mer frekventa orden. Någon given korrelation finns det emellertid inte för de allra högfrekventaste formorden, till exempel konjunktionen *och*, kan – och bör – beskrivningen göras skäligen generell och därför sannolikt rätt kortfattad. Och de av utomspråkligt sakliga skäl ovanligare orden kan i gengäld ha en rätt vildvuxen formvariation som kan kräva nog så vidlyftigt utrymme, till exempel ord för gamla medicinalväxter (som *tibast* i SAOB). I just en ordbok som SAOB, som grundar sin beskrivning på ett material och dessutom uttryckligen försöker reglera förhållandet mellan detta material och beskrivningen i ett så kallat relationstal (som anger förhållandet mellan antalet belägg i materialet och antalet spalter i den färdiga ordboksartikeln), borde ju ikoniciteten ha riktigt hyfsade förutsättningar att visa till och med positiv korrelation.

Den förväntade ikoniciteten förutsätter dock två saker. För det första måste förstås materialet vara någorlunda frekvensviktat, och det kan vi nog krasst slå fast att detta inte är förhållanden vad gäller SAOB:s material. De grå funktions- och formorden är uppenbart underföreträdda i SAOBArkiv, och det beror givetvis på hur materialet är insamlat. Genom excerpering av bara några få utvalda belägg per sida har excerptisterna haft en tendens att föredra litet färgstarkare ord i sitt urval. Man kan som jämförelse tänka på hur materialet hade sett ut om det istället hade varit totalexcerperingar av ett antal texter. Jag skall här bara ge ett enda exempel, från ord jag själv haft om hand i ordboken: verbet *tro* har 1200 belägg i SAOB:s material, prepositionen *till* 9000 och prepositionen *under* 2100, vilket låter som ett betryggande övertag för de båda formorden, men hade materialet istället utgjorts av en totalexcerpering av det tidningsmaterial som NFO 2 grundar sig på, skulle dessa deras övertag var för sig ha varit mer än dubbelt så stora.

För det andra skulle en materialgrundad ordboks tilltänkta ikonicitet förutsetta likabehandling av alla ord oavhängigt deras plats i beskrivningens strukturhierarki. Här drar SAOB:s redigeringspraxis åt ett annat håll. Självständiga artiklar tillåts normalt breda ut sig förhållandevis mer med avseende på beskrivningsprecision och inte minst antal citerade språkprov än vad sammansättningar eller avledningar behandlade inne i en annan artikel brukar få göra. Återigen skall jag bara ge ett enda exempel från SAOB-praktiken: substantivet *under*, som givetvis är en egen artikel, har i materialet 410 belägg och beskrivs i den tryckta ordboken på 3 spalter; verbet *underkasta*, som behandlas i en sammansättningsramsa inuti artikeln om prepositionen *under*, har 396 belägg och beskrivs på 2 spalter, 69 % av vad det borde ha haft med hänsyn till antalet belägg i jämförelse med substantivet *under*; avledningen *underkastelse*, behandlad i en sekundär avledningsramsa under nyssnämnda sammansättning, har 52 belägg och beskrivs på en tiondedels (0,1) spalt, 38 % av vad det borde ha haft i jämförelse med sitt huvudord och bara 26 % i jämförelse med substantivet

*under*. Denna ojämvikt gäller alltså i förhållande till faktisk beläggsrikedom i SAOB:s material. Om vi istället ginge till faktisk frekvens i språket, blir ojämvikten än påtagligare, eftersom *underkasta* är drygt dubbelt så vanligt som substantivet *under* enligt NFO 2.

