

Nordiske studier i leksikografi

12



# Nordiske studier i leksikografi

## 12

Rapport fra Konferanse om leksikografi i Norden  
Oslo 13.–16. august 2013

Redigert av

Ruth Vatvedt Fjeld  
Marit Hovdenak

Skrifter utgitt av Nordisk forening for leksikografi  
Skrift nr. 13



---

NOVUS FORLAG  
OSLO 2014

Rapporten kan bestilles fra:  
Språkrådet  
Postboks 8107 Dep  
0032 OSLO

© Nordisk forening for leksikografi 2014.

ISBN: 978-82-7099-763-3

Trykk: Interface Media as

Det må ikke kopieres fra denne boka i strid med åndsverksloven eller avtaler om kopiering inngått med KOPINOR, interesseorgan for rettighetshavere til åndsverk.

## Forord

Den første nordiske konferansen i leksikografi ble holdt i Oslo i mai 1991, og foredragene derfra ble i hovedsak presentert i *Nordiske studier i leksikografi 1992*. Denne første rapporten fikk ikke noe nummer, da vi jo ikke kunne vite at 1991-konferansen skulle bli den første i en lang rekke nordiske konferanser innen fagfeltet, men vi regner den selvsagt som nummer 1. Her foreligger den foreløpig siste, *Nordiske studier i leksikografi 12*.

Den tolvte nordiske konferanse i leksikografi ble arrangert i Oslo 13.–16. august 2013. Arrangør var Språkrådet i Norge i samarbeid med Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo. Hovedansvarlig arrangør var Jan Hoel, Språkrådet. Arrangørkomiteen for øvrig besto av Rikke E. Hauge og Øystein Baardsgaard, Språkrådet, Ruth Vatvedt Fjeld og Åse Wetås, Universitetet i Oslo.

Programkomiteen har bestått av Ruth Vatvedt Fjeld (Universitetet i Oslo) og Marit Hovdenak (Språkrådet), som også har redigert konferanserapporten.

På konferansen ble det holdt 49 faglige presentasjoner, av disse var 3 plenumsforedrag, 38 parallellforedrag, 1 fellesforedrag med demonstrasjon, 2 demonstrasjoner og 5 plakatpresentasjoner. De innsendte sammendragene ligger her: <http://språkrådet.no/Toppmeny/Aktuelt/Den-tolvte-konferansen-om-leksikografi-i-Norden-begynner-i-Oslo-i-morgen-tirsdag-13-august/>. Det var 88 deltakere på konferansen, fordelt slik etter nasjoner: Norge 49, Sverige 14, Danmark 11, Finland 9, Island 3, Polen 1,

## FORORD

Russland 1. Konferanserapporten inneholder 25 artikler. Alle artikler er lest og godkjent av to uavhengige lesere. Den enkelte forfatter er ansvarlig for det endelige innhold i artiklene.

Redaktørene har delt ansvaret mellom seg og har kommentert grundig de enkelte artiklene, i tillegg til en anonym spesialist i den enkelte artikkels emne. Kollegaer som har stilt sin tid og kompetanse til rådighet for kollegavurdering:

Lars Borin, Øystein Baardsgaard, Anne Engø, Piotr Garbacz, Anne Golden, Kjetil Gundersen, Kristin Hagen, Jan Hoel, Jon Haarberg, Tor Erik Jenstad, Janne Bondi Johannessen, Rune Lain Knudsen, Pia Lane, Patrick Leroyer, Erlend Lønnum, Astrid Nilsen-Nygaard, Anna Senje, Sverre Stausland, Arne Torp, Lars Trap-Jensen, Erik Velldal, Lars S. Vikør, Boye Wangensteen, Dagfinn Worren.

Masterstudent Kjersti Wictorsen Kola har gransket manus og rettet litteraturreferanser og grafisk oppsett. Alle takkes herved hjertelig for innsatsen!

Av de tre plenumsforedragene er bare to trykt i denne rapporten. Arleta Adamska-Salaciaks foredrag var en kortversjon av hennes artikkel ”Sameness of meaning in a bilingual dictionary”, som er trykt i *International Journal of Lexicography*, vol 27/2013, og det var dermed ikke noe poeng i å trykke en forenklet eller oversatt versjon i konferanserapporten. Victoria Roséns artikkel ”Språkteknologiens behov for leksikalsk informasjon ” behandler særlig de utfordringer som leksikografien står overfor som medspiller i generell data-maskinell utforskning og dokumentasjon av språk. Lars Trap-Jensens ”Leksikografisk tradition og fornyelse: tre revolusjoner på 100 år?” dekker et tema som passer godt som tilbakeblikk på de over 20 årene som er gått med nordisk samarbeid om utforskning av de nordiske språkenes ordforråd og utvikling av

ordbøker og annen dokumentasjon av nordisk ordforråd gjennom Nordisk forening for leksikografi. Disse to oversiktsartiklene innleder denne rapporten. De øvrige artiklene er ordnet alfabetisk etter forfatternavn, da det var vanskelig å gruppere dem innholdsmessig. Mange av artiklene omhandler flere relevante leksikografiske emner, og dermed kan en systematisk inndeling virke mer forvirrende enn organiserende. Vi følger derfor det anerkjente leksikografiske ordningsprinsippet: alfabetet!

Spredningen i temaene er god, med artikler om leksikografiens historie, leksikografisk terminologi og leksikografiens framtid, faglige utvikling og utfordringer i informasjonens tidsalder. Digitalisering både av materiale, metoder og produkter er blitt selvsagt, men de problemer vi strever med for å fange inn og beskrive forskjellige typer ordforråd, er langt på vei de samme. Ordenes flyktighet og foranderlighet gjør at leksikografi som fag er like interessant og utfordrende som før, til tross for alle nye hjelpemidler.

Av forskjellige grunner er det en del foredragsholdere som ikke har hatt anledning til eller har ønsket å publisere sitt innlegg som artikkel i konferanserapporten. Rapporten inneholder derfor 25 artikler. Det synes vi er et godt resultat, og spredningen mellom temaene er så stor at vi mener rapporten gir et godt innblikk i de siste to årenes leksikografiske arbeid og framgang i Norden. Dette er noen færre enn i de to største NFL-rapportene, med 53 fra konferansen i Lund i 2011 og samme antall fra konferansen i Oslo i 1991. En rask gjennomgang viser at konferanserapportene stort sett har hatt mellom 30 og 50 artikler. I en tid da tilbudet om konferanser og andre internasjonale møter har økt voldsomt og de fleste velger med omhu hva de prioriterer å delta på, sier det at de nordiske konferansene i leksikografi står seg godt, og det mener vi er uttrykk

## FORORD

for et spesielt godt og langvarig fagfellesskap. Det er da også en stor glede både å få arrangere og få delta på disse konferansene! Vi ønsker alle lesere, både konferansedeltakere og andre, gode stunder med konferanserapporten og ser fram til videre samarbeid om utforskning og dokumentasjon av ordforrådet i de nordiske språkene.

Konferansen har mottatt økonomisk støtte fra Nordpluss nordiske språk og kultur, fra Språkrådet og Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo. Konferanserapporten er bekostet av Språkrådet og Institutt for lingvistiske og nordiske studier.

Redaktørene ønsker også å takke Novus forlag for utmerket samarbeid om trykking og grafisk bearbeiding av manuset.

*Ruth Vatvedt Fjeld og Marit Hovdenak*



## **Innhold**

<b>Språkteknologiens behov for leksikalsk informasjon</b>	13
<i>Victoria Rosén</i>	
<b>Leksikografisk tradition og fornyelse: tre revolusjoner på 100 år?</b>	42
<i>Lars Trap-Jensen</i>	
<b>Grunden för projektet av en studieordbok över svenska fasta liknelser som beskriver människan</b>	69
<i>Alexej Alösjin</i>	
<b>Skjematiske fleirordskonstruksjonar i <i>Norsk Ordbok</i></b>	80
<i>Sturla Berg-Olsen</i>	
<b>Nydigitalisering av SAOB</b>	95
<i>Erik Bäckerud</i>	
<b>Automatisk identifisering av semantisk förändring med hjelp av lexikala distributionella mønster</b>	106
<i>Karin Cavallin</i>	
<b><i>Nordisk miniordbok</i> – et kritisk blick</b>	121
<i>Ruth Vatvedt Fjeld og Janne Bondi Johannessen</i>	

INNHold

<b>Samvirket mellom ordbank og ordbok</b> <i>Oddrun Grønvik og Christian-Emil Ore</i>	139
<b>Ivar Aasen og Universitetet i Oslo</b> <i>Kjetil Gundersen og Dagfinn Worren</i>	159
<b>Kvensk ordbok på nett – hvem har nytte av den?</b> <i>Mervi Haavisto, Kaisa Maliniemi, Leena Niiranen, Pirjo Paavalniemi, Tove Reibo og Trond Trosterud</i>	176
<b>Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi</b> <i>Anna Helga Hannesdóttir</i>	193
<b>Leksikografiens grunnenhet – hva heter det, dette vi lager ordbokartikler om?</b> <i>Petter Henriksen</i>	212
<b>Bokmålsordboka og Nynorskordboka gjennom ein generasjon</b> <i>Marit Hovdenak</i>	229
<b>Marcus Pauli Sadeleri – finsk medarbeitare i Schroderus <i>Lexicon Latino-Scondicum</i></b> <i>Petri Lauerma</i>	247
<b>Petter Dass' ordforråd – hans kilder og hans forfatterskap som kilde for ordbøker</b> <i>Hanne Lauvstad</i>	267

<b>Vinsmagningsordbogen <i>Ænolex Bourgogne</i>.</b> <b>En milepæl i pragmatisk fagleksikografi</b>	287
<i>Patrick Leroyer &amp; Asta Høy</i>	
<b>At have begreb skabt om noget</b> <b>– om Den Danske Begrebsordbog</b>	303
<i>Henrik Lorentzen, Sanni Nimb &amp; Thomas Troelsgård</i>	
<b>Om (bristende) bruklighetsstruktur i SAOB</b>	327
<i>Anki Mattisson</i>	
<b>Nja... En korpusbaseret undersøgelse af interjektioner i talesprog</b>	341
<i>Jessie Leigh Nielsen</i>	
<b>Att läsa ordboken baklänges.</b> <b>Om betydelseutveckling hos partikelinledda fasta sammansättningar och presentationen i SAOB</b>	353
<i>Pär Nilsson</i>	
<b>Från aspekt till övergripande</b> <b>– en ordlista över svensk akademisk vokabulär</b>	370
<i>Judy Ribbeck, Håkan Jansson &amp; Emma Sköldberg</i>	
<b>Från dilation till uppskov. Var ordböcker viktiga i etableringen av ett juridiskt ordförråd?</b>	385
<i>Lena Rogström</i>	
<b>Fypike gjorde selbureis i røytemåned.</b> <b>Om <i>Norsk-russisk ordbok</i> av Vladimir Arakin: utvalg, kilder, målgruppe og formål</b>	402
<i>Aud Anna Senje</i>	

INNHOLD

<b>Brugerne er allerede mobile!</b>	416
<i>Henrik Køhler Simonsen</i>	
<b>Tvåspråkiga morfemordböcker för inlärare</b>	430
<i>Jonas Winnerlöv</i>	
<b>Konferanseprogram</b>	448

# **Språkteknologiens behov for leksikalsk informasjon**

*Victoria Rosén*

The information to be included in a lexicon depends on its intended use. Automatic syntactic analysis of a large and varied text corpus requires detailed information on inter alia countability, compounds, spelling and inflectional variants, neologisms, and multiword expressions, information that cannot always be harvested from traditional dictionaries. In this article it is shown how deep syntactic analysis of a corpus can contribute to the enrichment of lexical resources.

## **1. Introduksjon**

Språkteknologiske applikasjoner trenger detaljert leksikografisk informasjon om flest mulig ord. Mange slike applikasjoner bruker leksikalske ressurser som har vært utviklet som tradisjonelle papirordbøker, der hensikten har vært å gi mennesker nyttige opplysninger om bruken av ordene i språket. Datamaskinelle programmer kan ha behov for andre typer leksikalsk informasjon, eller for at informasjonen organiseres på en annen måte.

Den leksikalske informasjonen som er nødvendig i oppbyggingen av en trebank for norsk, kan illustrere dette

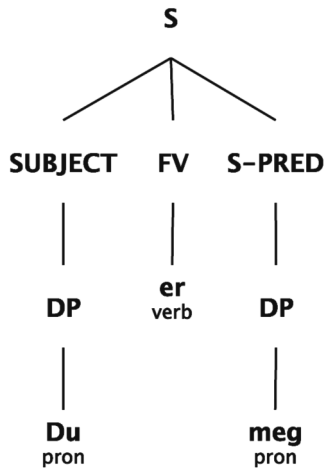
behovet. INESS lager en forskningsinfrastruktur for trebanker, som er korpora annotert med syntaktiske strukturer.<sup>1</sup> Prosjektet har to hovedmål: å lage en infrastruktur for trebanker og å lage en stor trebank for norsk. Det å lage en infrastruktur innebærer å gjøre trebanker lettilgjengelige for brukere; det skal være mulig å finne trebankene online og søke i dem gjennom å bruke en vanlig nettleser. Hovedtrekkene i hva trebanker er, og hvordan den norske trebanken i INESS utvikles, blir presentert i avsnitt 2 nedenfor. I avsnitt 3 vises det hvordan informasjon fra et språkteknologisk leksikon<sup>2</sup> som er utviklet fra en tradisjonell ordbok, kan suppleres og justeres slik at den syntaktiske analysen i trebanken blir mer treffsikker. Avsnitt 4 viser hvordan ord som ikke finnes i tradisjonelle ordbøker, og som man kanskje ikke engang ønsker der, legges til i de leksikalske ressursene som brukes i INESS.

## 2. Trebanker og syntaktisk analyse

Termen trebank kommer av at de første trebankene inneholdt frasestrukturtrær, som i figur 1. En slik trebank kalles gjerne en

- 
1. INESS (Infrastructure for the Exploration of Syntax and Semantics) er et prosjekt innenfor NFRs INFRASTRUKTUR-program (Rosén et al. 2012). INESS inngår i den norske delen av CLARIN-samarbeidet, CLARINO. Se: [clarino.uib.no/iness](http://clarino.uib.no/iness).
  2. I datalingvistikken brukes termen *leksikon* (engelsk *lexicon*) om en elektronisk ordbok.

konstituenttrebank (etter engelsk *constituency treebank*), siden trærne representerer setningenes analyse i syntaktiske konstituer.

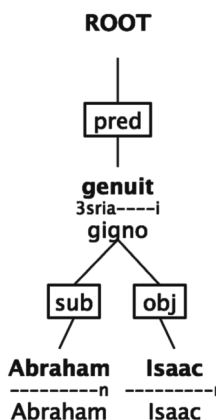


Figur 1: Syntaktisk tre for setningen *Du er meg* fra den norske Sofie-trebanken (Losnegaard et al. 2013).

Også andre typer syntaktiske strukturer forekommer i trebanker. Mange trebanker er dependenstrebanker, basert på dependensanalysen opprinnelig foreslått av Tesnière (1959). Det som står sentralt i en dependensanalyse, er avhengigheter mellom enkeltord i setningen. Et eksempel på en dependensanalyse er gitt i figur 2.

Noen trebanker er basert på bestemte lingvistiske formalismer, som for eksempel Head-driven Phrase Structure Grammar (HPSG; Pollard & Sag 1994) og Lexical Functional Grammar (LFG; Bresnan 2001, Dalrymple 2001).

Trebanker kan konstrueres på ulike måter. Det er mulig å lage en trebank helt manuelt, det vil si at annotatorer tildeler



Figur 2: Dependensstruktur for den latinske setningen *Abraham genuit Isaac* ‘Abraham ble far til Isak’ fra PROIEL-trebanken (Haug & Jøhndal 2008).

hver setning i et korpus en syntaktisk struktur. En ulempe med denne metoden er at den er svært arbeidskrevende. Det vanlige er derfor at korpuset parses automatisk. På grunn av leksikalsk og syntaktisk flertydighet vil denne metoden vanligvis resultere i flere mulige analyser for hver setning; av og til kan det være tusenvis av analyser for én enkelt setning. Derfor krever metoden at man entydiggjør, slik at man kan velge én analyse blant de foreslåtte. Entydiggjøring kan enten gjøres av annotatorer eller ved hjelp av statistiske metoder.

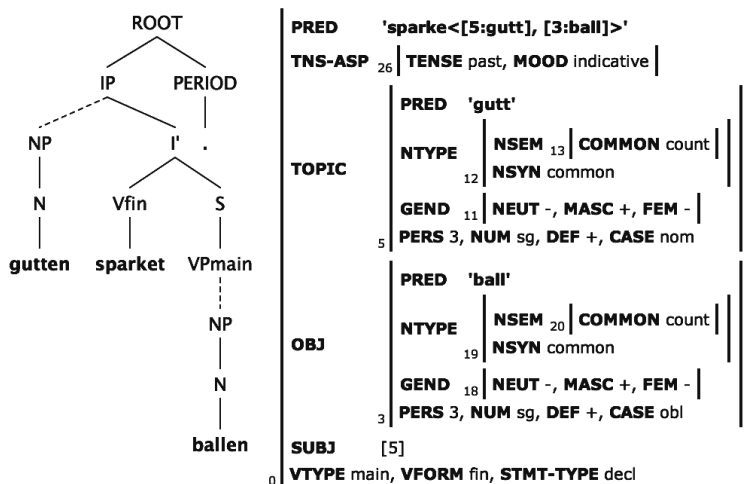
En viktig oppgave i INESS-prosjektet er utviklingen av en trebank for norsk. Den håndskrevne grammatikken som brukes til å parse korpuset, er NorGram (Dyvik 2000, Butt et al. 2002).<sup>3</sup>

3. Parsingplattform er Xerox Linguistic Environment (XLE), utviklet ved PARC (Palo Alto Research Center) i California (Maxwell & Kaplan 1993).



Strukturene i trebanken er basert på LFG, og det innebærer at hver setning får både en konstituentstruktur (c-struktur) og en funksjonell struktur (f-struktur). C-strukturen viser hvordan ordene i setningen er hierarkisk gruppert sammen i fraser. F-strukturen viser hva slags syntaktiske funksjoner frasene i setningen har, for eksempel subjekt og objekt, og også grammatiske trekk, som tall og tempus. I figur 3 vises c- og f-strukturene for setningen *Gutten sparket ballen*.

LFGs c- og f-strukturer inneholder omfattende grammatisk informasjon. Én av fordelene med trebanker, i motsetning til korpora som mangler syntaktisk annotasjon, er at det er mulig å søke ikke bare på bestemte ord, ordklasser og morfo-syntaktiske trekk, men også på bestemte grammatiske konstruksjoner som ord inngår i. For eksempel kan man lett finne hvilke verb som brukes i passiv. I et tagget korpus må man søke på perfektum partisipp og eventuelt også på et hjelpe-

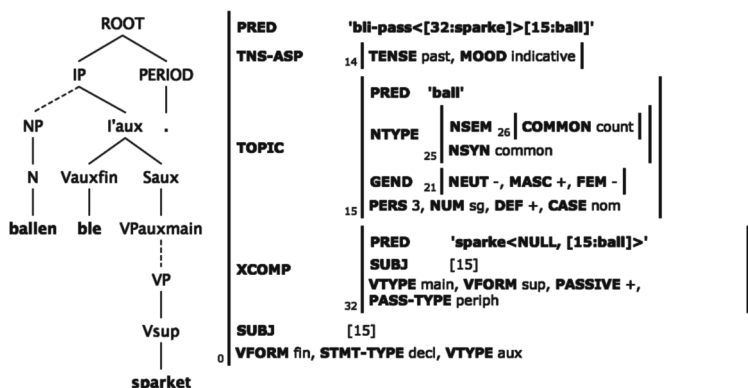


Figur 3: C- og f-struktur for setningen *Gutten sparket ballen*.

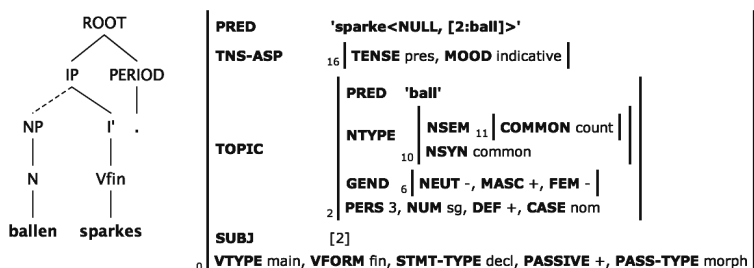
verb, eller på s-formen av verbet. Slike søk vil imidlertid gi mange treff som ikke gjelder passiv, siden både perfektum partisipp, hjelpeverbene og s-form av verb har andre funksjoner enn å danne passiv.

I en trebank kan man søke på passivkonstruksjoner på en mer direkte måte. Figur 4 viser en analyse av perifrastisk passiv, og figur 5 en analyse av morfologisk passiv. Begge analysene har verdien + for det grammatiske trekket *PASSIV* i f-strukturen. Den perifrastiske passiven har også trekket *PASS-TYPE* (passivtype) med verdi *periph* (perifrastisk), mens den morfologiske passiven har verdien *morph* (morfologisk) for dette trekket. Disse trekkene gjør det mulig å finne passiv på en enkel og treffsikker måte. For leksikografisk beskrivelse kan slike søk bidra til å klargjøre om et verb vanligvis brukes i perifrastisk passiv, morfologisk passiv eller begge deler.

Når man lager en trebank ved hjelp av en komputasjonell grammatikk, trenger man gode leksikalske ressurser. NorGram bruker flere ulike leksika i forbindelse med den syntaktiske



Figur 4: C- og f-struktur for setningen *Ballen ble sparket*.



Figur 5: C- og f-struktur for setningen *Ballen sparkes*.

analysen. De største leksikaene er basert på NorKompLeks (NKL; Nordgård 2000), et elektronisk leksikon som i sin tur er basert på *Bokmålsordboka* (BOB: Wangensteen 2005) og *Nynorskordboka* (NOB: Hovdenak et al. 2006). I tillegg til ordklasse og bøyning samt oppdelingen i lemmaer, som kommer fra BOB og NOB, inneholder NKL også informasjon om valens, altså informasjon om hvilke argumenter verb tar. Nettopp valensinformasjon er svært viktig for dyp syntaktisk analyse.

I tillegg til leksikaene som er basert på NKL, har NorGram et håndskrevet leksikon. Dette brukes bl.a. for å kode egenkapene til grammatiske ord. For at parsing skal fungere, må de syntaktiske reglene og leksikonet være samstemte med hensyn til morfosyntaktiske trekk. Som regel må de grammatiske ordene ha mer detaljert informasjon enn det som finnes i BOB og NOB.

Den norske trebanken som bygges i INESS-prosjektet, er en dynamisk trebank. Det første trinnet i arbeidet med å lage en slik trebank er å parse korpustekster automatisk. Blant de analysene som parseren foreslår, må annotatorene velge den riktige. Entydiggjøringsmetoden som brukes i INESS-trebanken, er at annotatorene velger mellom analyser gjennom å bruke diskrimi-

nanter, altså enkle egenskaper slik som entall vs. flertall, eller substantiv vs. verb. Når den ønskede analysen ikke finnes, må grammatikken og/eller leksikonet oppdateres. Etter slike oppdateringer vil en ny runde med parsing resultere i en bedre syntaktisk beskrivelse av korpuset. Denne metoden gir en dynamisk trebank der analysene kan videreutvikles parallelt med grammatikk og leksikon.

Selv med et stort leksikon og en velutviklet grammatikk får man ikke alltid en god analyse. En undersøkelse av de 255 første setningene i *Sofies verden* ble utført for å finne ut hvorfor noen setninger ikke fikk rett analyse (Losnegaard et al. 2012). Resultatet viste at 29 % av tilfellene skyldtes syntaktiske problemer, mens 71 % var leksikalske. Og blant de leksikalske problemene var de to vanligste typene flerordsuttrykk (41 % av tilfellene) og leksikalske kategorier (31 %). Resten av denne artikkelen handler om hva slags leksikalsk informasjon som er nødvendig for å få en fullverdig syntaktisk analyse.

### **3. Hvilken informasjon om ord er nødvendig for syntaktisk analyse?**

For å få gode syntaktiske analyser trenger man korrekt informasjon om bl.a. ordklasse, valens, tellelighet, sammensetningsstruktur og flerordsuttrykk. Mye av denne informasjonen er tilgjengelig i NKL (med utgangspunkt i *BOB* og *NOB*). I INESS-prosjektet viser det seg likevel ofte at disse leksikalske

ressursene mangler informasjon som er nødvendig for å analysere ganske vanlige ord.

### 3.1. Ordklasse

Å hente informasjon om ordklasse fra en ordbok kan kanskje virke trivielt, men det er ikke alltid problemfritt. Det er flere grunner til at den ordklasseinformasjonen som er nødvendig, ikke alltid er tilgjengelig. Det hender at ordklassen ikke finnes i NKL fordi den var «gjemt» inne i ordartikkelen i *BOB/NOB*. Noen ganger krever syntaktisk analyse en annen ordklasse enn den som finnes i *BOB/NOB*.<sup>4</sup>

#### 3.1.1. Ordklassen finnes inne i ordartikkelen

Hovedregelen for ordklassemarkering i *BOB* er at ordklassen angis rett etter oppslagsordet i artikkelhodet. I artikkelkroppen følger de ulike delbetydningene av ordet. Et eksempel gis i (1).

- (1) **I abstrakt** n3 (av *II abstrakt*, i bet. 2 fra eng.) **1** substantiv med abstrakt betydning [...] **2** sammendrag, referat  
**II abstrakt** a2 (fra lat., se *abstrahere*) **1** som bare kan fattes gjennom tanken [...] **2** språkv: som betegner en egenskap, tilstand el. handling [...]

---

4. Eksemplene i det følgende er tatt fra *BOB*, men forholdene er stort sett like i *NOB*.

Eksempel (1) viser at *abstrakt* har fått to oppslag markert med romertall, ett for substantivet og ett for adjektivet. Innenfor disse er ulike delbetydninger skilt fra hverandre med nummerring i arabiske tall.

I noen ordartikler står det en ny ordklassemarkering etter et arabisk tall, altså etter det som vanligvis markerer en ny delbetydning. Et eksempel gis i (2).

- (2) **I absolutt** n3 (av *II absolutt*) **1** absolutt krav **2** entydig begrep **II absolutt** a2 (fra lat. av *absolvere* ‘løse fra’) **1** hel, fullstendig, uinnskrenket [...] **2** adv: betingelsesløst, endelig, helt [...]

For *II absolutt* er ordklassen adjektiv markert på vanlig måte rett etter oppslagsordet, men det forekommer likevel en ny ordklassemarkering for adverb etter tallet 2. Konverteringen fra *BOB* til *NKL* har tydeligvis ikke tatt hensyn til slike ordklassemarkeringer siden *NKL* bare har kategoriene substantiv og adjektiv for ordet *absolutt*.

Det er imidlertid ikke bare etter et arabisk tall at en ny ordklassemarkering kan forekomme. Eksemplene (3)–(5) viser at slik markering også kan finnes inne i definisjonsdelen av artikkelen.

- (3) **jevnlig** el. **jamnlig** adv, med jevne mellomrom, stadig *besøke sine foreldre j-* / adj: hyppig *j-e besøk*<sup>5</sup>  
 (4) **juble** v1 (gj ty. fra mlat og lat., se *jubel*) rope, synge, le av glede *spillerne j-t over seieren / j- av glede / adv i pr pt: være j-nde glad*

---

5. I *NOB* er det omvendt: oppslaget for *jamleg* oppgir a2 som ordklasse rett etter oppslagsformen, og betegnelsen *adv* dukker opp senere i ordartikkelen.

- (5) **ofte** adv (norr *opt*) hyppig, mange ganger *møtes o-* / *oftest* i de fleste tilfeller / *så o-* (*som*) konj: hver gang

I (3) er det en ny ordklassemarkering, adj, inne i ordartikkelen. I (4) er det ikke bare en ny ordklasse, adv; denne ordklassen er også knyttet til en bestemt form av oppslagsordet, presens partisipp. I (5) er det ett av eksemplene i ordartikkelen som er forsynt med en ny ordklasse, konj. I dette tilfellet ser det imidlertid ikke ut som om ordklassen gjelder selve oppslagsordet, men hele uttrykket som er oppgitt som eksempel: *så ofte (som)*. Hvis det er hele uttrykket som får en ny ordklasse, bør det kanskje få et eget oppslag (se avsnitt 3.5).

Siden ordklassemarkeringene inne i ordartiklene er gjort på så mange ulike måter, kan det være vanskelig å finne alle ved konvertering til et elektronisk leksikon. Også brukere av ordboken i papirformat kan tenkes å ha problemer med å finne frem til relevant ordklassebetegnelse når den kan være så dypt innføyet i ordartikkelen.

### 3.1.2. Ordklassen er ikke optimal for syntaktisk analyse

Noen ord har en ordklasse som passer dårlig for automatisk syntaktisk analyse. Et eksempel finnes i oppslaget for verbet *unnskylde*.

- (6) **unnskylde** v2 (fra lty, sm o s ty. *entschuldigen*, eg ‘frita for skyld’) **1** beklage, be om unnskyldning for *u- sin glemsomhet* / ofte refl: *hun unnskylde seg fordi hun kom så sent* **2** som høflig innledning til et spørsmål e l: *unnskyld, kunne De si meg veien til byen?* [...]

Her er analysen at *unnskyld* som «høflig innledning til et spørsmål eller lignende» er imperativ av verbet *unnskyldde*. Dette er i og for seg en mulighet, men for automatisk syntaktisk analyse vil det være fordelaktig å også kunne analysere ordet som en interjeksjon. Grunnen er at *unnskyld* ofte har den syntaktiske distribusjonen til en interjeksjon, noe som ikke er tilfelle for imperativ av hvilket som helst verb. I denne bruksmåten er det også svært begrenset hvilke argumenter ordet kan ha. Objekt er en mulighet, slik som i *unnskyld meg* eller *unnskyld oss*, men ellers er det få mulige utfyllinger.

Noen ganger trenger man en mer finkornet inndeling enn den de tradisjonelle ordklassene gir. For eksempel er det i NKL bare en stor klasse adverb, som inneholder mange svært uensartede ord. Siden ulike typer adverb har ulik syntaktisk distribusjon, er det nødvendig å klassifisere dem i underkategorier.

- (7) *Foredraget var **temmelig** langt.*
- (8) *Han liker **dessverre** ikke ostekake.*
- (9) *Vi har **vel** hørt den før.*

Eksemplene (7)–(9) illustrerer henholdsvis gradsadverb, holdningsadverb og partikkeladverb. Partikkeladverb kommer før holdningsadverb, slik at *Vi har vel dessverre hørt den før* er mulig, mens *\*Vi har dessverre vel hørt den før* ikke er det. Når annotatorene finner at en setning med et adverb får en utilfredsstillende analyse, kan dette ofte løses gjennom at ordklassen endres fra default-kategorien adverb til en av de mer finkornete adverbkategoriene.



### 3.2. Valens

For å få en god syntaktisk analyse av en setning er det nødvendig å ha passende valensrammer for verbene i setningen. NKL angir valensrammer for alle verb, og disse utgjør grunnstammen i valensrammene som brukes i NorGrams leksikon. Når et korpus er parset, finner annotatorene ofte at rammer mangler, og disse legges da til i NorGrams leksikon. Eksempelene i (10) og (11) er setninger som ikke fikk korrekt analyse fordi den nødvendige valensrammen manglet.

(10) *Faren **mumlet** et farvel.*

(11) *Han **trengte** seg fram.*

Verbet i (10) manglet en transitiv ramme, og verbet i (11) manglet en ramme med refleksivt objekt og partikkel.

Også substantiv og adjektiv kan ta komplementer, som vist i (12) og (13).

(12) *Orker ikke **tanke** på mer drittsslenging.*

(13) *Han er **stolt** av datteren.*

En ramme der substantivet *tanke* tar en preposisjonsfrase er nødvendig for å analysere (12), og en lignende ramme for adjektivet *stolt* er nødvendig for (13).

Noen verb kan forekomme i såkalte inquit-konstruksjoner, der et sitat følges av finitt verb og subjekt (og eventuelt andre setningsledd). Forbausende mange verb forekommer i denne konstruksjonen; noen eksempler vises i (14)–(18).

(14) – *Jeg finner da veien hjem, **skrek** Fredrik etter ham.*

- (15) – *Ja, jeg leter etter en jobb, **lyver** jeg.*  
(16) «*Jeg vil se Farid,*» **jamret** Zoubida.  
(17) *Det dufter deilig, **skrøt** han mens hun bakte julekaker.*  
(18) «*Det er telegram fra doktern, fra Alex,*» **rettet** hun.

Inquit-verb er et subsett av verb som tar *at*-setninger som komplement; verb som *glemme, lære, oppnå* osv. er vanskelig å tenke seg i denne konstruksjonen. Siden NorGram-leksikonet ikke i utgangspunktet har informasjon om hvilke verb som kan brukes i inquit-konstruksjoner, må disse legges til når de dukker opp under entydiggjøringsarbeidet.

### 3.3. Tellelighet

Informasjon om tellelighet er relevant for syntaktisk analyse, siden tellelige og ikke-tellelige substantiv til dels har ulik syntaktisk distribusjon. For eksempel kan ikke-tellelige substantiv brukes som komplette nomenfraser uten determinativ i ubestemt entall, mens dette vanligvis ikke er mulig for tellelige.

- (19) *Øl/Kjøtt/Ertepuré er i kjøleskapet.*  
(20) *\*Drink/Kjøttkake/Ert er i kjøleskapet.*

Siden det ikke finnes informasjon om tellelighet i NKL, er alle substantiv blitt tildelt trekket tellelig som default i NorGram-leksikonet. Når ikke-tellelige substantiv oppdages under entydiggjøring, legges denne informasjonen til. Eksempler funnet i trebanken er gitt i (21) og (22).

- (21) *Piken var en skjensel, hun kastet **vanry** over dem.*  
 (22) *Han løftet forsiktig opp et smykke som glimtet i **gull**.*

Selv om informasjon om tellelighet ikke oppgis i *BOB*, er det noen substantiv som har entallskoder, som for eksempel *brie*, *øst*, *pomp*, *rytmikk* osv. Det er mulig at denne morfologiske informasjonen kan utnyttes til å finne en del ikke-tellelige ord i *BOB*, men informasjonen er ikke angitt systematisk.

### 3.4. Sammensetningsstruktur

I norsk er sammensetning en produktiv ordlagingsmåte. Nye sammensetninger lages i stor grad spontant i tale og i skrift, og disse kan naturligvis ikke føres opp i vanlige ordbøker. Før en tekst kan parses, må den analyseres morfologisk. Produktive sammensetninger kan da håndteres ved hjelp av automatisk sammensetningsanalyse.

Leksikaliserte sammensetninger har sin naturlige plass i ordbøker. I *BOB* skiller disse seg ikke ut fra usammensatte ord; det er ingen informasjon om den indre morfologiske strukturen til leksikaliserte sammensetninger. Men det kan være fordelaktig å ha slik informasjon, bl.a. for å kunne analysere elliptiske koordinasjoner av den typen som vises i (23)–(25).

- (23) *mun- og klovsyke*  
 (24) *Norsk Nærings- og Nytelsesmiddelarbeiderforbund*  
 (25) *vinter- og sommerdekk*

I (25) gjelder koordinasjonen leddene *vinter-* og *sommer-*, ikke

leddet *vinter-* og ordet *sommerdekk*. Hvis man skal kunne få en tilfredsstillende syntaktisk og semantisk analyse av disse, må man vite at *sommerdekk* består av leddene *sommer* og *dekk*.

### 3.5. Flerordsuttrykk

Flerordsuttrykk utgjør en stor utfordring for automatisk syntaktisk analyse. Hvis de behandles som sekvenser av enkle ord, får man ofte uriktige analyser. Det store problemet er at det er få kilder til kunnskap om hvilke flerordsuttrykk som finnes. Jackendoff anslo at antallet flerordsuttrykk i engelsk er av omtrent samme størrelsesorden som antallet enkle ord (1997: 156). Selv om det er vanskelig å anslå dette for norsk, er det liten tvil om at det også i norsk er et meget høyt antall flerordsuttrykk.

Det finnes flere forskjellige typer flerordsuttrykk; Sag et al. gir en god oversikt (2002). Én av de viktigste typene er faste uttrykk som er leksikalisert, og som ikke kan bøyes eller endres på andre måter. Disse kalles gjerne ord med mellomrom (*words with spaces* på engelsk), og det er lett å behandle dem som egne oppslag. *BOB* har en del oppslag av denne typen, for eksempel: *ad absurdum, for så vidt, i hende, lille julaften, tipp topp, world cup* osv. Det er nok mange flere slike uttrykk som ikke har egne oppslag, men som ordboken har informasjon om. Disse omtales inne i ordartiklene til ett eller flere av de grafiske ordene som inngår i dem. Et eksempel er uttrykket *dann og vann*, se (26)–(28).

(26) **I vann** el. **vatn** n1 (norr *vatn*) **1** klar, gjennomiktig væske  
[...]

- (27) **II vann** adv, se *dann*  
 (28) **dann** adv (ty. *dann und wann*) bare i uttr *d-* og *vann* nå og da

Ordet *vann* har altså to oppslag, ett som substantiv og ett som adverb. Oppslaget som adverb henviser til oppslaget for *dann*, som også er kategorisert som adverb. Men hverken *vann* eller *dann* kan brukes som adverb alene i norsk. Det faste uttrykket *dann og vann* fungerer som adverb, og det fortjener et eget oppslag. For automatisk syntaktisk analyse er det uheldig at *vann* er kategorisert som et adverb; dette betyr at når ordet forekommer i en setning, vil parseren ofte foreslå en helt uaktuell mulighet, nemlig at ordet *vann* fungerer som adverb. Heller ikke for menneskelige brukere av en papirordbok kan det ha noe for seg at enkeltordene *vann* og *dann* har egne oppslag som adverb.

Under oppslaget for *god* finnes eksempler som skal illustrere substantivisk bruk.

- (29) **god** a1 [...] **1** av høy kvalitet, bra, fin, gagnlig, tjenlig, skikket, dugende *g-t vær / g-e veier, forhold / en g- film, bok, debatt / ha g- helse, hørsel, samvittighet / et g-t spørsmål / ha g-e intensjoner / en g- kniv / i g-e, gamle dager [...]* / subst, med gl gen. etter til: *tvilen kommer tiltalte til g-e* blir godskrevet tiltalte / *(ha) til g-e* til overs, utestående; *(ha) til senere / g-t! fint!* / subst: *hva sier han til g-t?* hva har han å si? [...]

De formene som skal være substantiv, er altså *gode* i uttrykket *til gode* og *godt* i uttrykket *(hva sier han) til godt*. For at det skal være berettiget å kategorisere disse formene som substantiv, bør de kunne fungere i syntaktiske posisjoner som er typiske for substantiv. Disse eksemplene viser at de riktignok

kan forekomme etter en preposisjon, og dette er en typisk syntaktisk posisjon for nomenfraser. Men disse formene forekommer knapt etter andre preposisjoner enn *til*, og de kan ikke bygges ut med andre ord som er typiske for nomenfraser. Det virker mer nærliggende å betrakte dem som faste uttrykk som bør ha egne oppslag.

Noen flerordsuttrykk blir oppdaget under entydiggjøring fordi annotatorene merker at det er noe som ikke stemmer med analysen. Et eksempel på en slik setning gjengis i (30).

(30) *Men før Artur hoppet **over bord**, hadde Martin hatt et hav av tid.*

Problemet her er at det er noe rart med preposisjonsfrasen *over bord*, siden objektet for preposisjonen er et nakent substantiv. Man kan si *han hoppet over stolen/over relingen/over bordet*, men ikke *han hoppet over stol/over reling*. Når det går bra med *over bord*, er det fordi det er et stivnet uttrykk med en leksikalisert betydning. Uttrykket er med som eksempel i *BOB*, men det har ikke et eget oppslag.

Av og til oppdages flerordsuttrykk under entydiggjøring fordi et uttrykk som etter gjeldende normering skal skrives som flere grafiske ord, er blitt sammenskrevet. Et eksempel er *vær-sågod*. Flerordsuttrykket *vær så god* er markert som uttrykk i *BOB* under oppslagene for alle ord som inngår i det, både *god*, *så* og *være*.

Det kan være interessant å sammenligne med behandlingen dette uttrykket får i noen danske og svenske ordbøker. Oppslaget i (31) er fra *Politikens Nudansk Ordbog* (Becker-Christensen 1999).

(31) **værsgo** [...] udråbsord **1.** udtryk som bruges når man giver

el. rækker andre noget □ *værsgo, maden er serveret!* ·  
*værsgo, her er pengene* · *vil du række mig saltet?* - *ja selvfølgelig, værsgo!*

Her er uttrykket trukket sammen til ett ord. I alle eksemplene i denne ordartikkelen passer det fint med analysen som interjeksjon, eller «udråbsord», som det heter her. Det typiske for interjeksjoner er at de står for seg selv, egentlig utenfor integrerte syntaktiske konstruksjoner. Her er alle adskilt fra resten av setningen med komma, enten i begynnelsen eller slutten av setningen.

Oppslaget i (32) er fra *Dansk-dansk Ordbog* (Dissing & Helles 1992).

(32) **værsgo** (udråbsord) (el. *vær så god*). Værsgo maden er serveret! (forstærkende:) Vil du værsgo (el. vær så god) gøre rent efter dig.

I det første eksempelet i dette oppslaget har ordet *værsgo* en posisjon som er typisk for interjeksjoner, men i det andre, skrevet enten som ett ord eller som flere, er plasseringen heller typisk for adverb.

Oppslaget i (33) er fra *Norstedts stora svensk-engelska ordbok* (Sjödén 2000).

(33) **varsågod** se under *god I I*

Flerordsuttrykket *var så god* behandles utførlig i artikkelen under *god*, men med en henvisning fra et enkeltordsoppslag. Et slikt oppslag kan være spesielt viktig i en tospråklig ordbok, der brukere kan tenkes å ha vanskelig for å finne frem til ordet hvis det ikke har sitt eget oppslag.

Det ser ut til at dette ordet/uttrykket kan forekomme både

som enkeltord og som flerordsuttrykk i alle de skandinaviske språkene. I *BOB* er det kanskje slik at et enkeltordsoppslag mangler fordi normering har tilsagt at uttrykket skal skrives med flere ord. På den annen side er det god grunn til å behandle uttrykket komposisjonelt i visse tilfeller; ett av eksemplene under oppslaget for *god* er: *vær så g- å gå til bords*. Men tatt i betraktning at dette er et så høyfrekvent ord i dagligspråket, ofte uttalt med stor grad av fonologisk reduksjon, er kriteriene absolutt til stede for at det skal ha et oppslag som enkeltord. Dette virker som et tilfelle der det er høyst berettiget å ha begge typer oppslag.

#### 4. Hvilke ord er nødvendige for syntaktisk analyse?

Når man parser et korpus, vil det alltid være ord som er ukjente for den morfologiske komponenten og/eller for leksikon. Det kan være tale om nyord, men det kan også dreie seg om feilstavede ord, feil bøyingsmåte osv. Ett eneste ukjent ord i en setning vil kunne resultere i at parseren ikke finner rett analyse. For å få til en vellykket syntaktisk analyse må flest mulig av de ordene som forekommer i autentiske tekster, gjenkjennes, enten de er korrekte eller ikke.

INESS har et eget grensesnitt for tekstpreprosessering. Dette grensesnittet har to viktige funksjoner. Det første er å oppdage og korrigere feil som har oppstått på grunn av optisk tegngjenkjenning (OCR) i skannede dokumenter som leveres fra Nasjonalbiblioteket. Det andre er å registrere og klassifisere ukjente ord i de samme dokumentene.



Preprosesseringen av tekstene finner alle grafiske ord som ikke kjennes igjen av den morfologiske komponenten. Disse ordene blir så presentert for annotatorene. Hvis ordet er en OCR-feil, korrigeres den av annotatoren. For eksempel skjer det ganske ofte at ordet *lo* tolkes som tallet *10*. Hvis ordet er korrekt skannet, men likevel ikke gjenkjennes av den morfologiske analysen, behandles det på ulike måter avhengig av hva slags problem det gjelder.

Noen ganger er ordet en neologisme som ikke finnes i leksikon. Figur 6 viser grensesnittet som brukes for å legge til et nytt ord. I dette eksempelet er det ordet *usymmetri* som skal legges inn. Annotatoren legger til ordet gjennom å oppgi *usymmetri* som baseform. Systemet foreslår da et antall ord som ligner på denne strengen, nemlig de ordene det allerede har som ender på samme måte. Fra en rullegardinmeny velger annotatoren et ord som bøyes på samme måte som det nye ordet. Så kommer det opp et nytt vindu (til høyre i bildet) med et komplett bøyingsparadigme, slik at annotatoren kan sjekke at paradigmet er rett.

Noen ganger er ordet kjent, men bøyingsmåten er uvanlig. Et slikt eksempel er *nevnet* som preteritum av verbet *nevne*. Figur 7 viser grensesnittet under tillegg av en bøyingsform. Baseformen spesifiseres, og systemet viser bøyingsparadigmet. Annotatoren kan da velge hvilken av bøyingsformene i det vanlige paradigmet som den nye formen er en variant av. Denne formen legges så til bøyingsparadigmet for ordet.

Det er ofte feilstavede ord i tekstene som analyseres. For at også setninger med slike ord skal kunne parses, kan disse legges inn som varianter. I figur 8 er det et eksempel på det feilstavede ordet *alldeles* som legges inn.

En ordklasse med mange kreative nydannelser er interjeksjonene, både sammensetninger, som i (34), og infikser, som i (35).

**Store as:**

**Word:** usymmetri

**Correction:**

**Base form:** usymmetri  spelling error |  lect |  old

**Add to base form:**  (if different from base form) | Id:

**Inflects like:** asymmetri  or

**Verb frame:**  INTRANS |  TRANS |  COMP |  XCOMP |  special

**Name:**  Masc/C-m |  Fem/C-f |  Last/C-l |  Pers/C-n |  Title/C-t  
 Org/C-o |  Place/C-p |  Tax/C-r |  Loan/C-h |  Misc/C-e  
 has inflection

Stored as:

**Context(s):**

98 / 606 | øyet og ned langs kinnet , og skapte en besynderlig usymmetri , som om ansiktet var ved å blikke over til

*New paradigm(s):*

<b>usymmetri</b>	<i>subst mask appell ent ub</i>
<input checked="" type="checkbox"/> <b>usymmetrien</b>	<i>subst mask appell ent be</i>
<b>usymmetriene</b>	<i>subst mask appell fl be</i>
<b>usymmetrier</b>	<i>subst mask appell fl ub</i>

Figur 6: Skjermbilde av grensesnittet for et ukjent substantiv.

SPRÅKTEKNOLOGIENS BEHOV FOR LEKSIKALSK INFORMASJON

**Store as:**

**Word:**

**Correction:**

**Base form:**   spelling error |  lect |  old

**Add to base form:**  (if different from base form) | Id:

**Inflects like:**  or

**Verb frame:**  INTRANS |  TRANS |  COMP |  XCOMP |  special

**Name:**  Masc/C-m |  Fem/C-f |  Last/C-l |  Pers/C-n |  Title/C-t  
 Org/C-o |  Place/C-p |  Tax/C-r |  Loan/C-h |  Misc/C-e  
 has inflection

**Is a variant of:**

47704 nevnt – verb perf-part  
47704 nevnte – adj <perf-part> be ent  
47704 nevnte – adj <perf-part> fl  
47704 nevnte – verb pret

**Add to paradigm:** Inflected form:   
ID:   
Features:

Stored as:

**Context(s):**

70 / 510 hånd fortsatte ferden . \_ Han heter Richard . Han nevnet det ikke engang . A nel . Han fortalte meg

Figur 7: Skjerm bilde av grensesnittet for valg av ny bøyningsform.

**Store as:**

**Word:**

**Correction:**

**Base form:**   spelling error |  lect |  old

**Add to base form:**  (if different from base form) | Id:

**Inflects like:**  or

**Verb frame:**  INTRANS |  TRANS |  COMP |  XCOMP |  special

**Name:**  Masc/C-m |  Fem/C-f |  Last/C-l |  Pers/C-n |  Title/C-t  
 Org/C-o |  Place/C-p |  Tax/C-r |  Loan/C-h |  Misc/C-e  
 has inflection

Store as:

Already stored as:

Select the appropriate paradigm for the base form:  
1545  alldeles adv

**Context(s):**

52 / 374 om jeg ikke skjønnte det med englene . De var alldeles like .  
152 / 1069 \_ Betongkonstruksjoner , sa jeg til Frazer , \_ noen alldeles forjævlige betongkolosser , \_ spesialkonstruert for å kapsle Inn avfall

Figur 8: Skjerm bilde av grensesnittet under innlegging av alternativ stammeform.

(34) *Kan få fliser i tunga! Og kvae! **Dobbelt-æsj!** Verste som finnes!*

(35) [...] og hanen Hanibal gol: «**Kykkeli-gratuly-ky!**»!

Ofte er ordet i og for seg kjent, men skrivemåten er annerledes enn den normerte. I tabell 1 er det eksempler på interjeksjoner med stavemåten(e) som forekommer i korpuset, sammenstilt med det som ser ut til å være normert skrivemåte. Det ganske vanlige ordet *jippi* står hverken i *BOB* eller *NOB*, men Kunnskapsforlagets *Norsk-engelsk stor ordbok* har det som oppslag (Henriksen & Haslerud 2002). Særlig skrivemåten *fy til rakkeren* skiller seg ut; dette virker mer som en etymologi enn en avspeiling av hvordan dette muntlige uttrykket uttales.

Formene funnet i korpuset	Normerte former
<i>jovisst</i>	<i>jo visst</i>
<i>jaja</i>	<i>ja ja</i>
<i>jøssda</i>	<i>jøss da</i>
<i>javelja</i>	<i>ja vel ja</i>
<i>uffda</i>	<i>uff da</i>
<i>heisann/heisan</i>	<i>hei sann</i>
<i>ojsann/oisan</i>	<i>oi sann</i>
<i>fyttirakkern/fytterakker 'n/fytte rakker 'n</i>	<i>fy til rakkeren</i>
<i>jippiii/jippiiiii/jippiiii</i>	<i>jippi</i>

Tabell 1: Skrivevarianter av noen interjeksjoner.

Norske forfattere bruker ofte ord fra andre språk i tekstene sine. Engelske ord er spesielt populære, som i talespråket, men

også ord fra andre språk brukes flittig, for eksempel i skjønnlitterære dialoger. Noen eksempler som er blitt registrert under tekstpreprosessering er gjengitt i (36)–(40).

- (36) *Er han en slags, he-he, **boy-friend** eller noe sånt? spurte han.*
- (37) *Jeg dro rundt med mitt kamera og min «**business class**»-billett på maven.*
- (38) *«**Au contraire**, Nick,» lød Edmonds lyse, slepende stemme fra døren.*
- (39) – *Jeg har med frisisk **kruidkoek** til deg, sa Natasha og la en krydderkake på bordet.*
- (40) *Men dette greier du jo fint, det er jo **peanuts** for deg, Halvdan, [...]*

For at slike ord skal få plass i en vanlig ordbok, vil man kreve at de er godt etablerte som lånord i språket. Men for automatisk analyse er det avgjørende at alle ord gjenkjennes. Derfor legges også slike ord inn i morfologi og leksikon. Det siste eksempelet, *peanuts*, er så vanlig i norsk at det kan synes merkelig at det ikke har fått innpass i ordbøkene.

## 5. Konklusjon

For at en tekst skal kunne analyseres automatisk, må alle ord i teksten, enten de er en del av standardvokabularet eller de er lånord, feilstavinger, unormerte bøyingsformer e.l., være kjent

for systemet. I en vanlig ordbok pleier man å kreve et visst antall forekomster som et kriterium for et nytt leksikalsk oppslag, men for formålet parsing av autentisk tekst kan én forekomst være nok til å rettferdiggjøre et nytt oppslag. I denne artikkelen er det brukt eksempler som er relevante for bygging av en trebank, men dette er langt fra den eneste applikasjonen som har bruk for ord som ikke får plass i konvensjonelle ordbøker. For eksempel kan en liste over vanlige feilstavinger være nyttig for stavekontrollprogrammer.

I forbindelse med utviklingen av den norske trebanken i INESS utfører prosjektet et omfattende arbeid med leksikalske ressurser. Resultatene av dette arbeidet kan være av interesse for leksikografer. Fremtidsvisjonen bør være at vi kan utvikle et tettere samarbeid mellom de leksikografiske og språkteknologiske miljøene, til gjensidig glede og nytte.

## Litteratur

- Becker-Christensen, Christian (1999): *Politikens Nudansk Ordbog*. Aalborg: Politikens Forlag.
- Bresnan, Joan (2001): *Lexical-Functional Syntax*. Oxford: Blackwell.
- Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer (2002): The Parallel Grammar project. I: John Carroll, Nelleke Oostdijk & Richard Sutcliffe (red.): *Proceedings of COLING-2002 Workshop on Grammar Engineering and Evaluation, Taipei, Taiwan*.

- Stroudsburg, Pennsylvania: Association for Computational Linguistics.
- Dalrymple, Mary (2001): *Lexical Functional Grammar*, volume 34 of Syntax and Semantics. San Diego, California: Academic Press.
- Dissing, Børge & Sigrid Helles (1992): *Dansk-dansk Ordbog*. København: Gyldendal.
- Dyvik, Helge (2000): Nødvendige noder i norsk. Grunntrekk i en leksikalsk-funksjonell beskrivelse av norsk syntaks. I: Øivin Andersen, Kjersti Fløttum & Torodd Kinn (red.): *Menneske, språk og felleskap*. Oslo: Novus forlag.
- Haug, Dag T. T. & Marius L. Jøhndal (2008): Creating a parallel treebank of the old Indo-European Bible translations. I: *Proceedings of the Language Technology for Cultural Heritage Data Workshop (LaTeCH 2008), Marrakech, Morocco, 1st June 2008*.
- Henriksen, Petter & Vibecke C.D. Haslerud (2002): *Norsk-engelsk stor ordbok*. Oslo: Kunnskapsforlaget.
- Hovdenak, Marit, Laurits Killingbergtrø, Arne Lauvhjell, Sigurd Nordlie, Magne Rommetveit & Dagfinn Worren (2006): *Nynorskordboka: Definisjons- og rettskrivingsordbok*. Oslo: Det Norske Samlaget.
- Jackendoff, Ray (1997): *The architecture of the language faculty*. Cambridge, Mass.: MIT Press.
- Losnegaard, Gyri Smørdal, Gunn Inger Lyse, Anje Müller Gjesdal, Koenraad De Smedt, Paul Meurer, & Victoria Rosén (2013): Linking Northern European infrastructures for improving the accessibility and documentation of complex resources. I: Koenraad De Smedt, Lars Borin, Krister Lindén, Bente Maegaard, Eiríkur Rögnvaldsson, & Kadri Vider, (red.): *Proceedings of the workshop on Nordic language research infrastructure at NODALIDA 2013*.

- Linköping: Linköping University Electronic Press.  
 <<http://www.ep.liu.se/ecp/089/005/ecp1389005.pdf>> (mars 2014).
- Losnegaard, Gyri Smørdal, Gunn Inger Lyse, Martha Thunes, Victoria Rosén, Koenraad De Smedt, Helge Dyvik & Paul Meurer (2012): What we have learned from Sofie: Extending lexical and grammatical coverage in an LFG parsebank. I: Jan Hajič, Koenraad De Smedt, Marko Tadič & Antonio Branco, (red.): *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*. European Language Resources Association, 69–76.  
 <<http://link.uib.no/?rosen12lrec>> (mars 2014).
- Maxwell, John & Ronald M. Kaplan (1993): The interface between phrasal and functional constraints. I: *Computational Linguistics*, 19(4):571–589.
- Nordgård, Torbjørn (2000): NorKompLeks – A Norwegian computational lexicon. I: *COMLEX 2000*, Patras, Greece.
- Pollard, Carl & Ivan A. Sag (1994): *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik (2012): An open infrastructure for advanced treebanking. I: Jan Hajič, Koenraad De Smedt, Marko Tadič, & António Branco (red.): *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*. European Language Resources Association, 22–29.  
 <<http://link.uib.no/?rosen12lrec>> (mars 2014).
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger (2002): Multiword expressions: A pain in the neck for NLP. I: Alexander Gelbukh (red.): *Computational linguistics and intelligent text processing: third international conference, CICLing 2002, Mexico City, Mexico, February 17-23, 2002: proceedings*. Berlin: Springer, 189–206.



- Sjödin, Maria (2000): *Norstedts stora svensk-engelska ordbok*.  
Stockholm: Norstedt.
- Tesnière, Lucien (1959): *Éléments de syntaxe structurale*. Paris:  
Klincksieck.
- Wangensteen, Boye (2005): *Bokmålsordboka: Definisjons- og  
rettskrivningsordbok*. Oslo: Kunnskapsforlaget.

Victoria Rosén  
førsteamanuensis, dr.art.  
Universitetet i Bergen  
Sydnesplassen 7  
N-5007 Bergen  
victoria@uib.no

# **Leksikografisk tradition og fornyelse: tre revolutioner på 100 år?**

*Lars Trap-Jensen*

Viewed over the last 20 years, the development in lexicography has been overwhelming: many products have changed completely and so has the way of producing lexicographic products. The Internet, computer technology and publication forms are all different today from what they were just a few years ago, and like everyone else lexicographers have to deal with the situation. In many ways it is exciting times and many have probably asked themselves: is it really a paradigm shift we are going through, or is the whole thing just hot air? In this article I focus on three potential “revolutions” and assess in what way they have changed the conditions for practical lexicographic work and the way we think about dictionaries.

## **1. Indledning**

Hvis man betragter udviklingen på ordbogsmarkedet de seneste tyve år, bliver man hurtigt forpustet. De fleste produkter har ændret sig grundlæggende, og måden vi laver ordbøger på, har ændret sig. Internettet, computerteknologien, publiceringsformerne, stort set alt har ændret sig på få år, og som alle andre

må leksikografer og forlag forsøge at hænge på udviklingen så godt de kan. Det er på mange måder en spændende tid, men mange har sikkert også indimellem spurgt sig selv: Er det et reelt paradigmeskift vi oplever, eller er det hele bare varm luft? I denne artikel fokuserer jeg på tre perioder inden for de sidste 100 år som alle kan kaldes potentielle “revolutioner” inden for leksikografien. Jeg ser på hvilke spor de har sat sig, og på hvordan de har ændret det praktiske arbejde og måden vi tænker om ordbøger på.

## **2. Den første revolution: det deskriptive paradigme**

Hvis man går ca. 100 år tilbage og betragter ordbogsproduktionen omkring år 1900, er det ikke svært at få øje på mange forskelle. Det er omkring dette tidspunkt de nationale modersmålsordbøger er skudt frem og har fundet sin form med kartotekskasser fulde af excerpter og en hær af flittige leksikografer der producerer det ene bind efter det andet, sådan som vi kender det fra OED i England, SAOB i Sverige og ODS i Danmark. De er i dag kendte og klassiske værker hvis akronymer kendes langt ud over leksikografernes snævre kreds. Men selvom projekterne er gamle og hjælpemidlerne kan forekomme simple, adskiller metoden sig ikke principielt fra den vi kender og bruger i dag: beskrivelse af sproget og sprogbrugen på empirisk grundlag – korpusbaseret om man vil. Dengang bestod korpus af excerperede sprogprøver der blev opbevaret i kartotekskasser og analyseret for hvert ord og betydning. I Danmark

udgjorde denne praksis et virkeligt nybrud, et paradigmeskift i klassisk, kuhnsk forstand: ODS gjorde op med 1800-tallets fremherskende tradition og insisterede på at være deskriptiv. Man kan kalde det for *den deskriptive revolution*.

Traditionen før ODS var præget af det man har kaldt *akademiprincippet* med henvisning til den praksis som var gældende for Det Franske Akademis ordbog. I overensstemmelse med dette princip var ordbøgerne før ODS *præskriptive*, de var båret af et pædagogisk hensyn og et ønske om at opdrage befolkningen til det gode og smukke sprog med gode eksempler og derved opsætte en rettesnor for det. Jacob Langebek, en af pionererne bag Videnskabernes Selskabs Ordbog, mente at kun de «gode, rene, overalt brugelige og upaatvivlelige Danske» ord burde medtages, mens der omvendt ikke var plads til:

... Alle grove, plumpe og liderlige Oord og Talemaader, som stride imod Ærbarheden ... thi de behøves ikke at vides for dem, der ikke skiøtte derom, og de som have lyst til at vide dem, lære dem nok foruden des. (Langebek ca. 1738)

Ordbøgerne skal altså være selektive og kun medtage de gode ord, hvorimod de “grove, plumpe og liderlige” ikke har nogen plads i ordbogen.

Den samme tanke finder vi hos Molbech:

Selv den hyppigste Brug af et nydannet Ord, især i Tale-sproget, giver det ingen Auctoritet, og beviser intet for dets Brugbarhed i den rene tale og gode Stil, eller for dets Optagelighed i en Ordbog, saalænge det støder et for Sprog-brugen dannet Øre og den finere Sprogsands. (Molbech 1833:XIII, genoptrykt i Molbech 1859:VIII)

Ordbøgerne skulle kun indeholde «gode» ord, «sprogets skønneste blomster». «Det skal så at sige være en ære for et ord at blive optaget i ordbogen, ligesom det er en ære for et kunstværk at få plads i statens kunstsamlinger»— sådan udlægger ODS' grundlægger, Verner Dahlerup, Molbechs princip (Dahlerup 1907:68). Selv skriver Molbech at hans ordbog skal være «en Tolk for det rene, det dannede Skriftsprogs rigtige Brug i vor nærværende Alder».

Et sidste eksempel på akademiprincippet stammer fra den første danske slangordbog, *Ordbog over Gadesproget og saakaldt daglig Tale*. Forfatteren er V. Kristiansen, et pseudonym for professor Viggo Fausbøll, professor i indisk-østerlandsk filologi ved Københavns Universitet. Han skriver i forordet til ordbogen at formålet med ordbogen ikke er oplysende, men tværtimod at advare mod det vulgære sprog:

Dette vulgære sprog ... truer i den sidste tid med at trænge ind i familierne ... Ved her at have samlet noget af hvad der hører herhen, har jeg, næst efter at have haft et rent sprogligt formål, tillige i vort land villet henlede opmærksomheden på faren og søge at vække modstanden mod samme, og jeg antager, at når folk en gang have fået øjet åbnet for den utilbørlige overskridelse af grænsen, ville alle dannede mennesker være enige om at bandlyse gadesproget fra ethvert godt selskab og overlade det til gadedrenge og de Grundtvigianere, i hvis smag det falder. (Kristiansen 1866:V)

I førsteudgaven af ordbogen er de mest vulgære ord derfor sat med græske bogstaver så den brede befolkning ikke skulle blive fordærvet af ord de bestemt ikke behøvede at kende.

En anden ordbog man kunne nævne i samme forbindelse, er

afløseren for Molbechs ordbog, der udkom omtrent samtidig med at ODS gik i gang, nemlig Dahl og Hammers *Dansk Ordbog for Folket*, der bl.a. er kendt for at indeholde mange puristiske afløsningsord (fx *haandig* for *manuel*, *lemfaldsyge* for *spedalskhed* og *akselblad* for *skulderblad*), igen ud fra et ønske om at opdrage befolkningen og skåne den for uønsket ordstof, hvad enten det var fremmedord, vulgært sprog, dialekter eller andre former for lavsprog.

Det er denne præskriptive tradition ODS gør op med. Dahlerup lagde vægt på at ODS skulle være et videnskabeligt og praktisk hjælpemiddel til forståelse af sproget, og så nyter det ikke at udelade ord i sproget fordi man mener de er dårlige eller skadelige:

jeg kan ikke først og fremmest spørge: “bör det og det ord bruges?”, men: “bruges det, eller har det været brugt?”; hvis dette er tilfældet, optager jeg ordet, for så vidt hensynet til bogens omfang tillader det. (Dahlerup 1907:71)

Citatet stammer fra en artikel i *Danske Studier*, hvori Dahlerup redegør for sine overvejelser om den nye, store ordbog. Dahlerup var ikke den eneste der mente at den faktiske sprogbrug var nøglen til semantisk beskrivelse. Synspunktet var fremme i tiden med junggrammatikernes optagethed af samtidssproget som centralt for forståelsen af de undtagelsesløse lydlove og blev knæsat inden for filosofi og sprogvidenskab i løbet af det 20. århundrede. Princippet om empirisk analyse som grundlaget for en god beskrivelse følger vi stadig i dag, 100 år efter. Men det empiriske beskrivelsesgrundlag har ændret sig, ganske meget endda.

### 3. Den anden revolution: korpusrevolutionen

Korpusrevolutionen er det næste store spring der fundamentalt ændrede måden at lave ordbøger på. Teoretisk-metodisk er excerpering af tekster og belæg på kartotekskort ganske vist ikke grundlæggende forskelligt fra korpuskonkordanser og betydningsannotering – man kan sige at det mere er en forskel mellem analog og digital metode. Men den tekniske udvikling der fandt sted fra 1960'erne og fremefter, betød at en stor del af arbejdet kunne automatiseres med computerens hjælp. Det medførte en betragtelig udvidelse af beskrivelsesgrundlaget samtidig med at arbejdet kunne udføres for færre resurser.

The proper way to describe a word is to identify the grammatical constructions in which it participates and to characterize all of the obligatory and optional types of companions (complements, modifiers, adjuncts, etc.) which the word can have in such constructions, in so far as the occurrence of such accompanying elements is dependent in some way on the meaning of the word being described. (Fillmore 1995)

Potentialet i brugen af korpus er enormt hvis man tager det alvorligt – som Fillmore her udtrykker det – at det hører med til beskrivelsen af et ord at undersøge samtlige forbindelser ordet indgår med andre ord for at se om de på den ene eller anden måde betinger det pågældende ords betydning. Fillmore har selv gjort det i sit FrameNet-projekt, og også inden for leksikografien har vi set projekter der er optaget af tankegangen, mest direkte i den engelske DANTE-database der eksplicit bygger på tankerne fra FrameNet (jf. Atkins 2010), men også i en række

andre projekter er der en tendens til at fokus forskydes fra *lemmaet* som leksikografisk grundenhed til en mindre enhed bestående af en betydning og dens tilhørende lemmaform. Sådan er det i DANTE, og det er også sådan en udvikling som Det Danske Sprog- og Litteraturselskabs projekter har gennemgået, og som gør det muligt at knytte Den Danske Ordbogs (DDO) ord og betydninger til enhederne i afledte projekter som DanNet og Den Danske Begrebsordbog (jf. Lorentzen, Nimb & Troelsgård denne udgivelse).

Inden for leksikografien tog korpusrevolutionen for alvor fat omkring 1980 med COBUILD-projektet i Birmingham. I første omgang var den største fordel at eksempel materialet blev meget større i forhold til den hidtidige fremgangsmåde med excerpering og kartotekskasser. Hvis man alene ser på data-mængden målt i antal løbende ord, er korpusserne steget med noget der ligner en tidobling for hver 10-15 år: I 1960'erne og 70'erne var de første korpusser på omkring 1 mio. ord. Norden var her godt repræsenteret blandt de allerførste med Sture Alléns arbejde i 1960'erne og Press65. COBUILD's korpus var i 1985 på ca. 18 mio. ord og rummer i dag omkring 650 mio. ord under navnet Bank of English. DDO's korpus fra midten af 1990'erne var på 40 mio., mens British National Corpus nåede 100 mio. omtrent samtidig med DDO's i midten af 1990'erne. Til de største korpusser vi har i dag, hører det tyske COSMAS (eller DeReKo: Deutsches Referenzkorpus) fra IDS i Mannheim med over 5 mia. løbende ord og Googles korpus over tekster der er blevet scannet i forbindelse med projektet Google Books: over 500 mio. bøger fra år 1500 til i dag for en række af de større sprog. Om korpusset indeholder 155 mia. ord eller 175 eller 200, er ikke afgørende. Kvantitativt er det tæt på hvad de fleste bare regner for uendelig stort. Norden kan ikke helt være med størrelsesmæssigt, men Språkbanken i Göteborg har dog



mere end 1 mia. ord i deres korpussamlinger, og også det norske avis-korpus indeholder mere end 1 mia. ord. Danmark ligger i sammenligning hermed noget lavere. Hos Det Danske Sprog- og Litteraturselskab forsøger vi at indsamle lidt bredere og har vel i alt tekster med omkring ½ mia. ord, men ikke alle tekster er offentligt tilgængelige. Det vil de efter planen blive for en del teksters vedkommende i løbet af 2014.

Man kan spore en parallel udvikling i brugen af korpus til ordbogsarbejde svarende til væksten i de tilgængelige korpusser. I de tidlige år var den helt store gevinst at man hurtigt fik adgang til et langt større eksempelmateriale end hvad der havde været muligt ved hjælp af kartotekskort. For leksikografen bestod arbejdet i at kigge konkordanser igennem og ordne forekomsterne i homografer og betydninger, svarende til arbejdet med kartotekskort. Men enhver der har arbejdet med konkordanser, ved også at selv i mindre korpusser bliver opgaven en stor mundfuld når det drejer sig om at undersøge sprogets relativt almindelige ord fordi de mange konkordanslinjer hurtigt bliver uoverskuelige.

Det var derfor et fremskridt da man begyndte at få annoterede korpusser. Ved at begrænse søgningen til fx verbalforekomster af en homograf eller kun udvalgte bøjningsformer af et ord bliver konkordansen rensset for uønskede forekomster og arbejdet dermed lettere og hurtigere for leksikografen at udføre.

Omkring årtusindskiftet kom de syntaksopmærkede korpusser, hvad der yderligere gjorde det muligt at finde prægnante mønstre i teksterne, fx i form af såkaldte Word Sketches (Kilgarriff m.fl. 2004) eller andre former for leksikalske profiler. Og samtidig med at korpusserne voksede voldsomt i volumen, blev det i stigende grad nødvendigt at gøre noget ved den overvældende informationsmængde der fulgte med. Tendensen går derfor i retning af mere og mere præprocessering af materialet

ved hjælp af teknikker som forhåndsanalyserer teksterne efter forskellige parametre som gør det muligt for redaktørerne at finde netop de forekomster de har brug for på det aktuelle sted i redigeringsprocessen. Lad os se på nogle af de muligheder som udforskes.

At opdele korpusforekomster i homografer og betydninger står centralt i enhver leksikografs daglige arbejde, og selvom der endnu ikke – mig bekendt – er udviklet en operativ teknik til automatisk sortering af forekomsterne, skal det alligevel nævnes først da konkordansopstillingen fra første færd havde dette som formål. Dog har betydningsannotering indtil videre været noget som redaktørerne måtte foretage manuelt. I korpuslingvistikens nuværende fase kan leksikalske profiler hjælpe med til at afsløre nogle betydninger, idet der er en sammenhæng mellem semantisk beskrivelse og fx valensmønstre eller fagtilknytning. Begge dele kan ofte påvises automatisk ved hjælp af korpuslingvistiske metoder.

Brug af korpus som redskab i lemmaselektionen er en anden indlysende mulighed. Korpusfrekvens er et af de parametre der indgår når man skal afgøre hvilke ord der skal med i ens ordbog.

Statistiske metoder som Mutual Information, T-score m.fl. er metoder der er egnede til at påvise hvordan ord tiltrækker hinanden. Alt hvad falder inden for det fraseologiske område, kan man derfor få god hjælp til med korpuslingvistiske teknikker. Valensmønstre og leksikalske profiler bør nævnes i samme forbindelse da de findes med samme teknik som de fraseologiske kombinationer, blot med den forskel at det ikke er ordenes direkte tiltrækning af hinanden der måles, men udfyldningen af syntaktiske kategorier og ledfunktioner i det opmærkede korpus.

En forholdsvis ny og derfor mindre kendt anvendelse er at

bruge korpus til at overvåge sproglig udvikling, mest oplagt til at finde neologismer ved hjælp af et dynamisk monitorkorpus der analyseres diakront med henblik på at afsløre om der dukker ord eller samforekomster af ord op i de nyere tekster som er fraværende i et referencekorpus som der sammenlignes med (se fx Halskov 2010). Hvis det er tilfældet, kan det være tegn på en sproglig nydannelse. Metoden er især oplagt til at finde helt nye ord, mens det ikke er helt så enkelt at finde nye betydninger af eksisterende ord og flerordsforbindelser af eksisterende ord. Den samme metode kan bruges mere generelt: Ved at sammenligne et såkaldt *fokuskorpus* med et *referencekorpus* (jf. Cook m.fl. 2013) kan overhyppige forekomster i fokuskorpusset påvises og analyseres. Hvis der viser sig en ny kombination af eksisterende ord eller et ændret syntaktisk mønster, kan det være tegn på et nyt udtryk eller en ny betydning i sproget.

På samme måde kan sammenligning af et fagsprogligt fokuskorpus med et almensprogligt referencekorpus eksempelvis bruges til først at karakterisere fagsproget, dernæst identificere en bestemt type fagtekst blandt andre tekster og endelig påvise en udvikling af fagsproget. Hvis fx ord som *virus*, *orm* eller *sky* dukker op med en højere frekvens end normalt i tekster der handler om IT, er det et tegn på at der er tale om nye betydninger af de ord. Retningen kan også være den modsatte, fra det faglige til det almene, når man finder udtryk fra sportsverdenen i almensproglige tekster: *sænke paraderne*, *stå på mål for en sag* eller *skyde til hjørne* osv.

Mere generelt kan teknikken endvidere bruges som leksikografisk redskab i arbejdet med at karakterisere sprogbrugen ved hjælp af sprogbrugsmarkører. Hvis sammenligningen viser at et bestemt ord eller udtryk er overrepræsenteret i tekster fra et bestemt domæne, er det sandsynligt at der er tale om en faglig betydning. Hvis korpusteksterne er forsynet med metaoplys-

ninger om det, kan sammenligningen ligeledes vise om bestemte sprogbrugere eller teksttyper er overrepræsenterede.

Gode sprogbrugseksempler er et andet område hvor der har været bestræbelser på at præprocessere korpusmaterialet således at de bedst egnede citater kommer til at stå først i konkordansen. Som bekendt er det resursekrævende at finde frem til de mest velegnede citater, så her er et område hvor der kan spares tid og penge ved at præsentere materialet hensigtsmæssigt. Noget af det der kendetegner et godt citat, er at det består af en hel sætning, hverken for lang eller for kort (vel omkring 10-20 ord, måske 25), det bør ikke indeholde proprier (fordi navngivne personer eller andet navnestof kræver særlig viden); den bør ikke indeholde deiktiske udtryk eller pronominer med reference uden for sætningen; og det bør ikke indeholde svære eller sjældne ord, men derimod gerne en typisk kollokation (se også Lorentzen 1999). Den slags kriterier kan fastlægges på forhånd sådan at leksikografen som det første bliver præsenteret for de eksempler der opfylder kriterierne, og som regel vil der vise sig at være citater der egner sig til at komme i ordbogen. Denne måde at arbejde på kan fx bruges sammen med korpussøgeværktøjet SketchEngine, og et tilsvarende system er udviklet ved Berlin-Brandenburgische Akademie der Wissenschaften for tysk (Didakowski, Lemnitzer & 2012).

Dette er nogle af de muligheder der findes, men mulighederne for at udnytte korpus er dermed ikke udtømt. Et mål har længe været at kunne opdele konkordanserne i semantisk meningsfulde dele sådan at korpussøgeprogrammet præsenterer leksikografen for et automatisk forslag til betydningsopdeling. Mig bekendt er der dog ikke udviklet en lovende metode til semantisk tagging endnu, men der eksperimenteres med forskellige fremgangsmåder. Det Danske Sprog- og Litteraturselskab deltager eksempelvis i et treårigt forskningsprojekt i

samarbejde med Center for Sprogteknologi, Københavns Universitet, der har til formål dels at udvikle en maskinlæringsmetode, dels at udføre semantisk annotering af vores resurser, bl.a. ved hjælp af DanNet-resursen.

Sammenfattende kan korpusrevolutionen ikke betegnes som et paradigmeskift i kuhnsk forstand. Den videreudvikler den deskriptive tradition som blev grundlagt med de klassiske, store nationalordbøger omkring år 1900. Det var her idealet om ordbogen som et spejl af sproget blev fastlagt, men de store korpusser gjorde det muligt at komme gradvis tættere på idealet. Beskrivelsesfeltet blev udvidet fra at fokusere på det «gode» sprog hos de klassiske forfattere til at omfatte flere og flere andre tekstgenrer indtil vi med fx BNC og også DDO's korpus forsøgte at nærme os noget der lignede sproget i sin mangfoldighed, indfanget i et enkelt korpus. I 1990'erne mente vi at repræsentativitet var mindst lige så vigtigt som volumen, men den tanke er flere steder ved at blive forladt, formentlig mere af praktiske end teoretiske grunde. Balancerede korpusser er dyrere at udvikle, og det er kendetegnende for de store korpusser med milliarder af ord at de ikke er særlig godt balancerede. De består enten overvejende af avistekster (COSMAS, de store norske og svenske korpusser) eller er indhøstet fra nettet (fx Oxfords English Corpus, 2 mia.). Tilgængelighed er nøgleordet (se også Jakubiček m.fl. 2013).

Hvis man anskuer udviklingen udefra, er korpusrevolutionen formentlig ikke noget ret mange uden for den leksikografiske verden overhovedet har bemærket, for den har næppe ført til radikalt anderledes ordbøger. Den var til gengæld interessant for leksikograferne fordi den gav os et forbedret beskrivelsesgrundlag og dermed mulighed for at lave bedre ordbøger.

De voksende korpora har ført til ændrede arbejdsbetingelser og en anden rollefordeling for os der arbejder som redaktører. I

gamle dage stod redaktøren for næsten alt arbejdet med at redigere artiklen; det eneste præprocesserede materiale der fandtes, var kassen med excerpter, resten var redaktørens ansvar. Sådan er det ikke længere. I korpusleksikografiens tidlige dage var der stadig mange kreative opgaver for redaktøren: Korpus var et hjælpe-redskab, mens det var redaktørens rolle at gennemlæse konkordanserne og sortere og udvælge fra dem. Men heller ikke det er muligt når materialet bliver for stort. Redaktøren bliver i stigende grad præsenteret for en række halvfabrikata i form af forslag til oplysninger som computeren på forhånd har analyseret sig frem til: staveformer, bøjningsformer, valensmønstre, kollokationer, idiomer, morfologiske og syntaktiske begænsninger, måske citater og sprogbrugsmarkører for at nævne de vigtigste. Redaktørens rolle bliver at kontrollere, validere og vælge blandt det materiale som kommer ud af præprocesseringen. Der sker en ændring i redaktørens arbejdsbetingelser og -funktioner og dermed også i de kvalifikationer der kræves af en dygtig redaktør. Nogle færdigheder er ikke længere så vigtige, mens andre kommer til.

#### **4. Den digitale revolution**

Den tredje store begivenhed som jeg vil nævne, er den udvikling som vi befinder os midt i. Nogle kalder det for den elektroniske eller den digitale revolution: den udvikling der har ændret ordbogen fra papirprodukt til elektronisk produkt. Udviklingen er sket gradvis over de sidste 30 år. SAOB var blandt pionererne i denne udvikling da digitaliseringen begyndte i 1983 med

OSA-projektet. I 1990'erne fik vi cd-rom'er og PDA'er, men bortset fra at disse produkter tilbød bedre søgefaciliteter, var det grundlæggende de samme produkter. I løbet af 2000-tallet er næsten enhver ordbog med respekt for sig selv flyttet over fra papir til en elektronisk version, og især er onlineordbogen blevet populær. Navnlig er udviklingen dog gået virkelig stærkt de sidste 5-7 år, ikke mindst som følge af udbredelsen af smart-telefoner og tabletcomputere. Uanset hvordan årsagssammenhængen måtte være, vokser der en generation op som kommunikerer, læser og lærer langt det meste ved hjælp af en computer – især en lille, mobil én af slagsen. Det er derfor ikke overraskende at udviklingen også har påvirket ordbogsmarkedet.

Hvor korpusrevolutionen især havde stor effekt internt i det leksikografiske miljø, er den digitale revolution i høj grad noget der bemærkes af brugerne, ja, noget der påvirker alle. I papirordbøgernes tid var der en ret tæt forbindelse mellem indholdet i databasen og det færdige produkt. Fra redigeringen af DDO husker jeg således flere eksempler på at vi måtte vælge elementer i databasen efter hvordan de kunne gengives typografisk og ikke efter deres logiske indhold. Det trykte værk var hovedproduktet, databasen en mellemstation på vejen.

I dag er vi mere opmærksom på at databasen er det centrale element i arbejdet, og at basestrukturen skal være så tilpas veltilrettelagt og fleksibel at det kan lade sig gøre at publicere til forskellige medier og på forskellige platforme. Der er tale om en kæde af data, men i én retning: Redaktøren fylder indhold i databasen under redigeringen, mens databasens opbygning bestemmer hvad og hvordan der kan publiceres. Den eneste tovejskommunikation der foregår i processen, er redaktionens fælles møder hvor de træffer beslutninger om det redaktionelle arbejde. Om det kan gøres anderledes, tages op i afsnit 5.1 Interaktion og brugerinvolvering.

## 4.1. Overgang til digitale ordbøger og nye muligheder

De digitale medier giver nye muligheder for bedre ordbøger: lyd, billeder, videoer og anden form for multimedieanvendelse er noget de fleste redaktører gerne vil have med i ordbøgerne. Egentlig er det kun fantasien og pengepungen der sætter grænser, og det samme gælder muligheden for at linke internt mellem relevante dele af ordbogen eller eksternt til yderligere oplysninger på andre hjemmesider. Mange projekter befinder sig dog et andet sted fordi de er i en overgangsfase mellem papir og digital form. Det vil sige at ordbogsdata fra den trykte udgave først skal konverteres til en ny og mere hensigtsmæssig struktur før mediets muligheder kan udnyttes fuldt ud. Mange ordbøger bærer stadig, men i varierende grad præg af deres fortid som papirordbøger. Et godt eksempel er princippet om pladsøkonomi og den kondenserede stil som var karakteristisk for mange papirordbøger: Det gjaldt om at få mest mulig information ind på så lidt plads som muligt. Det princip er ophævet på skærmen og afløst af andre: Undgå at bruge forkortelser og symboler, men gør teksten mere læsevenlig; lad være med at sende brugeren på rundtur med henvisning efter henvisning, men giv klar besked på stedet.

Der består en risiko for at brugerne lider informationsdøden hvis de ikke kan finde den ønskede oplysning fordi den drukner mellem alle de mange oplysninger der nu er blevet plads til. Ti citater er ikke nødvendigvis bedre end to hvis de ekstra citater ikke tilfører nyt om ordet, og ti synonymmer er ikke bedre end to hvis ekstramaterialet er langt fra den pågældende betydning.

Der er også en sammenhæng mellem pladsøkonomi og lem-maselektion. Hvis en ordbog har en fast ydre grænse i form af et bestemt antal sider og bind, bliver det vigtigt hvilke ord der



beskrives på den tildelte plads. Under redigeringen af DDO var det korpusbaserede derfor vigtigt, og hvert eneste ord der blev optaget i ordbogen, var et resultat af en nøje fastlagt procedure hvor korpusudbredelse var vigtigst. Men reelt set ved vi ikke om korpusprincippet gavner brugerne. Det er i overensstemmelse med den videnskabelige, deskriptive tradition, og bag den ligger en antagelse om at hvis man beskriver de 50.000 eller 100.000 mest udbredte ord, har man også nogenlunde dækket behovet for hvad folk har brug for at slå op. Som en kollega engang bemærkede: «Når man som DDO går korpusbaseret frem, er der grænser for hvor galt det kan gå». Det er en god, pragmatisk indstilling, og formentlig er den også rigtig.

Men i stedet kunne man også interessere sig for hvad brugerne efterspørger. *Nisseghetto, glokø, mimrepenge, kampsvede, twitterflirte, pizzahak, kvalmsk* er eksempler på ord som brugere af DDO har indsendt til den nyordsbase som Det Danske Sprog- og Litteraturselskab driver sammen med Dansk Sprognævn, og man kan spørge om det snarere er den slags ord der skal med i ordbogen. Efter korpusprincippet skal de næppe, for nogle er for lavfrekvente, og andre knytter sig til emner der er for snævert knyttet til et dagsaktuelt emne, og som derfor kan forventes at forsvinde uden at have etableret sig i sproget. Omvendt må man spørge: Gør det noget når de ikke tager pladsen fra et andet ord? Nej, det gør de ikke, og argumentet mod at tage dem med er derfor mere praktisk end teoretisk. Det tager tid at oprette og redigere artikler, og derfor bør redaktionen have principper for hvordan man vælger det næste ord. På DDO's redaktion spørger vi indimellem os selv om vi skal blive ved med strengt at følge det korpusbaserede princip. For at besvare det spørgsmål ordentligt er der nogle oplysninger vi har brug for: Er der ord i DDO som aldrig bliver slået op? Hvis ja, tilhører de da frekvensmæssigt en bestemt del af korpus? Er de ord som brugerne

søger forgæves efter, frekvente, lavfrekvente eller sjældne ord? Nogle ordbogssites overlader det mere eller mindre til brugerne at styre lemmaselektionen, i Danmark fx ordbogen.com. Hvis en søgning ikke fører til en artikel, bliver artiklen oprettet og skrevet (Bergenholtz & Nordahl 2012:221). Det fører flere spørgsmål med sig: Er det mere sandsynligt at et ord der er blevet søgt efter én gang, bliver søgt efter igen frem for et andet ord med en bestemt korpusfrekvens. Gør det evt. en forskel om der er blevet søgt efter ordet to, tre eller flere gange?

Man har altså brug for at vide mere om sammenhængen mellem det brugerne selv indberetter, det de slår op, og udbredelsen af ord i korpus hvis man skal sige noget fornuftigt om disse principper som metode til lemmaselektion. Det er noget vi har taget fat på at undersøge, men resultaterne af undersøgelsen når jeg ikke at komme ind på i denne artikel.

## 4.2. Brugertilpasning

Da næppe to brugere har samme forudsætninger og behov, og da det teknisk er muligt at præsentere samme indhold i forskellige visninger, er det nærliggende at tilbyde en brugertilpasset version af databasen for den enkelte bruger. I løbet af de seneste år er forskellige muligheder blevet foreslået og afprøvet. I DDO tilbyder vi en kort og en lang visning, andre ordbøger skelner mellem forskellige brugerorienterede situationer som man har set det i ordbøger udgivet af Center for Leksikografi ved Aarhus Universitet. Senest har Centret dog forsøgt sig med at udgive dele af basen som selvstændige produkter i tillæg til den samlede netordbog (Den Danske Netordbog): en særlig skrive-

ordbog, en synonymordbog, en betydningsordbog og en grammatik- og skriveordbog. Selv har jeg også ved flere lejligheder argumenteret for brugertilpasning (jf. Trap-Jensen 2010), men er i dag mere skeptisk end jeg har været. Det skyldes at vores egne erfaringer ikke er positive, og indtrykket støttes af kolleger der har forsøgt sig med det samme (se Verlinde & Binon 2010:1150). Forudsætningen for at brugertilpasningen kan fungere, er at brugerne gør en aktiv indsats ved at specificere hvad de har brug for, og det sker desværre ikke. Problemet er at brugerne enten ikke opdager muligheden, eller at de ikke kan finde ud af at analysere deres behov og vælge rationelt. Hvad grunden end er, viser vores brugerundersøgelser at de ikke benytter sig af muligheden. I den nuværende udformning tror jeg derfor ikke på en fremtid for brugertilpasning – selvom jeg stadig anser det for at være en indlysende rigtig fremgangsmåde.

Noget af det man kan forestille sig ville virke mere effektivt, er en mere personlig brugertilpasning baseret på den enkelte brugers søge- og navigeringshistorik. Man kender det fra e-handelssider hvor et firma «anbefaler» andre produkter fra deres sortiment som kunne være interessant for kunden. Det kan virke grænseoverskridende i starten, men jeg må konstatere at de ofte er ganske gode til at foreslå relevante produkter, herunder ting som man ikke kendte til i forvejen. Forslagene er blevet til ved at koble søge- og købshistorik med et antal kunde profiler. Den samme tankegang kan formentlig godt overføres til en ordbogssituation hvor ordbogsvisningen tilpasser sig en bestemt brugerprofil på baggrund af hvad der tidligere er blevet søgt og navigeret efter. Uden at være ekspert på området forestiller jeg mig at det er nemmere at indføre funktionen for betalingsordbøger med en loginfunktion, men måske er en IP-adresse og en cookie tilstrækkeligt. Man kan også tænke sig at den personlige ordbog er tilknyttet den enkelte bruger og

udvikler sig enten som følge af brugerens adfærd eller i kombination med brugerens egne aktive specifikationer.

## 5. Fremtidens ordbøger

Hermed er vi fremme ved det der endnu ikke er realiseret, men hører fremtiden til. I den sammenhæng er det naturligt at spørge om ordbogen overhovedet behøver at være synlig som en selvstændig resurse som i dag. Det er der ikke nogen tvingende grund til, og personlig synes jeg ikke at tanken er skræmmende. Det er vel nærmest parallelt med situationen for papirordbogen i dag: Det er muligt at papirordbogen bliver her lidt endnu som et nicheprodukt for dem som foretrækker det, og sådan vil vi måske også have det med ordbogssites i fremtiden. Man kan konsultere en onlineordbog hvis man har lyst til at læse flere artikler i træk eller gå på opdagelse i ordbogen for fornøjelsens skyld, men ellers dukker ordbogen bare op dér hvor man har brug for den, i den situation som udløste behovet for at slå et ord op. Hvis man læser en tekst og møder et ord man ikke kender, er det unægtelig nemmere at få direkte besked ved fx at klikke på ordet og få forklaringen i et pop-op-vindue frem for at skulle gå hen et helt andet sted og slå ordet op dér – helst i den betydning ordet har i den aktuelle kontekst, men den løsning hører en endnu fjernere fremtid til.

## 5.1. Interaktion og brugerinvolvering

Et andet kendetegn ved moderne brugere er at de færdes på sociale medier, og ligesom dér vil de gerne deltage og bidrage, men har i øvrigt ikke stor tålmodighed. Flere ordbogssites forsøger at komme brugerne i møde på dette punkt og gør meget ud af de interaktive muligheder. Det er en måde at brande ordbogen på og dermed generere trafik på siden. Det sker også i et vist omfang på ordnet.dk. Brugere kan fx indsende forslag til nye ordbogsartikler, og de kan få svar på spørgsmål om sitet og sproglige spørgsmål. En del af dem publiceres i rubrikken Sprogligt hvis de vurderes at have mere almindelig interesse. Endelig findes der på siden et lille anagramspil, Krasser, der benytter ordene fra DDO. Dette er blot eksempler på nogle af mulighederne. Andre ordbøger har en mere udbygget sprogvejledning, laver særlige nyhedsbreve eller blogindlæg. Interaktionen kan også være overvejende i modsat retning så brugere bliver mere direkte involverede. De kan fx få mulighed for at redigere hele artikler eller udvalgte elementer af dem, fx udtale, ækvivalenter eller uploade multimediefiler.

Wiktionary er vistnok det eneste større projekt i det nordiske område som anvender brugerredigering i stor stil. I Danmark vokser Wiktionary kun langsomt og har ikke samme succes og opmærksomhed som søsterprojektet Wikipedia. I Sverige er Wiktionary væsentlig større: 54.293 opslagsord i grundform oplyser de på deres hjemmeside (pr. 31. oktober 2013).

Når det gælder kvalitetsbedømmelsen af brugergenereret indhold, kan det være problematisk at slutte fra internationale tendenser til vores område. Udover kulturelt bestemte faktorer og traditionen i det enkelte land betyder befolkningens størrelse helt oplagt noget, og hvad der er en succes i eksempelvis det

engelsksprogede område, er det ikke nødvendigvis hvis det overføres til de i den målestok små lande som de nordiske. Antallet af aktive brugere/redaktører betyder formentlig noget for kvaliteten af arbejdet. En generel erfaring fra Wikipedia synes at være at kvaliteten af artiklerne vokser jo længere historik de har. Det betyder nemlig at der er foretaget flere ændringer og har været flere personer til at rette faktuelle fejl eller tvivlsomme formuleringer. Observationen forekommer overbevisende, men jeg kender dog ikke til konkrete undersøgelser der har bekræftet den.

Hvis det er svært at overføre erfaringer fra store sprog med mange brugere, kan man til en vis grad sige det samme om finansieringen. Blandt kolleger fra det engelsktalende område er det nærmest blevet en sandhed at brugerne ikke vil betale for ordbøger, og at gratis ordbøger med reklamefinansiering er den eneste vej frem – i kombination med betalingsapps til mobile enheder. Jeg er dog ikke sikker på at den sandhed også automatisk gælder situationen i Norden. Det gør en forskel om der er over en milliard potentielle øjne på reklamen, eller om de skal måles i tusinder. Det er ikke nogen naturlov at brugere ikke vil betale, men måske kommer det an på hvem man spørger. I Danmark har forlagsbranchen luftet de samme synspunkter og i øvrigt reageret på den vigende betalingsvilje ved at skære ned på redaktionerne. Men måske har de bare ikke været dygtige nok? Måske er synspunktet for letkøbt, men som udenforstående kan jeg konstatere at mens Politikens Forlag har nedlagt ordbogsafdelingen, og Gyldendal meget længe har ligget underdrejet med en minimal redaktion, har et it-firma, ordbogen.com, formået at sætte sig solidt på det kommercielle ordbogsmarked i Danmark, og forretningen går tilsyneladende godt.

Den økonomiske afmatning som har ramt Europa generelt, har sat sit præg på de offentlige kilders og fondenes finansie-

ringsvilje, dog med visse undtagelser. Ikke mindst er den norske økonomi generelt i bedre form end det øvrige Nordens. Selvom økonomisk krise vanskeliggør finansiering, er det omvendt også paradoksalt, for digitale ordbøger er på mange måder et udtryk for effektivisering. Med en løbende digital vedligeholdelse undgår man at gentage unødigt dobbeltarbejde når ordbøgerne forældes og skal erstattes af helt nye projekter, måske allerede efter en generation sådan som man har oplevet det med de trykte ordbøger. For brugerne er det en klar fordel at indholdet altid er opdateret og aktuelt; for bevillingsgiverne er det en fordel at de får mere ordbog for de samme penge. Men forskydningsen fra afsluttet projekt til løbende drift kræver en mental erkendelse og omstilling hos bevillingsgivere, politikere og embedsmænd som måske endnu ikke har indfundet sig. Det betyder givetvis noget at der er mindre prestige forbundet med løbende drift end ved at søsætte nye og spændende projekter.

Når man skal vurdere den udvikling som vi befinder os midt i, er det uklogt at udtale sig alt for håndfast om hvor vi er på vej hen. Jeg tror dog at vi denne gang, i modsætning til hvad der skete som følge af korpuserevolutionen, kommer til at opleve at ordbøgerne ændrer sig meget. I den forstand er der mere revolution over det der er sket de sidste 5-6 år, men det er overvejende en teknisk omvæltning. En sandsynlig udvikling kunne være at ordbøger i den form vi kender dem, med tiden udvikler sig til at blive et nicheprodukt som de fleste mennesker måske slet ikke kommer til at møde i deres hverdag. Til gengæld vil de være integreret i mange flere forskellige stykker software rundt og dukke op når brugeren har brug for forklaringer, hjælp til skrivning eller oversættelse. Det betyder at der bliver større afstand mellem det indhold som leksikografen redigerer, og den visning som slutbrugeren kommer til at se. For leksikograferne kan det betyde at de ikke får noget med visningen at gøre, men alene

skal koncentrere sig om indholdet. Selvfølgelig kan man vælge at bevare en ordbogshjemmeside som leksikografer og ligesindede kan konsultere for fornøjelsens skyld – men altså som et udpræget nicheprodukt.

Til gengæld skal indholdet kunne udveksles med en lang række eksterne rekvirenter: med it-firmaer, med producenter af aviser, e-bogslæsere og hjemmesider samt andre udviklere og serviceleverandører som har brug for at kunne tilgå indholdet og kommunikere med de serversystemer hvor ordbogsindholdet ligger. Udvekslingen af data må derfor i højere grad forventes at komme til at ske ved hjælp af webservices og API'er til de eksterne rekvirenter. På den måde undgår dataindehaveren at udlevere fysiske data, mens rekvirenten får adgang til data, typisk via internettet, og kan i øvrigt tilpasse ordbogsindholdet efter deres formål, dvs. hvordan slutbrugeren ser indholdet, afhænger af de tilknyttede stylesheets.

Endelig bliver leksikografiske data i stigende grad interessante for andre end dem der laver ordbøger. Der er en stigende efterspørgsel efter leksikografiske data til andre typer af sprogteknologiske produkter, fx synonymlister, fuldformsleksika og domæneinventarlistes. Det er ting som indtil nu mest er blevet opfattet som spin-off af det leksikografiske arbejde med ordbøgerne, men det er sandsynligt at det bliver en mere central del af den leksikografiske aktivitet fremover. Der er med andre ord ikke tegn på at efterspørgslen efter leksikografiske data bliver mindre, og leksikografer kan derfor se fremtiden i møde med fortrøstning.



## Litteratur

### Ordbøger

- Dahl, B.T. & H. Hammer (1907–14): *Dansk ordbog for Folket I–II*. København og Kristiania: Gyldendalske Boghandel, Nordisk Forlag.
- DanNet, <<http://wordnet.dk>>
- DDO = *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab. <<http://ordnet.dk/ddo>> (oktober 2013).
- Den Danske Netordbog, <<http://www.ordbogen.com>>.
- Kristiansen, V. (1866): *Bidrag til en Ordbog over Gadesproget og saakaldt Daglig Tale*. Kjøbenhavn: Boghandler H. Hagerups Forlag.
- Molbech, Christian (1859): *Molbechs ordbog 1-2* (2. udg.). København: Gyldendalske Boghandlings Forlag.
- ODS = *Ordbog over det danske Sprog* (1918–56): København: Det Danske Sprog- og Litteraturselskab og Gyldendal. <<http://ordnet.dk/ods>> (oktober 2013).
- OED = *Oxford English Dictionary*, <[www.oed.com](http://www.oed.com)>.
- SAOB (1898–) = *Ordbok över svenska språket utgiven av Svenska Akademien* (Svenska Akademiens ordbok). Lund: Gleerups förlag.
- Videnskabernes Selskabs Ordbog* 1–8 (1793–1905). København.