## 2. ... och i praktiken

Jag skall nu gå över till en mindre stickprovsundersökning som jag gjort i några danska och svenska ordböcker, där jag försökt mig på att sätta siffror på ikoniciteten vad gäller ordfrekvens och ordboksutrymme. De här talen måste tas med en stor nypa salt, eftersom de är resultat av ett skäligen fyrkantigt tillvägagångssätt. Ordböckerna är DDO och ODS för danskan, SAOL, SO och SAOB för svenskan, och för ordfrekvenserna har jag använt Bergenholtz 1992 och NFO 2 (1971). Detta senare val är förstås inte så rättvisande för de historiska ordböckerna, men i gengäld har jag bara valt ut basala ord, dels de allra vanligaste i språket och de vanligaste substantiven, verben och adjektiven, dels ett antal lågfrekventa – och med tanke på SAOB endast ord på A–T. Jag har därvid utgått från NFO 2, som ger tal för lemmatiserade ord, och sedan tagit motsvarande ord i den danska frekvensordboken, som endast redovisar graford. För danskans del har jag därför fått lämna därhän vissa av orden, eftersom de uppenbart döljer homografer. De utvalda orden har sedan frekvensviktats inbördes, och så har jag som utgångspunkt för jämförelser tilldelat ord vars frekvens är en 1000-del eller mindre av det allra högfrekventaste ett ikoniskt omfång om en rad i SAOL och tre rader i DDO och SO, ord vars frekvens är en 5000-del eller mindre av det högfrekventaste sex rader i SAOB samt ord vars frekvens är en 10000-del eller mindre därav tre rader i ODS. Den sistnämnda särbehandlingen beror på att de ovanligare danska orden är förhållandevis ovanligare jämfört med de vanligaste än i svenskan, vilket sannolikt sammanhänger med att det är de svenska orden som varit utgångspunkt – med följd att vissa av de danska motsvarigheterna helt enkelt inte är normalsynonymen där.

Denna myckna sifferexercis har givit det utfall som redovisas i tabellen.

*Tabell. Relativ andel av ikoniskt ordboksutrymme för vanliga och ovanliga ord utifrån en tilldelning av en rad i SAOL, tre rader i DDO, ODS och SO och sex rader i SAOB åt de lågfrekventaste orden (taget som en snävare kategori i ODS och SAOB).*

De vanligaste orden						Några ovanliga ord					
Ord	DDO	ODS	SAOL	SO	SAOB	Ord	DDO	ODS	SAOL	SO	SAOB
<i>den</i>	3	5	3	2	27	<i>entenlantingen</i>	400	170	150	600	500
<i>og/och</i>	5	6	1	3	6	<i>førendlfförrän</i>	400	4000	50	100	130

<i>i</i>	12	11	1	3	10	<i>långs/längs</i>	650	400	<b>100</b>	216	<b>100</b>
<i>at/att</i>	8	8	1	2	19	<i>nogen-/någonting</i>	-	1000	<b>100</b>	233	<b>100</b>
<i>en</i>	23	9	2	3	9	<i>unden/undan</i>	-	1200	<b>150</b>	283	500
<i>på</i>	30	26	1	6	<b>105</b>	<i>somme, pl./somlig</i>	667	340	400	333	<b>130</b>
<i>till/till</i>	32	36	2	26	44	<i>hvorvidt/huruvida</i>	400	360	200	600	<b>130</b>
<i>med</i>	25	18	1	7	31	<i>inbyrdes/inbördes</i>	467	950	300	267	867
<i>aflav</i>	26	9	2	6	8	<i>ligesådan/likadan</i>	467	2600	200	300	333
<i>som</i>	25	38	2	5	13	<i>enhver/envar</i>	344	<b>133</b>	300	233	500
substantiv:											
<i>dag</i>	553	160	19	201	325	<i>kro/krog</i>	533	400	<b>100</b>	367	667
<i>del</i>	344	<b>117</b>	19	<b>96</b>	315	<i>provinsby/ landsortsstad</i>	267	<b>140</b>	200	33	<b>50</b>
<i>gang/gång</i>	234	79	20	142	173	<i>pistol</i>	767	280	<b>100</b>	467	1000
<i>land</i>	431	200	45	<b>103</b>	<b>91</b>	<i>silhouet/siluett</i>	700	900	300	367	1300
<i>tid</i>	455	206	33	<b>81</b>	171	<i>serart/särart</i>	-	400	<b>100</b>	200	300
verb:											
<i>ha</i>	...	...	2	16	21	<i>bemestre/bemästra</i>	233	2400	<b>100</b>	267	200
<i>kunne/-a</i>	24	15	1	16	13	<i>förudsigel/förutspå</i>	800	300	<b>100</b>	167	250
<i>skulle/skola</i>	54	32	3	16	61	<i>harmonisere/-a</i>	567	3000	200	367	500
<i>få</i>	...	...	2	34	57	<i>skimre/skimra</i>	-	2000	<b>100</b>	633	2000
<i>blive/bli</i>	<b>60</b>	17	5	12	<b>93</b>	<i>mala</i>	...	...	400	733	3000
adjektiv:											
<i>gammell/-al</i>	188	105	19	112	<b>88</b>	<i>barnlig/barnslig</i>	567	333	<b>100</b>	267	4000
<i>god</i>	141	109	80	220	573	<i>farverig/färgglad</i>	433	<b>100</b>	<b>100</b>	133	<b>100</b>
<i>ny</i>	121	<b>60</b>	5	35	34	<i>karrig/karg</i>	567	2000	200	333	1000
<i>stor</i>	143	148	4	41	131	<i>kynisk/cynisk stereotyp</i>	433	400	300	333	1000
							400	600	300	267	300