### Anden litteratur

- Atkins, Beryl T.S. (2010): *The DANTE Database: Its Contribu-*

- tion to English Lexical Research, and in Particular to Complementing the FrameNet Data. I: G.-M. de Schryver (ed.): *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hank*. Kampala: Menha Publishers, 267-297.
- Bergenholtz, Henning & Bjarni Norddahl (2012): Ordbogsartikler, som ingen læser. I: *LexicoNordica 19*, 207-223.
- Cook, Paul, Jey Han Lau, Michael Rundell, Diana McCarthy & Timothy Baldwin (2013): A lexicographic appraisal of an automatic approach for detecting new word senses. I: I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets & M. Tuulik (eds.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 49-65.
- Dahlerup, Verner (1907): Principer for ordbogsarbejde. I: *Danske Studier 1907*, 65-78.
- Didakowski, Jörg, Lothar Lemnitzer & Alexander Geyken (2012): Automatic example sentence extraction for a contemporary German dictionary. I: Ruth Vatvedt Fjeld & Julie Matilde Torjusen (eds.): *Proceedings of the 15th EURALEX international Congress 7-11 August, 2012, Oslo*. Oslo: University of Oslo, 343-349.
- Fillmore, Charles J. (1995): The Hard Road From Verbs To Nouns. I: M. Chen & O. Tzeng (eds.): *In honor of William S-Y. Wang*. Taipei, Taiwan: Pyramid press, 105-129.
- Halskov, Jacob (2010): Halvautomatisk udvælgelse af lemma-kandidater til en nyordsordbog. I: *LexicoNordica 17*, 73-98.
- Jakubiček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel (2013): The TenTen Corpus Family. I: *Lancaster, 7th International Corpus Linguistics*

- Conference CL 2013*, 125-127.
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell (2004): *The Sketch Engine. I: Geoffrey Williams & Sandra Vessier (eds.) Proceedings of the Eleventh EURALEX International Congress*. Lorient: Université de Bretagne-Sud, 105-115.
- Kilgarriff, Adam, Miloš Husák, Katy McAdam, Michael Rundle & Pavel Rychlý (2008): *GDEX: Automatically Finding Good Dictionary Examples in a Corpus. I: Elisenda Bernal & Janet DeCesaris (eds.): Proceedings of the XIII EURALEX International Congress*. Barcelona: Universitat Pompeu Fabra, 425-433.
- Langebek, Jacob (ca. 1738): Plan for Rostgaards ordbog. Manuskript. Her citeret fra: *Det Kongelige Danske Videnskabsbernes Selskab 1742-1942. Samlinger til Selskabets Historie*, v. Asger Lomholt, III, 1960, 219, med henvisning til *Nye Danske Magazin V. 4*, 1827, 271-76.
- Lorentzen, Henrik, Sanni Nimb & Thomas Troelsgård <denne udgivelse>: *Fra Begreb til Ord*.
- Lorentzen, Henrik (1999): Jagten på det gode citat. Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus. I: Martin Gellerstam, Kristinn Jóhannesson, Bo Ralph & Lena Rogström (udg.): *Nordiska studier i lexikografi 5. Rapport från Konferens om lexikografi i Norden, Göteborg 26–29 maj 1999*. Göteborg : Nordiska föreningen för lexikografi, 202-216.
- Rydstedt, Rudolf (1988): Creating a Lexical Database from a Dictionary. I: *Studies in Computer-Aided Lexicology*. Göteborg, 228-267.
- Trap-Jensen, Lars (2010): One, Two, Many: Customization and User Profiles in Internet Dictionaries. I: Anne Dykstra & Tanneke Schoonheim (eds.): *Proceedings of the XIV*

TRAP-JENSEN

*Euralex International Congress (Leeuwarden, 6-10 July 2010)*, Ljouwert: Fryske Akademy, 1133-1141.

Verlinde, Serge & Jean Binon (2010): *Monitoring Dictionary Use in the Electronic Age. I*: Anne Dykstra & Tanneke Schoonheim (eds.): *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*, Ljouwert: Fryske Akademy, 1144-1151.

Lars Trap-Jensen  
ledende redaktør  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
ltj@dsl.dk

# **Grunden för projektet av en studieordbok över svenska fasta liknelser som beskriver människan**

*Alexej Alösjin*

Basis of the project of an educational dictionary of Swedish regular comparisons which characterize people.

The article deals with the compiling of educational dictionaries of Swedish regular comparisons which characterize people. The dictionary will be aimed at Russian speakers. The author offers a model of the dictionary unit, analyses content of every zone in the dictionary unit. At the end of the article are given examples of dictionary units.

## **1. Inledning**

Fasta liknelser, som de flesta språkforskare inkluderar i fraseologin i vid bemärkelse, upptar en viktig plats i språkets fraseologiska system, eftersom de utgör en av de största grupperna bland fraseologiska enheter med likadana strukturella och semantiska egenskaper. Deras komparativa semantik uttrycks genom att jämförelsen görs med hjälp av att objekt införs med hjälp av jämförelsekonjunktioner. I svenskan används konjunktionerna *som* och *som om* i detta fall.

Det är viktigt att påpeka att med fasta avses här bara sådana liknelser som innehåller konjunktionerna *som* och *som om* och som finns i fraseologiska ordböcker och används regelbundet i tal och skrift, t.ex. *tjock som en gris, vacker som en gud, att se ut som om man har sett ett spöke* osv.

Som en fraseologisk enhet betraktas här bara konstruktioner som består av minst två ord, och sammansatta ord som *kol-svarta* eller *koögon* ska inte inkluderas i ordboken.

Det har hittills aldrig skrivits några större vetenskapliga verk om fasta liknelser i Sverige. De svenska forskarnas verk är ägnade snarare åt studier av idiom och ordspråk. Bland sådana kan nämnas artiklar av J. Allwood (1983), U. Clausén och E. Lyly (1995), monografier av J. Svanlund (2001) och E. Sköldberg (2004). Bland de första artiklarna om fasta liknelser och andra jämförelsekonstruktioner i ordböcker måste man nämna U. Clausén (2003) och E. Sköldberg (2010).

## 2. Syntaktiska typer av fasta liknelser

Svenska fasta liknelser kan man dela upp i följande syntaktiska typer.

1. Adjektiv + konjunktion + substantiv  
*mager som en skrika, hjälplös som ett litet barn*
2. Verb + konjunktion + substantiv (grupp)  
*klättra som en ekorre, känna sig som en sparv i trane-dansen, sova som en klubbad säl*

3. Verb + konjunktion + preposition + substantiv  
*sitta som på nålar; gå som i ett töcken*
4. Verb + konjunktion + particip + preposition + substantiv  
(grupp)  
*stå som fallen från skyarna, stå som träffad av blixten,  
vara som skuren ur ett enda block*
5. Verb + konjunktion + sats  
Denna typ förekommer bara i liknelser som inleds med  
 verbet ”se ut”  
*se ut som om man har sett ett spöke, se ut som om man  
har ramlat ner från månen, se ut som om man sålt  
smöret och tappat pengarna*

### **3. Teoretiska grunder för en lexikografisk representation av fasta liknelser**

Fasta liknelser brukar anges i fraseologiska ordböcker, särskilda liknelseordböcker och tvåspråkiga ordböcker över fasta liknelser. Men än så länge finns det varken någon ordbok över svenska liknelser eller någon svensk-rysk fraseologisk ordbok. Syftet med denna artikel är att utarbeta en modell för en ordboksartikel i en studieordbok över svenska fasta liknelser för rysktalande.

Som teoretisk bakgrund för detta arbete uppträder de ryska lexikografernas verk som lade grunden till studieordböcker i Ryssland (Denisov (1977), Morkovkin (1977)). I ett historiskt perspektiv brukar studieordböcker till skillnad från traditionella

akademiska ordböcker ha mindre volym och ha större fokus på användningen i undervisningen. En sådan ordbok riktar sig först och främst mot dem som studerar ett främmande språk. En viktig roll som förebilder spelade också ryska och europeiska ordböcker över fasta liknelser (Mokienko (2003), Fink (2006), Walter (2008).

Med en studieordbok menas en ordbok som är avsedd för dem som inte har språket (svenska i detta fall) som modersmål utan för dem som studerar det (rysktalande i detta fall), och en sådan ordbok ska kunna hjälpa dem att förstå liknelsernas bakomliggande bakgrund, alltså vilka förebilder och vilken motivering som ligger bakom dem, samt deras historiska och kulturella bakgrund. Därför kommer alla kommentarer inom ordboksartikelns att ges på ryska.

Ordlistan i denna planerade ordbok kommer att bestå av svenska fasta liknelser, som betecknar människan, som finns i svenska fraseologiska ordböcker och används i skönlitterära texter.

Ordboken ska vara uppbyggd enligt den ideografiska principen. Den ska bestå av flera tematiska avdelningar, såsom människans utseende, beteende, känslor etc. Varje avdelning ska innehålla alla ordboksartiklar som hör dit.

### 3.1. Ordboksartikelns struktur

Ordboksartikelns kommer att bestå av följande delar:



### 3.1.1. Uppslagsuttrycket

En fast liknelse skriven med stora bokstäver med fet stil. Om några delar av uttrycket har varianter, anges de i rund parentes.

T. ex. **SKIFTA (BYTA) HAMN SOM ETT TROLL**

Här kan också anges särskilda beteckningar vid uttrycket om det behövs, t.ex. *книжн., устар.*, (skriftspråkligt, ålderdomligt osv.)

### 3.1.2. En bokstavlig översättning till ryska

Denna del direkt efter uppslagsuttrycket anges med stora bokstäver, med en särskild beteckning *букв.* (bokstavligt) t.ex.

**TJOCK SOM EN GRIS**

*букв.* **ТОЛСТЫЙ КАК СВИНЬЯ**

### 3.1.3. Betydelsebeskrivning

Tolkningen anges med rak stil och gemena bokstäver. Den inleds på ryska med prepositionen som betyder ”om”

**TJOCK SOM EN GRIS**

**ТОЛСТЫЙ КАК СВИНЬЯ**

О чрезмерно толстом человеке. (om en för fet människa)

Om det finns några begränsningar för användningen av

ALÖSJIN

uttrycket eller några egenheter som gäller användningen, anges de också här.

### **VIT SOM EN LILJA**

#### *БЕЛЫЙ КАК ЛИЛИЯ*

О невинном человеке. Употребляется только по отношению к молодой невинной девушке. (Används bara om en ung oskuldsfull flicka)

#### 3.1.4. Historiska och kulturella kommentarer

Denna del är inte obligatorisk. Den behövs när uttrycket är helt obegripligt för rysktalande och kräver historiska eller etymologiska förklaringar. Vanligtvis saknar ett sådant uttryck ryska ekvivalenter. Kommentaren inleds med en ”ljus”-symbol och med rak stil.

### **VARA SOM KÄRRINGEN MOT STRÖMMEN**

#### *БЫТЬ КАК СТАРУХА ПРОТИВ ТЕЧЕНИЯ*

Об очень упрямом человеке, стремящемся сделать всё наперекор другим.

‡ Основание сравнения восходит к анекдоту XIII века о том, как во время праздника жена, несмотря на возражения мужа, из упрямства передвигала свой стул все дальше от стола и все ближе к реке. Когда она упала в реку и утонула, муж стал искать ее тело вверх по течению, заявив, что она и после смерти наверняка должна была делать все по-своему).

(Från sagor och anekdoter om hustrur som sätter sig upp mot sina män. I en anekdot, känd från 1800-talet, envisades hustrun under en trädgårdsfest med att flytta sin stol allt längre bort från bordet och allt närmare en å. När hon föll i vattnet och drunknade, letade mannen sedan uppströms efter henne och sa att hustrun även i döden borde gå mot strömmen) (Bevingat (2005))

### 3.1.5. Bruksexempel

Bruksexempel av fasta liknelser ska hämtas från skönlitterära verk eller internet. Denna del inleds med en symbol ”uppslagen bok”. Citaten anges med rak stil, liknelsen anges i fet kursivstil.

#### **SKINA SOM EN SOL**

букв. СВЕТИТЬСЯ КАК СОЛНЦЕ

О радостном, светящемся от положительных эмоций человеке.

📖 Mårten kom hem från sitt sommarparadis och en dag stod han i min tambur och *sken som en sol*. (H. Nordin, Sparsam mimik)

### 3.1.6. Ryska ekvivalenter

Denna del är inte obligatorisk heller. Här anges ryska ekvivalenter till den svenska liknelsen om de finns. De anges med kursivstil och inleds med beteckningen *русск.* (rysk).

ALÖSJIN

### **TJOCK SOM EN GRIS**

*Русск. Толстый как свинья, как боров, как кабан, как поросенок.*

#### 3.1.7. Svenska synonymer

Denna del hänvisar till svenska liknelser av samma betydelse som en given liknelse. De inleds med beteckningen Ср. (Jämför)

### **LÅNG SOM EN STÅNG**

*Ср. lång som en stör; lång som en stake.*

## **4. Exempel på ordbokens uppbyggnad**

**ПОВЕДЕНИЕ (BETEENDE)**

**ВЕСТИ СЕБЯ НЕУВЕРЕННО (BETE SIG OSÄKERT)**

**ГÅ SOM KATTEN KRING HET GRÖT**

букв. ХОДИТЬ КАК КОТ ВОКРУГ ГОРЯЧЕЙ КАШИ

Не решаться что-л. сделать, медлить с принятием решения

📖 Jag hade mina misstankar om en hel del vid det här laget, men vi gick runt varandra, *som katten kring het gröt* och försökte in i det sista att hålla masken. (Östergren K. Gentlemän)

Så antingen har politikerna inte insett att det där med att vara entreprenör och att vara företagare inte är något som alla samhällsklasser, eller så var deras debatt idag ett bra exempel på konsten att gå som katten kring het gröt.

(<http://kulturbloggen.com>)

### **ВЕСТИ СЕБЯ НЕТЕРПЕЛИВО (BETE SIG OTÅLIGT)**

### **GÅ OMKRING (RUNT) SOM EN ÄGGSJUK (LIGGSJUK) HÖNA**

букв. ХОДИТЬ КАК КУРИЦА, КОТОРАЯ ХОЧЕТ СЧЕСТИ ЯЙЦА

Беспокойно ходить из стороны в сторону

& Alltså *går* jag runt *som en äggsjuk höna* här hemma och försöker få något vettigt gjort. (<http://trivsamthem.blogspot.ru>)

## **Litteratur**

### **Ordböcker**

Fink Arsovski, Želka (2006): *Hrvatsko-slavenski rječnik poredbenih frazema*. Zagreb: Knjigra.

- Helsing Birgitta, Magdalena Hellquist & Anders Hallengren (2005): *Bevingat*. Stockholm: Albert Bonniers förlag.
- Mokienko, Valerij (2003): *Slovar russkih ustoitjivyh sravnenij*. St Petersburg: Norint.
- Walter, Harry (2008): *Wörterbuch deutscher sprichwörtlicher und phraseologischer Vergleiche*. Greifswald: Verlag Dr. Covač.

### Annan litteratur

- Allwood, Jens (1983): Naturen som metaforfält. I: Jens Allwood, Tore Frängsmyr & Uno Svedin (red.): *Naturen som symbol*. Stockholm: Liberförlag, 1–18.
- Clausén Ulla & Erica Lyly (1995): Idiom och lexikografi. I: *Språkvård 1/1995*, 23–31.
- Clausén Ulla (2003): Jämförelsekonstruktioner i Svensk konstruktionsordbok – struktur, variation och funktion. I: Zakaris Svabo Hansen & Anfinnur Johansen (red.): *Nordiske studier i leksikografi 6: rapport fra Konference om leksikografi i Norden, Tórshavn 21.-25. august 2001*. Tórshavn: Nordisk forening for leksikografi, 75–85.
- Denisov, Piotr (1977): Struktura komponentov učebnogo slovaria. I: V. Morkovkin (red.): *Aktualnye problemy učebnoj leksokografii*. Moskva, 23–42
- Morkovkin, Valerij (1977): Učebnaja leksokografija kak osobaja lingvometodičeskaja disciplina. I: V. Morkovkin (red.): *Aktualnye problemy učebnoj leksokografii*. Moskva, 28–37
- Sköldberg, Emma (2004): *Korten på bordet. Innehålls- och uttrycksmässig variation hos svenska idiom*. Göteborg: [E. Sköldberg].
- Sköldberg Emma (2010): Bland nötkärnor, fyrbåkar och skål-

- lade råttor. Om konventionaliserade liknelser i svenskan. I:  
K. Jóhannesson et al. (red.): *Bo 65: festskrift till Bo Ralph*.  
Göteborg: Meijerbergs institut för svensk etymologisk  
forskning, Göteborgs universitet, 363–371.
- Svanlund, Jan (2001): *Metaforer som konvention. Graden av  
bildlighet i svenskans vikt- och tyngdmetaforer*. Stockholm:  
Almquist & Wiksell.

Alexej Alösjin  
docent, ph. d.  
St. Petersburgs Statliga Telekommunikationsuniversitet  
Moika 61  
191186 St. Petersburg

# Skjematiske fleirordskonstruksjonar i *Norsk Ordbok*

*Sturla Berg-Olsen*

This article explores how four schematic multi-word constructions are covered in the monolingual documentary dictionary *Norsk Ordbok*. In accordance with the tenets of Construction Grammar it is argued that schematic constructions form a central part of speakers' knowledge of language and should be documented on a par with other linguistic units. A large dictionary database fed by a corpus provides a good basis for documenting such constructions. A structured database of constructions can in its turn provide valuable input to dictionaries.

## 1. Om konstruksjonar

Konstruksjonsomgrepet blir brukt av lingvistar og leksikografar med varierende presisjonsnivå og til dels ulikt innhald. For nokre formål kan det vere tilstrekkeleg å definere *konstruksjon* nokså rundeleg, slik det er gjort i *Norsk Ordbok*: «ordsamband el setning bygd opp på ein viss måte». Vitskapelege formål krev ein større grad av presisjon. Her vil eg halde meg til ein definisjon i tråd med det som er gjengs oppfatning innan den breie retninga av lingvistikken kjend som konstruksjonsgrammatikk.



Eg definerer ein grammatisk konstruksjon som *ei konvensjonalisert språkleg eining med ei viss form og ei viss tyding* (jf. til dømes Croft 2005:274ff og Goldberg 2006:5). Eit kjenneteikn på mange konstruksjonsgrammatiske teoriar er at dei ser på konstruksjonar – i denne tekniske forstanden – som dei grunnleggjande einingane i språket.

Ut frå denne definisjonen blir spekteret av konstruksjonar svært vidt. Ein kan gruppere dei ut frå fleire parametrar, til dømes kompleksitet og skjematistitet. Nokre døme på ulike typar konstruksjonar er sette opp i figur 1:

	SPESIFIKKE	SKJEMATISKE
KOMPLEKSE	hulter til bulter	SUBJ V OBJ
	skyte ei kvit pil etter NP	V- PRET N- PL
ATOMISKE	hest	ADJ
	sin	PREP

Figur 1: Døme på meir eller mindre komplekse og meir eller mindre skjematistiske konstruksjonar.

Til venstre i figuren står einingar med spesifikt leksikalsk innhald, både atomiske (dvs. ord) og komplekse (dvs. faste ord-samband med bunde innhald). Jo lenger til høgre ein kjem, dess meir skjematistiske blir konstruksjonane, dvs. at dei får fleire variablar eller opne plassar. Komplekse skjematistiske konstruksjonar er setningsmønster, mens ordklassar er døme på atomiske skjematistiske konstruksjonar.

Eit døme på *skjematiske fleirordskonstruksjonar* er ringa rundt i figur 1. Dette er komplekse konstruksjonar som har ein viss – ofte nokså høg – grad av skjematisitet. Kompleksiteten vil seie at konstruksjonen kan delast opp i fleire ord, men tydinga til konstruksjonen er ikkje føreseieleg ut frå tydinga til dei einskilde delane. Høg grad av skjematisitet vil seie at fleire av plassane i konstruksjonen kan fyllast med ulikt leksikalsk materiale. Skjematiske fleirordskonstruksjonar er meir eller mindre det same som det Farø & Lorentzen (2009) kallar *fraseoskabelonar*.

## **2. Skjematiske fleirordskonstruksjonar og *Norsk Ordbok***

*Norsk Ordbok* (NO) er eit vitenskapleg verk som set seg høge mål. På heimesida til prosjektet står det at ordboka «gjev ei uttømmende framstilling av ordtilfanget i dei norske dialektane og i nynorsk skriftspråk. Under oppslagsorda finn du informasjon om ordtydingar, dialektformer, grammatiske opplysningar og opplysningar om ordhistorie». Eit spørsmål som uvilkårleg melder seg frå ein konstruksjonsgrammatisk ståstad, blir da i kva grad og korleis ei stor allmennordbok som NO formidlar grammatiske opplysningar som ikkje utan vidare lar seg knyte til eit bestemt lemma. På bakgrunn av si undersøking av fleirordssamband i danske ordbøker konkluderer Farø & Lorentzen (2009:89) med at «de mere subtile og upåfaldende typer med fx tompladser får en noget stedmoderlig behandling». Lyngfelt &

Sköldberg (2013) analyserer korleis nokre utvalde skjematisk konstruksjonar er dekte i fire svenske ordbøker, og finn den same tendensen – jo meir skjematisk ein konstruksjon er, dess vanskelegare er det å gje han adekvat behandling i ei ordbok.

Her vil eg i liknande ånd som dei to nemnde studiane ta for meg fire skjematisk fleirordskonstruksjonar og analysere korleis dei tar seg ut i NO. Dei fire er syntaktisk ulike og gir høve til å belyse problematikken rundt konstruksjonar i ordbøker frå fleire synsvinklar. Eg avstår her frå å jamføre ulike ordbøker, men i vil i staden prøve å gjere eit djupdykk i éi bestemd ordbok. Det er verdt å merke seg at NO er meir vidfemnande enn dei ordbøkene som blei undersøkte i dei danske og svenske studiane, og også skil seg frå dei fleste av desse ved at ho òg dekkjer talemålsvariantar av norsk.

## 2.1. Konstruksjonen [x og x]

(1) og (2) nedanfor gir døme på konstruksjonen [x og x]:<sup>1</sup>

(1) eg måtte gå forbi Café Børs [...] Ja, **måtte og måtte**, men eg gjekk nå forbi (Fløgstad: *Dalen Portland*)

(2) det finst **jubileum og jubileum**, fru Blom (*Dag og Tid* 1998)

Elementet x, som blir gjentatt med *og* imellom, er i døma over høvesvis eit verb og eit substantiv, men andre ordklassar så vel

---

1. Uthevinga med halvfeite typar her og vidare i artikkelen er mi.

som frasar og vel i viss mon òg større einingar kan opptre i konstruksjonen. [x og x] blir brukt for å signalisere tvil om (gyldigheita av) noko som opptreer tidlegare i konteksten. Ein under-type av konstruksjonen med tillegget *fru Blom* ser vi i (2); denne undertypen har ikkje motstykke elles i skandinavisk, og kan etter alt å døme sporast til ein bestemd person, nemleg skodespelaren Per Aabel (NO1000 under *Blom*).

Det einaste faste leksikalske elementet i denne konstruksjonen er konjunksjonen *og*, og det er rimelegvis i artikkelen for dette lemmaet vi finn konstruksjonen omtalt i NO, under **II og**, tyding 2b:

**b** mellom to identiske ord for å syna at dei kan ha ulik tyding el at ein dreg i tvil om ordet ein bruker, er det rette: *artig og artig / det finst jubileum og jubileum, fru Blom (Dag og Tid 1998).*

Det er godt mogleg at [x og x] ligg føre (til dømes i bruksdøme) også andre stader i NO, men det er vanskeleg å finne slike døme ved søk. Verdt å merke seg er at databasen ikkje inneheld nokon lenker til tyding 2b i **II og** – det vil seie at redaktørane ikkje har sett behov for å kommentere ved noko bruksdøme at det har samanheng med denne bruken av *og*.

## 2.2. Konstruksjonen [din NP]

(3) – Vil du sjå til å rappa deg, **din rotagamp!** skreik Erling og smurde til han med vidja (Høydal: *Villstyring*)

(4) Høyrer du det du, **di gamle kattugle** du sitt der! (Fønhus: *Domedag i Skreppebrøtin*)

Konstruksjonen [din NP] inneheld eit possessivt pronomen i 2. person eintal saman med ein NP; pronomenet kongruerer med NP-en i genus. Fleirtal er ikkje mogleg (*\*dykkar idiotar!*). Konstruksjonen er brukt i tiltale og signaliserer ekspressivitet – oftast i form av utskjelling, men av og til òg i vennleg eller kjærleg tiltale til nokon.<sup>2</sup> I NO er [din NP] etter måten breitt omtala i artikkelen **din** under tyding 2:

2 kvard., adj knytt til subst (serl skjellsord) i tiltale (svarar logisk til eit *du* med apposisjon; jfr *II du 1*): *di gås / din kjeltring / er det liv i deg enno, ditt skrymt* (Ørj. Chr 134) / i venleg, kjælende tiltale: *din vesle krok* (Vo) / ofte med till. som *der du er (står, sit o l) / ditt barn der du er!* (Du.R 150) / i tilkn til eit frg *du: du, din arme stakkar* (Mor.Sv 89) [...]

Nett som ved førre konstruksjon er det heller ikkje her elektroniske lenker til denne tydingsbolken frå bruksdøme andre stader i ordboka, sjølv om [din NP] er rikt representert i eksempelaterialet i NO.<sup>3</sup> Eg har funne 108 instansar av denne konstruksjonen i NOs elektroniske artikkelbase. Interessant nok er det mellom desse ikkje eit einaste klart døme på vennleg eller kjærleg tiltale, sjølv om denne bruken altså er nemnd under **din**.

- 
2. Det finst ein nærskyld konstruksjon med pronomen i 1. person, brukt i utsegner retta til ein sjølv. Denne er omtala i NO under **II min**, tyding 5 med fleire døme av typen *eg, mi dumme tosse*.
  3. Eit konsekvent system der instansar av ein konstruksjon blei lenka til den plassen der konstruksjonen var omtala, kunne mellom anna ha gitt informasjon om produktiviteten til konstruksjonen.

Dei mange døma på [din NP] er truleg dels ein indikasjon på at svært mange NP-ar potensielt kan brukast som skjellsord i denne strukturen, og også på at slike vendingar er heller frekvente i kjeldematerialet til ordboka.

Døma på konstruksjonen er førte på fleire ulike måtar: Oftast står dei i ein eigen tydingsbolc, anten med eller utan bruksmerking (til dømes *nedsetj, skjellsord, sviv*. [for *svivyrdeleg*], *kvard.*, *skjemt.*), og stundom med funksjonsutgreiing av typen *nytta som skjellsord* (øg i kombinasjon med bruksmerking). Det hender også at døme på konstruksjonen er førte i tydingsbolc saman med andre døme, slik som her, i tyding 2b frå artikkelen **klåfinger**:

**b** ei(n) som er klåfingra [...]: *hald deg vekk, din 'klåfing'!* (YSandsvær) / overf: *ein klåfinger til å rive ned, gjera um, kaste burt det gamle* (Be.VR III,179).

### 2.3. Konstruksjonen [V seg (i)gjennom NP]

(5) direktøren **drikk seg gjennom perversitetane** (*Dag og Tid* 2000)

(6) billige erotiske scener der **skodespelarane onanerer seg gjennom tre verdsdelar** (*Firda* 2001)

[V seg (i)gjennom NP] inneheld eit verb med refleksivpronomen, preposisjonen *gjennom* eller *igjennom* og ein NP som er komplement til denne preposisjonen. Konstruksjonen blir brukt til å uttrykke ei bokstaveleg eller metaforisk rørsle som går føre seg gjennom noko. Fokuset kan anten ligge på årsaka til at

rørsla skjer eller måten ho skjer på. Dette er ein av ei rekke konstruksjonar i norsk med nærskyld form og tyding, og nett *gjennom*-konstruksjonen kan sjåast på som ein undertype av RCM-konstruksjonen (der RCM står for *reflexive caused motion*), analysert inngåande av Seland (2001) og også omtala i Sveen (2002).

Seland teiknar opp eit bilete der RCM-konstruksjonen er i slekt med og motivert av ei rekke andre konstruksjonar, som har visse trekk sams med, men på nokre punkt skil seg frå RCM-konstruksjonen. Dette gjeld til dømes den refleksive rørslekonstruksjonen (*åle seg gjennom mylderet*), den refleksive veg-konstruksjonen (*bane seg veg gjennom hopen*) og resultativkonstruksjonen (*trimme seg frisk*).<sup>4</sup>

*Gjennom*-konstruksjonen er relativt lett å finne ved søk, både i nynorskorpuset (som inneheld drygt 20 døme på han) og i bruksdømefeltet i NOs artikkeldatabase. Når det gjeld omtale av konstruksjonen i seg sjølv, er den naturlege plassen å leite artikkelen **gjennom** el **igjennom**. Her finn ein rett nok fleire døme på han, men dei er spreidde på fleire tydingsbolkar, og ei enskapleg framstilling av *gjennom*-konstruksjonen i seg sjølv manglar:

1) a) brukt om rørsle el noko oppfatta som rørsle for å uttrykkja at ho går føre seg i (innanfor, omslutta av) det som styringa nemner frå den eine enden, den eine sida heilt ut til (på) den andre, at ho omfattar det som er uttrykt i styringa frå byrjing til slutt; (inn) frå den eine og (ut) til den andre sida el enden av: [...] og *elden har vel ikkje kunne ete seg*

---

4. For mange formål kan det vere nyttig å skilje mellom fleire undertypar av resultativkonstruksjonen. For ei finare inndeling av tilsvarande konstruksjon i svensk sjå Jansson (2005).

*upp gjennom taket* (Up.TL 193) [...] *ei stjerne .. glødde seg gjennom skya* (Du.MM 253) [...]

2) saman med trykkveikt verb der prep og verb lagar eit meir el mindre fast samband som uttrykkjer rørsle gjennom (i tyd 1a) noko(n) (og prep såleis nærmar seg el går over til adv) [...] *arbeida seg (i)gjennom (noko)* [...] // bil., overf, ofte for å uttrykkja at nokon kjem til slutten av el på ein heldig el lykkeleg måte utstår el legg bak seg det styringa (el obj) nemner (t d tidbolk, sjukdom, vanskar, plager o l): *frysa seg, greia seg, gå, koma (seg), slita seg (i)gjennom (noko)* [...] // bil., overf, til å uttrykkja at nokon fylgjer skule, undervisning el fullfører eksamen o l [...] *stræva seg igjennom latinskulen* (Kl.LS 283) [...]

Det er vanskeleg å kritisere redaktøren som har skrive artikkelen **gjennom** for å ha hatt manglande fokus på konstruksjonar. Den primære eininga i ei ordbok er naturleg nok ordet, og artikkelen er meint å vere ei beskriving av eigenskapane til ordet (*i)gjennom*. I eit slikt perspektiv blir konstruksjonen [V seg (*i)gjennom* NP] meir perifer, og skil seg ikkje nødvendigvis så klart ut frå andre nærskyldte konstruksjonar. Også artikkelen **IV seg** inneheld døme på *gjennom*-konstruksjonen (og meir generelt på RCM-konstruksjonen), men handlinga er noko stuttare enn kva ein konstruksjonsgrammatikar kunne ønskt seg:

1 [...] **f** i konstruksjon med (intr) verb og adv som uttrykkjer retning, føremål, (ynskjeleg) resultat el sideverknad av v-handlinga: *dei tusla seg heim* (Håløygm 1962,111) / *dei streva seg fram på bakken heime* (Du.BA 207) / *så dansa dei seg gjennom gardar og tun langt innerter grenda* (Hordaland 1900) [...]



Døme på *gjennom*-konstruksjonen finst i ei rekke verbartiklar i NO, ofte skilde ut i eigne tydingsbolkar med grammatisk merknad *refl*, men av og til òg som nakne bruksdøme utan kommentar. Tre døme på ulike føringsmåtar er attgitte nedanfor.<sup>5</sup> Eg har ikkje funne døme på at instansar av denne konstruksjonen er førte som sublemma.

**gråta** v [...] **1 a)** (ved sterk sinnsrørsle) fella tårer (med klagande lyd attåt); grina (2a-b), sutra [...] // refl: [...] *han gret seg gjennom ei vise om mora* (T.Solh.IS 26) [...]

**sitja** v [...] **B** med obj. [...] **2** refl, med pred-ord el adv: *han sat seg gjennom dagane* (BrodalTMK 91) [...]

**IV stava** [...] **1** setja saman bokstavar til stavingar og ord; (serl:) lesa, nemna bokstav for bokstav i ord [...] *stava seg gjennom ei bok* [...]

#### 2.4. Konstruksjonen [NP VP<sub>1</sub> Ø VP<sub>2</sub>]

(7) Nei ho kjenner det ikkje. Ho er med i viktigare ting. **Ho ligg bøyer sine inste blad saman.** (Vesaas: *Signalet*)

Denne konstruksjonen skil seg på fleire måtar frå dei tre eg har tatt for meg fram til nå. VP<sub>1</sub> er ein verbfrase som uttrykkjer kva for stilling eller rørsle subjektet er i. Ø er ein tom plass, mens

---

5. I artikkelen **IV stava** er *stava seg gjennom ei bok* ført som redaksjonelt døme, derfor manglar det kjelde her.

VP<sub>2</sub> er ein verbfrase i same form som VP<sub>1</sub> som uttrykkjer ei handling som subjektet driv med, eventuelt ein tilstand som varer ved. Dette er ein regionalt avgrensa variant av den mykje vanlegare konstruksjonen [NP VP<sub>1</sub> og VP<sub>2</sub>], der dei to verbfrasane er bundne saman av konjunksjonen *og*. *Norsk referansegrammatikk* seier at varianten utan *og* mellom *anna* er vanleg «i de indre dialektene i Agder og Telemark» (NRG 535). Nettopp det dialektgeografiske aspektet er av interesse her, i og med at NO også er ei dialektordbok som gjerne heimfestar språklege data på form- og tydingsnivå i den mon kjeldegrunlaget gjer det mogleg. Å seie noko om kvar i landet ein viss konstruksjon er i bruk, er i det perspektivet også av interesse.

Eit problematisk trekk ved denne konstruksjonen i ein leksikografisk samanheng er at han ikkje inneheld noko fast leksikalsk element, men berre har opne plassar. Så mens den allmenne varianten [NP VP<sub>1</sub> og VP<sub>2</sub>] er omtala i artikkelen **II og** (under tyding 2o) i NO, er det ikkje éin openberr plass i ordboka å leggje inn informasjon om varianten utan *og*. Han er likevel nemnd i fleire artiklar for verb som typisk kan fylle plassen VP<sub>1</sub>, såleis i artiklane **I gå**, **sitja** og **IV stå** – men ikkje **liggja**, sjølv om kjeldematerialet inneheld døme på konstruksjonen med *liggja*, som vist i (7).

**I gå** [...] **A** [...] **5** [...] **d** (om person) fara meir el mindre føremålslaust ikring; driva (ofte brukt (durativt) føre *anna* verb for å syna at ein jamt, ideleg el over lengre tid gjer noko el er i ein tilstand som verbet uttrykkjer); sviva ikring, halda på [...] // utan inf-merke el konj: *son hass måtte gange gjæte* (R i Dølen 1870,21,3,2) [...]

**sitja** v [...] **A** [...] **4** (med avbleikt tyd i sideordna samband med *anna* v, i hovudsak for å uttrykkja durativ el iterativ

verbalhandling) halda på (støtt og stendig) [...] // utan bindeord mellom verba: [...] *dei såte då åte kvellsmaten* (GOS III,10) [...]

**IV stå** v [...] **A** [...] **2** med avbleikt tyd i sideordna samb med anna v, (serl om levande vesen i oppreist stode) brukt for å uttrykkja at ei handling el ein tilstand varer ved el (i fortid) varte ved (då noko anna hende); halda på (støtt og stendig) (jfr NoRef 648); jfr *sitja A4* [...] // utan bindeord mellom verba (jfr NoRef 535): *Siss står ser på dei med open munn* (Ves.Is 103) [...]

Konstruksjonen [NP VP<sub>1</sub> Ø VP<sub>2</sub>] har altså av naturlege grunnar ikkje fått noka samla behandling i NO, men er likevel omtala og eksemplifisert fleire stader. Det er noko påfallande at ordboka ikkje heimfestar konstruksjonen utan *og* i dei ulike verbartiklane der han kjem opp, sjølv om dei døma som blir førte, gir ein indirekte peikepinn om kvar denne uttrykksmåten er eller har vore gjengs.

### 3. Konstruksjonar i ordbøker og andre stader

Store ordboksverk som NO tar mål av seg til å beskrive ikkje berre ordforrådet til morsmålsbrukarane, men også dei grammatiske eigenskapane til einskilde leksikalske einingar. Kva for (meir eller mindre skjematisk) konstruksjonar eit gitt ord opptrer i, er ein sentral del av kompetansen til språkbrukarane, og

ein kan reint ut sagt spørje seg om det er mogleg å beskrive eit ord fullgodt utan å trekkje inn konstruksjonsperspektivet (Fillmore 1994:114). I alle tilfelle er det eit poeng at skjematisk fleirordskonstruksjonar bør kartleggast og dokumenterast i same grad som andre språklege fenomen. Dei får likevel ofte mangelfull behandling i ordbøkene, mellom anna fordi dei ikkje alltid kan knytast til eitt bestemt leksem. Jamført med dei danske og svenske ordbøkene som har vore analyserte i tidlegare studiar, kjem NO relativt godt ut av det: Dei fire stikkprøvene eg har gjort, tyder på at verket faktisk inneheld ein god del informasjon også om slike konstruksjonar. Likevel er det klart at ei ordbok – eller i alle fall ei ordbok strukturert på tradisjonelt vis – ikkje er optimal korkje for den som vil dokumentere skjematisk konstruksjonar eller for brukarar som ønskjer å finne informasjon om dei. Sjølv ei stor dokumentasjonsordbok som NO kan ikkje dokumentere alt, og leksikografar må nødvendigvis fokusere på det leksikaliserte og det typiske. Eit døme som *ho halvsov seg gjennom mattetimen* høyrer truleg ikkje heime i artikkelen **halvsøve**; Lyngfelt & Sköldberg (2013:87) åtvarar beint fram mot å ta med slike konstruksjonar i verbartiklar om konstruksjonen ikkje er klart leksikalt etablert med verbet det gjeld.

Som nemnt er det eit poeng i seg sjølv at konstruksjonane i eit språk blir kartlagde og dokumenterte. Ei ordbok som NO – kombinert med eit fyldig korpus – dannar eit godt utgangspunkt for å hente inn og systematisere kunnskap om form, tyding, utbreiing og produktivitet også når det gjeld etter måten skjematisk konstruksjonar som dei eg har kasta lys på her. Ein slik kunnskapsbase ( gjerne kalla *konstruktikon*, jf. Lyngfelt & Sköldberg 2013) kan i sin tur koplast til andre ressursar og mellom anna bli ei hjelp for leksikografar til å gjere konstruksjonane meir synlege og gje dei ei meir konsekvent behandling i ordbøkene.

## Litteratur

### Ordbøker

- NO = *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. 1966–. Oslo: Det Norske Samlaget.  
<<http://no2014.uio.no>> (oktober 2013).
- NO1000 = Tor Guttu (red.) (2005): *Norsk Ordbok med 1000 illustrasjoner*. Oslo: Kunnskapsforlaget.

### Annan litteratur

- Croft, William (2005): Logical and typological arguments for Radical Construction Grammar. I: Jan-Ola Östman & Mirjam Fried (red.). *Construction Grammars. Cognitive grounding and theoretical extensions*. Amsterdam/Philadelphia: John Benjamins, 273–314.
- Farø, Ken & Henrik Lorentzen (2009): De oversete og mishandlede ordforbindelser – hvilke, hvor og hvorfor?. I: *LexicoNordica 16*, 75–101.
- Fillmore, Charles J. (1994): The Hard Road from Verbs to Nouns. I: Matthew Y. Chen, Ovid J.L. Tzeng & William S-Y. Wang (red.). *In Honor of William S-Y. Wang: Interdisciplinary Studies on Language and Language Change*. Taipei: Pyramid Press, 105–129.
- Goldberg, Adele (2006): *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Jansson, Håkan (2005): *Har du ölat dig odödlig? En undersök-*

BERG-OLSEN

*ning av resultativkonstruktioner i svenskan*. Specialarbete, nordiska språk fördjupningskurs 2, Göteborgs universitet. <<http://hdl.handle.net/2077/19000>> (oktober 2013)

Lyngfelt, Benjamin & Emma Sköldberg (2013): Ett svenskt konstruktikon – ett konstruktionsgrammatisk perspektiv på lexikografi. I: *LexicoNordica 20*, 75–91.

NRG = Jan Terje Faarlund, Kjell Ivar Vannebo & Svein Lie (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Seland, Grete (2001): *The Norwegian Reflexive Caused Motion Construction. A Construction Grammar Approach*. Oslo: G. Seland.

Sveen, Andreas (2002): Konstruksjonens betydning for betydningen: å verbe seg til MÅL. I: *Norsk Lingvistisk Tidsskrift 20*, 3–26.

Sturla Berg-Olsen

redaktør, dr.art.

Norsk Ordbok 2014

Institutt for lingvistiske og nordiske studium

Universitetet i Oslo

Postboks 1021 Blindern

NO-0315 Oslo

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

# Nydigitalisering av SAOB

*Erik Bäckerud*

The Swedish Academy Dictionary, SAOB, has started a project with the aim to digitize all volumes of SAOB again to improve the quality of the text over the previous effort. The new version will first of all be correct as regards the actual characters in the text, also the typography will be identified correctly in most cases. In the second part of the project we attempt to automatically identify structures in the articles such as title, part of speech, etymologi, division of definitions and so on.

## 1. Syfte och bakgrund

### 1.1. Syfte

Digitaliseringen av SAOB är tänkt att tjäna två syften. Det första är för att kunna göra en version tillgänglig på internet som uppfyller högt ställda krav på korrekthet och läsbarhet och det andra syftet är att skapa en korrekt och strukturerad text som grund för det fortsatta arbetet med nästa upplaga av SAOB.

## 1.2. Bakgrund

Hittills har 36 band av SAOB tryckts omfattande A t.o.m. UT-SUDDA. Det första bandet kom ut 1898 och band 36 trycktes 2012. Från och med band 32 (1993) skrivs ordbokstexten på dator så förlagan till denna text finns i digital form. De första 31 banden förelåg däremot i traditionell tryckt form och måste digitaliseras för att kunna användas i en elektronisk publicering.

Varje band av SAOB innehåller drygt fem miljoner tecken, vi har alltså omkring 200 miljoner tecken ännu så länge. Den färdiga SAOB beräknas bestå av 38 band.

## 1.3. Tidigare digitaliseringsprojekt

SAOB-redaktionen har tittat på några andra projekt innan vi satte igång att nydigitalisera SAOB. Främst har vi sett på *Deutsches Wörterbuch* von Jacob Grimm und Wilhelm Grimm och på *Ordbog over det danske Sprog*. Båda dessa verk har digitaliserats med hjälp av Kompetenzzentrum vid universitetet i Trier.

### 1.3.1. OSA-projektet

Texten till de första banden av SAOB har digitaliserats vid ett tidigare tillfälle i det så kallade OSA-projektet (OSA = Om svar anhålles) som genomfördes vid Göteborgs universitet 1982–1996 (OSA 1996). SAOB-redaktionen ansåg dock att det behövdes bättre noggrannhet på texten, både vad gäller tecken-



rätthet och typografi, för våra planer. Det bedömdes att det snabbaste och billigaste sättet att få en teckenrätt digital version av texten var genom nydigitalisering hellre än att försöka korrigera den gamla texten.

## 2. Digitaliseringen

I februari 2011 gjorde en delegation från SAOB ett besök hos Kompetenzzentrum i Trier. Det beslöts då att Kompetenzzentrum skulle få uppdraget att digitalisera de 31 första banden av SAOB. Det beräknades ta två år att genomföra hela projektet. Senare reviderades leveranstiden till slutet av november 2013 för de sista banden. I själva verket levererades de sista sex banden redan den 3 september, tre månader tidigare än beräknat. Således har vi nu tillgång till SAOBs alla 36 band i digitalt format.

### 2.1. Processen

En utmaning för den som digitaliserar en text som SAOB är de många olika stilarna och storlekarna på text som förekommer. Det finns ett tiotal olika som t.ex. stor och liten rubrikstil (fet), stor och liten antikva och text som är kursiverad, spärrad eller både och. Dessutom förekommer förutom de vanliga latinska

skrivtecknen även grekiska, runskrift och många andra speciella tecken.

Vi är därför glada över den stora noggrannhet som digitaliseringsprojektet uppnått. Uppskattningsvis ett felaktigt tecken per 30.000. Även vad gäller identifiering av de olika sticlarna och graderna har god noggrannhet uppnåtts.

Här är ett exempel på hur en artikel kan se ut i den tryckta versionen av SAOB från 1903.

**AUTOTYPOGRAFI** *au<sup>1</sup>totyp<sup>1</sup>ografi<sup>4</sup>* l. -tå- l. -tō-, l. -ty<sup>1</sup>p-, äfv. -på- l. -pō-, äfv. *āw<sup>1</sup>-*, äfv. <sup>010104</sup>, äfv. **AVTOTYPOGRAFI** *av<sup>1</sup>-*, l. <sup>01-</sup>, r.; best. -en, äfv. -n. [jfr t. *autotypographie*, eng. *autotypography*, af gr. *αὐτός* o. *τύπος* (se **AUTOTYPI**) samt *γράφειν* (jfr **AUTOGRAF**)] (förr) reproduktion af manuskript, handteckningar o. d. gm (det å ett gelatinskikt med bläck af visst slag, tusch osv. tecknade) originalets öfverförande på metallklisché o. anv. af tryckpress; jfr **ZINKOGRAFI**. *NF* 19 (1895).

Figur 1: Inskannad artikel från SAOB.

När texten digitaliserats och kontrollerats i Trier levereras den till redaktionen i ett format som kallas TUSTEP (TUEbingen System of TExt Processing tools). TUSTEP är det format som används internt på Kompetenzzentrum. Artikeln ovan ser då ut så här när vi får den:

++<P> <A+1>**AUTOTYPOGRAFI**</A+1> <sup  
char="# (GES) ">au</sup>1totyplografi4 l. -t#;oa-  
l. -t#P+^W#P-- , l. -tylp-, <A-1>äfv.</A-1> -p#;oa-  
l. -p#P+^W#P-- , <A-1>äfv.</A-1> <sup  
char="# (GES) ">#P+a^W#P-</sup>1-, <A-1>äfv.  
010104, äfv. **AVTOTYPOGRAFI**</A-1> av1-, <A-1>l. 01--  
, r.; best. -en,  
äfv. -n. [jfr t. autotypographie, eng.  
autotypography,  
af gr. #G+% )ayt%/ow#G- o. #G+t%/ypow#G- (se  
**AUTOTYPI**) samt #G+gr%/afein#G-  
(jfr **AUTOGRAF**)]</A-1> (förr) reproduktion af  
manuskript,  
handteckningar o. d. gm (det #;oa ett gelatinskikt  
med bläck af visst slag, tusch osv. tecknade)  
originalets öfverförande p#;oa metallklisché o.  
anv.  
af tryckpress; jfr <A-1>**ZINKOGRAFI**. NF 19  
(1895).</A-1></P>

Figur 2: Artikeln AUTOTYPOGRAFI i TUSTEP-format.

Texten i TUSTEP-format omvandlas därefter till XML enligt ett schema som beskriver all nödvändig typografisk information. Samma XML-schema används för alla band av SAOB. Nedan visas hur början av ovanstående artikel kan se ut i XML:

```
- <STYCKE art="AUTOTYPOGRAFI" hänvisning="false">
<StorRubrik id="A2573_146156">AUTOTYPOGRAFI</StorRubrik>
<b id="A2573_146157" />
<StorKursiv id="A2573_146158">au</StorKursiv>
<StorAntikva id="A2573_146159" rend="upphöjd">1</StorAntikva>
<StorKursiv id="A2573_146160">totyp</StorKursiv>
<StorAntikva id="A2573_146161" rend="upphöjd">1</StorAntikva>
<StorKursiv id="A2573_146162">ogradi</StorKursiv>
<StorAntikva id="A2573_146163" rend="upphöjd">4</StorAntikva>
<b id="A2573_146164" />
<StorAntikva id="A2573_146165">l.</StorAntikva>
<b id="A2573_146166" />
<StorKursiv id="A2573_146167">-tå-</StorKursiv>
```

BÄCKERUD

```
<b id="A2573_146168" />
<StorAntikva id="A2573_146169">l.</StorAntikva>
<b id="A2573_146170" />
<StorKursiv id="A2573_146171">-tw-</StorKursiv>
<b id="A2573_146172" />
<StorAntikva id="A2573_146173">l.</StorAntikva>
<b id="A2573_146174" />
<StorKursiv id="A2573_146175">-ty</StorKursiv>
...
```

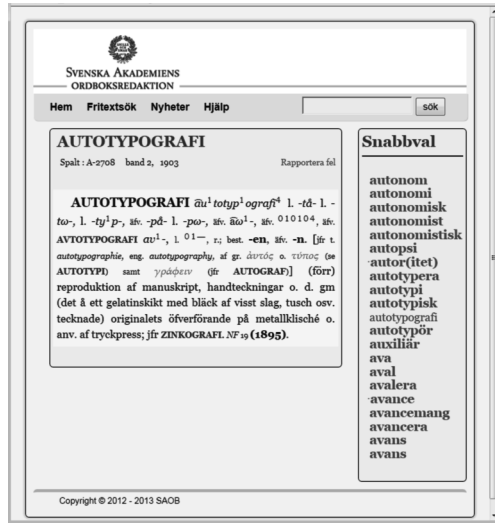
Figur 3: Artikeln som XML (avkortad).

Från denna text i XML kan vi sedan producera artikeltext i många olika format. T.ex som PDF-filer som kan se ut så här:

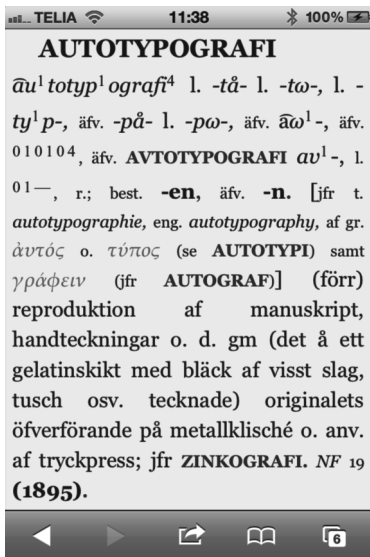
**AUTOTYPOGRAFI** *au<sup>1</sup>totyp<sup>1</sup>ografi<sup>4</sup>* l. -tå- l. -tω-, l. -ty<sup>1</sup>p-, äfv. -på- l. -pω-, äfv. *aw<sup>1</sup>*-, äfv. <sup>010104</sup>, äfv. **AVTOTYPOGRAFI** *av<sup>1</sup>*-, l. <sup>01-</sup>, r.; best. -en, äfv. -n. [jfr t. *autotypographie*, eng. *autotypography*, af gr. *αὐτός* o. *τύπος* (se **AUTOTYPI**) samt *γράφειν* (jfr **AUTOGRAFI**)] (förr) reproduktion af manuskript, handteckningar o. d. gm (det å ett gelatinskikt med bläck af visst slag, tusch osv. tecknade) originalets öfverförande på metallklisché o. anv. af tryckpress; jfr **ZINKOGRAFI**. *NF* 19 (1895).

Figur 4: Texten till artikeln AUTOTYPOGRAFI återskapad från det digitaliserade materialet.

Vi kan också från samma data göra webbsidor som kan visas på datorskärm eller mobila enheter. Så här kan texten se ut i redaktionens webbplats för korrekturläsning:



Figur 5: Artikeln AUTOTYPOGRAFI som webbsida.



Och så här ser samma artikel ut i en mobiltelefon:

Figur 6: Skärmbild från mobiltelefon.

## 2.2. Svårigheter

Även om hela processen att få SAOB digitaliserad har löpt på bättre än förväntat så har det naturligtvis varit några små bekymmer på vägen. Det första problemet vi ställdes inför var att tolka TUSTEP-kodningen som i vissa fall inte är helt konsekvent, och i ett fåtal fall rent felaktig. Tecknen % och # används t.ex. både för att inleda en sekvens som betecknar en symbol som saknas på tangentbordet och för att beteckna sig själva. Detta gör att det är litet svårt att avgöra om det är en speciell kodsekvens som börjar eller om det är ett enstaka tecken. Ett exempel på inkonsekvens är att ligaturen œ kan skrivas både som ”{oe}” och ”#.ö”.

När alla stilar är avkodade och alla specialtecken tolkats korrekt så översätts texten till Unicode. Detta går för det allra mesta bra även om det finns enstaka symboler och kombinationer av diakritiska tecken som inte återfinns i Unicode. Problemet med de saknade diakriterna har vi tills vidare löst genom att inte ta med dem utan endast grundsymbolen. Vi undersöker vidare om det finns bättre lösningar på detta problem.

Vissa tecken har man vid digitaliseringen inte lyckats tolka alls, t.ex. kan processen som används inte skilja på ligaturerna œ och œ (-ae-, -oe-) vilka blir mycket lika i kursiv stil. Här har SAOB-redaktionen fått hjälpa till med att identifiera vilket tecken som avses. Ett annat fall som vållat huvudbry är ord som är avstavade vid radslut. Här är det önskvärt att hålla samman orden så att typografin blir mer tilltalande, men också för att fulltextsökning skall fungera. I de allra flesta fall skall ord som bryts vid radslut hållas samman men ibland som t.ex. i vissa sammansättningar skall bindestrecket behållas. Även här har redaktionen fått rycka in för att kontrollera vilka ord som skall få behålla bindestrecket.

### 3. Indexeringen

Svenska Akademien har även startat ett projekt för indexering av SAOB. Detta projekt går ut på att med datorns hjälp identifiera flera viktiga element i varje artikel. T.ex. identifierar vi uppslagsord med stavningsvarianter, ordklass, etymologi och momentindelning. Till de uppslagsord som identifieras hör även avledningar, sammansättningar och särskilda förbindelser. I detta projekt använder vi den ovan beskrivna nydigitaliserade texten som indata.

Vi har medvetet valt att begränsa indexeringen till ett relativt litet antal strukturer som identifieras i texten. Saker som uttal, hänvisningar inom artiklar m.m. blir inte identifierade i denna första fas. Eftersom vi har en bestämd mängd resurser till vårt förfogande har vi valt att prioritera det nödvändigaste.

Denna struktur kommer att göra det möjligt för oss att göra en betydligt mer läsvänlig webbpresentation än vad som varit möjligt utan strukturering. Den kommer också att utgöra en god grund för ett redigeringsystem att användas vid en uppdatering då första upplagan av SAOB är färdigställd, vilket beräknas ske år 2017.

Arbetet med indexeringen gör redaktionen tillsammans med externa konsulter. Detta arbete är ännu inte avslutat men en del av vad vi åstadkommit hittills syns på nedanstående bild. Här har jag valt att visa uppslagsorden BARB i stället för AUTO-TYPOGRAFI eftersom det finns mer intressant struktur här:

## BÄCKERUD

Sök

Visar 4 av 4 resultat

### BARB

Huvudord  
Substantiv1  
Rapportera fel

<b>Biformer</b>	BARBE
<b>Formparentes</b>	(barbe Lex. Linc. (1640, under barbo o. mulkus), Schroderus Comen. 166 (1640, 1647), Fasciculus (1690) m. d.)
<b>Etymologi</b>	[jfr d. barbe, t. barbe, f., af lat. barbo l. barbuis, bildade af lat. barba, skägg; jfr afv. eng. barbel, fr.
<b>Moment</b>	-
	A. + benämning på två olika fiskarter, utmärkta bl. a.
	01 + <b>1</b> ) den af romarna ss. en läckerhet ansedda
	02 + <b>2</b> ) sötvattensfisken Barbus vulgaris Fleming
<b>Ssgr.</b>	-
	BARB-FISK -
	A. Ssgr: BARB-FISK <sup>3-2</sup> . (föga br.)
	01 + <b>1</b> ) = BARB <b>1</b> . Lex. Linc. (1640, under
	02 + <b>2</b> ) = BARB <b>2</b> . DELEEN (1814, under

### BARB

Huvudord  
Substantiv2  
Rapportera fel

<b>Etymologi</b>	[jfr d. barbe, t. barbe, f., eng. barb, af fr. barbe, af lat. barba (se föreg)]
<b>Moment</b>	+
<b>Ssgr.</b>	-
	BARB-KRAGE Ssgr: BARB-KRAGE <sup>3-20</sup> . N. jövern f. dam. 1855, s. 102.
	BARB-MÖSSA -MÖSSA <sup>20</sup> . N. jövern f. dam. 1858, s. 35.

### BARB

Huvudord  
Substantiv3  
Rapportera fel

<b>Etymologi</b>	[jfr eng. barb, af fr. barbe]
<b>Moment</b>	-
	(†) berbisk häst. Uppå en modig barb .. / Barbaren hotande på slätten tågar neder. ADLERBETH <i>Poet.</i> 2: 259 (1803). — jfr BARBER, sbst. <sup>1</sup> <b>2</b> , särsk. <i>anm.</i>

Figur 7: Exempel på strukturer som identifierats i BARB<sup>1-3</sup>.

## Litteratur

Cederholm, Yvonne, Mickel Grönroos, Susanne Manker & Lena Rogström (2000): *SAOB – en bok för hela folket*.



- Rapport 2 från projektet OSA*. GU-ISS-00-2. Research reports from the department of Swedish, Göteborgs universitet.
- DWB = *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*. <<http://woerterbuchnetz.de/DWB/>> (oktober 2013).
- Lundbladh, Carl-Erik (1992): *Handledning till Svenska Akademiens Ordbok*. Stockholm: Norstedts.
- ODS (1918–2005) = Dahlerup, Verner m.fl.: *Ordbog over det danske sprog inkl. supplement*. København: Det Danske Sprog- og Litteraturselskab. <<http://ordnet.dk/ods>> (oktober 2013).
- OSA (1996) = Sture Allén, Yvonne Cederholm, Sofie Kokkinakis Johansson, Lena Rogström, Rudolf Rydstedt & Lars Svensson (1996): *Om svar anhålles. Rapport från projektet OSA*. GU-ISS-96-4. Research reports from the department of Swedish, Göteborgs universitet.
- SAOB (1893–) = *Ordbok över svenska språket utgiven av Svenska Akademien (Svenska Akademiens ordbok)*, Lund. <<http://g3.spraakdata.gu.se/saob/>> (november 2013).

Erik Bäckerud  
systemansvarig  
Svenska Akademiens ordboksredaktion  
Dalbyvägen 3  
SE-224 60 LUND  
erik.backerud@svenskaakademien.se

# Automatisk identifiering av semantisk förändring med hjälp av lexikala distributionella mönster

*Karin Cavallin*

By looking at word relations in corpora from different time periods we aim to automatically find candidates of semantic variation. We start with the verbal predicate, nominal object relation. This is a syntactic relation that also has an obvious semantic relation. By using techniques from collocational analysis and other statistical measures, we convey differences in distribution that can be indications of semantic variation.

## 1. Introduktion

Syftet med föreliggande studie är att automatiskt identifiera semantisk förändring. Arbetet ska resultera i listor med ord som är goda kandidater för semantisk variation. Dessa listor kan underlätta för lexikografen då ett första urval görs, men det är sedan upp till lexikografen att bedöma och analysera kandidaterna. Med semantisk förändring avser vi *utvidgad* och *inskränkt* betydelse. Det vill säga ord som har en ökad eller minskad möjlig användning, oavsett av vilken anledning. Semantisk förändring kan ha flera olika skäl som exempelvis

metaforbildning och tabu, men vi är här endast intresserade av sådant som torde kunna mätas. Även *neologismer* och *arkaismer* kommer att ingå i denna studie, men mer som en bonus, då de kan ses som att dra en utvidgad eller inskränkt betydelse till sin spets. Dessutom har det visat sig att det som distributionellt beter sig som en arkaism eller neologism i många fall i själva verket är en inskränkt eller utvidgad betydelse.

Semantisk förändring har tills nyligen varit styvmoderligt behandlat vad gäller komputationella metoder. De befintliga ansatser som har applicerats på semantisk förändring har tillämpat metoder som tidigare använts inom komputationell semantik som exempelvis *word sense induction* (Lau et al., 2012) eller *vector space modelling* (Gulordava och Baroni, 2011). Vektormodeller och induktionsmodeller kräver stora lexikala resurser. Samtliga ansatser har hittills baserats på engelska. För engelska förekommer det förhållandevis stora historiska textresurser, men detta är få språk förunnat.

Ansatsen här är för svenska. Som huvudmaterial används Litteraturbankens copyrightfria texter från 1800-talet. Detta material jämförs dels med Parolekorpusen, dels med SUC-romanerna. Även semantiska förändringar i mer modern tid eftersöks. För dessa studier används pressmaterialet Press65 och Press95. Samtliga korpora distribueras av Språkbanken ([spraakbanken.gu.se](http://spraakbanken.gu.se)).

SUC-romanerna är en korpus som består av 58 skönlitterära böcker utgivna 1990-1994. Press65 och Press95 är korpora som består av pressmaterial. Parolekorpusen är en någorlunda balanserad korpus. Vi har dock valt att endast använda det material som är skönlitteratur, och viss presstext.

För att kringgå behovet av enorma korpora har vi valt en syntaktisk relation som även har en stark semantisk koppling:

relationen mellan predikat (verb) och deras nominala argument. För att konformera dessa data används endast nominalfrasens huvudord. Dessa så kallade *verb-objektpar* samlas i *lexikala mängder* (lexical sets). De lexikala mängderna används för att jämföra distributionella mönster.

Varje enskilt verb-objektpar tilldelas ett *log-likelihood*-värde (Dunning, 1993) som visar hur starkt associerat ett verb-objektpar är i jämförelse med de andra verb-objektparen. Detta används för att rangordna och sedan jämföra rangordningen i de lexikala mängderna.

Man kan även titta på skillnaden i hur många *olika* verbtyper ett substantiv förekommer som objekt till, respektive hur många olika substantivtyper verben väljer. Med typer avses distinkta varianter av verb, respektive substantiv, och hänsyn tas inte till frekvens.

Observationerna av distributionella skillnader kan man utnyttja och kvantifiera, och därigenom få listor på ord som är kandidater för semantisk förändring.

Den metod vi utvecklar här är i princip språkoberoende, men kräver att språket i fråga ska ha urskiljbara satsdelar, som utöver den syntaktiska relationen även har ett semantiskt beroende mellan sig.

Upplägget för denna artikel kommer vara att vi presenterar våra korpora, och sedan hur våra data ser ut. Därpå visar vi vilka metoder vi använder, följt av de resultat man får ut. Slutligen ges en sammanfattande diskussion av resultaten.

## 2. Korpora

De korpora vi använder oss av kommer alla från den svenska Språkbanken ([spraakbanken.gu.se](http://spraakbanken.gu.se)). Språkbanken tillhandahåller en stor mängd korpora. För detta projekt har fem korpora använts.

Parolekorpusen, SUC-romanerna, Press65 och Press95 har använts utan större manuell insats. Bortsett från att vi för Parolekorpusen har valt ut en delmängd av den. Vill man inte använda det givna gränssnittet Korp (<http://spraakbanken.gu.se/korp/>) kan man själv ladda ner korpusarna från Språkbankens hemsida (<http://spraakbanken.gu.se/swc/resurser/corpus>) och även delvis välja ut specifika delar ur korpusarna som exempelvis litterärt material, eller korpustexter från specifika datum. Den femte korpusen som använts, Litteraturbanken, har vi fått hantera annorlunda.

Litteraturbankens syfte är att vara en fri kulturhistorisk och litterär resurs för forskning, undervisning och folkbildning. Den är till för alla: forskare, lärare, studerande och litterärt allmänintresserade. Huvuduppgiften är att samla in och digitalisera skönlitteratur och viktigare humaniora samt tillgängliggöra materialet på sådant sätt att det blir möjligt för användare att arbeta med det. (<http://litteraturbanken.se>)

Numera är Litteraturbanken sökbar under det gängse Korpgränssnittet (<http://spraakbanken.gu.se/korp/>), men har tidigare inte distribuerats i samma form som övriga. Vi har därför i detta arbete en version av Litteraturbanken som skiljer

sig åt från den distribution som nu finns tillgänglig i Korp.

Litteraturbanken innehåller en stor mängd historiskt material, från så tidigt som 1200-talet, till material skrivet under senare delen av 1900-talet. För denna studie har det material som är från 1800-talet valts ut. Detta gjordes för att ha en någorlunda avgränsad epok att jämföra med, samt för att de språkteknologiska verktyg som behövs för att göra Litteraturbanken tillgänglig för språkteknologiska studier skulle kunna vara tillämpbara. De flesta språkteknologiska verktyg är skapade för att användas på modernt språk.

För att Litteraturbanken (med Litteraturbanken menas nu subkorporuserna med 1800-talsmaterial) skulle vara jämförbar med de övriga korpora från Språkbanken användes samma, eller snarlika, verktyg för ordklasstagning och parsning. Parsningen gjordes med MaltParser (Nivre et al., 2006) som är samma som används inom Språkbanken. För ordklasstagningen var vi tvungna att frånga HunPos som används inom Språkbanken och istället förlita oss på Trigrams and Tags tagger (T'n'T) (Brants, 2000). HunPos är *open source* och en reimplementation av TnT, som primärt endast skiljer sig åt vid hanteringen av okända ord, där T'n'T presterar bättre (Halaczy et al. 2007). Historiskt material har givetvis en större mängd, för taggaren, okända ord än ett modernare material, då taggaren är tränad på modernt material. För lemmatisering har vi fått materialet delvis lemmatiserat med hjälp av Språkbankens resurser, men även fått komplettera en stor mängd olemmatiserade ord manuellt.<sup>1</sup>

Storleken på korpusarna skiljer sig åt. Istället för att sampla korpusarna i lika storlek har vi valt att utvinna så mycket infor-

---

1. Det manuellt lemmatiserade materialet är unikt för detta projekt, men kan komma att infogas i Språkbankens distribution av Litteraturbanken framöver.

mation som möjligt ur de korpora vi har till vårt förfogande. Därför har vi normaliserat materialet. För Parolekorpusen har vi emellertid valt ut de delkorpusar som innehåller så litteraturlikt material som möjligt.

## 2.1. Verb-objektpar och lexikala mängder

En viktig ansats i detta projekt är att använda den uppenbara semantiska information som redan finns mellan vissa satsdelar. Idag utförs den mesta språkteknologin med lexikalsemantisk inriktning med hjälp av metoder som inte har någon lingvistisk grund (men som trots avsaknaden av lingvistisk grund ger ganska goda resultat). I denna studie har vi valt att utnyttja den semantiska relationen mellan verb (verbalt predikat) och objekt (det nominala huvudet i objektsfrasen). Vi kommer härnäst kalla dessa för verb och objekt, alternativt verb-objektpar. Skillnaden i distributionen mellan verb och objekt mellan olika tidsperioder torde visa på semantiska förändringar.

Verb-objektparen kan förhållandevis enkelt extraheras ur korporan då korporan är parsade, det vill säga uppmärksatta med satsdelsinformation.

Verb-objektparen samlar vi ihop, uppdelade på ursprung från respektive korpora. Därefter hanterar vi dem på olika sätt. Vi skapar lexikala mängder, vilket innebär att vi sorterar verben utifrån vilka objekt verben samförekommer med, och objekten utifrån vilka verb de samförekommer med. Nedan visas exempel på lexikala mängder, som utgår från en konstruerad text.

Det är sommar. Lisa är svettig. Hon köper en glass. Hon dricker även en svalkande läsk. Men den var inte särskilt god så hon slängde läsk. I eftermiddag tar hon bilen till landet. Hennes bror har köpt ett hus. Hon har köpt en ryggsäck till resan. Hon köpte även vin och tre liter vatten och en glass till sin bror. Ikväll ska de dricka vatten, vin och öl. Barnen får dricka mjölk. Hunden tog en promenad och kastade vatten innan de begav sig iväg.

Utifrån texten ovan extraheras lemmatiserade verb och objekt. Den extraherade datamängden redovisas nedan:

{köpa glass, dricka läsk, slänga läsk, ta bil, köpa hus, köpa ryggsäck, köpa vin, köpa vatten, köpa glass, dricka vatten, dricka vin, dricka öl, dricka mjölk, ta promenad, kasta vatten}

Från datamängden konstrueras sedan lexikala mängder som konstitueras, antingen utifrån det gemensamma verbet för objekten, eller utifrån det gemensamma objektet för verben.

köpa: glass, ryggsäck, hus, vatten

dricka: läsk, öl, vatten, vin, mjölk

ta: bil, promenad

glass: köpa

läsk: dricka, slänga

öl: dricka

vatten: dricka, kasta

bil: ta

promenad: ta



Från och med nu och fortlöpande i texten kommer vi inte längre att förhålla oss till korpusar, utan endast till datamängden som består av verb-objektpar. Datamängden kan vi även kalla lexikala mängder, särskilt då vi syftar på specifika lexikala mängder, men även som ett överbegrepp för innehållet i datamängden.

### 3. Metoder

Det finns ett antal grundantaganden som vi har gjort inför datahanteringen:

1. Det finns en semantisk, såväl som en syntaktisk relation, mellan ett verb och dess objekt
2. Skillnad i frekvens kan vara en första indikator på att ett ords betydelse skiljer sig åt mellan olika korpora.
3. Närheten, associationen, mellan verbet och objektet kan mätas med hjälp av metoder använda inom kollokationsforskning.
4. Rangordningen inom två motsvarande lexikala mängder kan påvisa en semantisk variation.
5. Antalet olika objekt- eller verbtyper ett verb, respektive objekt, förekommer med, kan vara en indikation på semantisk variation.

En viktig sak att peka på är att våra data är normaliserade. Detta görs för att de lexikala mängderna och datamängderna trots sin

olika storlek ändå ska vara jämförbara. Punkt 2 och 5 förhåller sig till normaliserade värden. I punkt 2 är normaliseringen gjord med avseende på antal *tokens* i den specifika lexikala mängden över antalet tokens i datamängden den lexikala mängden utgår från. I punkt 5 normaliserar vi antal *typer* i en specifik lexikal mängd över antalet tokens i den datamängd den lexikala mängder utgår från.

För att få fram kandidater för semantisk variation väljer vi att kombinera dessa antaganden och vaskar fram de verb eller objekt som uppfyller samtliga kriterier i punkt 2-5.

Punkt 1 är inte en metod, utan snarare en teoretisk grund för att övriga kriterier/punkter ska kunna vara behjälpliga för att hitta semantisk variation.

Punkt 2. Vi extraherar de verb eller substantiv som skiljer sig åt med avseende på frekvens, dvs. att ett ord förekommer betydligt fler gånger i en datamängd än i den andra

Det är viktigt att poängtera att vi här endast tittar på två datamängder i taget. Vi jämför Litteraturbanken med Parole, Litteraturbanken med Stor-SUC, samt Press65 och Press95 med varandra.

Punkt 3. Med hjälp av ett log-likelihood värde (Dunning, 1993) mäts associativiteten mellan ett verb och objekt och deras relation till övriga verb och objekt i datamängden. Par som är mycket närmare associerade i en datamängd i jämförelse med den andra datamängden ger en indikation på semantisk variation.

Punkt 4. Spearman ranking coefficient (Spearman, 1904) är ett mått som här mäter skillnader i rangordningen av verb-objektparen mellan två motsvarande lexikala mängder. Spearman ranking coefficient avger ett värde mellan -1 och 1. Ju lägre värdet är, desto mer olika är de motsvarande lexikala mängderna. De lexikala mängder som skiljer sig åt mest ger följaktligen en indikation på semantisk variation.

Punkt 5. En utvidgad eller inskränkt betydelse borde praktiskt påvisas med hjälp av att det förekommer andra typer, eller en större mängd typer i den ena än i den andra datamängden.

## 4. Resultat

Genom att kombinera metoderna listade ovan kan man identifiera distributionell variation som kan vara indikation på semantisk variation. Tanken är att dylika listor, som i tabellen nedan (Tabell 1), kan användas av någon som vill kartlägga förändringar eller variationer inom förslagsvis en specifik domän, jämföra terminologin i två närliggande discipliner, eller för att hitta generella förändringar.

I tabellen nedan väljer vi att kalla kategorierna för «Utvidgad bet.», «Inskränkt bet.», «Arkaism» och «Neologism». Orden är endast *kandidater* för dessa kategorier, det är alltså en *potentiell* utvidgning, arkaism och så vidare. Vi kommer här att redovisa en delmängd av de kandidater vi fått från en studie av Litteraturbanken och Stor-SUC och föra ett kort resonemang om intressanta fall i tabellen.

I punktlistan nedan sammanfattar vi de kriterier som spelar in för att få fram respektive variationstyp.

- «Utvidgad betydelse» ← utökad typ- och tokenförekomst, samt starkt varierad rangordning
- «Inskränkt betydelse» ← minskad typ- och tokenförekomst, samt starkt varierad rangordning.

- «Arkaism» ← minimal typ- och tokenförekomst i den ena datamängden (den yngre) men inte i den andra (den äldre).
- «Neologism» ← minimal typ- och tokenförekomst i den ena datamängden (den äldre) men inte i den andra (den yngre).

De ord som är homografer över ordklasser är markerade med frågetecken, «?», ord som är uppenbara fel vad gäller taggning är markerade med asterisk, «\*».

Utvidgad bet.	Inskränkt bet.	Arkaism	Neologism
?rätt	?för	konung	?far
råd	hjärta	anlete	statsminister
problem	natur	hop	skit
nummer	hustru	*själv	*sig
kurs	?vår	grevinna	Kollega
	syster	sällhet	Grej
	händelse	majestät	bomb
	?karl	Fröjd	statsråd
	?tro	tycke	värdering
	karaktär	skald	styck
	vagn	?före	relation

Tabell 1: Tabell över de ord som skiljer sig åt mest i Litteraturbanken och Stor-SUC m.a.p. de kriterier som beaktats.

De ord som står i kolumnen «Utvidgad bet.» är nog svårast att se om de är dugliga kandidater. Tidigare studier (Cavallin, 2012) har visat att *problem* är ett ord som har fått utvidgad betydelse, från ett mer abstrakt problem, till mer vardagliga och praktiska bekymmer. Övriga ord kräver mer ingående studier för att avgöra om de är dugliga kandidater. Under kolumnen «Inskränkt bet.» hittar vi fler kandidater som vi, med en generell språkkänsla för svenska språket, kan se som eventuellt goda kandidater. *Hustru* och *karl* är ord som drar åt det arkaiska, och i vissa grupper även behäftade med en anstrykning av negativ laddning. *Tro* är kanske inte förändrat i betydelse, men ett betydligt mindre använt ord i det numera sekulariserade Sverige. De ord som är kandidater under «Arkaism» är samtliga, utom möjligen *tycke* och *hop*, ord som har en arkaisk klang. Under «Neologism» ser vi att det knappast är helt nya ord, men det är ord som betecknar modernare företeelser.

Det måste nämnas att det är oändligt svårt att dra gränsen mellan samhällelig och semantisk förändring. De är givetvis inte oberoende av varandra, men något kan helt enkelt vara «inne», utan att ha förändrat betydelse. Aktuella och dramatiska händelser kan exempelvis påverka framförallt tidningstext, utan att ordet i sig har ändrat betydelse. Som exempel brukar *tsunami* lyftas fram. Få svenskar visste vad ordet betydde, eller ens att det förekom i svenskan förrän den stora tsunamin 2004, där bland tusentals andra, ett stort antal svenska turister förolyckades.

Utöver de mer lingvistiska vinningarna kan utdata i form baserad på syntaktisk och semantisk relation bidra till att exempelvis korrigera taggnings- och parsningsverktyg, och därigenom förbättra materialen.

De lexikala mängderna fungerar också som en minimal kontext, som kan göra det enklare att snabbt få överblick över vissa lingvistiska fenomen.

CAVALLIN

Dock måste alla resultat kritiskt granskas av en expert. Ansatsen här ger en första och ganska tydlig bild av vilka ord som uppvisar semantisk variation i de korpora man är intresserad av att studera.

## Litteratur

### Länkar

Litteraturbanken: <http://litteraturbanken.se/>  
Språkbankens hemsida för nedladdningsbara resurser:  
<http://spraakbanken.gu.se/swe/resurser/corpus>  
Korp: <http://spraakbanken.gu.se/korp/>

### Ordböcker

SAOL10 = *Svenska akademiens ordlista över svenska språket* (10:e utg.) (1976). Svenska Akademien.  
SAOL13 = *Svenska akademiens ordlista över svenska språket* (13:e utg.) (2006). Svenska Akademien.

### Annan litteratur

Brants, Thorsten (2000): TnT–A Statistical Part-of-Speech  
Tagger. I: *Proceedings of the Sixth Applied Natural Language  
Processing Conference ANLP-2000*. Seattle, WA, 224-231.

- Cavallin, Karin (2012): Automatic extraction of potential examples of semantic change using lexical sets. I: Jeremy Jancsary (red.): *Proceedings of KONVENS 2012 LThist 2012 Workshop*. Wien: ÖGAI., 370–377.
- Dunning, T. (1993): Accurate methods for the statistics of surprise and coincidence. I: *Computational Linguistics 19 (1)*, 61–74.
- Gulordava, Kristina & Marco Baroni (2011): A distributional similarity approach to the detection of semantic change in the google books ngram corpus. I: S. Pado & Y. Peirsman (red.) *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics, July 2011*. Edinburgh: Association for Computational Linguistics, 67–71.
- Halácsy, Péter, András Kornai & Csaba Oravecz (2007): Hunpos: an open source trigram tagger. I: *Proceedings of the 45th annual meeting of the ACL on Interactive Poster and Demonstration sessions, ACL '07*. Stroudsburg, PA: Association for Computational Linguistics, 209–212.
- Lau, Jey Han, Paul Cook, Diana McCarthy, David Newman, & Timothy Baldwin (2012): Word Sense Induction for Novel Sense Detection. I: M. Butt et. al. (red.) *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012), Avignon, France, April 2012*:. Stroudsburg, PA: Association for Computational Linguistics, 591–601.
- Nivre, J., J. Hall & J. Nilsson (2006): MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. I: N. Calzolari et. al. (red.) *Proceedings of the fifth international conference on Language Resources and Evaluation. (LREC2006), May 24-26, 2006, Genoa, Italy*, 2216–2219.
- Spearman, Charles (1904): The Proof and Measurement of

CAVALLIN

Association between Two Things. I: *The American Journal of Psychology* 15 (1)(January), 72–101.

Karin Cavallin  
forskarstuderande, fil.mag.  
Institutionen för filosofi, lingvistik och vetenskapsteori  
Göteborgs universitet  
Box 200  
405 30 Göteborg  
karin.cavallin@gu.se



## ***Nordisk miniordbok – et kritisk blikk***

*Ruth Vatvedt Fjeld og Janne Bondi Johannessen*

This paper discusses a new online multilingual dictionary between Norwegian, Swedish and Danish aimed at young people, the NMO. It is discussed w.r.t. lemma selection, types of definitions, and user interface. It is concluded that the dictionary fails to cover central lemmas, especially those coined as false friends and also those that are part of young people's vocabulary. The definitions are given in unsystematic ways both as regards the different languages and as regards the individual headwords, where some are too general and some are too specific or describe peripheral meanings.

### **1. Innledning<sup>1</sup>**

*Nordisk miniordbok* ble lansert på Internett 1.2.2013 av språkrådene i de nordiske landene. Dette er en flerspråklig ordbok mellom de nordiske språkene, og målgruppen er barn og unge.

---

1. Vi vil gjerne takke Per Erik Solberg og Pål Kristian Eriksen for noen av ordbokssøkene vi omtaler her. Vi vil også gjerne takke publikum på Nordisk konferanse for leksikografi 2013 for fine innspill og kommentarer.

Det er en begivenhet at vi har fått en gratis tilgjengelig nettordbok beregnet på unge skandinaver. Likevel, eller kanskje derfor, bør en slik bok tilfredsstillende grunnleggende kvalitetskrav på mange nivåer. Vi mener dessverre at denne ordboka ikke er god nok verken for målgruppen eller andre. Dette skal vi grunnngi nærmere i artikkelen.

Vi er klar over at det er vanskelig å lage en flerspråklig ordbok, og en firespråklig ordbok er selvsagt spesielt utfordrende. Likevel er vi kritiske til mange av de valg redaksjonen har tatt. Sentrale leksikografiske prinsipper er ikke ivaretatt i ordboka, og vi ønsker med denne gjennomgangen å peke på bedre og godt innarbeidede løsninger som kunne vært valgt.

Avsnitt 2 gir litt kort informasjon om ordboka, avsnitt 3 diskuterer lemmaseleksjon, inkludert falske venner og målet om et basisordforråd, mens avsnitt 4 diskuterer definisjonene, som i noen tilfeller er for spesifikke, i andre for generelle, og i atter andre beskriver perifere betydninger. Avsnitt 5 nevner kort noen andre sider ved ordboka, nemlig brukergrensesnitt og uttaleinformasjon, mens avsnitt 6 konkluderer, og avsnitt 7 gir informasjon om annen litteratur og ordlister.

## **2. Om *Nordisk miniordbok***

*Nordisk miniordbok* (NMO) er utgitt av språkrådene i Norge, Sverige og Danmark, slik det er oppgitt på hjemmesidene<sup>2</sup>. Prosjektleder og norsk redaktør er Torbjørg Breivik, mens Pia Jarvad er dansk hovedredaktør og Birgitta Lindgren er svensk redaktør.



Figur 1: Nordisk miniordbok.

NMO er en gratis nordisk nettdordbok for barn og unge. Både nettsiden og artikkelen som beskriver boka, Breivik (2012), legger vekt på at det er et mål for ordboka at den skal støtte opp om nabospråksforståelsen og -undervisninga i Danmark, Norge og Sverige. Nettsiden viser at ordboka skal vise hvordan mange ord betyr det samme eller nesten det samme på de ulike språkene, mens noen ord nettopp ikke betyr det samme. Dette er punkter vi kommer tilbake til utover i artikkelen.

Ordboka har 3200 oppslagsord på hvert språk, dvs. norsk bokmål (NB) og nynorsk (NN), dansk (DK) og svensk (SV). Grensesnittet har også med en begrenset bruk av finsk, samisk, grønlandsk, færøysk og islandsk, i den forstand at enkelte av de skandinaviske hovedoppslagsordene også er oversatt til disse, men ikke motsatt vei. Vi ser i det videre kun på de fire skandi-

- 
2. <http://miniordbok.org/> eventuelt <http://www.sprakradet.no/Toppmeny/Aktuelt/Nordisk-miniordbok-pa-nett/>

naviske skriftspråkene. NMO er finansiert av nordiske og nasjonale offentlige midler.

### 3. Lemmaseleksjon

Ordutvalget skal, ifølge Breivik (2012:150), dekke et basisordforråd, inkludert barnerelevante emner. Det skal videre gjøre rede for falske venner (jf. kap. 3.2.1), og det skal bøte på manglende ekvivalenter. I denne delen vil vi vise at NMO ikke gjør det den har satt seg fore å gjøre.

#### 3.1 Basisordforråd

Akkurat hvordan basisordforrådet er valgt ut, er ikke klart. Det er tydeligvis ikke brukt noen frekvensordliste, siden sentrale og høyfrekvente adverb som *kanskje* og *egentlig* mangler, i likhet med verbene *ta* og *huske* (i betydningen *minnes*). Men det finnes mange vanlige og høyfrekvente ord, som *hus*, *kaffe* og *vann* og mange lavfrekvente, som *meieri*, *mandarin* og *melon*. På den annen side kan man spørre hva som er poenget med å ta med ord som har lik form og betyr det samme i alle språkene. Det virker svært pussig at helt eller nesten like og vanlige ord som *kaffe*, *hus* og *vann* er tatt med i en slik definisjonsordbok. Kan det være skandinaviske barn som lurte på betydningen av *kaffe* i de andre språkene? Det mener tydeligvis redaktørene:

(1)

**NB kaffe:** kaffe er en varm drikk

**NN kaffi:** kaffi er ein varm drikk

**DK kaffe:** kaffe er en mørkebrun, lidt bitter, varm drik

**SV kaffe:** kaffe är en dryck

Om det derimot hadde vært tatt med vanlige uttrykksformer, som det svenske slanguttrykket *fika*, hadde det selvsagt gitt mye viktigere hjelp. Dette oppslagsordet er med, men bare med svensk inngang *fika*, ikke sammen med oppslagsordet *kaffe*. Ordboka er derfor bare å regne som resepsjonsordbok, utviklet for å forstå tekst på de andre språkene.

(2)

**SV:** att **fika** är att dricka kaffe

**NB:** vi drikker kaffe og hygger oss

**NN:** vi drikk kaffi og hyggjer oss

**DK:** vi drikker kaffe og hygger os

Problemet med et slikt ordutvalg er at ordboka blir lite informativ for de språkene som nå er dokumentert, da ordene stort sett er like eller nesten like på alle språkene, og til liten hjelp for nabo-språksforståelsen. Man kunne heller sørget for å få med ord og uttrykk som ikke er forståelige i nabospråkene. Hadde man tenkt tematisk ved lemmautvalget, kunne man også fått med det vanlige svenske *macka* ('smørbrød/matpakke'), men dette ordet er ikke med. Her kan man ha kommunikasjonsproblemer i Norden.

### 3.1.1. Barnerelevante emner

I tillegg er det satt opp som mål både i Breivik (2012) og på

nettsiden at det legges vekt på barnerelaterte emner som skolehverdag og fritid. Da er det påfallende at ord som gjerne kan trenge en ekstra forklaring hos nabospråksbarna, ikke finnes i NMO, som *sosiallærer, helsesøster, kontaktlærer, storefri, pennal, basket, ishockey, bandy, lærebok, speider, skolevei*. Til gjengjeld finnes norsk *korps* og *friminutt*, men da med forklaringer som ikke beskriver den mest typiske bruken, som nettopp er den som er knyttet til skole og fritid, som vi skal se senere.

Brevik (2012:150) opplyser at man blant lemmaene også vil finne fiksjonnymer (navn på personer i litteraturen) fra nordisk barnelitteratur. Det er en utmerket idé, dersom man også der tenker på fiksjonnymer som er forskjellige i de aktuelle språkene, som *Donald Duck / Anders And / Kalle Anka*. Men de fins ikke med i den versjonen som nå ligger på Internett. Heller ikke har vi funnet *Bullarebygden/Bakkebygrenda* og mange andre kjente navn i nordisk barnelitteratur. NMO mangler også navn som *Fætter Hidsigprop* og *Fætter Vims*. Slike lemmaer vil være særlig aktuelle dersom barne- og ungdomslitteratur skal leses på originalspråket i nabospråksundervisningen.

Det er ikke gjort rede for hvilket materiale som er brukt for å velge ut lemmaene, om det er laget et korpus av barne- og ungdomsbøker, om man har samlet inn ord og uttrykk blant skoleelever eller har tatt dem ut fra sitt eget hode. Den vanlige metoden i dag er å bygge et korpus, og siden NMO skal dokumentere det sentrale ordforrådet, ville det ikke vært en umulig oppgave å samle inn et korpus i hvert språk som kunne dekket det. Man kunne også gått systematisk og kritisk gjennom ordlister for skoletrinnet, som vanligvis inneholder rundt 10 000 oppslagsord. Her kunne man valgt ut de mest frekvente av disse, f.eks. ved å gjøre en frekvensstudie i norske lærebøker. Leksikografisk bokmålskorpus på 100 mill. løpeord er utmerket organisert for å trekke ut et subkorpus nettopp av slike bøker. I

Norge har vi også flere korpus over elevtekster (f.eks. KAL-korpuset) som kunne gitt hjelp til å finne et mer sentralt ordutvalg som fortjener plass i en ordbok myntet på barn og unge. Vi kan ikke se at lemmautvalget i NMO er gjort med hensyn til barns nabospråksforståelse når det gjelder basisordforråd.

### 3.1.2. Falske venner

Både Breivik (2012:150) og nettsiden understreker behovet for informasjon om falske venner, dvs. ord som er likelydende i de forskjellige språkene, men som har forskjellig betydning eller har spesielle bruksrestriksjoner i de forskjellige språkene. Det er ofte et prinsipp at det er disse ordene som først og fremst velges i ordbøker mellom nabospråk, og det er utarbeidet mange slike, se referanseliste bak. NMO har imidlertid ikke dette som hovedfokus. Ordboka hadde tjent enormt på å orientere seg i eksisterende ordbøker over falske venner og sørge for god dekning av slike. I stedet er deknningen dårlig.

Oppslagsordet SV *rolig* ('morsom') finnes, mens det ikke er noe oppslagsord for NB, NN, DK *rolig* ('stille'). Også SV *artig* ('høflig, veloppdragen') er med, mens NB, NN *artig* ('morsom') ikke er med. SV, DK *rar* ('snill') er oppslagsord, men NB, NN *rar* ('underlig') er det ikke, og *le* (med eller uten lyd, altså med betydningen til norsk 'skratte' eller 'smile', avhengig av språk) er ikke med som oppslagsord på SV, mens det er det på de andre. På dansk er *grine* med, og da definert med 'le'. Norsk *grine* er ikke med, så da får ikke ungdommene vite at det betyr det motsatte på norsk, nemlig lyd man lager når man er trist. Til og med den kjente falske vennen *frokost* blir ufullstendig forklart. Norsk og svensk *frokost/frukost* får den danske ekvivalenten *morgenmad*, men det er ikke oppgitt

samtidig at *frokost* også brukes i dansk med betydningen ‘formiddagsmat’.

Det burde vært en systematisk gjennomgang av disse falske vennene og kontroll på at de blir forklart for brukere uavhengig av hvilket språk som er inngangsspråk. Man kan ikke vente at en svensk bruker skal gå inn på det norske ordet *rar*; han eller hun bør få det som ekvivalent på *konstig*, et oppslagsord som for øvrig ikke finnes. Fatale konsekvenser i sosiale sammenhenger kan manglende kunnskap om ordet *anledning* gi. Det høres helt kjedelig og alminnelig ut, men om en nordmann svarer på en vennlig invitasjon fra en svensk venn at han eller hun *ikke har anledning til å komme* kan knuten heller ikke løses ved hjelp av NMO, siden man da har sagt at man ikke har noen *grunn* til lå komme.

NMO ville tjent på at man vurderte lemmautvalget fra alle de tre kildepråkene og ganske enkelt anga de forskjellige ekvivalentene i stedet for å komme med mer eller mindre likelydende definisjoner (jf. nedenfor). De mange definisjonene gir inntrykk av at betydningen er mer ulik enn den faktisk er.

Substantivet og verbet *bakke* har felles betydning i de tre språkene: substantiv: ‘en skråning i terrenget/jordhøy/væg el mark som lutar’, verb: ‘å rygge med båt’. Men substantivet har et homonym i dansk, som igjen har ekvivalenten ‘brett’ på norsk og ‘bricka’ på svensk. Det er altså bare dansk som har homonyme substantiv av *bakke*. Men den siste betydningen er ikke lemma i den danske ordboka. I prinsippet er det heller ikke det i norsk, der er *brett* kun definert som ‘et flatt trestykke’, som i dansk. Dette *brett* er selvsagt etymologisk beslektet med det typiske kjøkkenutstyret som tilsvarer det svenske *bricka*, men bør regnes som selvstendige lemmaer i moderne norsk. Homonymet *brett* for ‘fold i stoff eller papir’ er totalt utelatt.

Slike glipp i lemmautvalget ville man sluppet om man



hadde en bedre ordboksbasis å gå ut fra, helst et skandinavisk parallellkorpus eller flerspråklige skandinaviske ordlister, eventuelt lemmalister i vanlige skoleordbøker for de tre språkene, noe som ville gitt en mer systematisk oversikt over de vanligste lemmaene i de tre språkene. I og med at vi bare får se de direkte oversettelsene av oppslagsordene og ikke får direkte opplysning om falske venner, går vi glipp av det området der det virkelig kunne være noe å hente for nabospråksforståelsen.

### 3.1.3. Bøte på manglende ord

På nettsiden og i Breivik (2012) hevdes det at det skal bøtes på manglende ord, og de nevner *bestefar* som eksempel på noe som ikke finnes i svensk. Men mens *bestefar* er oppgitt for NB, NN og DK, er det bare en tom rute for svensk, der man burde finne *morfar* og *farfar*. Ordet *bad*, i betydningen baderom, er også et ord som står på tre språk, og med tom rute på svensk. Dette er pussig, siden det ikke er vanskelig å finne ekvivalenter til disse ordene i svensk.

## 3.2. Uvanlige lemmautvalg

Den porselensplaten som gjerne står under en kopp, er angitt som *tefat* i norsk (DK *underkop*, SV *kaffefat*). Det er nok noen dialekter hvor dette ordet brukes i norsk, men ved søk i et korpus (Leksikografisk bokmålskorpus, LBK) gir ordet *tefat* 20 treff, mens alternativet *skål* har hele 1819. Frasen *kopp* og *tefat* har 3 treff, mens *kopp* og *skål* har 44. Da er det ganske tydelig

at man ikke har valgt det mest utbredte lemmaet. Tilsvarende er det for NB valgt lemmaet *dagligstue* (223 treff), og ikke det langt vanligere *stue* (8631).

For en ordbok over basisordforrådet virker det også merkelig å ha med den høflige tiltaleformen *De* som oppslagsord på norsk, den brukes i dag neppe av noen under 80 år.

#### 4. Definisjonssystem og definisjoner

En ordbok bør ha et avklart forhold til hvordan definisjonene skal skrives. Ifølge Breivik (2012:152) følger NMO Cobuild-systemet:

...forklaringene skrives etter co-build-systemet, som betyr at oppslagsordet gjentas i den korte forklaringen som følger.

Nå er det ikke akkurat slik Cobuild-formatet er, jf. *Nordisk leksikografisk ordbok*:

Cobuildformat: definisjonsformat i to deler der lemmaet i den første delen settes inn i en typisk kontekst og i den andre delen får en betydningsbeskrivelse.

Ordet skal altså både ha en betydningsbeskrivelse og vises i en typisk kontekst, her eksemplifisert ved MyCobuild.com:

(3)

**Spoon:** A spoon is an object used for eating, stirring, and ser-

ving food. One end of it is shaped like a shallow bowl and it has a long handle. He stirred his coffee with a spoon.

Når vi så ser på praksis i NMO, er det ingen av definisjonene som tilfredsstillt deres eget mål. Ikke bare det, men de forskjellige redaktørene har tydeligvis valgt helt ulik strategi. Noen har valgt betydningsbeskrivelsen, andre eksemplet:

(4)

**SV glass:** glass är en fryst efterrätt eller godsak

**NB lunsj:** lunsj spiser du midt på dagen

**DK morgenmad:** morgenmad er dagens første måltid

**NN stille-1:** når det er stille, er det ingen lydar

Definisjonene er usystematiske mellom språkene, også der samme ord beskrives. Se for eksempel definisjonsdelen for *gaffel*, der DK bruker genus proximum (altså et overordnet begrep, hyperonym, hvor definiendum inngår som et hyponym), mens SV og NN får klare seg med en eksempelsetning, hvor det er en slags forklaring innebygd. Definisjonene er ikke samordnet verken når det er form eller funksjon som trekkes fram:

(5)

**DK** en gaffel er et redskab til at spise med

**SV** en gaffel äter man med, den har flera spetsar

**NN** ein gaffel blir brukt til å eta med

Når det gjelder selve innholdet i definisjonene, spriker det også svært mye. Noen er svært generelle, og gir knapt noen pekepinn om betydningen:

(6)

**SV frukost:** frukost är det du äter på morgonen

**BM fotball:** fotball er et spill  
**BM is:** is er et godteri  
**NB adel:** adel finnes ikke lenger i Norge  
**NB vær:** vær er hvordan det er ute  
**NN skei:** ei skei er til å eta med  
**NN gaffel:** ein gaffel blir brukt til å eta med

Det er mye man kan spise om morgenen, som egg eller grøt. Det er mange spill, som monopol og ludo. Det er mange godterier, som kamfersukkertøy og karameller. Det er mye som ikke finnes lenger, som dinosaurer og hestedrosjer. Det er mange måter som ting kan være på ute, for eksempel ørken eller park. Og det er mye man kan spise med, som smekke eller med gode venner. Og hva er forskjellen mellom skei og gaffel?

Noen definisjoner er sirkulære:

(7)  
**NN berg:** eit berg er eit lite fjell  
**NN fjell:** eit fjell er eit stort berg

Andre definisjoner er altfor spesifikke:

(8)  
**SV sked:** en sked äter man soppa med  
**SV vardagsrum:** i ett vardagsrum tittar man på tv eller umgåå med varandra  
**NN stove:** ei stove er daglegrommet i huset du bur i

Man spiser vel mye annet enn suppe med en *sked*. Tv kan man gjerne titte på i ulike rom, ikke bare i *vardagsrummet*, dessuten kan man vel også sitte og lese i avisen der. *Stova* er dagligrommet i hus som andre bor i også.

Der ordene har polyseme betydninger, har NMO gjort til dels svært underlige valg. Her er noen eksempler:

(9)

**BM ballett [sic]:** en ballett er en danseforestilling

**BM friminutt:** et friminutt er en pause

**BM korps:** et korps har en felles oppgave, f.eks. musikkorps

*Ballett* er vel først og fremst en danseform som kan bedrives både med og uten forestilling. *Friminutt* er ikke en generell pause, det brukes mellom skoletimer. For de fleste vil *korps* være synonymt med musikkorps, som finnes på så å si hver eneste barneskole, selv om ordet brukes spredt også i andre sammenhenger (Røde Kors Hjelpekorps).

Det er en god del definisjoner som nok ikke deles av resten av språksamfunnet eller av autoritative ordbøker. Noen eksempler fra NB: *bak er et finere ord for rumpe, is er et godteri*. Disse finnes ikke i Bokmålsordboka.

NMO inneholder mange pussige valg når det gjelder både samme ord og nære synonymer. Her er noen eksempler på uforståelige forskjeller i definisjonene av *stue* og *dagligrom* i de tre språkene: *dagligrommet i huset du bor i, en oppholdsstue er der hvor sofa og tv er, dagligstuen er oppholdsrommet i et hus, i ett vardagsrum tittar man på tv eller umgås med varandra*.

Pronomener har underlige og forvirrende ordforklaringer, som dessuten også spriker mellom språkene:

(10)

**NN du:** Tomas og du får eit eple kvar, han er større enn deg

**DK du:** du skal gå nu, du og Tobias får et æble hver

**SV du:** Vill du ha ett äpple?

(11)

**NN eg:** du bruker eg når du snakkar om deg sjølv

**DK jeg:** jeg er det ord du bruger, når du taler om dig selv

**SV jag:** man använder jag när man talar om sig själv

(12)

**NB han:** han heter Karl, det var ham vi hjalp, hans sko ble ødelagt

**NB vi:** inkluderer deg selv og andre

**NB de:** se på kattungene, er de ikke søte?

Også mellom ord av samme type, som ballspill, er det store og usystematiske forskjeller mellom definisjonene, her eksemplifisert med *håndball* og  *fotball*, hvor det i tillegg er noen forskjeller innenfor samme ord: *håndball er et ballspill med to lag, i handboll slår och kastar man bollen med händerna, fotball er et spill, fodbold er en sportsgren, fotboll är en bollsport där spelarna ska sparka bollen.*

For ett og samme begrep er det også helt ulike definisjoner, som ved *is/glass*: *is er et godteri, en is er en sød, frossen matvare, glass är en fryst efterrätt eller godsak.*

Det samme illustreres også ved forskjellene i definisjon mellom begrepet *høflig/velopdragen/artig* (hvor det for øvrig også er merkelig at NB har oppslagsordet *velopdragen*, mens NN har *høfleg*): *den som är artig uppför sig helt korrekt, en velopdragen person opptrer høflig, ein høfleg person oppfører seg fint, den som er artig, opfører sig helt riktigt.*

Et tredje eksempel på de svært forskjellige definisjonene av samme ord er av *brett/bræt/bricka*: *et brett er et flatt trestykke, et bræt er et fladt stykke træ, på en bricka bär man fram exempelvis kaffekoppar till ett bord.* Og til slutt tar vi med definisjonene av *sukker*, som også varierer veldig i form og spesifiseringsgrad: *sukker er en søt matvare, sukker er sødt og slik er lavet af sukker, socker är sött.*

## 5. Brukergrensesnitt, uttale o.a.

Brukervennlighet går ikke bare på innhold, men også på tilgjengelighet. NMO er lett å finne (ved å google *Nordisk miniordbok*), og det er lett å skrive inn ord i søkefeltet. Mens man skriver, kommer det opp en liste av de ordene som passer til det man har skrevet i søkefeltet, men lista er presentert hulter til bulter, ikke alfabetisk sortert, pussig nok, noe som minsker nytteverdien.