Siffrorna i tabellen anger procentuell andel av det i förhållande till frekvensen i faktiskt språkbruk ikoniska ordboksutrymmet (givet det tilldelade radutrymmet för de lågfrekventaste orden); idealet är med andra ord 100 – då är utrymmet helt ikoniskt. (Tal mellan 50 och 150 = någorlunda ikoniska värden utmärks med fet stil i tabellen.) Nu skall inte dessa tal läsas som ”hur väl man lyckats” i detta hänseende, eftersom detta slags ikonicitet knappast har varit något medvetet mål alls i någondera ordboken (annat än möjligtvis ytterligt medelbart i SAOB genom dess relationstal (se avsnitt 1 ovan)); det man inte har någon tanke på att eftersträva kan man inte heller lyckas eller misslyckas med att uppnå. Närheten till eller avståndet från det ikoniska 100 är istället huvudsakligen ett utfall av andra omständigheter och överväganden kring redigeringen av de utvalda orden.

Som väntat finner vi överlag låga för att inte säga mycket låga tal för språkets allra vanligaste ord och lika väntat allra lägst i ordlistan SAOL, högst i SAOB.



Också ODS och i än större utsträckning DDO uppvisar ett antal tvåsiffriga tal, och särskilt DDO mycket jämnare fördelade sådana än de svenska. Möjligen återspeglar detta senare att DDO bygger på en datoriserad korpus med totalexcerperade texter (där givetvis faktisk frekvens mycket tydligare får fullödigt genomslag). I SAOB återfinns dessutom en ytterst påtaglig avvikare i sällskapet: prepositionen *på* med nästintill ikoniskt omfång, en följd av den enastående detaljrikedom som denna artikel uppvisar (i sin tur väl delvis en följd av de tilläggsexcerperingar som redaktören uppges ha gjort i arbetet med detta ord). Av de vanligaste innehållsorden håller sig sedan substantiven överlag väl framme, också i ordlistan (jämförelsevis i alla fall). I SAOB har *dag* och *del* till och med tre gånger större utrymme än vad frekvensen borde lett till, och i DDO *dag*, *land* och *tid* hela fyra-fem gånger större utrymme. De högfrekventa adjektiven har påfallande jämnt fördelat och rätt ikoniskt utrymme i de två danska ordböckerna. Här närmar sig rentav också SAOL full ikonicitet för *god* (detta tack vare ovanligt frikostig redovisning av idiomatiska uttryck samt en särupplysning om komparativens och superlativens användning). I SAOB har *god* fått nästan sex gånger ”för stort” utrymme. Bland de ovanligare orden har förstås många lika oikoniskt stort utrymme, eller ännu mycket större. Detta gäller särskilt i de två stora ordböckerna, ODS och SAOB, med SAOB:s *barnslig* som allra ”värst”: 40 gånger större än sitt ikoniska omfång.

Ordet *barnslig* är avledning av *barn* men behandlas i SAOB i en egen artikel (liksom *barnlig* i ODS). Å andra sidan är det, mot bakgrund av vad som sades i avsnitt 1 ovan, värt att lägga märke till att det enda ordet bland de ovanligare som underskrider sitt ikoniska omfång är flerledssammansättningen *landsortsstad* i SAOB, där redovisat i en sekundär sammansättningsramsa (under *landsort*, i sin tur i ramsa under *land*). Också i ODS gynnar sådan osjälvtändig ställning i beskrivningen ikonicitet bland de lågfrekventa orden: både *provinsby* och *farverig* är beskrivna i sammansättningsramsor inuti respektive förleds artikel.

Om vi jämför ordböckerna parvis över språkgränserna, kan vi se att DDO uppvisar påtagligt bättre ikonicitet än SO för de vanligaste formorden men å andra sidan har oikoniskt stort utrymme för de vanligaste substantiven, där SO överlag ligger nära det ikoniska idealet. Den viktigaste iakttagelsen vad gäller ODS och SAOB är kanhända den att ODS överlag har en litet jämnare nivå i respektive grupp av vanliga ord – ingalunda alltid ikonisk men jämförelsevis jämn, där SAOB:s variation däremot är rätt påfallande. I gengäld har SAOB en jämnare nivå i gruppen lågfrekventa formord, där det till och med är så att hälften har fått ett någotsånär ikoniskt utrymme.