Alle ordene og ordforklaringene er lest inn på alle språkene. En pussighet i den norske utgaven er det at mens NB leses med en typisk østlandsk Oslo vest-uttale av både den mannlige og kvinnelige oppleseren, leses NN av to personer med forskjellig dialekt, begge fra Vestlandet. Det virker veldig tilfeldig. De svenske og danske oppleserne bruker samme språkvariant.

## 6. Konklusjon

Målet med NMO er å støtte opp om nabospråksforståelsen og nabospråksundervisninga i Danmark, Norge og Sverige. Den kan lastes ned på mobil og lesebrett, og det er fint, siden den retter seg mot ungdom.

Imidlertid har vi pekt på alvorlige problemer med ordboka når det gjelder lemmaseleksjon og definisjonskonsistens. Når det gjelder lemmaseleksjon, er det vanskelig å forstå at man har valgt ut helt ekvivalente, vanlige ord i alle de tre forskjellige språkene,

som *kaffe*. Særlig forvirrende er det når slike ord får forskjellige definisjoner i de tre språkene. Dette antar vi er en arbeidsulykke og resultat av manglende koordinering mellom redaktørene. Det mest sentrale, nemlig falske venner og manglende ordekvivalenter av grunnleggende ord, er usystematisk fremstilt.

Definisjonene følger ingen mal, verken innenfor et språk eller på tvers av språkene. Det ser grelt ut når definisjonene er plassert ved siden av hverandre, slik de er i NMO. Noen definisjoner er for generelle, andre for spesifikke, og også formmessig varierer de sterkt.

Til tross for prosjektets gode forsetter og de språkfaglige institusjonene som står bak satsingen, mener vi dessverre at Nordisk miniordbok viser en del svakheter, og at redaksjonen ville tjent på å sette seg bedre inn i leksikografisk teori og metode, samt eksisterende ordbøker, frekvensordlister og korpus, før de gjorde sine valg.

Vi mener at NMO ikke har nådd sine uttrykte mål. Ungdommene som bruker den, vil ikke få vite om falske venner, og definisjonene vil stort sett være overflødige for dem som allerede er skandinaver. Den manglende systematikken i alle ledd vil gi dem en svært feilaktig forståelse av hvordan ordbøker skal se ut, og sannsynligvis føre til en manglende respekt for innholdet. De vil også få et uheldig inntrykk av det faglige nivået hos de nordiske språkrådene og nordiske språkforskere.

## Litteratur

Liste over noen flerspråklige skandinaviske ordlister



- Börestam, Ulla & Ella Olaug Rekdal (1989/2006): *Svensk-norsk ordliste*. Lund/Uppsala: Nordisk språksekretariat och Bokklubben Norden.
- Cramer, Jens & Peter Kirkegaard (1996): *Dansk sproglære for nordmænd*. Oslo: Ad notam Gyldendal.
- Dansk-norsk ordbok*: <http://www.ordbok.com/dansk-norsk.html>
- Dansk-norsk-svensk ordliste*: [http://www.nordisk-sprakrad.no/folder\\_att\\_forstaa/ordliste\\_ord.htm](http://www.nordisk-sprakrad.no/folder_att_forstaa/ordliste_ord.htm)
- Lindgren, Birgitta, Skirne Helg Bruland, Allan Karker & Ståle Løland (1994): *Skandinavisk ordbok (svensk-norsk-dansk)*. Oslo: Kunnskapsforlaget og Nordisk språksekretariat.
- Mårtenson, Per & Anton Fjeldstad (1999): *Svenska för norrmän*. Oslo: Ad notam Gyldendal.
- Rekdal, Ella Olaug (1988/2010): *Norsk-svensk ordlista*. Uppsala: Nordisk språksekretariat och Bokklubben Norden.

#### Liste over noen ordbøker rettet mot barn

- Ernby, Birgitta, Martin Gellerstam, Sven-Göran Malmgren & Per Axelsson (2001): *Norstedts första svenska ordbok*. Stockholm: Norstedts.
- Lunde, Knut (2003): *Norsk ordbok for grunnskolen*. Oslo: Kunnskapsforlaget.

#### Noen frekvensordlister og korpus

##### Frekvensordlister:

<http://www.hf.uio.no/iln/om/organisasjon/tekstlab/tjenester/frekvenser.html>

KAL-korpuset (Kvalitetssikring av læringsutbyttet i norsk

skriftlig):

<https://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/kal/>

Leksikografisk: bokmålskorpus: <http://www.hf.uio.no/iln/tjenester/kunnskap/samlinger/bokmal/veiledningkorpus/>

Nordisk dialektkorpus: <http://www.tekstlab.uio.no/nota/scandiasyn/index.html>

#### Andre referanser

- Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén (1997): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Breivik, Torbjørg (2012): Nordisk barneordbok på nett. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi. Rapport från Konferensen om lexikografi i Norden Lund 24.-27. maj 2011.*[Lund]: Nordiska föreningen för lexikografi.
- Wangensteen, Boye (red.) (2005): *Bokmålsordboka. Defini-sjons- og rettskrivningsordbok*. Oslo: Kunnskapsforlaget. <http://www.nob-ordbok.uio.no>

Ruth Vatvedt Fjeld  
professor  
Universitetet i Oslo  
ILN  
P.b. 1102 Blindern  
0317 Oslo  
[r.e.v.fjeld@iln.uio.no](mailto:r.e.v.fjeld@iln.uio.no)

Janne Bondi Johannessen  
professor  
Universitetet i Oslo  
ILN/MultiLing/Tekstlab  
P.b. 1102 Blindern  
0317 Oslo  
[j.b.johannessen@iln.uio.no](mailto:j.b.johannessen@iln.uio.no)

## Samvirket mellom ordbank og ordbok

*Oddrun Grønvik og Christian-Emil Ore*

In August 2012 the web edition of *Nynorskordboka* was launched with a revised orthography for Nynorsk and many new functions, directed at improving user friendliness and information transmission about the Nynorsk standard language. Until then, part of speech information and (abbreviated or coded) inflection patterns were hard coded as part of the dictionary text. In the new web edition, the dictionary database is linked to the Word Bank for Nynorsk. The entry head is generated from the Word Bank, as are schema showing inflected forms written in full. The Word Bank for Nynorsk has become an efficient tool for maintaining and transmitting orthographic information on Nynorsk. Use of the revised web edition has increased considerably in the last year.

### 1. Oppgradering, oppkopling og kvifor

Handordbøkene Bokmålsordboka (BOB) og Nynorskordboka (NOB) har ei lang fortid som ordbøker på nett. BOB og NOB kom i andre utgåve i 1993, og vart publiserte som gratis-resursar<sup>1</sup> i 1994 frå INL.<sup>2</sup> Nettsida for handordbøkene fekk noverande utsjånad i 2009,<sup>3</sup> med ei viss oppgradering for funk-

sjonar (Hoel 2010:10–11). Teksten frå 2005- og 2006- utgåvene stod uendra.

Frå 2009 vart det arbeid med å leggje inn teksten i handordbøkene i det databasesystemet som Norsk Ordbok nyttar. Det vil sin tur gje betre ordbøker både på papir og på nett. Arbeidet vart utført ved ILN, UiO,<sup>4</sup> med medverknad frå Språkrådet.<sup>5</sup> Oppgraderinga omfatta (1) korrekte fullformer i bøyingsskjema i artikkelhovudet, (2) fullskriving av forkortingar i definisjonar og bruksdøme og (3) endringar i databasestrukturen for BOB, NOB og Norsk ordbank. Endringane i databasestrukturen omfatta inndeling av ordartiklar etter artikkeltype, kopling av tilvisingar, og komplettering av informasjon i artikkelhovudet (oppslagsdelen). Alle ordartiklar skulle ha med dei same kategoriane for informasjon om ordklasse og bøying.<sup>6</sup>

Føremålet med oppgraderinga var (1) standardisering av normeringsformatet for einskildord (same opplysningar, betre systematikk) og (2) ei meir fullstendig og opplysende formidling til folk flest om rettskriving og bøying i norsk.

Språkrådet hadde i tillegg eit ønske om arbeidssparing. Informasjon om gjeldande rettskriving var i 2009 spreidd på Norsk ordbank, handordbøkene, og årsmeldingar og sakspapir frå Språkrådet. Det var stor overlapping, men også ein del divergens. Om all rettskrivingsinformasjon for norsk kunne

- 
1. På [www.edd.uio.no/ordboksoek/ordboksoek.html](http://www.edd.uio.no/ordboksoek/ordboksoek.html).
  2. Institutt for nordistikk og litteraturvitskap, UiO (1990–2002).
  3. <http://nob-ordbok.uio.no/perl/ordbok.cgi>.
  4. ILN, som INL vart del av frå 1.1.2003.
  5. Språkrådet og UiO er felles eigarar av BOB og NOB på nett.
  6. Jf. brev 22.12.2010 frå styringsgruppa for Norsk ordbank til Språkrådet.

samlast på ein stad, og hentast derifrå av t.d. utgjevarar av språkverktøy (ordlister, retteprogram osv.), ville det spare Språkrådet for tidkrevjande kontrollarbeid på uferdige produkt.

Frå språkvitskapleg og leksikografisk side var det ønskeleg med ei grundig oppgradering både av Ordbanken for nynorsk og NOB. Gjennomarbeiding av kategorisystem, paradigme-register og innhald ville gje eit heilt anna oversyn over nynorsk-rettskrivinga enn det som låg føre i 2009, og dermed eit mykje betre grunnlag for å vurdere føresetnader for og verknader av den påtenkte rettskrivingsrevisjonen for nynorsk.

Parallelt med oppgraderingsarbeidet i ordboksdatasane førebudde Språkrådet ei ny rettskriving for nynorsk. Den vart sanksjonert av Kulturdepartementet 20.9.2011.<sup>7</sup> Språkrådet fekk i oppgåve å få den nye rettskrivinga kunngjord via NOB før skulestart 2012. Oppgradering av NOB vart derfor prioritert framfor BOB.

## 2. Kva er Norsk ordbank

Norsk ordbank (bokmål og nynorsk) er ein database over grunnord, bøyingsmønster og kopling mellom dei. Koplingane kan innehalde informasjon om normering og tidsrommet normeringa gjeld for. I tillegg inneheld Norsk ordbank ein full-formsgenerator. Norsk ordbank har eksistert sidan 2001<sup>8</sup> og vart

---

7. Brev frå Kulturdepartementet til Språkrådet 20.9.2011.

8. <http://www.edd.uio.no/prosjekt/ordbanken/>.

i si tid laga slik at den i tillegg til annan bruk (Hagen og Nøklestad 2010:55 f.) kunne eigne seg til å vere lager og vedlikehaldspunkt for rettskrivinga av bokmål og nynorsk. Mens BOB og NOB alltid har vore eit offisielt talerøyr for Språkrådet, har ikkje Norsk ordbank vore det før i 2012. Dersom Norsk ordbank også skulle spare Språkrådet for dobbeltarbeid, og alle andre for avskrift, måtte Ordbanken koplast til ordboksdatabasane slik at artikkelhovudet på nett vart riktig generert for kvar artikkel. I neste runde må dette formatet tilpassast ein ny prenta versjon av handordbøkene.

Samankopling av ordbank og ordboksdatabase var dermed sett som den mest rasjonelle måten å (1) kunngjere den nye nynorskrettskrivinga, (2) rasjonalisere vedlikehaldet av rettskrivinga, (3) leggje til rette for generering av oppslagsdelen i ordbøker og ordlister frå Ordbanken. I resten av denne artikkelen vil vi konsentrere oss om nynorskdelen av Norsk ordbank. Termen «Ordbanken» vil derfor bety nynorskdelen av Norsk ordbank dersom ikkje anna er sagt. Datateknisk er dei to delane like og har lik oppbygging, jf. avsnitt 2.1 nedanfor.

## 2.1. Ordbankens oppbygging

Ordbanken er basert på ei utviding av eit stavekontrollsystem som IBM laga på 1980-talet (Engh 1993). Hovudtanken er å lenke kvar grunnform til eitt eller fleire bøyingsmønster som i sin tur kan generere alle moglege bøyde former av grunnforma. Denne prosessen inneber at det blir generert teoretisk moglege, men udokumenterte bøyde former. Mellom dei udokumenterte formene vil det vere somme former som verkar

så avvikande at dei kan vekkje reaksjonar mellom språkbrukarane<sup>9</sup>. Desse formene treng ein i språkteknologiske analyseverktøy, medan dei ikkje trengst i t.d. system for skrivestøtte. Der vil ein helst sjå berre dei allment godtekne formene. For å unngå moglege, men uakseptable former, og sikre at alle sett med bøygde former svarer til offisiell rettskriving, er det lagt tilleggsinformasjon til lenka mellom grunnform og bøyingsparadigme.

Tilleggsinformasjonen gjeld kategoriane normeringsstatus og tidsperiode då normeringsstatusen er eller var gyldig. Det er dermed mogleg å nytte Ordbanken til å lage eit historisk register over norsk rettskriving. Ordbanken har no informasjon for tidsperioden 1996 til 2013. Ein kan, om ein vil, generere ei normert rettskriving for kvart av åra i perioden.

Figur 1 viser redigeringsgrensesnittet i Ordbanken. Topp-linja inneheld identifiserande informasjon. Under til venstre er tabellar for normeringsstatus og paradigme med status. Fullformstabell er nedst til venstre. Til høgre frå toppen ser ein tabell med lenke til ordboksdatasane, under tabell for grammatisk merking i IBM-tilfanget og tidlegare utgåver av NOB.

Kvar artikkel i Ordbanken har ein unik kombinasjon av ei grunnform pluss eitt eller fleire bøyingsparadigme pluss statushistorie i høve til grunnform og per paradigme. Ordbankartiklane er ikkje skilde på homografar etter dette punktet, slik dei kan vere i ei ordbok (t.d. ut frå ordhistorie). Norsk er som kjent rikt på homografe grunnformer. Dermed er det mogleg at éin ordbankartikkel kan vere lenka til fleire ordartiklar i ei ordbok. (Grønvik og Ore 2013:250 f.).

---

9. Døme på slike former er gradbøygde former av somme adjektiv, som *\*interessantare* – *\*interessantast*.

Universitetenes samlingsdatabaser, Språk - [Norsk ordbank, nynorsk]

Arkiv Rediger Vis Søk Eksportere resultat Vindu Hjelp

Nynorskordboka 2013 Norsk ordbank, nynorsk

Søking Objektvisning

Lemma-id: 56435 Funksjon: lemma Grunnform: lyge  Skjul for tagg

Registrer | Tekst | Kommentar/logg

Normering

Normering	Fra dato	Til dato
normert	1996	4000

Paradigmer

Pa...	Ordklasse	Normering	Fra dato	Til dato
197KL	verb	klammeform	1996	2012.07.31
197KL	verb	unormert	2012.08.01	4000
198	verb	normert	1996	4000
198K1	verb	klammeform	1996	2012.07.31

Bøyingsdata | Taggermerker | Genererte fullformer

P...	N...	Fullform	Tag	Fra dato	Til dato
198	1	lyge	verb inf <intrans1> <trans11/for> <trans: 1996	4000	
198	2	lyga	verb inf <intrans1> <trans11/for> <trans: 1996	4000	
198	3	lyg	verb pres <intrans1> <trans11/for> <tran: 1996	4000	
198	4	lygast	verb inf pass <intrans1> <trans11/for> <t: 1996	4000	
198	5	laug	verb pret <intrans1> <trans11/for> <tran: 1996	4000	
198	6	loge	verb perf-part <intrans1> <trans11/for> <: 1996	4000	
198	7	loge	adj <perf-part> nøyt ub eint <intrans1> <: 1996	4000	
198	8	logen	adj <perf-part> m/f ub eint <intrans1> <t: 1996	4000	
198	10	logne	adj <perf-part> bu eint <intrans1> <trans: 1996	4000	
198	11	logne	adj <perf-part> fl <intrans1> <trans11/fo: 1996	4000	

Kilder

F...	Kildnavn	Eleme...
	NN_ORDBOK	46245
	NOB_2013	46245
	NOB_2012	46245
	NN_ORDBOK	47030

Alternative pa

Kilde	Alternativ kode
IBM	199
NN_O	Vs

Ny ortoklump

Ortoklump

oddrung@EDDF

Figur 1: Ordbanken – redigeringsgrensesnittet

I både Ordbank og ordboksdatabasane er det lagt stor vekt på dataintegritet. Det er knytt identitetsnummer til artiklar, i NOB også til definisjonar og døme. Bøyingsparadigma i Ordbanken har egne kodar, tre tal eller bokstavar.



## 2.2. Ordbankens omfang og innhald

Når det gjeld realinnhald, har Ordbanken for nynorsk om lag 113 000 artiklar, noko meir enn NOB, som har om lag 100 500 oppslagsord fordelte på rundt 85 000 innhaldsartiklar (normalartiklar). Dette er den delmengda av nynorsk ordtilfang som er eksplisitt normert. Ordbanken for nynorsk har om lag 657 000 fullformer i alt. Av desse er 509 000 fullformer etter 2012-rettskrivinga. Fullformene er genererte på grunnlag av 1578 paradigme. Etter 2012-rettskrivinga er 518 paradigme med i rettskrivinga. Rettskrivingsrevisjonen som tok til å gjelde 1.8.2012, har dermed ført til at meir enn 1000 ulike bøyingsparadigme og 150 000 ordformer (som er variantar) er gått ut or nynorskrettskrivinga.

I 2005 vart ordklassenemningane for norsk reviderte <sup>10</sup> og dei endra ordklassenemningane tilrådde til i bruk i skulen (Johannesen og Hagen 2007). Ein del ord vart omklassifiserte etter dei nye nemningane, som byggjer på ordklassesystemet i Norsk referansegrammatikk (NRG:20 f.). Etter oppgraderinga av NOB i 2012 kjem ordklassenemningane frå Ordbanken. I Ordbanken var systemet i NRG nytta fullt ut, og det måtte dermed justerast noko for å samsvare med 2005-vedtaket. I tillegg måtte det først inn nemningar på ordklasseplass for forkortingar, avleiingsprefiks, forledd i samansetningar og symbol. Dette er språklege einingar som har artiklar i alle allmennordbøker for norsk, men som fell utanom det vanlege ordklassesystemet, av di dei er ordlagingselement (forledd i samansetningar, avleiingsprefiks), eller av di dei representerer ein særskild, konvensjonalisert bruk av ord og ordlag (forkortingar,

---

10. Utdanningsdirektoratet og Språkrådet (2005).

symbol). Ordbøker tek med desse artikkeltypane av di det ofte er spørsmål etter korrekt skrivemåte.

### 2.3. Paradigme og vising av bøyingsmønster

Revisjonen av nynorskrettskrivinga 2011–2012 føresette ein full gjennomgang av bøyingsparadigma i Ordbanken før endringane vart innførte. Det vart avgjort at all variasjon måtte liggje på paradigmenivå eller over. Variantar i høve til bøyingsskjema innanfor eit og same paradigme vart fjerna.<sup>11</sup> Dersom eit verb t.d. har to fortidsformer, føreset dette to sjølvstendige paradigme. Går den eine paradigmet ut i samband med rettskrivingsendring, kan dette uttrykkjast ved sluttdato for status «normert», medan det andre står uendra.

Figur 2 syner eit døme på korleis Ordbankens paradigmeskjema – som vart utarbeidde som del av fullformsgeneringa omkring midten av 1990-talet – er nytta til å gje bøyingsinformasjon i NOB på nett. Skjemaet har både fullformer og informasjon om det grammatiske systemet. Har eit ord fleire bøyingsparadigme, blir desse viste på kvar si linje i skjemaet. Merk at standardskjemaet for nynorsk verbøyings har to infinitivformer, både e-infinitiv og a-infinitiv. I staden for ein ekstra kolonne har vi valt å setje dei under kvarandre.

---

11. Skjemaet for verbparadigme hadde t.d. separat femininumform for verbaladjektiv og sams form for maskulinum og femininum.

**Bøying i samsvar med 2012-rettskrivinga:**

lyge	Infinitiv	Presens	Preteritum	Presens perfektum	Imperativ
v.	å lyga å lyge	lyg	laug	har loge	lyg

lyge	Perfektum partisipp				Presens partisipp
	Hankjønn/hokjønn	Innkjønn	Bunden form	Fleirtal	
v.	logen	loge	logne	logne	lygande

Figur 2: Bøyingsinformasjon frå Ordbanken om verbet *lyge*, inkludert skjema for samsvarbøying av verbaladjektivet.

Ei eittspråkleg papirordbok er berar av mykje implisitt informasjon. Redaksjonen kan spare plass ved å sløyfe det som synest opplagt for morsmålsbrukarar. Plasskrevjande informasjon som gjeld store kategoriar kan stuvast inn i forord og brukarrettleiingar. Men i ei elektronisk ordbok er dette korkje naudsynt eller ønskeleg. Det er best at informasjonen kan bli vist saman med ordartikkelen. Alle paradigme må vere eksplisitt registrerte på alle relevante grunnformer. Det vart naudsynt å lage nokre få nye paradigme for nynorsk. Kategorien «blanda bøyning» av svake verb har t.d. fått eigne paradigme knytte til dei rette grunnformene, med kode «v3». <sup>12</sup> Paradigmesystemet i Ordbanken byggjer på rettskrivingssystemet til IBM. Føremålet med IBM-systemet var praktisk, det byggjer ikkje på nokon spesifikk morfologisk teori.

12. v1 = a-verb, t.d. *rope - ropar - ropa - ropa*. v2 = te-verb, t.d. *rope - roper - ropte - ropt*. Verb med bøyning etter båe desse mønstera kan også ha eit kombinert paradigme, *rope - ropar - ropte - ropt*. Dette paradigmet er gjort eksplisitt i 2012-utgåva av NOB på nett og har fått koden v3.

Når ein set opp eit system av bøyingsmønster for norsk, står hovudvalet mellom å føre fleire alternativ på kvar paradigme-linje (komplekse mønster) eller ei einskild form for kvar linje (enkle mønster). I Ordbanken er alle paradigme i utgangspunktet enkle mønster. Men sidan ei grunnform kan vere tilordna fleire paradigme, kan ein sjå på kombinasjonar av paradigme knytte til ei grunnform som komplekse mønster.

Om ein definerer ord som paret <grunnord, bøyingsmønster>, vil talet på ord auke når talet paradigme aukar. I tabellen i figur 3 viser vi korleis det blir for nynorsk. Dei komplekse paradigma er laga ved å setje saman enkle paradigme. Til dømes kan «700» (m1)<sup>13</sup> og «800» (n1)<sup>14</sup> setjast saman til eit felles (m1, n1) med fleire alternativ på kvar linje.

Talet på paradigme stig når ein tel med komplekse paradigme (ein får både kvart einskilt og kombinasjonar av dei), men berre til ein viss grad. Talet på paradigme aukar kraftig når

	Alle		2012-rettskrivinga	
	tal ulike paradigme	tal ord	tal ulike paradigme	tal ord
Komplekse	755	111 545	580	103 650
Enkle	1 578	269 556	518	112 409

Figur 3: Paradigme og grunnformer i Ordbanken for nynorsk.

13. Paradigme nummer 700 = bøyingsmønsteret uttrykt i *hund* – *hunden* – *hundar* – *hundane* med NOB-kode m1.

14. Paradigme nummer 800 = bøyingsmønsteret uttrykt i *hus* – *huset* – *hus* – *husa* med NOB-kode n1.

alle sideformene skal gje opphav til egne paradigme. Men talet på grunnformer fell når ein tel med komplekse paradigme, for då blir oppslagsformer med meir enn eitt paradigme talde berre ein gong i staden for fleire.

Tabellen i figur 3 viser at 1 060 bøyingsparadigme og meir enn 150 000 rettskrivingsvariantar fekk sluttdato 2012 i samband med rettskrivingsrevisjonen for nynorsk. Han syner også at det store fleirtalet av paradigme som er unormerte no, var tillegg til hovudparadigma.

Endeleg må som nemnt alle grunnformer ha paradigme i Ordbanken, endå om grunnforma er ei einaste form. Føremålet er få vist rettskrivingsinformasjonen så ryddig som mogleg i nettutgåva av NOB, og bruke einsarta skjema så langt råd er.

### 2.3. Leddanalyse i Ordbanken og NOB

Tidlegare utgåver av NOB og BOB markerer deling av oppslagsord etter typografiske behov, i høve til avsnittsinndelinga i den prenta ordboka. Oppslagsord i same avsnitt er forkorta, og delestreken i første oppslagsord syner kva som er sløyfa i dei følgjande oppslagsorda.

Delemerke kan falle saman med komposisjonspunkt, men treng ikkje gjere det. Samansette ord i eine avsnitt har ikkje delemerke, medan avleiingar kan ha det. Deling av avleidd ord gjeld særleg importord. Døme på delingspraksis i NOB 2006 er vist i figur 4.

type	NOB 2006	NOB 2012
grunnord	rip el	ripel
avleiing	ralj ere	raljere
avleiing	ragg ete	raggete
	ransak e	ransake
samansetning	ramp estrek	rampe strek
samansetning	rakletre	rakle tre

Figur 4: Døme på delemerke i den trykte ordboka i 2006 og i den elektroniske ordboka i 2012.

I samband med oppgraderinga av NOB og Ordbanken var det ønske om å kunne vise leddeling etter språklege kriterium, det vil seie forledd og etterledd med ordklasse, og eventuell fuge. Det vart derfor laga ein leddelingstabell, som tillegg til Ordbanken for nynorsk. Leddelingstabellen byggjer på ei grov leddeling gjord med Paul Meurer sitt leddelingsprogram.

Fugeregisteret for norsk omfattar i tillegg til -e-, -s- osv. også nullfuge (figur 5, døme 1) og minussfuge (døme nr. 3). Ord som har leddeling med minussfuge, blir viste i ordartikkelen med sine to godkjende oppslagsformer, i dette tilfellet «trafikkaos el. trafikk-kaos».

Når ei samansett grunnform frå Ordbanken er registrert og tagga i leddelingstabellen, vil orddelinga følgje med til oppslagsordet for ordartikkelen i NOB. Førebels må kvar samansetning registrerast med leddeling, men eit framtdsmål er å gjere denne prosessen (halv)automatisk.

I ein leddelingsanalyse vil det alltid kome for dagen ein god del samansetningar der førsteleddet ikkje kan knytast sikkert til berre eitt anna oppslagsord, jf. tabellen i figur 5 siste linje, som

Grunn- form	For- ledd	For- ledd gram	Fuge	Etter- ledd	Etterledd gram.	Oppslags- form
raudvin	raud	Adj		vin	Noun+M+ Sg+Indef	raud*vin
raudvins- toddi	raudvin	Noun	s	toddi	Noun+M+ Sg+Indef	raudvins* toddi
trafikkaos	trafikk	Noun	- k	kaos	Noun+N+ Sg+Indef	trafikkaos
trillebag	trille-	Føre- ledd		bag	Noun+M+ Sg+Indef	trille*bag

Figur 5: Leddeling i nynorskordbanken

både kan knytast til verb og substantiv.<sup>15</sup> I praksis spelar ordklasse på forledd lita rolle, for funksjonen til forleddet i ei samansetning er å modifisere eller spesifisere etterleddet, same kva ordklasse forleddet har opphav i.<sup>16</sup>

15. Verb < ‘bag som kan trille’, substantiv < ‘bag som er (som) ei trille’.

16. I denne typen samansetningar er forleddet før ofte rekna som verb, jf. Beito (1986) s. 141 og NRG s. 75. I leddelegssystemet for handordbøkene står spørsmålet ope, i staden for å vere bunde til ein usikker hypotese.

### 3. Oppslagsdelen i Nynorskordboka (NOB)

Oppslagsdelen i NOB 2012 omfattar minst eitt sett av kategoriane (1) oppslagsord, (2) ordklasse, (3) bøyning. Oppslagsdelen i ein ordartikkel kan ha meir enn eitt oppslagsord med formopplysningar. Det inneber at to eller fleire ordbankartiklar kan vere kopla til same artikkel i ordboksdatabasen.

I NOB 2006 mangla ordklasse og bøyning for alle samansette oppslagsord. Det mangla også merking tilsvarende ordklasse for kategoriane forkorting, avleiingsprefiks, forledd i samansetning og symbol. I Ordbanken er desse kategoriane utfylte for kvar grunnform, men det mangla eit system for å føre denne informasjonen over i NOB. Det systemet er no på plass i og med at artikkelhovudet blir generert frå Ordbanken.

#### 3.1. Frå eitt artikkelformat til fleire artikkeltypar

NOB 2006 har eitt artikkelformat. Det er ikkje skilt formelt mellom tilvisingsartikkel på den eine sida og normalartikkel med definisjon på den andre. Ved førstegongsredigeringa på 1970–80-talet var det eit overordna mål å gje lik status til alle former av oppslagsord, ettersom ein den gongen ikkje visste sikkert kva former som var mest i bruk. Eit oppslag kunne omfatte meir enn fire oppslagsformer.

I Ordbanken førte formatsidestillinga mellom hovudartikkel og tilvisingsartikkel til eit høgt tal mange-til-mange-koplingar, med fire eller fleire ordbankartiklar kopla til like mange ord-



artiklar. Det kunne vere opp til 25 koplingar utan sikker peikar til den eine ordartikkelen som inneheldt definisjon og bruksdøme. Det var ein etter måten liten del av ordartiklane i NOB som hadde parallellkoplingar, men det var mange nok til å gjere sambandet mellom ordboksdatabase og ordbank uoversiktleg.

I NOB 2012 er det skilt mellom (1) normalartikkel, (2) tilvisingsartikkel, (3) forleddsartikkel (oppslagsform endar med bindestrek) og (4) etterleddsartikkel (oppslagsform tek til med bindestrek). Normalartikkelen har alle rettskrivingsvariantar kopla til seg som oppslagsformer. Tilvisingsartiklar skal no ha ei oppslagsform, og er knytte til rett normalartikkel med «hyperlenking» frå definisjonsfeltet. Dette gjev ryddigare tilhøve mellom ordbank og ordboksdatabase.

### 3.2. Samsvar mellom oppslagsdel og tydingsdel i NOB

I NOB 2012 må alle ordartiklar ha ein definisjon. Dette kan settast opp som ei likning:

Oppslagsdel = Tydingsdel [tyding 1, tyding 2 .. Tyding n]

NOB har frå før mange skjelettartiklar for samansetningar, som består av oppslagsordet og ingen definisjon. Dei står som type-døme. Dei blir ståande i NOB på nett med ordklasse og bøyingsparadigme, men dei blir også lenka til ei samansetningsrekke, anten under ordartikkelen for forleddet eller i eigen ordartikkel for samansetningar (forleddsartikkelen).

Alle oppslagsformer for ein artikkel skal kunne brukast for alle tydingane som er definerte i artikkelen. Kan dei ikkje det,

er dei ikkje rettskrivingsvariantar, men nærsynonym. Eventuelle restriksjonar i form – t.d. i bruk av fleirtalsformer for konkret tyding (t.d. kjærleik [på pinne]) og eintal (kjensla kjærleik) som indikasjon på abstrakt tyding – er knytte til definisjonen, ikkje til oppslagsforma. I lys av prinsippa bak ordnett vil det kanskje vere best å knyte bøyning til dei ulike definisjonane. Men det vil vere eit særst omfattande arbeid som ein ikkje har resursar til no.

### 3.3. Splitting av artiklar i NOB

I NOB 2006 har ein del ordartiklar jamstilte oppslagsformer som ikkje er ortografiske variantar, men fullsynonym eller delsynonym. Døme er samansetningar med *lege* m1 el. *lækjar* m1 (ulik avleiing og etymologi), *sjuke* f2 el. m1 el. *sykje* f2 (ulike genus), eller *verd* n og *verdi* m (ulike genus, ulik etymologi).

I NOB 2012 er det eit krav at ordformer som er alternative oppslagsformer til same ordartikkel, må vere rettskrivingsvariantar. Dei må ha same ordklasse(r), same etymologi og ei forlikeleg normeringssoge. Samansette oppslagsord til same artikkel må ha same leddeling. Elles blir generering av oppslagsdel frå Ordbanken altfor innfløkt. Tanken om generering av oppslagsdelen frå Ordbanken til ordboksdatatabasen har såleis også ført til ein langt strengare lingvistisk stringens i hand-saminga av ordtilfanget for nynorsk.

#### 4. Ordbanken, ordboka og språket

Rettskrivingsordlister føreset ein omfattande implisitt kunnskap om tyding. Mellom morsmålsbrukarar og innanfor ramma av ein stabil skriftkultur har det vore rekna som sjølvsgatt at alle språkbrukarar er morsmålsbrukarar, og at alle derfor kjenner tydingane på ord i ordlista, eller har folk rundt seg som veit kva orda tyder. Men slik er det ikkje lenger. Det svært store talet med andrespråksbrukarar i skole og arbeidsliv gjer at sambandet mellom ord(form) og tyding må gjerast eksplisitt for det normerte ordtilfanget. Ein fullformgenerator som manglar lenke på artikkelnivå til ein ordboksdatabase med definisjonar og bruksdøme, fortel ikkje meir enn ei gloseliste gjer. Dermed blir fullformsgeneratoren utilfredsstillande som språkverktøy. Ordbanken kan brukast til å lage tydingsberande språkverktøy av di han er knytt til ordartiklane i ei gjennomarbeidd ordbok.

Det er koplinga til NOB som ligg til grunn for at ein kan skilje mellom I *glis* m ‘person’, II *glis* m1, n1 ‘hånflir’ og III *glis* n1 ‘stor, fin bil’. Sambandet til tydingsdelen gjer det klart at Ordbanken her må ha tre «glis»-artiklar, og ikkje to eller ein. Det same gjeld skiljet mellom I *rett* m ‘(tillaga) mat’ og II *rett* m1 ‘privilegium, fridom’. Utan tyding, bruk og ordhistorie kunne ein ikkje sagt sikkert at dette skulle vere to artiklar i Ordbanken, og ikkje ein artikkel med to paradigme.

## 5. Resultat eitt år etter

Handordbøkene på nett er svært mykje brukte i Noreg, men bruken er førebels lite kartlagd. Trass i at første nettversjon var lite for auga, kunne Språkrådet i årsmeldinga si i 2008 skrive at «I 2008 vart det på denne adressa gjort 14,8 millionar oppslag i Bokmålsordboka og 10,7 millionar oppslag i Nynorskordboka. Bruken er tredobla på fire år og aukar stadig.» (s. 17).

Den nye formatet frå 2009 var utvikla med skolen for auga, og resulterte i auka bruk. Forbetra funksjonalitet kom i 2012 for NOB, og har blitt godt motteke. Det var 28 millionar oppslag i NOB i perioden november 2012 til november 2013. Det er såleis ikkje tvil om at innsatsen har svara seg når det gjeld å gjere formidlinga av norma for nynorsk standardspråk meir tilgjengeleg og effektiv.

## Litteratur

- Beito, Olav (1986): *Nynorsk grammatikk*. Oslo: Det Norske Samlaget.
- BOB= Wangensteen, Boye (2005): *Bokmålsordboka. Definisjons- og rettskrivingsordbok*. (1. utg. 1986, 2. utg. 1993, 3. utg. 2006, ny nettgåve 2013). Oslo: Kunnskapsforlaget.
- Engh, Jan (1993): *Linguistic Normalisation in Language Industry. Some Normative and Descriptive Aspects of Dic-*

- tionary Development. I: *Hermes, Journal of Linguistics* 10, 53–64. <<http://download2.hermes.asb.dk/>>
- Grammatiske termer til bruk i skoleverket* (mai 2005) (oppdatert 17.03.2006). Tilråding fra Norsk språkråd og Utdanningsdirektoratet.  
<<http://www.sprakradet.no/upload/152/Gramterm.pdf>>
- Grønvik, Oddrun & Christian-Emil Smith Ore (2013): What should the electronic dictionary do for you – and how?. I: I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets & M. Tuulik (red.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 243–260.
- Hagen, Kristin & Anders Nøklestad (2010): Bruk av et norsk leksikon til tagging og andre språkteknologiske formål. I: *LexicoNordica* 17, 55–72.
- Hoel, Jan (2010): Edle årganger i ny tapning. I: *Språknytt* 2010 nr. 3, 10–11.
- Johannessen, Janne Bondi & Kristin Hagen (2007): Nye grammatiske termer i skoleverket. I: *Språknytt* 2007 nr. 2, 24–30.
- NOB = Marit Hovdenak, Laurits Killingbergtrø, Arne Lauvhjell, Sigurd Nordlie, Magne Rommetveit & Dagfinn Worren (2006): *Nynorskordboka* (1. utg. 1986, 2. utg. 1993, 3. utg. 2001, 4. utg. 2006, ny nettutgåve 2012). Oslo: Samlaget.
- NRG = Jan Terje Faarlund, Kjell Ivar Vannebo & Svein Lie (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Språkrådet (2008): *Språkrådet 2008*. [Årsmelding]. Oslo: Språkrådet.

GRØNVIK OG ORE

## Lenker

Norsk ordbank: <http://www.edd.uio.no/prosjekt/ordbanken/>  
Bokmålsordboka og Nynorskordboka på nett:  
<http://nob-ordbok.uio.no/perl/ordbok.cgi>

## Arkivmateriale

Brev 22.12.2010 frå styringsgruppa for Norsk ordbank til  
Språkrådet.

Brev 20.9.2011 frå Kulturdepartementet til Språkrådet.

Christian-Emil Smith Ore  
førsteamanuensis, dr.scient.  
faglig leder Eining for digital dokumentasjon  
Institutt for lingvistiske og nordiske studium  
Universitetet i Oslo  
Postboks 1123 Blindern  
NO-0317 Oslo  
[c.e.s.ore@iln.uio.no](mailto:c.e.s.ore@iln.uio.no)

Oddrun Grønvik  
hovudredaktør, Dr. Litt. H. C. University of Zimbabwe  
Norsk Ordbok 2014  
Institutt for lingvistiske og nordiske studium  
Universitetet i Oslo  
Postboks 1021 Blindern  
NO-0315 Oslo  
[oddrun.gronvik@iin.uio.no](mailto:oddrun.gronvik@iin.uio.no)

## Ivar Aasen og Universitetet i Oslo

*Kjetil Gundersen og Dagfinn Worren*

With his grammar of 1848, *Det norske Folkesprogs Grammatik*, and his dictionary of 1850, *Ordbog over det norske Folkesprog*, Ivar Aasen (1813–1896) provided the first systematic account of Norwegian dialects and initiated the era of modern Norwegian lexicography. These publications constituted the scientific basis for a national written language for Norway. Aasen had no formal academic education, but Det Kongelige Norske Videnskabers Selskab in Trondheim provided Aasen with the funding he needed to collect linguistic material from the dialects. In the final stages of the work on the grammar and the dictionary, Aasen sought advice from leading academics at the University of Oslo to ensure that his account of the language was scientifically sound and up to date. The article features the collaboration between Aasen and the university academics in this period, and particular attention is given to a disagreement that Aasen and the academics had regarding how the closed and open vowels should be handled in the written code. The article also addresses how the same environment of academics who had been important for Aasen's breakthrough as a linguist and lexicographer turned against his project when it was no longer just a theoretical linguistic project, but a practical project to establish a dialect-based written Norwegian language as an alternative to Danish. Finally the focus is on how the University of Oslo in the years after Aasen has taken care of and developed the constitutionally grounded project to implement a national language.

## Emne

Med *Ordbog over det norske Folkesprog* (1850) innleidde Ivar Aasen (1813–1896) den etablerte leksikografiske tradisjonen i Noreg. Som språkvitskapleg grunnlag hadde han da allereie gjeve ut *Det norske Folkesprogs Grammatik* (1848). I oppspelet til denne etablerte tradisjonen var det svært sparsamt med prenta ordbøker. Det fanst berre fire slike ordbøker, og dei må karakteriserast som dialektordbøker, ettersom dei ikkje hadde eit eige norsk standardmål til normeringsinstans (jf. Vikør 2000:119).

Eit særdrag ved Aasen var at han skaffa seg kunnskapane sine ved sjølvstudium (Venås 1996:38). Jamvel om han ikkje hadde utdanning frå akademiske institusjonar, søkte han aktivt kontakt og råd hos fagfolka ved universitetet i Christiania (no Oslo) da han gav ut desse to første vitskaplege hovudarbeida sine. I denne artikkelen skal vi for det første syne korleis dette samarbeidet med fagfolka arta seg og særskilt omtale drøftingane omkring beskrivinga av dei opne og lukka vokalane. Vi vil også sjå på korleis nokre sentrale fagfolk på universitetet stilte seg til Aasens arbeid da det seinare ikkje berre var eit teoretisk språkprosjekt, men hadde utvikla seg til eit bruksmål, kalla *landsmålet*. Til slutt skal det dragast nokre liner til vår tid om landsmålet og den nynorske leksikografien på universitetet.



## Bakgrunnen

Grunnlova frå 1814 la rammene for nasjonalstaten Noreg. Men spørsmålet om eit eige nasjonalspråk var ope og påtrengjande. Ikkje minst var ideologiseringa av folkemålet i romantikken ein av grunnane til det (Hyvik 2009:37f.). I det offentlege ordskiftet vart det sett fram mange løysingsframlegg om eit eige veldokumentert norsk skriftspråk. Dei tre store aktørane i språkdebattane på 1830-talet var Jonas Anton Hielm, P.A. Munch og Henrik Wergeland (Hyvik 2009:206). I eit innlegg i *Almindeligt norsk Maanedsskrift* i nr. 2, 1831 etterlyser Hielm eit samarbeid mellom språkkunnige for å framstille «af Landets virkelige Sprog dets virkelige Grammatik» og ei «videnskabelig Ordbog» (etter Hyvik 2009:230). Ivar Aasen (1813–1896) sat på Sunnmøre og formulerte tilsvarande tankar i 1836 i «Om vort Skriftsprog». Han såg føre seg at eit selskap skulle få i oppdrag å «udarbeide en fuldstændig norsk Ordbog, med tilsvarende Grammatik» (Aasen 1912:11). Det er ikkje dokumentert at Aasen kjende til Hielms artikkel, men i Aasen-biografien sin ser Venås (1996:327) dette slik at Aasen følgde vake med i den nasjonale språkdebatten.

Høvet til å gjennomføre eit slikt prosjekt fekk Aasen med prosjektplan og stipend frå Det Kongelige Norske Videnskabers Selskab (DKNVS) i Trondheim i 1842. Prosjektplanen galdt både grammatisk og leksikalsk beskriving av dialektane, og var ei samanarbeiding av tankane til Aasen og planar om emnet som vitskapselskapet hadde (jf. Walton 1996:73–78). Universitetet som institusjon såg derimot ikkje ut til å ha morsmålet som akademisk emne på denne tida (Collett 2011:290), men da prosjektplanen vart laga, konfererte preses i vitskapselskapet, Frederik Bugge, med språkprofessorane Rudolf Keyser og Chris-

topher Andreas Holmboe ved universitetet i Christiania om det faglege innhaldet. Med utgangspunkt i denne planen gjennomførte Aasen språkferdene sine frå 1842 til 1847 rundt om i heile Sør-Noreg og så langt nord som til Vefsn i Nordland. På denne bakgrunnen dokumenterte han den språklege tilstanden for det norske folkemålet i 1848-grammatikken og 1850-ordboka.

### **Første møte med Christiania 1845 og arbeidsmelding til DKNVS i 1846**

Første gongen Aasen var i personleg kontakt med folk i fagmiljøet ved universitetet i Christiania, var sommaren 1845. Aasen var da i Christiania for første gong, og der talte han med Holmboe og Keyser, dei to professorane som DKNVS hadde rådført seg med om Aasens prosjektplan. Varig fagleg samarbeid med universitetsfolka kom derimot først i stand i oktober 1847. Da kom Aasen til Christiania igjen for å ferdigstille og gje ut grammatikken og så ordboka. Frå denne tida slo han seg også meir eller mindre ned i hovudstaden.

Før vi tek for oss det samarbeidet som utvikla seg der, skal vi stoppe litt ved ei arbeidsmelding som Aasen skreiv til DKNVS 31.12.1846, medan han var i Trondheim. Meldinga gjev ein peikepinn om kvifor Aasen ønskte å komme i kontakt med universitetsfolka. Aasen skriv at han ser seg ferdig med innsamlingsarbeidet, og at han er klar til å fullføre grammatikken og ordboka. I utdraget gjer han greie for korleis han ser for seg den siste bearbeidinga av materialet:

For at denne Bearbejdelse kunde blive saa fuldkommen som muligt, var det vistnok at ønske, at man maatte være saa vel bevandret i den gamle Literatur og de vigtigste Sprogværker, at man kunde vide saa meget, som for nærværende Tid er muligt, om Ordenes oprindelige Form og Betydning, og om Formernes historiske Udvikling. Ordene have nemlig i vore Dialekter en saadan Forskjellighed i Betydning og Former, at der behøves megen Granskning til at kunne adskille det ægte fra det uægte, og at kunne give et klart Begreb om Formernes indbyrdes Forhold. (Djupedal 1960:134f.)

Det er tydeleg at Aasen her meinte at det ville bli vanskeleg å analysere dialektmaterialet med tanke på å velje ut normalformer. For å gå vitskapleg til verks med analysen var det naudsynt med meir kunnskap om gammalnorsk og andre beslekta språk enn det han hadde fått høve til å skaffe seg til da. Miljøet han kontakta i Christiania, hadde nettopp den kunnskapen Aasen trong.

## **Aasen i Christiania**

Aasen drog altså til Christiania att seinhausten 1847. Eit brev Aasen skreiv til vennen Maurits Aarflot på Sunnmøre 4.11.1847, viser at han da raskt kom i kontakt med fagfolka:

Efter Ankomsten hertil henvendte jeg mig til DHr. Lange, R. Keyser, C. Holmboe, Munch og Unger, for at faae et eller

andet Raad med Hensyn disse Sprogarbeider og især Grammatiken, som nu først skulde bringes i Trykken[.] (Djupedal 1957:143f.)

Den første som blir nemnd her, Lange, var riksarkivar, så han ligg utanfor emnet for denne artikkelen, men han hjelpte særleg til med å få i stand trykking av grammatikken og ordboka. Dei fire andre var tilknytte universitetet og blant dei fremste historikarane, filologane og språkvitarane i landet. Rudolf Keyser var den første professoren i gammalnorsk ved universitetet, P.A. Munch var både historikar, språkforskar og folkeminnegranskar, Carl Richard Unger var stipendiat og førelesar i gammalnorsk, og Christopher Andreas Holmboe var professor i orientalske språk.

Av desse var det Unger og Munch som etter kvart involverte seg mest i Aasens arbeid. I brev til Olaus Vullum, ein venn frå tida i Trondheim, datert 9.11.1847 skriv Aasen om Unger og Munchs rolle, og han gjev til kjenne at eit føremål med å kontakte fagfolka har vore å få dei til å lese arbeidet og gje tilbake-meldingar som kan gjere det betre og meir vitskapleg:

De fleste Sammenkomster har jeg havt med Munch og Unger, som nu begge have gennemseet min Grammatik og meddeelt mig adskillige Oplysninger angaaende enkelte Stykker af dette Arbeide. Det gik, som jeg ventede, at jeg her paa Stedet snart fik Anledning til at gjøre Rettelser og Forbedringer i Arbeidet; jeg har allerede opnaaet, hvad jeg saa længe ønskede, nemlig at høre kyndige Mænds Bemærkninger over de mest iøinefaldende Mangler ved Arbeidet, førend dette skulde fremlægges for Almeenheden. Begge de nævnte Mænd have meget venligt og skaansomt gjort mig opmærksom paa adskilligt, som kunde gjøres bedre, eller som ved en vis Forandring vilde komme til at

stemme mere overeens med de nyere Sprogforskeres  
Anskuelse af Sagen. (Djupedal 1957:146)

Grammatikken vart trykt vinteren 1847–48, og Aasen byrja deretter på fullføringa av ordboka. I eit brev til DKNVS 25.4.1848 legg Aasen vekt på at han i denne fasen har stor nytte av den faglitteraturen som han får tilgang til i Christiania:

Jeg har saavidt som muligt søgt at benytte den Leilighed, som man her paa Stedet kan have til at blive bekjendt med de forskjellige beslægtede Sprog, da en saadan Kundskab forekom mig at være nødvendig ved Behandlingen af et Dialektsprog, hvori man deels finder samme Ord i forskjellige Former, og deels finder samme Form ved forskjellige Ord. I denne Henseende har især Hr. Candidat Unger ydet mig megen Hjælp saavel ved at lade mig benytte hans betydelige Samling af Sprogværker som ved at underrette mig om de Iagttagelser, som han selv har anstillet. (Djupedal 1957:154)

Eit anna poeng i dette utdraget er at i denne perioden var Unger den som Aasen samarbeidde tettast med. Aasen og Unger utvikla etter kvart eit vennskap som vara livet ut.

## **Striden om dei opne og lukka vokalane**

Utdraga frå Aasens brev viser at fagfolka ved universitetet, og særleg Unger og Munch, gav både støtte og råd til Aasen i

arbeidet med grammatikken og ordboka. Spørsmålet som da melder seg, er kva for konkret innverknad dei hadde på desse verka. Aasen skriv i det heile ikkje mykje om dette, men i brevet til Vullum 9.11.47 nemner han at han har omarbeidd bolken om vokaluttale etter drøftingar med Unger og Munch (Djupedal 1957:146). Dette kan ha samanheng med ei særskild sak som Aasen og dei han rådførte seg med ved universitetet, var usamde om, nemleg korleis dei opne og lukka vokalane skulle markerast typografisk. Opne og lukka vokalar var vokal-lydar som Aasen forklarte som uttalevariantar av visse monoftongar, eller visse «enkelte Vokaler» (Aasen 1848:7), som var Aasens omgrep. Det var i utgangspunktet ni monoftongar i Aasens lydssystem (a, e, i, o, u, y, æ, ø og å), men seks av desse (i, e, y, u, o og ø) hadde i tillegg to uttalevariantar, open og lukka. I Aasens system var det dermed til saman femten monoftongar, som vist i figur 1.

i	y	u	ù
ì	ỳ		o
e	ø		ò
è	ø̇		
		å	
æ		a	

Figur 1: Ivar Aasens system over monoftongar i det norske folkemålet, her med venstrevend aksent over opne vokalar som i Aasens opphavlege plan.

Dette systemet kjende Aasen frå sitt eige sunnmørsmål, og han fann nok skiljet mellom opne og lukka vokalar i alle fall delvis gjennomført også i andre vestlandske målføre. Skiljet hadde samanheng med kvantitets- og lydtilhøve i norrønt og dei

endringane som skjedde i lydsystemet etter kvantitetsomlegginga. Dei opne vokalarne hadde utvikla seg frå vokalar som var korte i norrønt, men som hadde blitt forlengde under kvantitetsomlegginga. Lukka vokalar svara på si side til lange vokalar i norrønt (sjå t.d. Skjekkeland 1997:50f. og 56f.). Gjennom å vektleggje skiljet mellom opne og lukka vokalar fekk Aasen synt fram at målet i samtida stod i samband med gammalnorsk, noko som var eit kjernepunkt i Aasens historiske språksyn (jf. Aasen 1848:III f.).

Aasen ønskte i utgangspunktet å markere open/lukka-skilnaden i skrift ved å setje venstrevend aksent over dei opne vokalarne, som i figur 1. Eit argument for å markere dei opne, og ikkje dei lukka vokalarne, var at dette ville vere mest til hjelp for lesarane. Det var nemleg dei opne vokalarne som hadde ein uttale som skilde seg frå det folk kjende, og som dei dermed trong å bli gjort merksame på. Uttalen av lukka vokalar svara derimot om lag til vanleg leseuttale av bokstaveikna. (Jf. Aasen 1846:75)

Aasens plan møtte motstand frå fagfolka i Christiania. Dei peika på at aksent over opne vokalar braut med skrivemåten for gammalnorsk. Der brukte ein høgrevend aksent, og denne vart plassert over lukka vokalar. Aasens måte ville dermed vanskeleggjere jamføringa med gammalnorsk. Dei drog mellom anna fram den tyske grammatikaren Jacob Grimm, som også berre brukte aksent over lukka vokalar, til støtte for synet sitt. Saka enda med at Aasen gav opp planen sin, noko han gjorde greie for i brevet til Vullum 9.11.1847:

Med Hensyn til Vokalernes Betegning har jeg maattet  
bekvemme mig til en Omvendelse; man forestillede mig  
klart og tydeligt, at det ikke gik an at sætte Tegn over de  
aabne Vokaler, men kun over de lukte; man understøttede

sin Paastand ved samtlige nyere Sprogforskeres Autoritet, og forelagde mig den sidste Udgave af Grimms Grammatik, hvoraf der nemlig er udkommet et stort Bind, som handler alene om Vokalernes Lydforhold. Mod slige Kjæmper nytter det ikke at stride; jeg ansaae det raadeligst at give de aabne Vokaler en god Dag, og saaledes vil man da være den Uleilighed kvit. (Djupedal 1957:146)

I grammatikken lét Aasen vere å bruke aksentar i det heile i språklege døme, men i framstillinga av sjølv lydssystemet retta han seg etter Grimms system med høgrevend aksent over lukka vokalar. I ordboka kom derimot Aasen fram til ein annan måte å markere open og lukka vokal på. I 1850-ordboka introduserte han nemleg ein eigen informasjonskategori for open og lukka vokal. Desse vokaltypene vart markerte i parentesar rett etter oppslagsordet i tilfelle der han meinte det kunne vere tvil om uttalen, som i dei to homografane av verbet *liva* (der dobbel vokal i den andre homografen indikerer lukka uttale):

**liva** (aab. *i*), v.n. (*e-de*), leve. Udtales paa mange Steder *leva* og *læva*, ogsaa *lava* (Namd.) og *laavaa* (Indr. og Ørk.). G. N. *lifa*. - *leva av* (med Vægten paa *av*): overleve En; ogsa leve over en vis Tid. - *leva ihop*: leve sammen; især om Ægtefolk. - *liva mæ*: leve med; ogsaa drive meget Væsen med; f. Ex. *Dei livde mæ han!* dvs. de drev et svare Spektakel med ham. B. Stift. *liva utav (læva ta)*: leve af. *Dei livde væl*: de vare friske og befandt sig vel.

**liva** (*ii*), v.a. (*e-de*), spare, skaane. G.N. *hlifa*. I Særdeleshed: 1) beskytte, dække, f. Ex. mod Vind og Uveir. *Dær æ inkje som live*: der er aabent paa alle Sider. - 2) skaane, spare for Umage eller Ubhagelighed. *live seg*: tage det



mageligt, søge at undgaae Besværligheden. *Han æ 'kje te live*: man skal ikke spare ham. - 3) spare paa noget, tage sparsomt til; ogsaa være til Besparelse. *liva paa*: spare paa. *Dæ live paa Brø'e*: det er til Besparelse for Brødet. *Dei ha livt dæ av*: de have sparet det, faaet det tilovers ved at spare. - Meget udbredt Ord, og især meget brugeligt vestenfjelds. (Aasen 1850)

## Munch melder grammatikken og ordboka

Både grammatikken og ordboka vart melde av P.A. Munch i *Norsk Tidsskrift for Videnskap og Litteratur*. Meldingane er i hovudsak svært positive, og i baa beskriv Munch Aasens verk som nasjonalverk (etter Myhren 1975:24 og 40). Særleg dei første linene i meldinga av grammatikken er ofte dregne fram som prov på Munchs høge verdsetjing av Aasens arbeid: «Dette Skrift som vi her anmelde, er ikke alene en Prydelse for vor Litteratur, men det er et Nationalverk, af hvilket det hele folk kan være stolt». Munch la særleg vekt på at Aasen hadde fått fram einskapen i folkemålet og prova sambandet mellom gammalnorsken og det norske samtidsmålet. Men i baa meldingane kom også Munch inn på ei sak der han og Aasen hadde ulike syn. Munch meinte at gammalnorsk burde vere normgrunnlag for skrivemåten av orda i folkemålet, også i tilfelle der dei gammalnorske formene ikkje fanst i nokon dialekt i samtida. Ei slik radikal etymologisk line ville Aasen følgje korkje i 1848-grammatikken eller 1850-ordboka. I desse verka la Aasen vekt på at

skriftformene skulle spegle det faktiske samtidsmålet, og han ville i hovudsak skjegle til gammalnorsk når det var snakk om å velje mellom ulike eksisterande former (jf. Venås 2000:51f.).

Arbeidet med grammatikken frå 1848 og ordboka frå 1850 var ei vitenskapleg sveineprøve for Aasen, og universitetsfolka engasjerte seg altså som rådgjevarar og støttespelarar. Ikkje minst var meldingane til Munch ei definitiv anerkjenning av Aasen som vitenskapsmann. Etter utgjevinga av grammatikken og ordboka konsentrerte Aasen seg i større grad om å utvikle eit standardisert norsk skriftmål, *landsmålet*, som alternativ til dansk. Dette tiltaket, som kunne komme til å velte fullstendig om på den norske språksituasjonen, var naturleg nok kontroversielt, også i det akademiske miljøet. Eit sentralt spørsmål i det følgjande er kva haldningar fagfolka ved universitetet hadde til denne delen av Aasens prosjekt.

## **Landsmålet frå teoretisk prosjekt til bruksmål**

I dei kommande tiåra etablerte Aasen *landsmålet* gjennom tekstsamlinga *Prøver af Landsmaalet i Norge* (1853) og nyttgjeving av grammatikken, *Norsk Grammatik* (1864), og ordboka, *Norsk Ordbog* (1873), med status som nasjonale vitenskaplege standardverk. Aasen skapte dessutan litterære tekstar i landsmålet. Her var det særleg skodespelet *Ervingen* (1855) og diktsamlinga *Symra* (1863) som synt fram landsmålet som bruksmål. Hit kan også reknast *Norske Ordsprog* (1856).

## Universitetsfolka i møte med landsmålet som bruksmål

Språkbygginga der Ivar Aasen i utformingsfasen allierte seg med universitetsfolk, kan sjåast på som ei form for språkplanlegging. Etter Vikør (2007:109) har språkplanlegginga tre hovudfasar. Det er val av normgrunnlag, det er kodifisering og endeleg implementering eller iverksetjing med sin særlege samsfunnsrelevans. Aasen involverte universitetsfolka mest i kodifiseringsfasen. Spørsmålet er no korleis Munch og andre universitetsfolk stilte seg når implementeringa av landsmålet utfordra deira daglege arbeids- og privatspråk.

Allereie i 1853 vende P.A. Munch tommelen ned. Dette nasjonalmålet hadde enda ikkje nådd eit utviklingssteg som «passer for den moderne Videnskabelighed», hermer Venås (2000:53) i *Ivar Aasen og Universitetet* etter Munch frå ein polemikk i *Morgenbladet*.

Professor Sophus Bugge (1833–1907) reknar Venås (2000:68–69) som ein av dei største humanistiske granskarane som har vore ved Universitetet i Oslo. Bugge sa desse minneorda under ei førelising:

Vi kjender alle den store mand, som gik bort fra os i gaar – Ivar Aasen. Vi mindes ham ikke bare som den store videnskabsmand, men ogsaa for hans store betydning for vor nasjonalitets udvikling. Derfor er vi ham alle i endrægtighed taknemlige. (Etter Venås 2000:71)

Tre år seinare var Bugge hovudtalar på eit riksmåls møte 29.11.1899. Der framførte han dette avvisande synspunktet, som var mynta på landsmålet:

Jeg vilde aldrig kunne billige Kravet paa to Sprog med lige Ret i Norge; men jeg vilde kunne finde et slikt Krav naturligt, hvis dets Opnaaelse var mulig. *Det er den ikke.* (Etter Venås 2000:72)

Johan Storm (1836–1920), som frå 1873 var professor i romansk og engelsk filologi, interesserte seg også for beskrivinga av norske målføre. *Kortere Ordliste* med tillegget *Norsk Lydskrift med kort Forklaring* (publisert i 1884) inneheld valde døme på ord som til saman skal få fram så vel den fonologiske som den morfologiske strukturen i norske målføre. Storm la ordlista fram for Aasen, som tykte at dette tiltaket var «en heldig Tanke» (Venås 2000:79). Men jamvel om dei hadde denne faglege respekten for kvarandre, var Storm negativ til bruksverdien av landsmålet. I «Det norske Maalstrev» (1878) klaga han på at landsmålet i bruk ikkje hadde den einskap og konsekvens som eit skriftmål skulle ha. Imot landsmålet sette han dertil opp dansk-norsken som stetta grunnleggjande krav, nemleg tradisjon, klassisitet og kontinuitet (Venås 2000:81).

Desse tre framståande universitetsfolka stod for ein tradisjonalisme som nok var representativ for fleirtalet av dei i språk- og kulturspørsmål på denne tida (jf. Vikør 2007:184). Godtakinga av språknormer er i stor grad avhengig av faktorar utanfor det språkvitskaplege feltet (jf. Vikør 2007:126), og i denne samanhengen ville ikkje desse universitetsfolka gje prestisje til det normsynet som låg til grunn for landsmålet, nemleg eit skriftspråk som bygde på vanleg talemål og ikkje på ein litterær tradisjon. Men denne avvisande haldninga hindra likevel ikkje at landsmålet vart forskingsemne ved universitetet.

## Aasen-prosjektet på universitetet etter Aasen

Med landsmålet som ein realitet i det norske samfunnet kunne det også verte forskingsemne, og det aktualiserte dermed spørsmålet om eit professorat. Ein delegasjon av to leiande studentar skal ha oppmoda Aasen om å stille seg til rådvelde for ei slik stilling: «Synes De jeg ser ud til aa staa paa et kateder, jeg?», var svaret. Det skal ha skjedd i 1877 (Venås 2000:36).

Professorat vart det likevel. Etter førehavingar i Stortinget og på universitetet vart det vedteke å skipe «et bevægeligt professorembede i landsmaal og dets dialekter» frå 1899. Det stod ved lag til 2013. Stortinget sørge i 1972 for at også den leksikografiske tradisjonen frå Aasen vart ført vidare på universitetet i Norsk leksikografisk institutt. Prosjektet *Norsk Ordbok*.

*Ordbok over det norske folkemålet og det nynorske skriftmålet* var ein av grunnpilarane da instituttet vart skipa, og byggjer på Ivar Aasens leksikografiske pionerarbeid. I 2002 vart dette prosjektet fornya gjennom ein eigen prosjektorganisasjon med status som grunnlovsjubileumsprosjekt og med omfattande økonomisk støtte frå Kulturdepartementet, der universitetet er garantist både for fagleg kvalitet og nyutvikling. Utgjevinga har gått planfast mot slutføring av det redaksjonelle arbeidet i 2014 og publisering av det tolvte og siste bandet.

Denne utgjevingsmåten har ikkje endra seg dramatisk etter utgjevinga av 1850-ordboka. Venås (1996:119) slår jamvel fast at denne ordboka ser ut slik som vi er vane med at ordbøker skal vere med eigne artiklar om kvart ord i alfabetisk rekkjefølgje. I Noreg har det såleis vore ein etablert leksikografisk tradisjon i meir 160 år. Men med 2002-prosjektet kom *Norsk Ordbok* over på ei digital plattform. Med språkteknologien som

medspelar i leksikografien har vilkåra for materialtilgang og presentasjonsformer av informasjonskategoriane i ordartikkelen endra seg radikalt. I botnen er det likevel ein grunnleggjande føresetnad, som for *Ordbog over det norske Folkesprog*, at beskrivinga i den einskilde ordartikkelen byggjer på inngåande kunnskap i språkvitskap, og forskingsansvaret på dette feltet er det no som før universitetet som forvaltar. Ivar Aasen var fram-synt her.

## Litteratur

- Collett, John Peter (2011): *Universitetet i Oslo 1811–2011. Bok 1. 1811–1870. Universitetet i nasjonen*. Oslo: Unipub.
- Djupeadal, Reidar (1957): *Ivar Aasen. Brev og Dagbøker. Band 1*. Oslo: Det Norske Samlaget.
- Djupeadal, Reidar (1960): *Ivar Aasen. Brev og Dagbøker. Band 3*. Oslo: Det Norske Samlaget.
- Hyvik, Jens Johan (2009): *Språk og nasjon 1739–1868*. Oslo: Det Norske Samlaget.
- Myhren, Magne (red.) (1975): *Ei bok om Ivar Aasen. Språkgranskaren og målreisaren*. Oslo: Det Norske Samlaget.
- Skjekkeland, Martin (1997): *Dei norske dialektane: Tradisjonelle særdrag i jamføring med skriftmåla*. Kristiansand: Høyskoleforlaget.
- Venås, Kjell (1996): *Då tida var fullkomen. Ivar Aasen*. Oslo: Novus forlag.
- Venås, Kjell (2000): *Ivar Aasen og Universitetet*. Oslo: Universitetet i Oslo.

- Vikør, Lars S. (2000): Ordbøker over norske bygdemål før 1814. I: *LexicoNordica* 7, 119–132.
- Vikør, Lars S. (2007): *Språkplanlegging. Prinsipp og praksis*. Oslo: Novus forlag.
- Walton, Stephen (1996): *Ivar Aasens kropp*. Oslo: Det Norske Samlaget.
- Aasen, Ivar (1846): Om en Ordbog over det norske Almuesprog samt en dertil hørende Grammatik. I: *Det Kongelige Norske Videnskabers Selskabs Skrifter i det 19de Aarhundrede, nr. 4*, 55–96.
- Aasen, Ivar (1848): *Det norske Folkesprogs Grammatik*. Kristiania: Werner & Comp.
- Aasen, Ivar (1850): *Ordbog over det norske Folkesprog*. Kristiania: Carl C. Werner & Comp.
- Aasen, Ivar (1912): *Skrifter i Samling. III*. Kristiania og København. Gyldendalske Boghandel. Nordisk Forlag.

Kjetil Gundersen

redaktør

Norsk Ordbok 2014, Institutt for lingvistiske og nordiske studium, Universitetet i Oslo

Gaustadalleen 30 D

0373 Oslo

kjetil.gundersen@iln.uio.no

Dagfinn Worren

hovudredaktør

Norsk Ordbok 2014, Institutt for lingvistiske og nordiske studium, Universitetet i Oslo

Gaustadalleen 30 D

0373 Oslo

dagfinn.worren@iln.uio.no

## **Kvensk ordbok på nett – hvem har nytte av den?**

*Mervi Haavisto, Kaisa Maliniemi, Leena Niiranen, Pirjo Paavalniemi, Tove Reibo og Trond Trosterud*

The article presents a co-operation project between the University of Tromsø, the Kven Institute and Halti Kvenkultursenter, with a goal to create a digital dictionary with a morphological analyser, which is an important tool for language revitalization. Even though the digital dictionary is not finished yet, it can already help language users to read and write Kven. We discuss how such a dictionary can be made efficient, and which groups of language users will benefit most from it. We will try to explain how this project contributes to Kven language planning.

### **1. Innledning**

I denne artikkelen presenterer vi et samarbeidsprosjekt mellom UiT, Kainun institutti – Kvensk institutt og Halti kvenkultursenter. Vi lager digitale hjelpemidler for brukere av kvensk skriftspråk. Artikkelen gir først en kort oversikt over det kvenske språket og etableringen av skriftnormen i kvensk. Vi vil ha spesielt fokus på hvordan digitale verktøy i kvensk kan gjøres effektive, og hvilke brukergrupper som vil få den største nytten av utviklingen av slike verktøy.



Vi vil også drøfte hvordan arbeidet med digitale hjelpemidler forholder seg til korpusplanlegging og statusplanlegging (Kloss 1967; Vikør 1994). Begge disse anses for å være viktige i språkrevitalisering. Kort definert kan det bety at man snur et pågående språkskifte: et truet språk læres på nytt, det etableres lese- og skriveferdigheter i et slikt språk, og nye prestisjefulle bruksdomener tas i bruk (Hiss 2012, se også Fishman 2001; Huss 1999).

Imidlertid er språkrevitalisering en komplisert prosess, der blant annet språkideologi spiller en rolle på grunn av at den påvirker hvordan man oppfatter hva begrepet ”språk” inkluderer (Gal 2006). Gal beskriver språkideologi som politiske, moralske og estetiske holdninger hos aktører (f.eks. stat, skole, media, eller forskjellige kulturinstitusjoner) som former språklige holdninger og praksis hos språkbrukere, og som samtidig former deres kulturelle forventninger til hvordan begrepet ”språk” oppfattes. Tidligere språkideologi som var knyttet til nasjonalstat og nasjonalspråk, verdsatte verken minoritetsspråk eller bruk av dem, og var derfor en av årsakene til at mange kvener skiftet til majoritetsspråket (Lane 2010). Revitalisering representerer en ny språkideologi, der man ønsker å øke bruken av truede minoritetsspråk. Tross at ideologien har endret seg, og språklige holdninger til minoritetsspråk er blitt mer positive enn tidligere, er det nødvendigvis ikke noen direkte sammenheng mellom positive holdninger og endret språklig atferd med økt bruk av minoritetsspråk (King 2000). Vi drøfter både muligheter og problemer knyttet til å få økt bruk av kvensk ved hjelp av digitale hjelpemidler.

## 2. Hva er kvensk språk?

Kvensk er et av minoritetsspråkene i Norge. Det fikk status som nasjonalt minoritetsspråk i 2005 gjennom ratifiseringen av Europarådets minoritetsspråkpakt. Kvensk er nært beslektet med nordfinske dialekter og med meänkieli i Tornedalen, og er i likhet med disse et agglutinerende, morfologirikt språk (Lane 2006). Kvensk kan deles i vestlige og østlige dialektvarianter som skiller seg fra hverandre spesielt når det gjelder en del morfologiske bøyningsformer (Lindgren 1993).

Forskjellene mellom østlige og vestlige dialekter reflekterer den kvenske bosetningen på tilsvarende områder i Nord-Norge. Den vestlige kvenske bosetningen er eldre enn den østlige, og kan dateres tilbake til 1600- og 1700-tallet. Disse bosetterne kom hovedsakelig fra Tornedalen, et geografisk område delt mellom Sverige og Finland. Mens denne bosetningen var en bondekolonisasjon som spredte seg nordover, var den østlige kvenske bosetningen primært et resultat av en arbeidsvandring. De sistnevnte kom fra hele Nord-Finland; i tillegg flyttet det folk også fra andre deler av Finland til Nord-Norge (Niemi 2010). Kvenene som allerede var bosatt i Norge, kunne også flytte videre østover; et eksempel på dette er kvenske gruvearbeidere i Kåfjord i Alta som flyttet til Øst-Finnmark, etter at gruvevirksomheten i Kåfjord tok slutt på slutten av 1800-tallet (Nielsen 1995).

Forskjellene mellom vestlige og østlige dialekter kan derfor forklares med forskjell mellom bosetninger i disse områdene ikke bare når det gjelder flyttetiden, men også stedet kvenene flyttet fra. I tillegg spiller språkkontakter som ble etablert i Norge, en rolle for hvordan kvenske talemål utviklet seg videre.

I Øst-Finnmark holdt bosetningene kontakt med hjemlandet lenge, mens på vestlige områder, spesielt i områder som ikke hadde grense mot Finland, ble slike kontakter tidlig brutt (Maliniemi 2010). Kvenene kom i kontakt med både norsk og samisk språk, og det er stor lokal variasjon i utviklingen av tospråklighet og eventuelt trespråklighet (Lindgren 2009).



© KAINUN INSTITUTTI - KVENSK INSTITUTT

Bilde 1: Kvenske dialekter. Dagens ordbok og analysator bygger på de vestlige dialektene.

I dag regnes det med at det finnes ca. 4000–8000 brukere av kvensk talespråk (Lane 2006). De fleste som bruker kvensk daglig, er eldre mennesker som snakker østlige dialekter. Likevel har vi begynt vårt arbeid med å lage en digital ordbok basert på den vestlige varianten fordi den er brukt mer i skrift enn de andre variantene. Ikke minst har Alf Nilsen-Børsskogs

forfatterskap økt den skriftlige bruken av Porsanger-varianten (Söderholm 2008). Vi planlegger å ta hensyn til den østlige dialektvarianten senere i vårt arbeid.

### **3. Korpusplanlegging – utvikling av kvensk skriftspråk**

Etter at kvensk fikk status som nasjonalt minoritetsspråk i Norge, begynte arbeidet med å utvikle det til et skriftspråk. Språk som bare brukes muntlig, har problemer med å overleve i dagens samfunn der bruk av skriftspråk er viktig i mange funksjoner. Etablering av skriftspråk hører til korpusplanlegging av et språk, og sammen med statusplanlegging anses korpusplanlegging å være viktig for revitaliseringen av et språk (sml. Huss 1999).

Kvensk språkting ble etablert i 2008. Språktinget administreres av Kvensk institutt, som har ansvar for standardiseringen av kvensk. Språktinget har etablert tre ulike normer for skriftlig kvensk. En av disse er basert på dialekter i Porsanger og Nordreisa (vestvarianten), en på dialekter i Varanger (austvarianten), og en er basert på dialektene mellom disse områdene (jokivarsi/elvedal-varianten). Den sistnevnte varianten er den som f.eks. er i bruk i Alta og Tana, og den ligger nærmest tonedalsfinsk/meänkieli (Söderholm 2009).

Hvorfor har det kvenske språktinget ønsket å etablere flere parallelle skriftnormer for et språk som har få talere? Hadde det ikke vært enklere å satse bare på én skriftlig variant? Den

nyeste forskningen innen språkrevitalisering har imidlertid problematisert holdninger der fokuset ligger ukritisk på språk samtidig som man glemmer språkbrukere. Monica Heller & Alexandre Duchêne påstår at diskursen om truede språk er konsentrert om å se språk som «an organic, systematic whole which has a life of its own outside of social practice» (2007: 6). Nettopp her ligger en fare for essensialisering og språklig puritanisme der man konstituerer et standardisert språk tilpasset til en viss kultur eller/og en viss nasjon.

Hvis man glemmer språkbrukerne, kan standardiseringen bare øke avstanden mellom språkbrukerne og den normerte skriftlige varianten, og den kan til og med påvirke språklige holdninger til det negative. Om man derimot presenterer flere normer, er det lettere for språkbrukerne å akseptere en skriftlig norm for sitt talemål når de kan finne flest mulig trekk som ligner deres muntlige språk. Avvisning av en skriftnorm har også vært registrert i tilfeller der minoriteter har blitt servert en skriftnorm som brukes som nasjonalspråk i nabolandet (Wright 2007). Derfor er bruk av finsk som skriftnorm problematisk for mange kvener, selv om andre igjen mener at finsk kunne ha vært brukt som skriftnorm i Norge (Niiranen 2011).

Vårt arbeid med å lage digitale verktøy for kvensk baserer seg på arbeidet i Kvensk språkting og på Söderholms grammatikk (2012). En fornyet og større grammatikk skrevet av Söderholm vil komme ut i slutten av 2013. Denne grammatikken vil til en viss grad vise til språklig variasjon mellom de tre variantene som Kvensk språkråd har valgt. Selv om vi bruker hennes grammatiske analyse som utgangspunkt for den morfologiske analysen som brukes i analysatoren, må vi ofte finjustere reglene eller endre dem slik at de kan brukes i databasert språkanalyse. I tillegg vil arbeid med en digital ordbok i fremtiden også handle om terminologiutvikling for

kvensk. På denne måten bidrar vårt arbeid til kvensk korpus-planlegging.

#### 4. Digitale verktøy for kvensk

Det fins en kvensk nettbasert ordbok (Aronsen 2010), men som de fleste ordbøker inneholder den bare grunnformen av ordene. I løpende tekst blir kvenske ord bøyd i 90 % av tilfellene, og for å finne grunnformen til et bøyd ord behøver man hjelp av et analyseprogram. Vi setter ordboka og analyseprogrammet sammen i grensesnittet Neahttadigisánit (Johnson et al. 2013).

Den kvenske morfologiske analysatoren inneholder foreløpig få ord, ca. 2800 substantiver og 1000 verb. I tillegg er pronomen lagt inn, og arbeidet fortsetter med å legge inn adjektiver. Vi arbeider også med å skaffe mer korpus for å utvide ordforrådet i analysatoren.

Bilde 2 viser hvordan analysatoren fungerer. Den morfologiske analysatoren er en såkalt endelig tilstandstransduser (finite state transducer, FST). Grammatikken blir modellert som et avgrenset (endelig) sett av symbolpar. En slik maskin kan f.eks. sammenligne og kjenne igjen to strenger med forskjellige symboler (Koskenniemi 2004). I praksis vil disse strengparene være par av lemma + grammatisk analyse og ordformer. Et eksempel kan være 'stupe+V+Pret : stupte'. Bilde 2 viser hvordan kvensk morfologi og fonologi danner hver sin transduser, som til sammen utgjør det grammatiske analyseprogrammet. (jf. Beesley & Karttunen 2003).

kala+N+Sg+Par	Grammatisk representasjon.
kala>^A	Stamme + suffiksgrense + ^A (fellesvokal for a/ä).
kala>^A	Output fra leksikalsk automat er input til den fonologiske.
kallaa	Vokalharmoni: ^A -> a etter a. Geminering: Dobbel l foran lang aa.

Bilde 2: Hvordan analysatoren fungerer

#### 4.1. Morfologisk analysator som hjelpemiddel for å lese kvenske tekster

Ordbok i kvensk kan tas i bruk ved lesing av digitale tekster. Ved ordklikk kan man få fram ordets betydning (se bilde 3). En slik digital ordbok vil være lett tilgjengelig for brukere, og etter at ordboka blir utviklet til også å vise ordets bøyninger, vil den dekke et mye større antall ord i løpende tekst sammenlignet med en ordbok basert bare på grunnformer.

)

onflikthiin kaksi vuotta aikkaa. Sen **tähden** se joutui ottamha nyt progammit oon taas netissä. Paitsi **^RK**:n suomenkieline

ratkaisun  
ti met uuðistimma  
Met häy?ymä keks  
isteele.  
lee keskelä työpäiv  
nän sannoo.

**tähden**

*tähti* (s.) — stjerne

*tähden* (Po) — på grunn av

Bilde 3: Løpende tekst med ordklikk og popup-vindu fra ordboka

I analysatoren ligger alle ordene fra ordboka, og i tillegg har vi lagt inn mange flere ord. Om man sammenligner to teksttyper (se tabell 1), dekker analysatoren mer av ordforrådet i en tradisjonell tekst enn i en moderne tekst. Ordforrådet i ordboka og analysatoren mangler altså fortsatt mange moderne ord. For at leseren kan forstå en tekst, må 95 % av ordene være forstått ifølge Nation & Waring (1997). Selv om dekningsgraden foreløpig ikke er så høy, er tekstens lesbarhet likevel vesentlig forhøyet med analysatoren sammenlignet med at man bare bruker en ordbok uten bøyingsformer, slik tabell 1 viser.

Tradisjonell tekst (eventyr)	680	80,6 %	64,4 %	25,7 %
Moderne tekst (Ruijan Kaiku)	5 633	70,0 %	52,9 %	33,4 %
Dekningsgrad for ulike tekstsjangre	<b>Totalt</b>	I analysator	I ordbok med analysator	I ordbok uten analysator

Tabell 1: Ordbokas dekningsgrad for to teksttyper

#### 4.2. Hvordan kan en digital ordbok gjøres effektiv?

Analysatoren kan allerede hjelpe leseren til å forstå ord i løpende tekst, selv om den ikke er ferdig. Ordboka håndterer også dynamiske sammensetninger som ikke er lagt inn i den, og på denne måten økes antall ord som kan oversettes (se tabell 2). Forleddet i kvenske sammensetninger kan stå i nominativ eller genitiv, og analysatoren kan lese begge typer.



Analysatoren gjør det også mulig å behandle ortografisk variasjon: Den kvenske analysatoren aksepterer *d* for *ḏ* (saada/saada ‘å få’) og *æ/ø* for *ä/ö* (mæki/mäki ‘bakke’). Det er mulig å legge inn også f.eks. morfologisk variasjon. Som nevnt er variasjon noe som preger kvensk på grunn av at normeringen av skriftspråket ikke har eksistert tidligere.

Ortografisk variasjon	Dynamisk sammensetning	%
Nei	Nei	54,2 %
Ja	Nei	54,6 %
Ja	Ja	55,9 %

Tabell 2: Dekningsgrad med og uten ortografisk variasjon og dynamisk sammensetning. Forskjellene er små fordi teksten inneholder få dynamiske sammensetninger og lite ortografisk variasjon.

### 4.3. Framtidige muligheter for analysatoren

Analysatoren gjør det mulig å utvikle interaktive oppgaver for språklæring. Selv om noen oppgavetyper vil kreve en syntaktisk maskinbasert analyse av kvensk, er det med utgangspunkt i den nåværende analysatoren mulig å utvikle oppgaver om ordbøyninger, et av de krevende læringsmålene i et morfologirikt språk.

Morfologisk analysator gir også grunnlag for å lage et retteprogram i kvensk. De fleste majoritetsspråk har en slik ressurs, men mange mindre språk mangler et slikt program. For minoritetsspråklige er slike ressurser særlig viktig, fordi de ikke har lært skriftspråket på skolen.

I tillegg kan analysatoren brukes i fremtidig kvensk terminologiutvikling. Dette forutsetter at korpus av forskjellige teksttyper kan etableres. Det er et behov for å utvikle ny terminologi i kvensk for å dekke områder som har med det moderne samfunnet å gjøre (Söderholm 2007).

Kvensk språkforskning vil i tillegg kunne benytte et digitalisert korpus for kvensk. Digitaliseringen innebærer at store mengder data kan analyseres for å undersøke ulike fenomener i språket. For eksempel kunne man analysere kollokasjoner eller ord som ofte brukes sammen. Kollokasjoner skaper ofte problemer for de som ikke behersker språket som sitt morsmål (Jantunen 2009).

## **5. Statusplanlegging – utvidet bruk av kvensk**

Ordboka er nyttig for språkbrukere som kan snakke kvensk, men ikke skrive det. Det er selvfølgelig problematisk at mange av disse er eldre personer som ikke er vant med å bruke datateknologi. Størst utbytte av en digital kvensk ordbok har derfor de som vil lære seg kvensk og ta språket i bruk, og som mangler både ordforråd og kunnskap om ordbøyning. Ordboka vil gi dem verdifull hjelp til å kunne forstå kvensk og ikke minst til å kunne produsere egne tekster. Fordi ordboka kan forventes å utvide bruken av kvensk i samfunnet, bidrar den til statusplanleggingen for kvensk.

Selv om utvidet bruk av kvensk i samfunnet er målet for vårt arbeid, kan vi ikke være sikre på om alle grupper av språk-

brukere ønsker å ta elektroniske hjelpemidler, slik som ordboka, i bruk. Elever i grunnskolen og i den videregående skolen kan velge kvensk som andrespråk i Troms og Finnmark, men det er ikke sikkert at disse hjelpemidlene inspirerer ungdommer til å lære mer kvensk. Selv om mange ungdommer er opptatt av moderne elektroniske kaneler, som internett, møter de der norsk og engelsk. Samtidig påvirker moderne teknologi deres språk, språkbruk og språklige holdninger, og ungdommer er gjennom denne teknologien utsatt for en språkideologi som presenterer hvilke språk det er viktig å beherske i en globalisert verden. I en slik situasjon kan forfedres språk, som ikke lenger er deres morsmål, virke gammeldags og umoderne. Liknende problemer finner man også hos andre etniske minoriteter der ungdommer ønsker å være urbane og moderne, og derfor heller velger å lære et av verdensspråkene i skolen i stedet for forfedrenes eller til og med foreldres språk (Gal 2006).

Dette prosjektet ønsker å være en del av revitaliseringsprosjektet som presenterer en motbevegelse mot andre språkideologier. Til tross for at prosjektet kan møte problemer i strid mellom forskjellige språkideologier, er det viktig å satse på språkteknologiprojekter samtidig som man må aktivt kjempe mot negative eller fordomsfulle holdninger mot minoritetsspråk. Språkteknologi kan hjelpe det kvenske språket til å bli oppfattet til å høre til den moderne verden og kan derfor løfte dets status i samfunnet.

## 6. Konklusjon

En digital ordbok for et minoritetsspråk er et verktøy som kan brukes på sentrale områder av språkrevitaliseringa. Den er et verktøy som kan hjelpe minoritetsmedlemmer å etablere lese- og skriveferdigheter. Bruk av teknologi betyr i tillegg at nye og prestisjefulle domener tas i bruk. Vi har argumentert for at en digital ordbok er en del av både korpusplanleggingen og statusplanleggingen for kvensk. En grammatisk analysator, selv med dårlig dekningsgrad, bidrar til en fungerende elektronisk resepsjonsordbok. For morfologirike språk kan intelligente ordbøker som denne derfor spille en sentral rolle i språkrevitaliseringa.

Dette prosjektet har også som mål å være et holdnings- og skapende uttrykk i forsvar for et minoritetsspråk i et samfunn der det eksisterer strid mellom forskjellige språkholdninger og -ideologier. Vi ønsker å få kvensk til å bli sett som et interessant og levende språk som har framtid.

## Litteratur

### Ordbøker

Aronsen, Terje (2010): *Kvensk-norsk-kvensk elektronisk ordbok*. Redigert av Verena Schall og Trond Trosterud. Universitetet i Tromsø.

## Annen litteratur

- Fishman, Joshua A. (2001): *Can threatened languages be saved? Reversing language shift revisited. A 21st century perspective*. Clevedon [England]: Multilingual Matters.
- Gal, Susan (2006): Contradiction of standard language in Europe: Implications for the study of practices and publics. I: *Social Antropology 14. Vol. 2*, 163–181.
- Heller, Alexandre & Monica Duchêne (2007): Discourses of endangerment: Sociolinguistics, globalization and social order. I : A. Heller & M. Duchêne (red.): *Discourses og endargement: Ideology and interest in the defence of languages*. London: Continuum.
- Hiss, Florian (2012): *Encountering Complexity. Hybrid Discourse and Individual Management of Multilingualism and Social Meaning*. [Tromsø]: University of Tromsø, Faculty of Humanities, Social Sciences and Education, Department of Language and Linguistics.
- Huss, Leena (1999): *Reversing Language Shift in the far North: Linguistic revitalization in Northern Scandinavia and Finland*. Uppsala: Almqvist & Wiksell.
- Jantunen, Jarmo Harri (2009): Minulla on aivan paljon rahaa - fraseologiset yksiköt suomen kielen opetuksessa. I: *Virittäjä 3/2009*, 356–381.
- Johnson, Ryan, Lene Antonsen & Trond Trosterud (2013): Using finite state transducers for making efficient reading comprehension dictionaries. I: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013), May 22–24, 2013, Oslo University, Norway*, 59–71.
- King, Kendall A. (2000): Language Ideologies and Heritage Language Education. I: *International Journal of Bilingual*

- Education and Bilingualism, Vol.3, No.3, 167–184.*
- Kloss, Heinz (1967). Abstand languages and Ausbau languages. I: *Anthropological Linguistics 9 (7)*, 29–41.
- Koskenniemi, Kimmo (2004): *Käsittehakemisto*. FST  
<<http://www.ling.helsinki.fi/kit/2004s/terms-en.shtml>>
- Lane, Pia (2006): *A Tale of Towns: A Comparative Study of Language and Culture contact*. [Oslo]: Faculty of Humanities, University of Oslo.
- Lane, Pia (2010): “We did what we thought was best for our children”: a nexus analysis of Language. I: *International Journal of Language, Vol. 2010, Issue 202*, 63–78.
- Lindgren, Anna-Riitta (1993): *Miten muodot muuttuvat: ruijan murteiden verbintaivutus Raisin, Pyssyjoen ja Annijoen kveenyhteisöissä*. [Tromsø]: Universitetet i Tromsø.
- Lindgren, Anna-Riitta (2009): Kvensk i Norge. I: T. Bull & A.-R. Lindgren (red.): *De mange språk i Norge. Fler-språklighet på norsk*. Oslo: Novus forlag, 107–124.
- Maliniemi, Kaisa (2010): *Hva arkivene skjulte: en undersøkelse av kvensk og samisk i offentlige arkiver i Kistrand (Porsanger) og Nordreisa 1865-1948*. Oslo. ABM-media.
- Nation, Paul & Robert Waring (1997): Vocabulary size, text coverage and word lists. I: N. Schmitt og M. McCarthy (red.): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge Language Teaching Library. Cambridge: Cambridge University Press, 6–19.
- Nielsen, Jens Petter (1990): *Det arktiske Italia. Atlas historie, bind 2*. Alta: Alta kommune.
- Niemi, Einar (2010): Kvenene – Nord-Norges finner. En historisk oversikt. I: A. Bonnevie Lund & B. Bolme Moen (red.): *Nasjonale minoriteter i det flerkulturelle Norge*. Trondheim: Tapir akademisk forlag, 33–52.
- Niiranen, Leena (2011): Finsk som andrespråk i norsk skole –

- fagets historie og dagens utfordringer. I: *NOA norsk som andrespråk, årgang 27, 1/2011*, 59–85.
- Söderhom, Eira (2007): Planlegging av kvensk språk - utvikling av ordforråd. I: A.R. Lindgren, E. Niemi, M.A. Hauan, L. Niiranen & T. Thuen (red.): *Kvener og skogfinner i fortid og nåtid. Rapport fra seminaret "Kvener og skogfinner i fortid og nåtid - identitetsforvaltning og straegier", Vadsø oktober 2005*. Tromsø: [Institutt for historie, Universitetet i Tromsø]. 43–50.
- Söderholm, Eira (2008): Som perler i et kjede. Språksplittelse på Nordkalotten og eget skriftspråk for kvener. I: E. Niemi & M. A. Hauan (red.): *Kvenene – i går og i dag*. Tromsø: Tromsø Museum – Universitetsmuseet. 20–26.
- Söderholm, Eira (2009): *Kielitingan ensi kokkouksessa vuona 2008 ehdotetuitten kielivariantiitten muutamii trekkii, mitä ei ole suomen kirjakielessä*. Upublisert notat.
- Söderholm, Eira (2012): *En liten kvensk grammatikk. Vesta – Porsanger – Nordreisa -varianten*. Ms, UiT.
- Vikør, Lars (1994): *Språkplanlegging. Prinsipp og praksis*. Oslo: Novus forlag.
- Wright, Sue (2007): The right to speak one's own language: Reflections on theory and practice. I: *Language Policy vol. 6 No. 2*, 203–224.

Mervi Haavisto  
informasjons- og kulturmedarbeider  
Kvensk institutt  
Kventunet  
N-9716 Børselv  
mervi.haavisto@kvenskinstitutt.no

HAAVISTO, MALINIEMI, NIIRANEN, PAAVALNIEMI, REIBO OG TROSTERUD

Kaisa Maliniemi  
prosjektleder, dr. art.  
Kvensk institutt  
Kventunet  
N-9716 Børselv  
kaisa.maliniemi@kvenskinstitutt.no

Leena Niiranen  
professor  
Universitetet i Tromsø  
Hansine Hansens veg 18,  
N- 9019 Tromsø  
Leena.Niiranen@uit.no

Pirjo Paavaliemi  
språk- og kulturmedarbeider  
Kvensk institutt  
Kventunet  
N-9716 Børselv  
pirjo.paavaliemi@kvenskinstitutt.no

Tove Reibo  
språkmedarbeider  
Halti kvenkultursenter IKS  
Hovedveien 2  
N-9151 Storslett  
tovereibo@gmail.com

Trond Trosterud  
professor  
Universitetet i Tromsø  
Hansine Hansens veg 18,  
N- 9019 Tromsø  
Trond.Trosterud@uit.no



## **Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi**

*Anna Helga Hannesdóttir*

The source language–target language dichotomy is not valid in the electronic dictionary, where information units in the target language are just as accessible as units in the source language. In this article, the consequences of multiple access in bilingual lexicography and its implications for dictionary typology are discussed. Focus lies on the representation of the target language vocabulary – more precisely on the coverage of the Swedish target language in the online Icelandic–Swedish *Islex* dictionary. The equivalent lacunas observed are partly due to discrepancies in the lexical structure of the languages involved, partly to the outdated role played by Swedish as target language in the dictionary.

### **1. Bakgrund**

Ordbokens underhållningsvärde är underskattat. Ett undantag i det sammanhanget utgör dr Johnsons *A Dictionary of the English Language* (1755). Den och dess upphovsman spelar en viktig roll i ett avsnitt av den brittiska TV-serien *The Black Adder*. Det är ordbokens ordförråd som står för de komiska

effekterna, eller snarare lexikografens bryderi när han konfronteras med det ena påhittade ordet efter det andra som uppenbarligen saknas i hans ordbok. När han till slut konstaterar att han faktiskt har missat ordet *sausage* bryter han mer eller mindre ihop. Hela scenen är ganska bisarr.

Även om dr Johnsons förtvivlan nog överdrevs är det försmädligt att konstatera att ord som borde ha behandlats i ordboken faktiskt saknas. Lemmalakuner hör emellertid till lexikografin, ingen ordbok omfattar hela ordförrådet. Vissa ord bedöms av lexikografen som för triviala, speciella eller marginella för att beredas plats i ordboken, andra förbises. I den färdiga ordboken lyser de hur som helst med sin frånvaro. Så har det alltid varit.

\*

I det följande diskuteras de teoretiska och metodiska konsekvenserna för den tvåspråkiga lexikografen av de utökade sökmöjligheter och den åtkomlighet som den elektroniska ordboken erbjuder. Relationen mellan källspråk och målspråk har förändrats och kategorier som bruksexempel, idiom och kollokationer kan utnyttjas i nya funktioner. Problemet illustreras i en liten undersökning av täckningsgraden av en avgränsad lexikalisk subgrupp inom det svenska ordförråd som finns som sökbara enheter i den isländsk-svenska nätordboken i *Islex*.

## 2. Källspråk och målspråk

I den elektroniska ordboken är det inte nödvändigtvis bara lemmarna som är sökbara och som därmed bildar ingångar i ordbokstexten. Störst konsekvenser får denna sökbarhet för den tvåspråkiga lexikografin. Nu kan inte bara lemmat utan även ekvivalenten bilda en ingång i ordbokstexten. Den tvåspråkiga ordboken kan göras såväl bidirektionell och bifunktionell som biskopal. Bidirektionella och bifunktionella ordböcker, dvs. ordböcker som är avsedda både för källspråkstalare och målspråkstalare och både för källspråkstalets produktion på målspråket och målspråkstalets avkodning av källspråket (figur 1), är långt ifrån ovanliga – särskilt mellan marknadsmässigt små språk som de nordiska. Dessutom är de som regel avsedda att tjäna som hjälpmedel vid all typ av avkodning och produktion, dvs. även vid all slags översättning av text.

<b>Funktion</b>	<b>Källspråk</b>		<b>Målspråk</b>
produktion:	L1	>	L2
avkodning:	L2	>	L1

Figur 1: Den konventionella tvåspråkiga ordbokens språk och dess funktioner i relation till källspråkstalare (L1) och målspråkstalare (L2).

I biskopala ordböcker, dvs. ordböcker «der hvert av språkene er både kildespråk og målspråk» (NLO 1997:86), vanligen ordnade i två separata delar, har emellertid en strikt rollfördelning rätt mellan de båda språken, eftersom «oavsett om ordboken är mono- eller bidirektionell, går den ju i enbart en språkriktning, från ett källspråk till ett målspråk» (Svensén 2004:37, not).

### 3. Primärspråk och sekundärspråk?

De kriterier som ligger till grund för den av Kromann et al. etablerade typologin över tvåspråkiga ordböcker räcker inte längre till (Kromann et al. 1984 passim). I och med att målspråksenheterna blir åtkomliga framstår främst uppdelningen i källspråk och målspråk som en svaghet. Dikotomin källspråk–målspråk är relevant enbart i den tryckta, monodirektionella ordboken, där lemmat utgör den enda ingången i artikeltexten. Inom den elektroniska lexikografen är det rimligare att tala om *sekundärspråk* än *målspråk*. Skillnaden mellan målspråk och sekundärspråk ligger dels i deras respektive åtkomlighet, dels i deras relation till källspråket. Målspråksenheten är inte sökbar och dess förekomst i ordboken motiveras helt och hållet av att den aktualiseras av en källspråksenhet. Sekundärspråkets enheter är däremot sökbara och de utgör i sig en avgränsad delmängd av sekundärspråkets totala ordförråd. Den lexikografiska beskrivningen av uppslagsenheternas lexikaliska egenskaper, tidigare enbart aktuell för ordbokens källspråk, är relevant även för sekundärspråkets enheter (figur 2). Dessas morfologiska, kombinatoriska och pragmatiska egenskaper kan nu redovisas för de användare som väljer sekundärspråket som ingång i ordboken.

Även källspråkets funktioner blir mindre entydiga i den elektroniska lexikografen. Medan det fortfarande bildar utgångspunkt för den kontrastiva beskrivningen av de två aktuella språken, och på så sätt är primärt i förhållande till målspråket/sekundärspråket, utgör det inte längre den enda utgångspunkten för användarens sökningar. Tvärtom kan det aktualiseras som resultatet av en sökning som utgår från sekun-

därspråket. I den elektroniska tvåspråkiga lexikografin är därför beteckningen *primärspråk* lämpligare än *källspråk* för det språk som bildar utgångspunkt för beskrivningen av relationerna mellan de båda språken.

<b>Funktion</b>	<b>Primärspråk</b>		<b>Sekundärspråk</b>
produktion:	L1	< >	L2
avkodning:	L2	< >	L1

Figur 2: I en elektronisk ordbok kan båda språken göras sökbara och därmed båda fungera som ingångar i ordbokstexten.

Den tvåspråkiga lexikografin står här inför en både teoretisk och metodisk utmaning. Det ställs nu helt andra krav på täckningsgraden och relevansen av sekundärspråkets ordförråd än på det ordförråd som tidigare enbart var indirekt åtkomligt via ett källspråk. Nu framträder även sekundärspråkets samlade ordförråd, och förutom de redan bekanta lemmalakunerna blir nu också avsaknaden av centrala ord i sekundärspråket uppenbara och man kan även tala om ekvivalentlakuner (Rauset et al. 2012:522).

#### 4. Språkproven i tvåspråkiga ordböcker

I lexikografiska kretsar råder consensus beträffande relationen mellan källspråk och målspråk i den tvåspråkiga lexikografin: «källspråket måste beskrivas utifrån målspråkets perspektiv» (Svensén 2004:196). När det gäller urvalet av bruksexempel

konstaterar Atkins & Rundell att «the great bulk of examples in a bilingual entry are chosen entirely on the basis of their translation» (Atkins & Rundell 2008:507). Bruksexemplens funktion menar de är «complementing (or replacing) the direct translation. This is the principal raison d'être of the example in the bilingual dictionary ...» (Atkins & Rundell 2008:509).

Bruksexempel som tillkommit utan att det kontrastiva perspektiv som här förespråkas anlagts, är inte alltid effektiva när det gäller att illustrera de egenskaper som är relevanta just i relationen mellan de inblandade språken. Dessutom kan särskräkliga formuleringar och förekomsten av kulturella företeelser ibland ställa till onödiga problem för målspråksredaktionen (se Rauset et al. 2012:520 f.).

I den elektroniska ordboken kan ordförrådet i alla slags språkprov, dvs. förutom bruksexempel även idiom och kollokationer, göras tillgängligt genom fritextsökningar i hela ordbokstexten. De ekvivalentluckor som framträder vid en sökning bland ekvivalenterna kan sålunda fyllas genom att den sökta enheten förekommer i översättningen av ett bruksexempel, ett idiom eller en kollokation. Därigenom har språkproven fått en helt ny funktion för den användare som tar sekundärspråket som utgångspunkt för sina sökningar.

## 5. Källspråk och målspråk i *Islex*

I det följande redovisas resultaten av en liten undersökning av täckningsgraden av en avgränsad lexikalisk subgrupp inom det

svenska ordförråd som finns som sökbara enheter i den isländsk–svenska nätordboken i *Islex*. De teoretiska och metodiska konsekvenser som den nya åtkomligheten har för lexikografin diskuteras också kort. Här ges först en skissartad beskrivning av *Islex*.