All denna sifferexercis skall nu tas för vad den är, och den tillåter knappast att dra alltför stora växlar på. Dels är dess utfall tillkommet med en grovhuggen metod vars validitet kan ifrågasättas, dels bevisar den något som vi i dess huvud-

drag resonemangsvis kunde räkna ut på för hand, dels mäter den ju något som bara är en sidoverkan av helt oavhängiga ställningstaganden i den lexikografiska praktiken. Huruvida den inte bara borde vara denna oavsedda sidoverkan får bli ämnet för det avslutande avsnittet.

### 3. Ikonicitet alls önskvärt?

Är då det större mått av ikonicitet som ändå legat som ett slags uttalat eller rentav halvt uttalat idealtillstånd bakom den hittillsvarande framställningen alls något att sträva efter, ens i en stor och uttömmande lexikografisk beskrivning som den i ODS eller SAOB? Fullständig ikonicitet kan vi genast avfärda som rätt eller snarare alldeles orimlig (liksom den givetvis är det också i en grammatisk framställning – något annat menade naturligtvis inte heller Staffan Hellberg). Det verkar minst sagt långsökt att SAOB:s artikel om konjunktionen *och* skulle göras 20 gånger längre bara för att motsvara ordets enastående vanlighet och väl än orimligare att språkets någorlunda grundläggande men ändå inte så väldigt frekventa ord, av typen *mala*, *karg* och *barnslig*, skulle behöva avstå från en fullödig beskrivning bara för att hovsamt hålla sig inom sina ikoniska skrankor. Likafullt vill jag hävda att det kan finnas skäl att emellanåt ha med detta slags ikonicitet i bakhuvudet; annars hade jag förstås inte alls tagit till orda i ärendet. Användningen av detta som argument för en fylligare, eller för den delen mindre fyllig, beskrivning måste givetvis lämpas efter läglighet, men ett ords centralitet i ordförrådet borde vara en självständig faktor, bland många andra förvisso (och delvis samspelande), när man avgör hur mycket man vill avsätta av ordbokstexten åt det. Att utrymmet i förhållande till frekvens är så väldigt olika inom var sina grupper med frekvensmätigt snarlika typer av ord som det ändå är i SAOB bör till exempel trots allt kunna verka som ett litet memento i sammanhanget. (Det återstår ju dock att jämföra dessa förhållanden också med materialets omfång för respektive ord.) Åtminstone vill jag så här till sist slå ett slag för att ett ords beskrivning såsom en självständig artikel eller såsom en underordnad del av en annan artikel inte i sig borde ha någon betydelse för hur stort utrymme det tillåts få, det vill säga hur ikoniskt dess omfång blir.

#### KÄLLOR OCH LITTERATUR:

Bergenholtz, Henning. 1992: Dansk frekvensordbog baseret på danske romaner, ugeblade og aviser 1987–1990. København.

- DDO = Den danske ordbog, hovedred.: Ebba Hjorth & Kjeld Kristensen, udg. af Det Danske Sprog- og Litteraturselskab, sex band. København 2003–2005.
- Liszka, James Jakób, 1996: A General Introduction to the Semeiotic of Charles Sanders Peirce. Bloomington & Indianapolis.
- NFO 2 = Allén, Sture, 1971: Nusvensk frekvensordbok baserad på tidningstext 2. Lemman. Stockholm.
- ODS = Ordbog over det danske sprog, grundlagt af Verner Dahlerup, udg. af Det Danske Sprog- og Litteraturselskab, bind 1–27 + supplementbind 1–5. København 1919–1954, 1992–2005.
- SAOB = Ordbok över svenska språket, utg. av Svenska Akademien, 1893 ff. Lund.
- SAOL = Svenska Akademiens ordlista över svenska språket, trettonde uppl. Stockholm 2006.
- SO = Svensk ordbok, utg. av Svenska Akademien, utarbetad vid redaktionen för Svenska Akademiens samtidsordböcker, Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet, två band. Stockholm 2009.

*Bo-A. Wendt*

Svenska Akademiens ordboksredaktion  
bo.wendt@svenskaakademien.se