### 5.1. *Islex*

*Islex* är en flerspråkig, elektronisk ordbok mellan å ena sidan isländska och danska, norska och svenska å den andra (*Islex.is*, *Islex.dk*, *Islex.no*, *Islex.se*). Arbetet har utgått från isländskan, den har tjänat som ett traditionellt källspråk, och de andra språken fungerar som målspråk i förhållande till den. Nu pågår arbetet med att införliva även färöiska och finska i *Islex*.

*Islex* är i första hand avsedd att vara en produktionsordbok för islänningar (L1>L2), men den skall även kunna fungera som avkodningsordbok för dansk-, norsk- och svensktalande användare (L2>L1). På så sätt är den en traditionell, tvåspråkig lexikografisk produkt mellan två «små» språk, där så många funktioner som möjligt bör fyllas av en och samma produkt.

Ordböckerna är gratis tillgängliga på Internet. De innehåller närmare 50 000 isländska uppslagsord, omkring 30 000 bruks-exempel samt en mängd idiom och kollokationer. Allt sådant material översätts till målspråken. *Islex* har beskrivits ingående bl.a. av Úlfarsdóttir (2013), Jónsdóttir & Úlfarsdóttir (2012) och Sigurðardóttir et al. (2008) samt recenserats av Sanders (2013).

## 5.2. Språkens inbördes relationer i *Islex*

Även om *Islex* i praktiken fungerar som en tvåspråkig ordbok är det isländska språkmaterialet i första hand avsett att utgöra en avgränsad och väl beskriven delmängd av isländskans centrala ordförråd. Denna delmängd skall kunna användas som källspråk i tvåspråkiga ordböcker med i princip vilket som helst språk som målspråk. Den analys och den beskrivning av isländskan som bildar utgångspunkt för målspråken är därför strikt sarspråklig. I det nära samarbete som rådde mellan de olika *Islex*-redaktionerna fanns det ett visst utrymme för modifieringar av det isländska materialet efter önskemål från de skandinaviska redaktionerna. I stort sett var dock det sarspråkliga perspektivet förhärskande.

Avsaknaden av en kontrastiv analys av isländskan i relation till respektive målspråk kunde medföra problem för målspråkens redaktioner, framför allt när det gällde språkproven. Bruksexempel som är avsedda att illustrera uppslagsordets egenskaper utan att relateras till det aktuella målspråket är inte alltid optimala (se Rauset et al. 2012:516 ff.). Det skall emellertid framhävas att när det gäller *Islex*-arbetet kompliceras naturligtvis relationen mellan källspråk och målspråk av att målspråken redan från början var avsedda att bli flera.

Genom de generösa sökmöjligheter som användargränssnittet i *Islex* erbjuder, framstår de skandinaviska språken för användaren snarare som ett sekundärspråk än som målspråk. Såväl lemnarna som ekvivalenterna är sökbara, liksom ordförrådet i alla typer av språkprov och i deras översättningar.



## 6. Det svenska ordförrådet i *Islex*

Den delmängd av källspråkets ordförråd som presenteras i den tvåspråkiga ordboken är genomtänkt och väl motiverat. De lemmalaker som trots allt brukar kunna konstateras är för det mesta slumpmässiga och de gäller enstaka ord. Risken är liten att en hel lexikalisk subgrupp av misstag skall saknas. När det gäller målspråkets representation i ordboken är ordförrådet sekundärt och reaktivt i förhållande till källspråket. I den traditionella tryckta ordboken är målspråkets representation en irrelevant entitet.

I den elektroniska ordboken framträder däremot sekundärspråkets sökbara ordförråd som en egen urskiljbar del i ordboken. Sekundärspråkets representation framstår då som en kvalitetsaspekt. I *Islex* består det svenska ordförråd som genereras som ekvivalenter av färre unika enheter än de isländska uppslagsorden: 41 000 unika svenska ekvivalenter mot de ca 50 000 isländska uppslagsorden (Úlfarsdóttir 2013:68). Trots att många av uppslagsorden förses med flera ekvivalenter uppträder var och en av dem ofta som ekvivalent till flera uppslagsord. De 41 000 ekvivalenterna utgör heller inte en väl avvägd, noggrant sammansatt delmängd av det svenska ordförrådet; de är en uppsättning med ord som etablerats i sin egenhet av ekvivalenter till det isländska lemmabeståndet. Det är därför väntat att även centrala och frekventa ord som förekommer i målspråket saknas i ordboken – ord som inte upprättats som ekvivalenter, även om vissa av dem kunde ha gjort det.

Genom att enheterna i sekundärspråket är sökbara framstår de tidigare osynliga ekvivalentlaker nu som en brist i ordboken (Rauset et al. 2012:522). Och till skillnad från de slump-

mässiga lemmalakunerna kan lakunerna i sekundärspråket bilda strukturella luckor där vissa lexikala subgrupper till stor del saknas.

### 6.1. Det internationella svenska ordförrådet

Inlån av ord och ordbildningsmaterial är, allmänt sett, betydligt vanligare i svenskan än i isländskan. Sålunda förekommer exempelvis romanska eller grekiska termer mycket sparsamt i isländska (Jónsson 1997:167). Den delmängd av det svenska ordförrådet i *Islex* som här har undersökts utgörs av verb bildade med det romanska eller latinska avledningssuffixet *-era* med varianterna *-iera*, *-fiera*, *-ficera* m.m. Suffixet förekommer visserligen i isländskan men i avsevärt mindre omfattning än i svenskan. I *Islex* finns 30 isländska *era*-verb bland uppslagsorden.

De begrepp som i svenskan har lexikaliserats som ett *era*-verb betecknar ofta abstrakta processer. Dessa processer kan uttryckas som verbfraser i isländskan. Eftersom dessa verbfraser inte nödvändigtvis registreras som uppslagsenheter i isländskan, genereras inte svenskans *era*-verb i en utsträckning som motsvarar deras ställning i det svenska ordförrådet. Resultatet är en lång rad med ekvivalentluckor i den svenska som nu är åtkomlig i *Islex*.

## 6.2. Svenska *era*-ord som lemma respektive sekundärspråksenhet

I *Svensk baklängesordbok* (2007) förekommer 2 038 verb som slutar på *-era*. Långt ifrån alla dessa är lexikaliserade till den grad att de beskrivs i enspråkiga svenska ordböcker. Bland annat förekommer 11 sammansättningar med *taxera* som efterled och förutom simplex verbet *placera* ytterligare sju som har verbet som efterled.

För att få en utgångspunkt för en jämförelse av täckningsgraden av de internationella svenska *era*-verben i *Islex* har de 2 038 verben i *Svensk baklängesordbok* (SB) sökts i den tryckta *Svensk-isländsk ordbok* (1983).<sup>1</sup> Det svenska källspråket i *Svensk-isländsk ordbok* (SIO) antas vara en noggrant utvald och väl avvägd delmängd av det svenska ordförrådet, anpassad till ordbokens avsedda målgrupp: islänningar som använder den som avkodningsordbok och svenskar som använder den vid produktion av text på isländska.

Av de 2 038 verben i SB återfinns 1 071 i SIO. Eftersom makrostrukturen i SIO är nischalfabetisk, är det långt ifrån alla dessa verb som utgör ingång i en artikel. Sålunda påträffas *omdatera* i en nischartikel som inleds av *ombunden*. Där föregås det av de kondenserade *-byggnad*, *-byte*, *-bytlig*, *-bytlighet*, *-bytt*, *-checka*, *-dana* och *-daning* och följs av ytterligare sju nischlemman.

Enligt baksidestexten innehåller SIO omkring 60 000 uppslagsord, 20 % fler än de nästan 50 000 uppslagsorden i *Islex*. Det är därför rimligt att anta att den även innehåller fler *era*-verb än

---

1. Båda dessa ordböcker har excerperats manuellt. Enstaka förbiseenden eller dubblingar kan därför föreligga.

*Islex*. Så är också fallet. Av de 1 071 *era*-verben i SIO förekommer 360 som ekvivalenter i *Islex*. För den som använder *Islex* som stöd vid produktion av text på svenska är de luckor i det svenska ordförrådet som de drygt 700 utelämnade verben utgör ett överkomligt problem om ordboken anvisar andra sätt att formulera det avsedda innehållet. Svårare är det för den som använder svenskan som ingång i ordboken, vare sig det är vid avkodning av svensk text eller vid produktion av text på isländska.

### 6.3. Språkproven som resurs

Genom att även sekundärspråkets enheter är sökbara i *Islex* avslöjas alltså luckor i den uppsättning ekvivalenter som bildar stommen i sekundärspråkets ordförråd. Möjligheten att söka i all artikeltext, dvs. i bruksexemplen och bland idiom och kollokationer, gör att vissa av de synbara luckorna bland ekvivalenterna täpps till i och med att en lexikalisk enhet som inte genereras som ekvivalent kan förekomma i översättningen av ett språkprov av något slag.

En sökning i all artikeltext i *Islex* på de drygt 700 svenska *era*-verb som finns i SIO men som inte utgör ekvivalenter i *Islex* resulterade i ytterligare 132 verb. Av de 1 071 *era*-verben i SIO finns sålunda 492 i *Islex*, vilket också det är betydligt färre än förväntat med tanke på den relativt lilla skillnaden i ordböckernas totala ordförråd. Att det ordförråd som tillhandahålls i SIO kanske inte är helt representativt för 2000-talets svenska ordförråd framgår av att sökningen på all artikeltext i *Islex* också genererade 31 *era*-verb som inte finns i SIO. Vid sökningen i all artikeltext har även participformer räknats som

träffar medan participformer i SIO inte har räknats som verb.

#### 6.4. Förekomst i språkproven

En tredjedel av de svenska *era*-verben i *Islex* förekommer alltså i översättningar av bruksexempel eller som motsvarigheter till idiom och kollokationer. Dessa verb genereras av någon typ av flerordsenhet i källspråket. Ett exempel är verbet *installera*. I *Svensk ordbok utgiven av Svenska Akademien* (SO) redovisas tre betydelser av verbet: 'högtidligt sätta in i ämbete', 'utföra teknisk installation av' och 'låta flytta in och ge tillfälle att inrätta sig' (SO 2009 s.v. *installera*). I *Islex* förekommer verbet i två artiklar. I den stora verbartikeln *koma ges installera sig* som motsvarighet till reflexivfrasen *koma sér fyrir* (figur 3).

##### *koma sér fyrir*

■ göra sig hemmastadd

■ installera sig

*hún er að koma sér fyrir í nýju íbúðinni*

■ hon håller på att installera sig i den nya lägenheten

Figur 3: Förekomsten av *installera* i *Islex*.

*Installera* förekommer också i ett bruksexempel som illustrerar användningen av adjektivet *einsætt* (ex. 1).

(1) <honum> þykir einsætt að <leggja hita í gólfín>

<han> tycker det är självklart att <installera golvvärme>

Av de tre betydelserna i SO är det sålunda två som framgår av översättningar av språkproven i *Islex*.

Ett annat exempel är verbet *absorbera*. Det råkar förekomma i översättningen av bruksexempel dels i artikeln *nærfatnaður* (ex. 2, konkret betydelse), dels i artikeln *altaka* (ex. 3, överförd betydelse).

- (2) nærfatnaðurinn þarf að draga í sig svita og þorna fljótt  
underkläderna skall kunna absorbera svett och torka snabbt
- (3) áhugamálið altók huga hans  
han var helt absorberad av sin hobby

Ett tredje exempel är verbet *inkludera* som förekommer som ekvivalent i sin participform, *inkluderad*, till uppslagsordet *innifalinn*. Dessutom förekommer verbet i verbartikeln *telja* som motsvarighet till verbfrasen *telja með* (ex. 4).

- (4) telja <tímaritin> með  
inkludera <tidskrifternar>

Många av de verb som endast förekommer i översättningarna betecknar begrepp som antingen inte är särskilt etablerade eller väl avgränsade i isländskan. En del av dessa har lexikaliserats som verbfraser, *koma sér fyrir*, *telja með*. I SIO ges exempelvis tre isländska motsvarigheter till *absorbera*: «drekka (sjúga) e-ð í sig, [...] draga e-ð í sig, [...] taka e-ð í sig». Andra begrepp har inte alls etablerats som ett lexikaliserat uttryck i isländskan – eller i varje fall inte tillräckligt för att ta plats bland de 50 000 uppslagsenheterna i *Islex*.

## 7. Ekvivalentlakuner

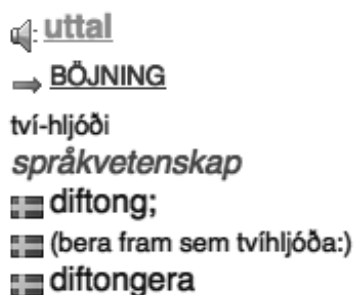
En del av de drygt 700 svenska *era*-verb som finns i SIO men inte i *Islex* är relativt marginella och därför inte motiverade i en ordbok som i första hand redovisar det samtida allmänspråket. Några av orden saknas exempelvis i SO (*advocera, omarrondera*), andra är märkta med «mindre brukl.» (*absolvera, kuvertera*) och ytterligare andra har en tydlig termkaraktär (*lemmatisera, nitrera*). Men en rad av de svenska *era*-verb som inte förekommer eller som i dagsläget inte är åtkomliga i *Islex* utgör i svenskan etablerade uttrycket för väl avgränsade och etablerade begrepp (se Hannesdóttir 2012:48 f.). Bland dessa verb finns *abstrahera, associera, cirkulera, exemplifiera, figurera, individualisera, internationalisera, optimera, precisera, realisera, socialisera* och ytterligare en lång rad.

När ekvivalentlakunerna väl har konstaterats i den publicerade ordboken är det avhängigt av databasens flexibilitet om och i så fall hur sekundärspråkets ordförråd kan kompletteras. Just i fallet *Islex* erbjuder det elektroniska formatet och databasens struktur en rad möjligheter att konstruera enheter som fyller ekvivalentlakuner, främst med hjälp av bruksexempel och kommentarer. Verbet *diftongera* får illustrera en av möjligheterna (figur 4).

Verbet *diftongera* har ingen lexikaliserad motsvarighet i isländskan och det förekommer sålunda inte som ekvivalent i *Islex*. Det förekommer inte heller i någon översättning. För att ändå inlemma det svenska verbet har artikeln *tvíhljóði* kompletterats med en kommentar: «bera fram sem tvíhljóða:» 'uttala som diftong' «diftongera». Verbet är sökbart och den svensk som söker den isländska beteckningen på den aktuella pro-

cessen blir hjälpt, liksom islänningen som använder *Islex* för avkodning.

## tvíhljóði mask.



Figur 4: Verbet *diftongera* i *Islex*.

## 8. Avslutning

Så länge tvåspråkiga ordböcker framställs med utgångspunkt i det ena av språken kommer ekvivalentlökuner troligen att vara avsevärt mycket vanligare än lemmalakunerna och dessutom endast i liten utsträckning resultatet av lexikografens aktiva ställningstagande. För att den elektroniska, tvåspråkiga ordboken ska bli så multifunktionell och biskopal som möjligt, bör även täckningen av sekundärspråkets ordförråd beaktas. Därigenom får publiceringen av tvåspråkiga ordböcker i elektronisk



form sådana teoretiska och metodiska konsekvenser för den tvåspråkiga lexikografin att det föranleder den revidering av ordbokstypologin som här har förts fram. Den tvåspråkiga ordboken är nu betydligt mera komplex än den var för några decennier – och århundraden – sedan. Dr Johnson's ordbok var enspråkig och tryckt. Därför lyste lemmat *sausage* med sin frånvaro – om man får tro TV-sketchen i *The Black Adder*. Hade det rört sig om en missad ekvivalent i en tvåspråkig ordbok, hade nog inte ens dr Johnson upptäckt sin fadäs.

## Litteratur

### Ordböcker

*Islex* = *Islex-orðabókin*. Stofnun Árna Magnússonar í íslenskum fræðum: [www.islex.is](http://www.islex.is), [www.islex.dk](http://www.islex.dk), [www.islex.no](http://www.islex.no) och [www.islex.se](http://www.islex.se).

*Islex.se* = <http://www.islex.se> (sökningarna kontrollerade i oktober 2013).

NLO (1997) = Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.

SB (2007) = Allén, Sture & Christian Sjögren: *Norstedts svenska baklängesordbok*. Stockholm: Norstedts Akademiska Förlag.

SIO ([1983] 2005) = *Norstedts svensk-isländska ordbok* (4. uppl.). Stockholm: Norstedts Akademiska Förlag.

HANNESDÓTTIR

SO (2009) = *Svensk ordbok utgiven av Svenska Akademien*.  
Stockholm: Norstedts (i distribution).

### Övrig litteratur

- Atkins, B.T. Sue & Michael Rundell (2008): *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.
- Eaker, Birgit, Lennart Larsson & Anki Mattisson (red.) (2012): *Nordiska studier i lexikografi 11*. Lund: Svenska Akademiens ordboksredaktion.
- Hannesdóttir, Anna Helga (2012): Ekvivalensrelationer i tvåspråkig lexikografi. I: *LexicoNordica 19*, 39–58.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2012): ISLEX – en flersproget nordisk ordbog. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi 11*. Lund: Svenska Akademiens ordboksredaktion, 353–366.
- Jónsson, Baldur (1997): Isländska språket. I: Allan Karker, Birgitta Lindgren & Ståle Løland (red.): *Nordens språk*. Oslo: Novus forlag, 161–176.
- Kromann, Hans-Peder, Theis Riiber & Poul Rosbach (1984): Überlegungen zu Grundfragen der zweisprachigen Lexikographie. I: H.E. Wiegand (red.): *Studien zur neuhochdeutschen Lexikographie V*. Hildesheim/New York: Georg Olms, 159–238.
- Rauset, Margunn, Anna Helga Hannesdóttir & Aldís Sigurðardóttir (2012): Ein-, to- eller fleirspråkleg ordbok? I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi 11*. Lund: Svenska Akademiens ordboksredaktion, 512–523.
- Sanders, Christopher (2013): *Islex* foråret 2013. (Recension av *Islex*.) I: *LexicoNordica 20*, 259–277.

- Sigurðardóttir, Aldís, Anna Helga Hannesdóttir, Halldóra Jónsdóttir, Håkan Jansson, Lars Trap-Jensen & Þórdís Úlfarsdóttir (2008): ISLEX. An Icelandic-Scandinavian Multilingual Online Dictionary. I: Elisenda Bernal & Janet DeCesaris (red.): *Proceedings of the XIII Euralex International Congress. Barcelona 2008*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 779–789.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska förlag.
- Úlfarsdóttir, Þórdís (2013): ISLEX – norræn margmála orðabók. I: *Orð og tunga 15*. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum, 41–71.

Anna Helga Hannesdóttir  
docent, fil. dr  
Göteborgs universitet  
Box 200  
SE-405 30 Göteborg  
anna.hannesdottir@svenska.gu.se

## Leksikografiens grunnenhet – hva heter det, dette vi lager ordbokartikler om?

*Petter Henriksen*

A dictionary article has a word (often called *entry* or, in Norwegian, *oppslagsord*) as its title, but normally also deals with other forms of that word: inflection forms, spelling variants etc., more specifically a group of words that the lexicographer deems to be sufficiently related to merit a dictionary article under a common title. This word group – and not the entry itself – determines the article's range, and the criteria for this kind of grouping are fundamental to lexicography's way of systemizing a language's word mass. The author suggests that Nordic lexicography seems to lack a common designation for this kind of word group, and invites colleagues in Scandinavia to join in on selecting an appropriate term.

I denne artikkelen skildres en jakt vi har drevet med i det siste i redaksjonen for Det Norske Akademis Store Ordbok, etter en bestemt leksikografisk term *x*. I dagligspråket pleier de som ikke er leksikografer, å kalle termen *x* for *ord*, for eksempel i følgende utsagn:

- I Shakespeares skriftlige produksjon forekommer rundt 30 000 ulike *ord*.
- *Ordet* «være» er et av de hyppigste i norske lovtekster.
- «Syklus», «syklusen» og «cyklus» er egentlig samme *ord*.
- En artikkel i en ordbok handler om et *ord*.

Vi har ønsket å vite hva leksikografene kaller det, for konsistent termbruk i vår redaksjonshåndbok.

Vi har undersøkt nærmere i faglitteraturen og hos kolleger i andre miljøer, uten å få et entydig svar. Denne artikkelen er et forsøk på å komme nærmere en avklaring.

## 1. Nærmere om begrepet

Deskriptive disipliner baseres på at man forenkler det universet man studerer, at man gjennom abstraksjoner løfter blikket fra å beskrive individer til generisk å beskrive individgrupper som har fellestrekk.

I botanikkens floraer beskriver man ikke enkeltteksemplarer av hestehov, men *arten* hestehov, som ikke finnes fysisk i naturen, men er en abstraksjon. *Art*, hovedgrupperingen av individer på laveste plan, har status som en grunnenhet i plantesystematikken. En grunnenhet i fargelæren kan sies å være *farge*; mange ulike fargevarianter persiperes intersubjektivt som abstraksjonen *blått*, som kan beskrives kollektivt: «blått er en kjølig farge». Fonologien har en grunnenhet *fonem*; fonemet /r/ er ulike måter som det er mulig å uttale r på uten at det får betydningskillende effekt.

Termen *x* som vi er på jakt etter, er tilsvarende den abstraksjonen som gjør at man kan si at *syklus*, *syklusen*, *cyklus* og *cykluser* er manifestasjoner av samme *x* og hører hjemme i samme artikkel i en ordbok, leksikografiens analogi til botanikkens flora. Det er nærliggende å kalle dette *x* for leksikografiens eller ordsystematikken grunnenhet.

Vi ser at *x*-ens manifestasjoner fordeler seg langs to akser:

1. **Bøyningsvarians** – at ulike bøyingsformer – *syklusen*, *sykluser*, *syklusene*, *syklusens* – regnes som samme *x*. Leksikografien har sin godt innarbeidede term for akkurat denne samlingen: De ulike bøyingsformene sies å være varianter av samme *leksem*.
2. **Formvarians** – at ulike variantformer (stavemåter), f.eks. *syklus* og *cyklus*, kan regnes som samme *x* når bestemte kriterier er oppfylt. Disse kriteriene, gjerne kalt lemmatiseringskriterier, varierer en del mellom de ulike ordbokredaksjonene, men består typisk i at det er semantisk sammenfall, at det er etymologisk avvik bare i nyere språkhistorie, kanskje også at det – slik tilfellet vil være for norsk – har vært truffet en avgjørelse hos en normeringsmyndighet om alternativ stavemåte.

Formvarians, ikke minst den historisk betingede, forekommer i skriftspråk flest. I *The New Oxford Dictionary of English* (Oxford 1989) handler artikkelen om verbet *organize* også om formvarianten *organise* og den historiske formen *organyse*. Men formvarians er kanskje særlig karakteristisk for bokmålsnorsk, med dets frie bruk av de alternativene som har oppstått gjennom det danskspråklige innslaget i arven direkte fra det norrøne. NAOBs artikkel med overskriften *såpe* (*subst.*) rommer bl.a. følgende manifestasjoner av *x*-termen, inkludert historiske former: *såpe*, *såpes*, *såpen*, *såpens*, *såper*, *såpers*, *såpene*, *såpenes*, *sepe*, *sepes*, *sepen*, *sepens*, *seper*, *sepers*, *sepene*, *sepenes*, *sæpe*, *sæpens*, *sæper*, *sæpers*, *sæpene*, *sæpenes*, *såpa*, *såpas*, *sæbe*, *sæben*, *sæbens*, *sæber*, *sæbers*, *sæbene*, *sæbenes*.

Leksikografiens *x* har videre en avgrensning mot et søster-

begrep, som vi i vår redaksjon gjerne kaller *oppslagsord*, nemlig den av alle  $x$ -ens former som redaktøren velger som overskrift for ordbokartikkelen, f.eks. infinitivformen for verb. De to termene er beslektet; vi har å gjøre med et termpar, beskrevet analogt etter ulike karakteristika i tabell 1.

<b>Det artikkelen egentlig omhandler,</b>	<b>Artikkeloverskriften,</b>
<b>dvs. samlingen med former,</b>	<b>dvs. formen som representerer samlingen,</b>
<b>dvs. abstraksjonen,</b>	<b>dvs. artikkelens konkrete inngang,</b>
<b>dvs. den generiske enheten,</b>	<b>dvs. enhetens betegnelse</b>
<b>dvs. <math>x</math></b>	
såpe, såpes, såpen, såpens, såper, såpers, såpene, såpenes, sepe, sepes, sepen, sepens, seper, sepers, sepene, sepenes, sæpe, sæpens, sæper, sæpers, sæpene, sæpenes, såpa, såpas, sæbe, sæben, sæbens, sæber, sæbers, sæbene, sæbenes, ...	såpe

Tabell 1

## 2. Kandidater for $x$

For oversiktens skyld gis innledningsvis (tabell 2) samtlige alternative betegnelser for  $x$  som er brakt for dagen i løpet av

vår undersøkelse, sammenstilt med sammenhørende betegnelser for søsterbegrepet, artikkeloverskriften.

Alt.	Det artikkelen egentlig omhandler,	Artikkeloverskriften,
	dvs. samlingen med former,	dvs. formen som representerer samlingen,
	dvs. abstraksjonen,	dvs. artikkelens konkrete inngang,
	dvs. den generiske enheten,	dvs. enhetens betegnelse
	dvs. <i>x</i>	
<i>I autoritativ faglitteratur</i>		
1	lemmategn	lemma
2	lemmategn	oppslagsord
3	lemmategn	inngangsord
4	lemma	lemmaform
<i>I bruk i praksis (usus)</i>		
5	lemma	lemma
6	leksem	lemma
7	oppslagsord	oppslagsform
8	former	oppslagsord
9	–	oppslagsform
10	ortoklump	lemma
<i>Kreative forslag</i>		
11	lemma	lemmaoverskrift
12	leksikalsk enhet	lemma
13	ord	lemma
14	leksikalsk ord	lemma
15	lemmem	lemma

Tabell 2



### 3. Kandidater for *x* i autoritativ faglitteratur

Tidlig i vår jakt har vi konstatert at det finnes en regulær fasit, nemlig det som står svart på hvitt i nordisk leksikografis sentrale referanseverk, Bo Svenséns *Handbok i lexicografi* (Svensén 2004:115): **lemmategn**:

**LEMMAT** [*footnote*: en vanlig synonym till “lemma” er **UPPSLAGSORD**] står i ordboken som representant för ett språkligt tecken, nämligen den lexikalska enhet som ordboksartikeln handlar om (**LEMMATECKNET**). Lemmat har till syfte att identifiera lemmatecknet, i en ordbok med alfabetisk makrostruktur också att bestämma artikelns placering i ordboken.

Sammen med lemmategn opererer Svensén altså med betegnelsen *lemma* for oppslagsordet (**alternativ 1**).

Termen *lemmategn* føres tilbake til tyskeren Herbert Ernst Wiegands metaleksikografiske systematikk fra 1980-tallet:

Strictly speaking, the information contained in the microstructure does not refer to the lemma but to the linguistic sign which the lemma arbitrarily represents in the microstructure. The sign may be called the **lemma sign** (G. Lemmazeichen, F. signe-lemme). Lemmatisation will thus refer to the selection of one single morphological form whose function in the macrostructure is to represent the total set of grammatical and morphological forms of the linguistic sign treated in the microstructure. (Wiegand 1989–91:329)

Wiegand knytter altså enheten som en ordbokartikkel omhandler, til Saussures og lingvistikkens *språktegn*, den dualistiske entiteten med en uttrykksside og en innholdsside, det som betegner og det som betegnes, *signifiant* og *signifié*.

Også *Nordisk leksikografisk ordbok* (NLO) understøtter bruken av *lemmategn* om  $x$  og *lemma* om artikkelens overskrift:

**lemmategn** språklig tegn som et lemma representerer (NLO 173)

**lemma** ... er den metaleksikografiske betegnelse for det som i verkstedsspråket og i ordbøkers omtale kalles oppslagsord eller inngangsord. Gjennom lemmaet finner man fram til den ordboksartikkelen man ønsker, idet lemmaet samtidig fungerer som en slags overskrift for artikkelen. (NLO 172)

Dette gir oss to nye termpar-alternativer (**alternativ 2** og **3**), der *lemmategn* brukes sammen med hhv. oppslagsord og inngangsord.

Med dette burde vi ha kunnet si at vår jakt på  $x$  allerede var slutt:  $x$  kalles *lemmategn*.

Ett forhold har allikevel fått oss til å nøle: Det er tilsynelatende ingen leksikografer som kaller  $x$  for *lemmategn*. Vi har valgt å ta noen stikkprøver før vi slo oss til ro med *lemma*-tegn.

Først har vi konsultert nok en velrenommert faglitterær kilde, Lars Vikørs og Ruth Vatvedt Fjelds lærebok i leksikografi *Ord og ordbøker* (Fjeld & Vikør 2008)). Ifølge forfatter Lars Vikør søker man der gjennomgående å bruke betegnelsen *lemma* om samlingen med former og *lemmaform* om formen som brukes som overskrift (**alternativ 4**), f.eks.:

lemmaforma blir i ordbøker vanlegvis sett med halvfeit skrift og ofte markert med innrykk eller utrykk (Fjeld & Vikør 2008:163)

*Ord og ordbøker* gir altså ingen støtte til lemmategn som betegnelse for *x*. Videre er det suggestivt at NLO selv ser ut til å unngå *lemmategn* bortsett fra i selve artikkelen *lemmategn*. *Lemmatisering* har f.eks. følgende definisjon:

**lemmatisering** det at ordformer som hører til samme leksem, blir samlet under samme lemma (NLO 173)

På det stedet man kunne ha brukt *lemmategn*, brukes altså *leksem* om *x*, enda *leksem* slik det ellers er definert, ikke omfatter formvarians.

Videre har vi interessant nok fått vite at termen *lemmategn* også til en viss grad unngås av Bo Svensén selv, som forekommende nok har respondert på en forespørsel med å gi utdypende kommentarer til teksten i *Handbok i lexikografi* (Svensén 2004). Han bekrefter at han i *Handbok i lexikografi* velger termen *lemmategn* for *x* enda den kan være tungvint å bruke og er mindre vanlig, og han utdyper tankevekkende at han i annet opplag av den engelske utgaven av håndboken (Svensén 2009) har forsøkt å styre unna problemet gjennom å føye til en note som ikke står i den svenske versjonen:

For the sake of simplicity, the term ‘lemma sign’ is used in this book only when there is a special need to distinguish the linguistic sign described in an entry from the representative of that sign heading the entry. Where no such distinction is thought to be necessary, the simple term ‘lemma’ is used throughout. (Svensén 2009:93)

Svensén tilføyer at han erkjenner at dette ikke er spesielt tilfredsstillende fra teoretisk synsvinkel, men at han tror at det er ganske vanlig at de som anvender termen *lemma* i praktisk ordbokarbeid, bruker den i begge betydningene uten egentlig å reflektere så mye over det. Og han legger til at man like gjerne kunne velge eller skape en annen term, hovedsaken er at man skiller mellom de to begrepene når det behøves.

I sum bekreftes inntrykket av at *lemmategn* møter en viss motvilje også i det autoritative fagbokmiljøet. Det er egentlig ikke så uventet. Vellykkede termer er intuitivt forståelige. *Lemmategn* er det kanskje ikke i tilstrekkelig grad. For folk flest – som ikke er bevandret i Saussure – dreier et tegn seg om noe på uttrykkssiden, om en betegnelse og ikke et begrep. Om noe, skulle *lemmategn* da betegne selve artikkeloverskriften, altså noe annet enn den foreslåtte definisjon.

#### **4. Kandidater for $x$ ut fra usus**

Gitt dette har vi ikke slått oss til ro med *lemmategn* som  $x$ , men fortsatt jakten ved å foreta en rundspørring om termbruken hos noen av de tyngre leksikografamiljøene i Norden. Vi har sendt en forespørsel til fire miljøer i Norge, to i Sverige og to i Danmark. Forespørselen har hatt form av en e-post, med noe varierende utforming og detaljeringsgrad, men med innhold essensielt som følger: «En ordbokartikkel har en overskrift i form av ett enkelt oppslagsord (gjernes også kalt *lemma*), men handler egentlig om en samling med former: bøyningsformer, variant-

former o.a. I en del av faglitteraturen ser det ut til at denne samlingen kalles for *lemmategn*. Er denne termen i bruk i deres miljø? Hvis ikke, har dere en annen betegnelse som dere bruker om denne samlingen?»

Det understrekes at rundspørringen har vært uformell og på ingen måte metodologisk vanntett, så svarene (tabell 2) er behørig anonymisert. Resultatet gis nedenfor:

- I to av miljøene kalles samlingen med former for **lemma** samtidig som man også kaller artikkelens overskrift for **lemma**, uten at man problematiserer det – **alternativ 5**. Dette sammenfaller for øvrig med Bo Svenséns inntrykk.
- I et miljø brukes betegnelsen **leksem** om samlingen mens man bruker **lemma** om artikkelens overskrift, slik også NLO gjør i sin definisjon av lemmatisering – **alternativ 6**.
- I et miljø brukes betegnelsen **opplagsord** om samlingen med former og **opplagsform** om artikkeloverskriften – **alternativ 7**.
- I et tilfelle brukes **opplagsord** om artikkeloverskriften, og **former** om samlingen med former – **alternativ 8**.
- I et miljø brukes **opplagsform** om artikkeloverskriften, mens man ikke opererer med noen benevnelse for samlingen med former – **alternativ 9**.
- I et miljø er man kjent med **lemmategn** som betegnelsen på samlingen med former og **lemma** som betegnelse på overskriften, men bruker ikke spesielle betegnelser som dette i praksis.
- I et miljø fortelles det halvt i spøk at man – fordi man reserverer **lemma** for å betegne artikkeloverskriften – i mangel av en bedre betegnelse kaller samlingen med former for **ortoklump** – **alternativ 10**.

Dette bekrefter for det første vårt inntrykk av at *lemmategn* ikke er i bruk om  $x$  i praksis, og det kan tilføyes at ved nærmere spørsmål om man kunne tenke seg å ta *lemmategn* i bruk som term, har svaret gjennomgående vært negativt.

Videre har vi måttet konstatere at det ikke forekommer noe utbredt, innarbeidet alternativ til *lemmategn*. Tvert imot – mellom de forskjellige miljøene er det omtrent ikke sammenfall i termbruken.

Videre har vi nokså uventet sett at sentrale betegnelser som *lemma* og *oppslagsord* opptrer på begge sider av skillelinjen mellom søsterbegrepene.

Dessuten har vi – også uventet – sett at det er stor variasjon på hva man kaller noe så konkret som artikkeloverskriften. Dette som har begynt hos oss som en jakt på et  $x$ , utvikler seg til å bli en jakt også på et  $y$ , altså en omforent betegnelse på artikkeloverskriften.

## 5. Evaluering av de eksisterende alternativene

Vi har kommet til at ingen av usus-termparene (tabell 2) peker seg ut som bedre løsninger enn dem fra den autoritative faglitteraturen: **Alternativ 5** ( $x = \text{lemma}$ ,  $y = \text{lemma}$ ) bryter med terminologilærens krav om at termer ikke skal være polyseme. Ulike begreper skal ha ulike betegnelser. **Alternativ 6** ( $x = \text{leksem}$ ,  $y = \text{lemma}$ ) er valgt bort da *leksem* defineres klart både hos Svensén (Svensén 2004:117–124) og NLO (167) som noe annet enn  $x$ , nemlig samlingen av bøyingsformer alene. **Alter-**

**nativ 7** ( $x = \text{oppslagsord}$ ,  $y = \text{oppslagsform}$ ) er valgt bort da Svensén og NLO så klart definerer *oppslagsord* som betegnelse på  $y$  og ikke  $x$ . **Alternativene 8, 9 og 10** er valgt bort da de ikke gir noen fagmessig presis betegnelse for  $x$ .

I lys av dette har vi vurdert på nytt de autoritative alternativene: **Alternativ 1** ( $x = \text{lemmategn}$ ,  $y = \text{lemma}$ ) har vi beholdt som et gangbart alternativ i kraft av autoritetenes tale, mens vi har valgt bort de mer muntlige **alternativ 2** ( $x = \text{lemmategn}$ ,  $y = \text{oppslagsord}$ ) og **alternativ 3** ( $x = \text{lemmategn}$ ,  $y = \text{inngangsord}$ ) da vi har sett det som en fordel å bruke den internasjonale termen *lemma* for  $y$ , og da sammensetningsdannelsen *lemma*  $\rightarrow$  *lemmategn* kaster nyttig lys over slektskapet mellom de to begrepene.

**Alternativ 4** ( $x = \text{lemma}$ ,  $y = \text{lemmaform}$ ) har vi sett på som problematisk da det strider mot en svært innarbeidet måte å definere *lemma* på, nemlig som artikkelens overskrift  $y$ , og i praksis innebærer noe så drastisk som at termen *lemma* nydefineres. Men alternativet har sine tiltalende sider. Det griper den tilbøyelighet som mange leksikografer inkludert Svensén har til å la *lemma* betegne den generiske enheten, det er brukt i den respekterte publikasjonen *Ord og ordbøker* (Fjeld & Vikør 2008), og ved å benevne  $y$  med en sammensetning med *lemma* som førsteledd, demonstrerer man slektskapet mellom de to termene. Vi har beholdt alternativet som kandidat, selv om det ikke er optimalt.

I sum har ingen av de eksisterende alternativene pekt seg klart ut. Vi har avrundingsvis unnet oss å tenke på fritt grunnlag.

## 6. Kreative forslag til kandidater

Hvis vi først skal gå inn på premissene for alternativ 4 med et redefinert *lemma* som  $x$ , er det ikke sikkert at *lemmaform* er den mest velvalgte sammensetningen for å betegne  $y$ . *Lemmaform* er en semantisk treffende sammensetning når førsteleddet *lemma* betyr 'artikkeloverskrift', men her er det faktisk ikke tilfellet – *lemma* er definert som den generiske enheten  $x$ . Da hadde nok en sammensetning som *lemmaoverskrift* vært vel så treffende. Paret  $x = \mathbf{lemma}$ ,  $y = \mathbf{lemmaoverskrift}$  er ført opp som et **alternativ 11**.

Videre har vi lekt med Svenséns tanke om at man like gjerne kunne skape en annen term for  $x$  enn *lemmategn*, men da fastholde *lemma* som betegnelse for  $y$ . Det siste er det nemlig sterke grunner til: at det er godt innarbeidet og at det smidig danner meningsfylte sammensetninger.

For det ene har vi vurdert en noe defensiv løsning, å kalle  $x$  for **leksikalsk enhet (alternativ 12)** ettersom det er definert slik i NLO:

**leksikalsk enhet** ord, orddel eller ordgruppe sett fra leksikologisk eller leksikografisk synspunkt (NLO 168)

Vi har ikke gått videre med alternativet. *Leksikalsk enhet* er et fruktbart begrep, men dekker mer enn den enheten man lager ordbokartikler om. Dessuten ville betegnelsen gi svak gjenkjennelse som term; det ville nesten bli som å kalle leksikografiens grunnenhet for «leksikografiens grunnenhet», helt generisk.

Videre har vi vurdert den folkelige betegnelsen **ord** som kandidat for  $x$ , slik det er brukt muntlig i foredragets innledning (**alternativ 13**). Men NLO definerer *ord* som noe desidert annet, likt med termen *token* eller *løpeord*:



**ord** kombinasjon av språklyder eller bokstaver som utgjør en selvstendig innholdsmessig og uttryksmessig enhet som i tale kan skilles ut av innfødte språkbrukere ved hjelp av forskjellige grensesignaler, og som i skrift markeres med avstand foran og etter (NLO 196)

Dessuten legges *ord* effektivt dødt som anvendelig leksikografisk term i *Ord og ordbøker*, i kapittelet “Kva er eit ord?” (Fjeld & Vikør 2008:27–38), som påviser at begrepet er uklart og flertydig. *Ord* alene har vi da valgt bort som *x*. Derimot har vi føyet til **leksikalsk ord** som et **alternativ 14**, altså et «ord» i leksikografisk betydning.

En annen innfallsvinkel er å skape en avledning ved hjelp av suffikset *-em*. Lingvisten Hadumod Bussmann definerer det slik:

**-Em** ... Aus dem Griech. entlehntes Wortbildungselement zur Bezeichnung funktionaler (d.h. distinktiver) “Einheiten” auf der Ebene der →Langue; vgl. →Phonem, →Morphem usw. (Bussmann 2002:187)

Suffikset brukes altså på *langue*-nivå, i abstraherende betydning, i ord som *fonem*, *morfem*, *grafem* og *leksem*, til å betegne generisk kategori.

Hvis vi på fritt grunnlag skulle konstruere en universell term for leksikografiens grunnenhet, ville det vært nærliggende å ta utgangspunkt i det greske ordet for det vi beskriver i faget vårt, *lexis*, ‘ord’, og lagt til endelsen *-em*, altså **leksem**, identisk med **alternativ 6**. Nå er som kjent *leksem* allerede innarbeidet som leksikografisk term i mer snever betydning (gruppen av bøyingsvarianter) og dermed vurdert som uaktuelt for *x*.

Derimot finnes ingen sedvane som hindrer en annen

kreativ bruk av suffikset *-em*: å legge suffikset til den innarbeidede termen for artikkeloverskriften, *lemma*: altså *lemmem* (**alternativ 15, x = lemmem, y = lemma**). Dette ville ikke kreve noen nydefinering av *lemma*, og ville gi en sammensetning som kan leses nokså intuitivt som den generiske abstraheringen av lemmaet, hvilket er en treffende måte å beskrive *x* på.

Utfallet av vår jakt kan leses i tabell 3, der våre egne foretrukne kandidater står igjen.

Med dette er vår kreativitet uttømt. Vi håper våre kolleger i andre miljøer vil bli med og vurdere alternativene.

## 7. Behøves omforente termer?

Til slutt, hvis vi skulle heve blikket fra NAOBs interne pedagogiske behov til den nordiske leksikografien generelt: Man ser altså ut til å være i den situasjonen at man ikke har noen standardbetegnelse for grunnenheten *x*, og heller ikke for dens overskrift *y*. Til dét kan man spørre om det egentlig er så viktig innen et så praktisk innrettet fag som vårt.

Et overfladisk argument for å standardisere terminologien på dette punkt kunne være at alt annet må gå på den faglige stolthet løs, innen et fag som har til oppgave å bringe klarhet i bruk av ord. Et mer tungtveiende argument kunne være at mangelen på en standardbetegnelse for grunnenheten svekker grunnlaget for en viktig faglig diskurs – om kriteriene for inndeling og systematisering av ordforrådet. Innenfor biologi-

Alt.	Det artikkelen egentlig omhandler,	Artikkeloverskriften,
	dvs. samlingen med former,	dvs. formen som representerer samlingen,
	dvs. abstraksjonen,	dvs. artikkelens konkrete inngang,
	dvs. den generiske enheten,	dvs. enhetens betegnelse
	dvs. x	dvs. y
<i>I autoritativ faglitteratur</i>		
1	lemmategn	lemma
4	lemma	lemmaform
<i>I bruk i praksis (usus)</i>		
–	–	–
<i>Kreative forslag</i>		
11	lemma	lemmaoverskrift
14	leksikalsk ord	lemma
15	lemmem	lemma

Tabell 3

faget kjenner fagfolk seg umiddelbart igjen i betegnelsen *art*, som benevner biologiens grunnenhet. Dette legger til rette for en løpende og essensiell drøftelse av kriteriene for inndeling i arter, som dreier seg mye om reproduksjonsevne, men slett ikke unyansert, hvilket bevitnes av at det finnes nærmest en egen gren innen biologien som kalles *artsproblemet* – *the species problem*. I leksikografifaget har vi ennå mye u diskutert om kriteriene for å inndele ordforrådet i enheter som allmennheten

HENRIKSEN

persiperer som koherente enda de manifesteres på høyst ulike vis. Og den diskusjonen vil hemmes så lenge grunnenheten ikke trer skarpt frem terminologisk.

## Litteratur

- Bussmann (2002) = Hadumod Bussmann: *Lexikon der Sprachwissen-schaft*. Stuttgart: Alfred Kröner Verlag.
- Fjeld & Vikør (2008) = Ruth Vatvedt Fjeld & Lars S. Vikør: *Ord og ordbøker*. Kristiansand: Høyskoleforlaget.
- NLO (1997) = Henning Bergenholtz et al.: *Nordisk leksiko-grafisk ordbok*. Oslo: Universitetsforlaget.
- Oxford (1989) = John Simpson & Edmund Weiner: *The Oxford English Dictionary*. Oxford: Clarendon Press.
- Svensén (2004) = Bo Svensén: *Handbok i lexicografi*. Stockholm: Norstedts Akademiska Förlag.
- Svensén (2009) = Bo Svensén: *A Handbook of Lexicography*. Cambridge: Cambridge University Press.
- Wiegand (1989–91) = Herbert Ernst Wiegand, Franz Josef Hausmann et al.: *Wörterbücher / Dictionaries / Dictionnaires: ein interna-tionales Handbuch zur Lexikographie*. Berlin: de Gruyter.

Petter Henriksen  
prosjektleder for Det Norske Akademis Store Ordbok (NAOB)  
Petter.Henriksen@Kunnskapsforlaget.no

## ***Bokmålsordboka og Nynorskordboka gjennom ein generasjon***

*Marit Hovdenak*

The standard one-volume monolingual dictionaries *Bokmålsordboka* and *Nynorskordboka* for the two official forms of Norwegian were launched in 1986. They were edited as a joint project of the Language Council of Norway and the University of Oslo. The focus is on both orthography and definitions. The dictionaries have been available on the web since 1994, and are now both linked to word banks with full forms of updated official orthography. The last years the website has been updated, and there has been technical and also lexicographic upgrade. During 2012 there were 70 million hits. Further technical and lexicographic revision ought to be done for both books.

### **1. Innleiing**

Dei to definisjons- og rettskrivingsordbøkene *Bokmålsordboka* (BOB) og *Nynorskordboka* (NOB) kom ut første gongen i 1986 og vart lagde ut på Internett i 1994. Bøkene kom til som eit samarbeidsprosjekt mellom Norsk språkråd og Universitetet i Oslo. Som «likeverdige allmennordbøker for dei to offisielle

målformene av norsk skriftmål» (Worren 1998:59) er dei eit tydeleg uttrykk for norsk språkpolitikk.

Grensensnittet til nettutgåvene har vorte endra dei siste åra, og det har truleg medverka til aukande bruk. I 2012 var det så mykje som 70 millionar treff på nettutgåvene.

## 2. Bakgrunn

Etter fleire store rettskrivingsreformer for bokmål og nynorsk frå tidleg på 1900-talet av tok språkpolitikken ei ny vending i 1964 då regjeringa oppnemnde ein komité til å «vurdere språksituasjonen og i tilfelle kome med framlegg til lovreglar og andre tiltak som kan tene til å samle krefter om å verne og utvikle norsk språk» (sitat frå den kongelege resolusjonen ved oppnemninga). Komiteen fekk namn etter leiaren Hans Vogt, rektor ved Universitetet i Oslo. Vogtkomiteen la i 1966 fram ei innstilling i seks punkt. Særleg tre av punkta var viktige i denne samanhengen. Det første punktet var stimulering av utforsking av norsk språk, med utgjeving av «gode ordbøker» og oppretting av eit «Norsk leksikografisk institutt» ved Universitetet i Oslo til dette. Det andre punktet var at prinsippet om likeverdig status for bokmål og nynorsk skulle vedtakast i lovs form. Det sjette punktet var oppretting av eit «Råd for språkvern og språkdyrking». Det vart til Norsk språkråd, ei omdanning av det dåverande Norsk språknemnd.

Desse framlegga vart sette i verk. Norsk leksikografisk institutt og Norsk språkråd kom begge i funksjon i 1972.

Når det gjaldt «gode ordbøker», var det opplagt at det som trongst mest, var moderne allmennspråklege ordbøker som forklarte orda og viste korleis dei blir brukte. I det norske ordboklandskapet i 1960- og 1970-åra fanst det på den eine sida mange skuleordlister. På den andre sida hadde vi dei store vitskaplege ordbokverka: *Norsk Riksmålsordbok* (1937–57) låg føre i fire band. *Norsk Ordbok* tok til å komme ut i 1950, men hadde berre komme eit lite stykke ut i alfabetet. For bokmål fanst dessutan *Tanums store rettskrivningsordbok*, med heile 300 000 oppslagsord, men snautt med forklaringar. Dei to nye institusjonane sette i januar 1973 ned ei nemnd som la planar for ei eittbands ordbok for nynorsk og ei for bokmål. Eittbands definisjonsordbøker, lenge kalla *handordbøkene* og *Norsk handordbok* som samla prosjekt, skulle fylle eit tomrom i ordboklandskapet. Det spela inn at dansk og svensk hadde nyare morsmålsordbøker av liknande type, og i prosjektfasen kom det ut ei tilsvarande for den uoffisielle normalen riksmål, *Riksmålsordboken*.

### 3. Prosjektet *Norsk handordbok* kjem i gang

Etter planlegging og prøveredigering kom redigeringa i gang i 1974. Då var dei viktigaste redigeringsprinsippa fastlagde. Arbeidet gjekk føre seg frå 1974 til 1985, det vil seie i elleve år. Arbeidsstaden var Norsk leksikografisk institutt ved Universitetet i Oslo, og instituttet stod for den største delen av arbeidssinnsatsen. Kvar av dei to redaksjonane hadde òg ein redaktør

lønt av Norsk språkråd. Redaksjonane utarbeidde eit førsteutkast for kvar sin halvdel av alfabetet og bytte sidan, og såleis kunne dei byggje på arbeidet til kvarandre. Det vart orientert om progresjonen i bladet *Språknytt* og i årsmeldingane frå Norsk språkråd.

Dei to ordbøkene skulle ha dei same informasjonskategoriane: oppslagsord med eventuelle rettskrivingsvariantar, ordklasse og bøyning, uttale for visse ord, ordhistorie (etymologi) og tydingar med bruksdøme og eventuelt faste uttrykk. Redigeringsreglane for ordklasse og bøyning blir kommenterte nedanfor.

### 3.1. Redigeringsreglar: Bøyingsopplysningar

Ordbøkene skulle gje fullstendig informasjon om den offisielle norma i rettskriving og bøyning. Slike opplysningar kan ta stor plass sjølv om ein forkortar, ikkje minst på grunn av valfrie former. I ordbøkene var det viktig å spare plass. Difor vart det tidleg bestemt at for dei vanlegaste bøyingskategoriane av substantiv, verb og adjektiv skulle det brukast bøyingskodar, som bygde på statistiske opplysningar (Kulbrandstad 1976:8). Det vart laga kodenøklar med bøyingsmønster som vart prenta på innsida av permen. Del av kodenøkkel i BOB (før 2005):

m1	stol, stolen, stoler, stolene pumpe, pumpen, pumper, pumpene
m3	bever, beveren, bevere [beverer] el. bevrer, beverne el. beverene



Mange ord har to kodar fordi dei har valfri bøyning, mens andre ord med valfridom har ein kombinasjon av kode og bøyingsendingar eller fullformer (her frå BOB):

**I pumpe** fl el. m1

**penny** m1 el. – el. *pence, pencene* i fl

Det var noko tvil alt i planleggingsfasen om kodesystemet var ein god idé. Kodar stiller nokså store krav til brukarane, som må slå opp i kodenøkkelen om dei ikkje kjenner ein kode, eller dei overser koden. Somme har peikt på at kodane har eit språkpedagogisk poeng fordi dei viser systematikken i språket med regelrett bøyning for store kategoriar. Kodesystemet har opplagt både pluss- og minussider. I nettutgåvene får ein no fullskrivne bøyingsformer, jf. pkt. 9. Studien til Kola (2012) tyder på at elevlar synest bøyingsopplysningane er vanskelege å forstå.

#### **4. Kva er ulikt – kva skil ordbøkene frå kvarandre?**

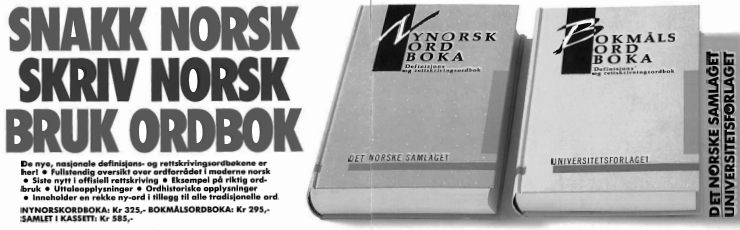
Éin skilnad er grunnleggjande: Det er ordbøker for kvar målform, bokmål og nynorsk. Norsk har to skriftspråksnormalar som har svært mykje sams, men som òg har kvar sine skriftkulturar.

Den mest synlege skilnaden elles er talet på oppslagsord og dermed sidetalet. BOB har ca. 65 000 oppslagsord, NOB ca. 90 000. Dei siste bokutgåvene har sidetal på høvesvis 1218 og 1478. I dei første planane var det meininga at ordbøkene skulle

bli like store; eit utgangspunkt var at dei skulle ha minst dobbelt så mange ord som større ordlister. I forordet i BOB står det at boka skal ha med «det sentrale ordforrådet i moderne bokmål». For nynorsk skulle ein ha med noko fleire samansetningar, fordi det manglar eit verk som Tanum for nynorsk. Det er likevel ikkje heile forklaringa på den store skilnaden i talet på oppslagsord. Worren (1998) er ein forvitneleg artikkel om målføreorda i *Norsk handordbok* og om kvifor ordbøkene vart så ulike: Svaret ligg mykje i at NOB i endå større grad enn BOB var eit nybrottsarbeid, sidan det ikkje fanst noka moderne ordbok for nynorsk. I nynorsk er det heller ikkje noko skarpt skilje mellom dialektar og standardmål, og redaksjonen ville ha god dekning av dialektane. Som ein praktisk regel for å velje ut målføreord (dialektord) kom redaksjonen i NOB fram til at målføreord som er registrerte i tre fylke eller meir, burde med sjølv om dei er lite brukte i litteraturen. Ordsamlingane til Norsk Ordbok vart nytta til å velje ut desse orda. Det var altså to faktorar, noko fleire samansetningar og den ovannemnde trefylkesregelen, som gjorde at NOB vart større enn BOB.

## 5. Utgjeving i 1986 og seinare utgåver

Ordbøkene kom ut i februar 1986 og vart lanserte på ein pressekonferanse i Oslo. Dei kom på kvar sitt forlag, BOB på Universitetsforlaget og NOB på Det Norske Samlaget. Forlaga samarbeidde om presentasjon og marknadsføring. Ordbøkene var nesten heilt like å sjå til, i to nyansar av gult:



Illustrasjon 1: Forlagsbrosjyre 1986.

Forlaga framheva ordbøkene som eineståande: «De nye, nasjonale definisjons- og rettskrivningsordbøkene er her! Fullstendig oversikt over ordforrådet i moderne norsk. Siste nytt i offisiell rettskriving. Eksempel på riktig ordbruk. Uttaleopplysninger. Ordhistoriske opplysninger. Inneholder en rekke ny-ord i tillegg til alle tradisjonelle ord». Bøkene kunne kjøpast enkeltvis eller i samlekassett, og dei kom som ekstrabøker i Bokklubben Nye Bøker. Det vart lagt merke til at bøkene kom, og det var oppslag i mange aviser. Opplaget det første året var 30 000 for BOB og 22 000 for NOB. BOB kom òg ut i to band i *Hjemmets store leksikon* (1991), med tittelen *Bokmålsordbok*.

Nye utgåver av ordbøkene har komme etter større rettskrivingsendringar, og når opplaga har vore utselde. I 1993 kom 2. utgåva av begge to. Dei hadde no heilt ulik farge – BOB blå perm, NOB raud – som òg seinare utgåver har. I 2001 kom 3. utgåva berre av NOB; det var gjort ein del rettskrivingsendringar sidan førre utgåva. I 2004 kom 3. utgåva av BOB med ajourført rettskriving og redaksjonelle tillegg og endringar. Tillegga og endringane skreiv seg ikkje minst frå meldingar som nettbrukarar av ordboka hadde sendt inn. Det var no Kunnskapforlaget som gav ut ordboka. Alt året etter, i 2005, kom BOB i 2. reviderte opplag av 3. utgåva, og no var det ei rettskrivingsreform som var grunnen. Norsk språkråd hadde i fleire år arbeidd

med revisjon av rettskrivinga for bokmål og nynorsk, og ei reform vart godkjend for bokmålet sin del av Kulturdepartementet i 2005. Det viktigaste i reforma var at det ikkje skulle vere sideformer lenger. I 2006 kom 4. utgåva av NOB. Det var lagt til om lag 500 nye oppslagsord, som bokmålsredaktør Boye Wangenstein hadde stilt til rådvelde for NOB-redaksjonen. I 2012 kom det ei ny utgåve berre på nett av NOB, og i den var rettskrivingsreforma for nynorsk gjennomført i oppslagsdelen av ordartiklane.

Bortsett frå tilvekst i nyord har det ikkje vore gjort store endringar i ordbøkene. BOB er noko meir oppdatert i innhaldet enn NOB.

## **6. På Internett frå 1994**

Utviklinga i i språkteknologien kom tidleg inn og endra bruken av BOB og NOB. Ordbøkene vart lagde ut på Internett av Universitetet i Oslo alt i 1994.

Ein føresetnad for publisering på nett var at ordbøkene fanst i elektronisk form. Manuskripta til førsteutgåvene var skrivne inn på terminal og vart sette frå magnetband. Dei var lagra i eit format utvikla av Håvard Hjulstad, som hadde ansvaret for databehandlinga av ordbøkene. Hjulstad stod òg for dataopplegget til andreutgåvene. For dei seinare utgåvene har universitetet (Dokumentasjonsprosjektet / Eining for digital dokumentasjon) hatt ansvar for dataopplegget.

Det kan nemnast at ordbøkene ikkje har vore utjevne i anna

elektronisk format, dei har ikkje funnest på diskett eller cd-rom, heller ikkje som betalingsteneste på Internett. Eigarane har tenkt slik at sidan ordbøkene er utvikla for offentlege midlar, skal dei vere gratis å bruke for publikum.

## **7. Oppbygging av Norsk ordbank – fullforms ordlister**

Norsk ordbank har hatt mykje å seie for den vidare utviklinga av ordbøkene. Ordbanken, som er enkle fullformsordlister for bokmål og nynorsk, var eit resultat av Taggarprosjektet ved Universitetet i Oslo. Dette prosjektet skulle utvikle ein morfologisk og syntaktisk taggar for bokmål og nynorsk, no kjend som Oslo-Bergen-taggar. Som eit ledd i prosjektet vart det i 1996 laga morfologiske databasar for bokmål og nynorsk (Hagen og Nøklestad 2010). Desse databasane bestod av

- BOB og NOB
- ordlister utvikla av IBM Norge
- diverse tillegg

Universitetet hadde kjøpt materialet frå IBM Norge, som hadde lagt ned språkavdelinga si. Materialet frå ordbøkene og frå IBM måtte tilpassast eit sams opplegg, og det vart supplert med diverse tillegg, m.a. vanlege feiltypar i norsk og vanlege namn, ord som er vanlege i tekster.

Ordbanken har etter kvart vorte ein sjølvstendig ressurs som

er brukt i mange språkteknologiske applikasjonar. Ordbanken er kopla til BOB og NOB for å gje rettskrivingsinformasjon, og er den viktigaste kanalen Språkrådet bruker for å vise publikum den offisielle rettskrivinga. Det er markert i Ordbanken om ei ordform er normert eller unormert (dvs. utanfor den offisielle norma).

## 8. Utvikling av Ordbanken og ordbøkene

Det er Språkrådet og Universitetet i Oslo som saman eig Ordbanken og ordbøkene. Eigarane har sidan 2001 hatt eit styre for Ordbanken med representantar for Språkrådet og dei aktuelle miljøa ved universitetet: Tekstlaboratoriet, Eining for digital dokumentasjon (EDD) og ordbokredaksjonane. Ordbankstyret har behandla saker og sett i gang tiltak som har med drift og utvikling å gjere:

- oppdatering og korrekturlesing av Ordbanken
- sal av fullformsordlister og anna ordbokmateriale
- nytt søkjegrensesnitt for ordbøkene
- utviklingsprosjekt (jf. pkt. 9)

Oppdatering av rettskrivinga og korrekturlesing av Ordbanken har vore nødvendig på grunn av rettskrivingsendringar og for å tilpasse oppføringane frå dei ulike kjeldene til eit eins system.

Ordbanken var ein periode til sals til kommersiell bruk, mens han var gratis for språkforskarar. Det var då ein del firma

som kjøpte Ordbanken, og ordbankstyret la ned mykje arbeid i kontraktar o.a. I dag kan alle laste ned ordlistene gratis frå Norsk språkbank.

I det første søkjegrensesnittet for ordbøkene var søk i BOB førstevalet, nok ut frå tanken at dei fleste brukarar var ute etter opplysningar på bokmål. Det nye grensesnittet frå 2009 fekk ei ny grafisk utforming. Standardsøket vart no enkelt søk i begge ordbøkene på ein gong, av pedagogiske og språkpolitiske grunnar (Hoel 2010). Ved å klikke på bøyingskoden fekk ein no opp det generelle bøyingsmønsteret som oppslagsordet høyrde til.

## 9. Teknisk oppgradering 2012–2013

Universitetet og Språkrådet har dei aller siste åra arbeidd med oppgradering av Ordbanken og dermed ordbøkene. Med oppgraderinga har ein oppnådd at informasjonen som ligg i ordartiklane og Ordbanken, blir vist betre fram. Ei stor forbetring er at vi har fått bøyning også av samansetningar, ikkje berre av grunnord. Før hadde *balkong* bøyingskode, men ikkje *balkongkasse* og *luftebalkong*. Denne endringa føreset at det er rette koplingar mellom ordbok og ordbank. Særleg ved homografar må ein vere merksam på dette. Ei anna forbetring er at ein kan klikke på bøyingskoden og få opp full bøyning av oppslagsordet i eit eige vindauge. Før oppgraderinga fekk ein berre fram det generelle bøyingsmønsteret.

For å gjere ordartiklane i ordbøkene lettare å lese blir ein del

## HOVDENAK

forkorta skrivemåtar løyste opp, iallfall i nettutgåva. I døma har regelen vore at oppslagsord skal skrivast forkorta, t.d. under *amerikansk*: *ein a- bil*. No står det *ein amerikansk bil*.

Å finne ordgrupper, t.d. faste uttrykk, i ordbøkene er ikkje lett, verken på papir eller nett. Difor er det lagt inn søk på uttrykk med % (prosentteikn), t.d. «i %» og «% ryggen». Det første søket gjev tilslag på oppslagsord som «*i dag*» og «*i stand*» i begge ordbøkene. Søk med % gjev flest resultat i NOB, fordi det der er plukka ut fleire tusen faste uttrykk som søkbare underoppslag, t.d. *dolke i ryggen*. Leddeling blir lagd inn i Ordbanken og kjem då inn i ordbøkene (t.d. slik at *villand* blir delt *vill|and*).

The screenshot shows the Bokmålsordboka search interface. The search bar contains 'balkongkasse'. Below the search bar, the results for 'balkongkasse' are displayed, including the word form 'balkongkasse m1, f1 vorandakasse'. A detailed inflection table is shown for 'balkongkasse', with columns for 'Entall' and 'Flertall', and rows for 'Ubestemt form' and 'Bestemt form'.

balkongkasse	Entall		Flertall	
	Ubestemt form	Bestemt form	Ubestemt form	Bestemt form
m1	en balkongkasse	balkongkassen	balkongkasser	balkongkassene
f1	ei balkongkasse	balkongkassa	balkongkasser	balkongkassene

Illustrasjon 2: Skjermbilde med bøyingsmønster i eige vindauge.



## 10. Omtale og bruk av ordbøkene

Alt året etter lanseringa kom det ein grundig fagleg omtale (Nes 1987). Nes var jamt over positiv til både ordutval, definisjonar og døme. Derimot var han kritisk til ein del uttaleopplysningar og ikkje minst til mange av etymologiane.

Worren (1998) drøfter mange sider ved emnet målføreorda i ordbøkene (jf. pkt. 4 ovanfor) og er i det heile ein viktig internrapport.

Nygaard (2006) er ei hovudoppgåve der emnet er å konstruere eit norsk ordnett basert på BOB. Det finst òg andre hovudoppgåver, masteroppgåver og andre studentarbeid som baserer seg på ordbøkene i større eller mindre grad.

Fleire redaktørar og andre norske leksikografar har skrive om emne som kjem inn på ordbøkene, m.a. Killingbergtrø (1994). Engø (1999) og Wangensteen (2010) kommenterer m.a. bruk av søkjeloggar frå ordbøkene. Fjeld (1995) peiker m.a. på skeiv representasjon av tradisjonelle mannlege og kvinnelege domene i ordbøker, og Fjeld (2009) og Vikør (2009) drøftar behandlinga av fleirordssamband.

Akø (1991) har skrive om gråsoner i norske (bokmåls)ordbøker, blant dei BOB. Trosterud (2010) stiller seg kritisk til definisjonane i ordbøkene.

Minst like viktig er bruken blant det store publikum. Oppslaga på Internett har auka mykje. Det kjem respons med forslag og rettingar frå publikum. Brukarundersøkinga til Kola (2012) seier noko om kva nytte elevar har av bøyingsopplysnigane, men ei større undersøking kunne fortelje oss meir om korleis folk bruker ordbøkene, og kva dei ønskjer seg av dei.

## **11. Prøveperiode med revisjon av Nynorskordboka**

NOB kom som nemnt i ei ny nettutgåve i 2012 med ny rettskriving i oppslagsdelen av ordartiklane, jf. punkt 5 over. I teksta elles er det forelda rettskriving (men endringane er ikkje svært store; sideformer som gjekk ut av rettskrivinga, førekom ikkje i tydingsdelen). Denne situasjonen med berre delvis oppdatert tekst er klart uheldig. Difor prøvde universitetet og Språkrådet i 2012–2013 ut ein innhaldsrevisjon. Det skulle vere ein avgrensa revisjon med tanke på ny papirutgåve og deretter ny nettutgåve. Prøveperioden tyder på eit stort revisjonsbehov, både behov for tekniske justeringar knytte til Ordbanken og revisjon av innhaldet (Grønvik 2013). Det er svært tydeleg at verda og språket har endra seg sidan 1970- og 80-åra.

## **12. Ordbøkene og framtida**

BOB og NOB er utvikla for offentlege midlar, og det forpliktar. Dei to eigarinstusjonane vil at BOB og NOB og Ordbanken skal vere ei stabil, fritt tilgjengeleg og gratis publikumsteneste over Internett (forslag til rammeavtale 2012). Oppdatering må vere ei fast oppgåve, og det er betre med kontinuerleg vedlikehald enn skippertak. Ordbøkene bør etter mitt syn mest mogleg følgjast åt i utviklinga.

Nettutgåvene kan gjerast meir funksjonelle og lesarvennlege

på ulike måtar. Det er viktig å byggje ut brukarkontakten og utnytte han godt. Eit viktig framsteg er at vi har fått store korpus til støtte for redigeringa, både for bokmål og nynorsk, Leksikografisk bokmålskorpus og Nynorskkorpuset. BOB og NOB har vunne plass som dei standardordbøkene dei skulle vere, og brukarane fortener at dei blir godt oppdaterte.

## Litteratur

### Ordbøker

BOB = *Bokmålsordboka. Definisjons- og rettskrivningsordbok* (1986–). Oslo: Universitetsforlaget/Kunnskapsforlaget, <<http://www.nob-ordbok.uio.no/>>

1. utg. 1986: Redaksjon: Marit Ingebjørg Landrø og Boye Wangensteen

3. utg. 2005, medverkande (frå forordet): Boye Wangensteen, Ruth Vatvedt Fjeld, Christian-Emil Ore, Tone Adele Bjørnstad

NOB = *Nynorskordboka. Definisjons- og rettskrivningsordbok* (1986–). Oslo: Det Norske Samlaget, <<http://www.nob-ordbok.uio.no/>>

1. utg. 1986: Redaksjon: Marit Hovdenak, Laurits Killingbergtrø, Arne Lauvhjell, Sigurd Nordlie, Magne Rommetveit, Dagfinn Worren.

4. utg. 2006, medverkande (frå forordet): Anne Engø, Kristin Bakken, Marit Hovdenak, Jostein Stokkeland,

## HOVDENAK

- Andreas Tandberg, Dagfinn Worren, Christian-Emil Ore, Tone Adele Bjørnstad, Lars Jørgen Tvedt.  
Guttu, Tor (hovedred.) (1977): *Riksmålsordboken*. Oslo: Kunnskapsforlaget.  
Tanum = *Tanums store rettskrivningsordbok*. (5. utg. 1979 og seinere utg). Oslo: Tanum – Norli / Kunnskapsforlaget.

## Annan litteratur

- Akø, Jørn-Otto (1992): Gråsoner i norske ordbøker. I: Ruth Vatvedt Fjeld (red.): *Nordiske studier i leksikografi. Rapport fra Nordisk konferanse i leksikografi i Oslo, mai 1991*. [Oslo]: Nordisk forening for leksikografi, 65–75.
- Engø, Anne (1999): Bokmålsordboka og Nynorskordboka på Internett. I: *Ord om ord. Årsskrift for leksikografi nr. 5*, 9–12.
- Fjeld, Ruth Vatvedt (1995): Kvinneordboka – et kjønnsperspektiv på leksikografien. I: *Ord om ord. Årsskrift for leksikografi*, 25–30.
- Fjeld, Ruth Vatvedt (2009): Leksikografisk dokumentasjon av flerordsenheter i norsk. I: *LexicoNordica 16*, 103–118.
- Grønvik, Oddrun (2013): *Rapport om innholdsrevisjon av Nynorskordboka 1.11.2012–30.6.2013*. Oslo: Universitetet i Oslo.
- Grønvik, Oddrun & Christian Emil Ore (2014): Samvirket mellom ordbank og ordbok. I: Ruth Vatvedt Fjeld & Marit Hovdenak (red.): *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.–16. august 2013*.
- Gundersen, Dag (1967): *Fra Wergeland til Vogt-komiteen*. Oslo: Universitetsforlaget.
- Hagen, Kristin & Anders Nøklestad (2010): *Bruk av et norsk*

- leksikon til tagging og andre språkteknologiske formål. I: *LexicoNordica* 17, 55–72.
- Helset, Stig Jarle (1999): Definisjonar i norske ordbøker. *Ord om ord. Årsskrift for leksikografi nr. 5*, 21–31.
- Hoel, Jan (2010): Edle årganger i ny tapning. I: *Språknytt* 3, 10–11.
- Killingbergtrø, Laurits (1994): *Språket i kryssord*. Oslo: Institutt for nordistikk og litteraturvitskap.
- Kola, Kjersti Wictorsen (2012): A study of pupils' understanding of the morphological information in the Norwegian electronic dictionary Bokmålsordboka and Nynorskordboka. I: *Euralex proceedings 2012*. [s.l.]: Euralex International Association for Lexicography, 672–675.  
<[http://www.euralex.org/elx\\_proceedings/Euralex2012/pp672-675%20Kola.pdf](http://www.euralex.org/elx_proceedings/Euralex2012/pp672-675%20Kola.pdf)>
- Kulbrandstad, Lars Anders (1976): Norsk handordbok. I: *Språknytt* 2, 7–8.  
*Leksikografisk bokmålskorpus*. <<http://www.hf.uio.no/iln/tjenester/kunnskap/samlinger/bokmal/veiledningkorpus/>>
- Nes, Oddvar (1987): [Melding av] *Bokmålsordboka. Definisjons- og rettskrivningsordbok. Nynorskordboka. Definisjons- og rettskrivningsordbok*. I: *Maal og Minne*, 227–239.
- Norsk språkråd: *Årsmelding 1974–2004*  
<<http://språkrådet.no/Toppmeny/Om-oss/Aarsmeldingar/>>
- Nygaard, Lars (2006): *Frå ordbok til ordnett*. [Hovedoppgave]. Oslo: L. Nygaard.
- Nynorskorpuset*. <<http://no2014.uio.no/korpuset/>>
- Universitet i Oslo og Språkrådet (2013): *Forslag til rammeavtale om vedlikehold og vidareutvikling av Bokmålsordboka og Nynorskordboka og Norsk ordbank (bokmål og nynorsk)*.
- Språkrådet: *Årsmelding 2005–2012* <<http://språkrådet.no/Topp>

HOVDENAK

[meny/Om-oss/Aarsmeldingar/](#)>.

- Trosterud, Trond (2010): Felles leksikalske ressursar for språkteknolog og leksikografi. I: *LexicoNordica* 17, 211–223.
- Vikør, Lars Sigurdsson (2009). Behandling av ordsamband i norske einspråklege eittbandsordbøker. I: *LexicoNordica* 16, 241–253.
- Wangensteen, Boye (2010): Ord som mange lurer på. I: *Språknytt* 2010, 1, 22–23, <[http://www.sprakradet.no/Toppmeny/Publikasjoner/Spraaknytt/Arkivet/Spraknytt-2010/Spraknytt\\_12010/Ord-som-mange-lurer-pa/](http://www.sprakradet.no/Toppmeny/Publikasjoner/Spraaknytt/Arkivet/Spraknytt-2010/Spraknytt_12010/Ord-som-mange-lurer-pa/)>
- Worren, Dagfinn (1981): Norsk handordbok. I: *Språknytt* 1, 17–18.
- Worren, Dagfinn (1998). Om å avgrense eit ordtilfang – saga om målføreorda i Nynorskordboka. I: Ruth E. Vatvedt Fjeld & Boye Wangensteen (red.): *Normer og regler. Festskrift til Dag Gundersen 15. januar 1998*. Oslo: Kunnskapsforlaget, 59–70.

Marit Hovdenak  
seniorrådgjevar  
Språkrådet  
Postboks 8107 Dep  
0032 Oslo  
[marit.hovdenak@sprakradet.no](mailto:marit.hovdenak@sprakradet.no)

## **Marcus Pauli Sadeleri – finsk medarbetare i Schroderus *Lexicon Latino-Scondicum***

*Petri Lauerma*

Marcus Pauli Sadeleri (ca 1610–1662), a clergyman born in Pori and educated in Uppsala and Turku, helped Ericus Schroderus in Oulu to collect and edit Finnish material for *Lexicon Latino-Scondicum* (1637), a dictionary containing Latin, Swedish, German and, for the first time, also Finnish words. The Finnish material (ca 2400 words) includes words from Sadeleri's own Satakunta dialect, but also Ostrobothnian lexical material, probably reflecting his own interests and connections to this area. There are also neologisms (both derivatives and compound words) probably formed by Sadeleri, as well as words from folk poems. Though Schroderus was considered a purist, also loan words with non-Finnish initial consonant clusters and voiced stops occur. Sadeleri also wrote some historical studies on Finland and its literature, but these have not survived. Later in life Sadeleri worked in several Finnish parishes and schools, but due to his difficult character he invariably found himself in trouble, and he finally died in poverty.

## 1. Inledning

Den allra första ordboken som innehåller finska ord är *Lexicon Latino-Scondicum* (1637), ett fyrspråkigt parlörlexikon sammanställt av Ericus Schroderus (ca 1608–1639), som var en pionjär också inom svensk lexikografi. (Om Schroderus och hans ordböcker se t.ex. Hesselman 1929, 1941; Romppanen 2001:106–114.)

Schroderus föddes i Uppsala troligen år 1608. Han inskrevs i Upplands nation vid Uppsala Akademi år 1625 och blev filosofie magister tio år senare. Han hade redan börjat sammanställa en ordbok med svenskt, tyskt, latinskt och grekiskt material, då han i december år 1635 mottog en kunglig skrivelse där det stod att han skulle resa till Uleåborg och där leta upp ett manuskript, som historikern Johannes Messenius hade skrivit i fängelse.

Schroderus arbetade i Uleåborg nästan ett år. Arbetet med Messenius tog inte hela hans tid i anspråk, och eftersom han erfor att det i Finland fortfarande saknades en ordbok över det finska språket började han skriva ett sådant verk själv. Hovdhaugen m.fl. (2000:61) antar att Schroderus redan hade publicerat en latinsk-svensk ordbok år 1631, men enligt Romppanen (2001:106–107) har ett sådant verk inte dokumenterats och tryckningen av *Lexicon Latino-Scondicum* redan år 1632 tycks vara bara ett bibliografiskt tryckfel. Det är mera sannolikt, att Schroderus avbröt sitt tidigare lexikaliska arbete och började sammanställa ett fyrspråkigt parlörlexikon, där det i stället för grekiskt material för första gången ingick finskt material.

Schroderus var troligen inte kunnig i finska när han inledde sitt arbete i Uleåborg. I förordet till *Lexicon* (Epistola dedicata, a 4) berättar Schroderus att han assisterades av sin vän



Marcus Pauli Björneburgensis, som lärde honom finska och hjälpte honom att samla ihop det finska materialet för ordboken. Schroderus-forskare har nästan glömt Marcus Pauli (t.ex. nämner Romppanen 2001:110 honom bara i en fotnot, 133), men vi kan bygga upp en bild av hans liv med hjälp av olika källor (Strandberg 1834:285–287; Setälä 1898:315–316; Melander 1941:10–11; Jokipii 1954:221–222; Palola 1999; Väänänen 2011; jfr också Hakulinen 1960:118–119).<sup>1</sup>

Marcus Pauli föddes cirka 1610 i Björneborg. Hans far var sadelmakare, vilket avspeglas i släktnamnet Sadeleri eller Sadelerus, som Marcus Pauli så småningom började använda med varierande stavningar (ofta med Z-). År 1631 inledde han sina studier vid Uppsala universitet. Senare blev han kaplan i Uleåborg, där han mötte Schroderus år 1636.

Vid den tid då Schroderus och Sadeleri inledde sitt samarbete hade Schroderus redan samlat in lexikalskt material för sitt ordboksprojekt. Det är alltså troligt att Sadeleri utgick från Schroderus ordlistor och använde dem när han samlade in finskspråkigt material för ordboken.

## **2. De finska uppslagsorden i Schroderus *Lexicon* och deras dialektala drag**

*Lexicon Latino-Scondicum* består av 45 kapitel, men ordartik-

---

1. I det följande nämner jag inte dessa individuella källor när de är samstämmiga i fråga om Sadeleris liv och verksamhet.

larna i den fyrspråkiga ordlistan är uppdelade i 36 ämnesgrupper (de sista nio kapitlen består av svenska avledningar och sammansättningar med översättningar till latin, men inte till finska och tyska). Ordboken innehåller cirka 2400 uppslagsord, men där finns inte lika många finska ekvivalenter. Uppslagsorden har i regel bara en finsk ekvivalent, men det finns också undantag.

Redan Ojansuu (1907:47 fotnot 2) – mer detaljerat i Sainio (1959:156) och Häkkinen (1987:21) – påpekar att namnen på vissa växter (och ett djur) tycks ha varit okända för Schroderus och hans finska språklärare. Därför blir *Hoxl* (äppelrönn, *Sorbus domestica*), *Alm* (alme, *Vlmus*) och *Lind* (*Tilia*) bara *Se pw* 'ett träd (ordagrant: det trädet)', *Nata* 'Anagallis', *Molla* 'Atriplex' och *Summill* 'Filipendula' blir bara *Se ruoho* 'ett gräs', *Meeran* (*Majorana*), *Matram* (*Matricaria*), *Myskia* (*Matri-sylvia*), *Amor* (*Melilotus*) och *Timian* (*Thymus*) blir bara *Se yrtti* 'en ört' och *Lämmel* (*Mus noricus*) blir bara *Se hijri* 'en mus'.

Man kan fråga sig om forskarna bara ser fel i de gamla lexi-kografernas arbete, för endast Lehtivaara (1964:73) påpekar att det också finns ord som har två finska ekvivalenter. Inte ens Lehtivaara studerar dock dessa fall närmare, trots att sådana ord inte alls är sällsynta. Det finns nästan hundra stycken, så bara huvudtyperna kan presenteras här.

I några fall är förklaringen till att det finns flera finska ekvivalenter att det finns flera ekvivalenter också i de andra språken i *Lexicon* (t.ex. Pars – Deel / Part / Stycke – Thets / Stycke – Kappale / Osa). Orsaken kan också vara att det ingår lånord från olika språk (t.ex. *Cantherius equus* – Rone / Wallack – der Wallach – Rwna / Walacka). Det finns också fall, där både ett lånord och det ursprungliga finska ordet ges som ekvivalenter (t.ex. *Cattus* – Katt / Katta – die Katze – Kissa / Katti). Men i de

flesta fallen verkar det vara så att de olika finska orden kommer från olika dialekter (t.ex. Aestas – Sommar – Sommer – Kesä / Suvi). Detta visar att den person som samlat in de finska orden för Schroderus kände till flera dialekter – eller att Schroderus hade flera finska språklärare eller andra finska källor från olika dialektområden.

Även om man inte har fäst uppmärksamhet vid de finska synonymerna, har man i den tidigare forskningen noggrant förklarat den dialektgeografiska spridningen av det finska materialet i Schroderus *Lexicon*. Redan Lehtivaara (1964:70–71) påpekar att det finns ord i *Lexicon* som är kända i olika dialekter i Satakunta, men Nummila (2010:193) framhåller att sådana ord (t.ex. *Allicko* [*allikko*] ’vattenpöl’, *Haapio* ’eka’) vanligtvis också är kända i åtminstone några andra dialekter. Både lägger märke till att *Lexicon* också innehåller ett stort antal dialektala ord, som är typiska för syd- och mellanösterbottniska dialekter (t.ex. *Hillku* [*hilkku*] ’bahytt’, *Capaloxet* [*kapalokset*] ’linda’, *Klunti* [*kluntti*] ’skolpojke’ < sv. *glunt*). Där finns också ett fonetiskt drag, som numera är typiskt för dessa dialekter: de ord som i riksspråket slutar på *e* + aspiration slutar i *Lexicon* ofta på *et* (t.ex. *Hamet* ’kjol’); detta dialektala drag var tidigare också typiskt för dialekter i norra Satakunta (Rapola 1933:233; Lehtivaara 1964:63–64). I *Lexicon* finns inga ord som bara är kända i Uleåborgsdialekten, som Lehtivaara (1964:72) tänker sig, men det finns några ord som bara hör hemma i de nordligaste finska dialekterna (*Marsio* ’matportion för husdjur’, *Tarvi* ’kracka’, *Tohtaia* [*tohtaja*] ’storlom’) (Nummila 2010:193–194). Enligt Lehtivaara (1964:73) finns det också ett litet antal ord, som är typiska för dialekterna i östra Finland (t.ex. *Ajelus* ’tiggare’, *Linda* [*linta*] ’mjölkvälling’); några av dem är särskilt typiska för Kajanaland. Även dessa ord är emellertid också kända i några västfinska dialekter.

Om vi noggrant läser vad Schroderus (1637) skriver i sitt förord (*Epistola Dedicatoria*, a4), märker vi att han inte förnekar att han även har haft andra finska assistenter. Marcus Pauli är dock den enda som uttryckligen nämns och sannolikt den viktigaste av dem alla. Han var själv intresserad av det österbottniska området, för det är känt att Henrik Gabriel Porthan, den finska historieforskningens fader, i slutet av 1700-talet försökte hitta «Zadeleri Manuscript om Österbotten» (Forsman 1900:155). Schroderus kan ha haft en assistent från det österbottniska dialektområdet som hörde till Sadeleris egen familj, för hans hustru var dotter till kyrkoherden Simon Ericus Frosterus. Denne var född i Kalajoki, som numera hör till det mellanösterbottniska dialektområdet men som på 1600-talet låg mycket närmare det sydösterbottniska dialektområdet. Intressant är också att Frosterus hade arbetat i Paltamo i Kajanaland, i det östfinska dialektområdet, som det också finns inslag av i vokabulären i *Lexicon*. Lehtivaara (1964:73) tänker sig att det var just Frosterus som införde de östfinska synonymerna till de västfinska orden i *Lexicon*, om de inte var kända i Uleåborgsdialekten.

### 3. Neologismer och ord ur folkdiktningen

De flesta finska orden i *Lexicon* representerar den vokabulär som kan kontrolleras med hjälp av texter från 1500- och 1600-talet. En del ord förekommer för första gången i Schroderus ordbok. Rapola (1960:88) har räknat ut att 16 av dessa ord ännu

används, men det finns tidigare förekomster av tre av orden (*Cantzli* 'kansli', *Paino* 'tyngd' och *Pyörä* 'skiva', Jussila 1998:86, 184, 210).

Materia	Wärcke / ämne	die Materia	Aine
Testudo	Hwalff	das Gewelb	Hålfvi [holvi]
Turbo	Ruta	der Kreusel	Cartio [kartio]
Pulvis tormen- tarius	Krwt /	das Pulver	Kruuti [ruuti]
Classis	Class	die Class	Luocka [luokka]
Cotyla	Fierding	ein Viertheil	Nelikoi [nelikko]
Horologium	Senare	die Vhr	Tundi [tunti] kello
Nicosianum	Toback	der Taback	Tobacki [tupakki]
Inquisitio	Räffst	die Straff	Tutkimus
Collimare	Måtta till	Nach zielen	Tähdätä
Protervia	Tredska	der Mutwill	Vppiniskaisus [uppiniskai- suus]
Dies Civilis	Dygn	Tag vnd Nacht	Vuorocausi [vuorokausi]
Severitas	Alfwar	der Ernst	Yxtotisus [yksi- totisuus]

Tabel 1. Finska neologismer i Schroderus *Lexicon*.

I tabellen ingår bara sådana nya ord som fortfarande används – inte alla nya ord.

I förordet till *Lexicon* uppmanar Schroderus (1637, Epistola Dedicatoria a3) läsaren att undvika lånord och att i stället använda egna stamord, avledda ord och sammansättningar. *O-*avledningarna som *Saardo* (*saartaa* 'belägra' + *o* -> *saarto* 'förskansning') är särskilt vanliga i det finska materialet i *Lexicon*, men där finns också andra avledningar som t.ex. *Werhotus* (*verhota* 'täcka' + *us*) 'täkt'. Sammansatta ord är vanligare än avledningar. Några av sammansättningarna är ganska roliga: 'torp' är *Köyhäkylä*, ordagrant 'fattig by', 'bordell' är *Huorahuonech*, ordagrant *hora* + *hus*, 'arkiv, skattkammare' är *Parhan calunkätkö*, ordagrant 'gömma för det bästa materialet'. I det finska materialet i *Lexicon* finns enligt Ojansuu (1907) nästan 40 avledda ord (39) och mer än 100 (101) sammansättningar som inte använts tidigare, men dessa avledningar och sammansättningar är vanligen okända också i senare källor, särskilt i dialekterna (Lehtivaara 1964:75). Några ord har kanske bara försvunnit, men största delen av neologismerna har aldrig blivit allmänna.

Men alla ord som inte kan hittas i tidigare tryckta böcker eller i dialekter är inte nyskapelser. Man har också använt folkdiktningen som källa för den finska vokabulären i *Lexicon*. Det finns t.ex. ganska många mytologiska figurer i ordboken. Sarajas (1956:32–33) förstår att ord som *Köpeli* 'spöke', *Liekiö* 'rå', *Menningäiset* 'skepnad' och *Tyrä* 'troll' inte är nyskapelser av Sadeleri, trots att man bara har kunnat hitta några få av dem i tidigare litterära källor. Även sammansatta ord som *Calevan Poica* 'jätte' (ordagrant «Son till Kaleva»), *Lindu codonmies* 'dvärg' ('Man av Fågelland') *Vaaka lintu* (enligt *Lexicon* 'Griip', men i dialekterna också 'örn') och *Wedenhaltia* 'vattenande' kommer naturligtvis från folkdiktningen (jfr också Toivonen 1944:7–56, 110–128).

Hakulinen (1960:119) påpekar att det finns sådana icke-mytologiska ord i *Lexicon* som är vanliga i folkdiktningen, men

som exempel ger han bara ett ord, *Rauta paita* 'pansar'. Man kan hitta också andra armétermen i den (numera) elektroniska databasen för gamla finska folkdikter (SKVR).

Officianus	Officerare	der Befehls- haber	Käskyläinen
Signifer	Fendrick	der Fendrich	Lipun candaja
Lorica	Pantzar	das Pantzar	Rauta paita
Tuba	Trometa	die Trommete	Vaski torvi
Strategus	Öfwerste	der Oberster	Ylimäinen

Tabel 2. Armétermen i *Lexicon* som är kända från folkdiktningen.

Orden som *Käskyläinen* (jfr motsvarande tyska ordet) och *Ylimäinen* (ordagrant: 'den som är högst uppe') kan vara lånöversättningar, som är vanliga i *Lexicon* (Ojansuu 1907:47). Men de andra orden (som ordagrant betyder 'fanbärare', 'järnskjorta' och 'mässinghorn') kommer troligen från folkdiktningen. Sadeleri var med säkerhet intresserad av denna, för enligt Warmholtz (1782:n:o 750; jfr också Melander 1941:10) skrev han några forskningar om Finlands litteratur och historia.

#### 4. Svenska lånord i finskan

Det finns också lånord i det finska materialet i *Lexicon*, trots att Schroderus var emot dessa. I förordet till *Lexicon* citerar Schroderus tankar av Anders Bureus (1626) ur dennes kommentar till

en nordisk atlas. Bureus skriver att finska ord inte börjar med *b*, *d*, *g*, *f* eller med två konsonanter och att finskspråkiga personer inte kan uttala dessa ljud när de talar andra språk. Men Schroderus (1637, Ad Lectorem b3 – b4) märkte att det finns sådana svenska lånord som *Duomio*, *Fijkuna*, *Psaltari* och *Bläckki* i det finska språket. Sådana ord är inte särskilt ovanliga i *Lexicon*.

Det finns sju ord som börjar med *b* (det finska ordet *Brukaus* är en *us*-avledning av det svenska *Bruk*).

Bläckki	Bleck
Bojorti	Bojort
Bolstari	Bolster
Brukaus	Brwk
Bunet	Boonat
Buodhi	Bodh
Böfweli	Böfwel

Tabel 3. Finska ord som börjar med *b* och motsvarande svenska ord i *Lexicon*.

Det finns tretton ord som börjar med *d* – ett av dem (*denga*) kommer från ryskan (< денга, SSA under ordet *tenka*).



Daartti	Dart
Damaski	Damesk
Dantzi	Dantz
Denga	Denning
Desmo	Desma
Dickuri	Decker
Dircki	Dyrckel
Doctari	Dochtare
Drotzi	Drotz
Drötingi	Drottning
Duomio	Doom
Duomitä	Döma
Dusina	Dussin

Tabel 4. Finska ord som börjar med *d*.

Det finns fem ord som börjar med *g* – ordet *gaffeli* kommer från svenska dialekter (i *Lexicon* även *Gafwel*, jfr också SSA under ordet *kahveli*).

Gaffeli	Pijk
Galleijo	Galeya
Glasi	Glaas
Greifvi	Grefwe
Gryni	Gryyn

Tabel 5. Finska ord som börjar med *g*.

Det finns arton ord som börjar med enkelt *f*-ljud.

LAUERMA

Faderma	Hållon
Fangi	Fånge
Fangius	Fängsle
Fati	Faat
Fechteri	Kempe
Fedla	Fedla
Feldi	Fäll
Fijla	Fijla
Fijlata	Fijla
Fikuna	Fijkon
Focka	Focka
Formu	Form
Foudhi	Fogde
Funtti	Funt
Fuori	Foder
Färnisia	Fernissa
Fästi	Fäste
Förstinna	Fröken

Tabel 6. Finska ord som börjar med *f*.

Ordet *Faderma* 'hallon' är inte ett lånord, men det är ett exempel på hur *f*-ljudet har kommit in i den naturliga vokabulären i några finska dialekter, särskilt i Satakunta (Rapola 1933:96, 1966:18). Ordet *Fangius* 'fångenskap' är en *us*-avledning av den svenska *Fånge*. Ordet *Fechteri* kommer från tyskans *der Fechter*, men *Förstinna* 'Principissa, die Fürstin' är ett direkt lån av det svenska ordet *Förstinna* (SSA under ordet *Försti*, VKS under ordet *Förstinna*).

Det finns också sex ord som börjar på en konsonantförbindelse som börjar med *f* (det finska ordet *Flasku* är en *u*-avledning av det svenska ordet *Flaska*).

Flammiska	Flammesk
Flasku	Flaska
Fleini	Fleen
Flitteri	Flitter
Fläski	Fläsk
Frouwa	Frw

Tabel 7. Finska ord som börjar med *fl* och *fr*.

Finska ord som börjar med konsonantförbindelser är ganska vanliga i *Lexicon*. Von Fieandt (1907:35–38) listar inalles 89 sådana ord (inklusive 13 ord som börjar med *bl*, *br*, *dr*, *gl*, *gr*, *fl* och *fr*). Här följer ett urval av ord som börjar med andra konsonantförbindelser (det deskriptiva ordet *Snippu* kommer möjligen från det svenska ordet *snibb*, SSA under ordet *nippa*).

kl-	Clasi	Glass
kr-	Kraana	Kraan
kn-	Knappi	Knöpar
pl-	Pladhi	Blaad
pr-	Prami	Prååm
sk-	Skantzi	Skanz
skr-	Skrwfi	Skrwff
sm-	Smacko	Smacka
sn-	Snippu	Preen
sp-	Spanni	Spann
spr-	Spröthi	Spröte
st-	Stalli	Stall
str-	Strutzi	Strutz
sv-	Swetskyt	Swetzkon
tr-	Traani	Tråån

Tabel 8. Finska ord som börjar med konsonantförbindelser *kl*, *kr*, *kn*, *pl*, *pr*, *sk*, *skr*, *sm*, *sn*, *sp*, *spr*, *st*, *str*, *sv* och *tr*.

Ord som börjar med en tonande klusil, *f* eller en konsonantförbindelse var vanliga i äldre finskt skriftspråk (Rapola 1933:96–100). Det finns ord som börjar med en konsonantförbindelse eller med *f* även i de finska dialekter som gränsar till det svenska språkområdet i Finland, men ordinitiala konsonantförbindelser som börjar med *s* (obs. även *skr-*, *spr-* och *str-*) är möjliga bara i några finska dialekter i Österbotten (särskilt i Laihela), på några ställen i Egentliga Finland och framför allt i Vittisbofjärd nära Björneborg (Von Fieandt 1907:15–16; Kettunen 1969:karta n:o 1). Det är alltså möjligt att Sadeleris hemdialekt (han var ju född i Björneborg) också i någon mån avspeglas i lånordens fonologi.

Det är också intressant att Schroderus accepterade att det fanns så många lånord i det finska materialet i *Lexicon*, trots att han i grunden hade puristiska åsikter. Kanske hade varken Sadeleri eller Schroderus andra finska assistenter tid att konstruera flera finska ekvivalenter till dessa lånord. Schroderus hann i alla fall inte göra det efter att han hade återvänt hem till Uppsala, eftersom de båda förorden i *Lexicon* är daterade redan i början av år 1637, samma år som ordboken trycktes. Efter publiceringen av *Lexicon* kunde Schroderus återgå till att arbeta med sin andra ordbok *Dictionarium Quadrilinguae*, men tyvärr bara under två år, eftersom han omkom i en druckningsolycka i Stockholm redan den 14 januari 1639.

## 5. Sadeleris senare verksamhet

Sadeleri levde längre. Han stannade inte i Uleåborg utan lämnade sin anställning som kaplan och återvände till Uppsala för att studera. År 1640 blev han konrektor för Björneborgs trivialskola, och senare samma år blev han anställd som konrektor vid Åbo katedralskola. Han blev också kaplan i den finska församlingen i Åbo.

Sainio (1959) frågar sig om Sadeleri kunde ha haft någonting att göra med *Variarum rerum Vocabula*, den näst äldsta finska ordboken som publicerades år 1644, men svarar själv att det inte är sannolikt, eftersom ortografin i *Vocabula* är betydligt modernare än den i *Lexicon*. Sådana äldre skrivsätt som *Haria* (jfr *Harja*) ’borste’, *Händhä* (jfr *Händä*) ’svans’, *Kaula* (jfr *Caula*) ’hals’ och *Pw* (jfr *Puu*) ’träd’ förekommer inte längre i *Variarum*; ordartiklarna i *Variarum* är också längre och fyllda med synonymer och det finns bara tusen gemensamma ord (Sainio 1959:153–156). Nummila (2010:194–196) påvisar, att det även i *Variarum* finns ett stort antal ord från norra Satakunta, men redaktören för denna ordbok är fortfarande okänd.

Ett annat anonymt verk, den latinspråkiga finska grammatiken *Rudimenta linguae finnicae breviter delineata*, som publicerades förra året (Lauerma 2012), innehåller några ortografiska arkaismer av samma typ som *Lexicon* (Palkki 2012:45). /k/ är transkriberat också med *k* även för bakvokaler (t.ex. *Akana(t)* ’agn(ar)’), /h/ är transkriberat även med *ch* i slutet av stavelsen (t.ex. *Hochtimet* ’hovtänger’). /v/ är ibland transkriberat med *fw*, men mera sällan i *Rudimenta* än i *Lexicon* – i *Lexicon* skrivs adjektivet *hyvä* ’bra’ med *fw* (*Hyfwä* eller *Hyfve*), men i *Rudimenta* bara med *w* (*hywä*). Ordet *kuka* ’vem’ stavas *Cuka* i *Lexicon*, men *cuca* eller *kuca* i *Rudimenta*; enligt *Rudimenta* saknas

*ff* helt i finskan, men i *Lexicon* finns ordet *Safframi* 'saffran'. På basis av ett brev som har bevarats i Riksarkivet (Marieberg, Stockholm)<sup>2</sup> kan det även konstateras att Sadeleris handstil avviker från handstilen i *Rudimenta*. Tyvärr har manuskripten till Sadeleris historiska forskningstexter försvunnit – det enda egna verket av Sadeleri som med säkerhet finns bevarat är ett latinspråkigt tal om retorik, som hölls på Uppsala universitet och trycktes i Stockholm 1642.

Efter grundandet av Kungliga Akademien i Åbo fortsatte Sadeleri sina studier där, men studierna drog ut på tiden på grund av att han var stridslysten och orsakade trubbel av olika slag. Han blev också avstängd från sin tjänst under ett år. År 1647 blev han äntligen filosofie magister.

Samma år utnämndes han till kyrkoherde i Uleåborg. Men det blev problem och rättegångar också där (om dessa se Palola 1999:32) och år 1655 blev han avsatt. Efter några stökiga år i Åbo utnämndes Sadeleri till rektor för Nykarlebys trivialskola år 1659, men redan följande år avsattes han även från denna tjänst. Sadeleri tillbringade sina sista år i Åbo men var vid den tiden redan så alkoholiserad att han inte längre fick någon befattning. Han dog år 1662.

Sadeleri var en lärd man men hans karaktär och levnadsvanor gjorde att han inte var lämpad för kyrkliga tjänster. Om en enda av hans egna historiska forskningar hade bevarats, skulle hans eftermäle ha kunnat vara totalt annorlunda. Men *Lexicon* försvann inte, och vi borde komma ihåg att det var framför allt Sadeleri som samlade och även skapade det finska materialet för Schroderus och att Sadeleri var den första finländare som sysslade med lexikografiskt arbete i Finland.

---

2. Brevet är inte katalogiserat i brevskrivarregistret, men det finns i Skoklostersamlingen II, band E 8168.

## Litteratur

### Ordböcker

- Schroderus, Ericus (1637/1941): *Lexicon Latino-Scondicum*. Utgivet med alfabetiska register över de svenska och de finska orden av Bengt Hesselman. Uppsala: Lundequitska Bokhandeln.
- Schroderus, Ericus (1929): *Dictionarium Quadrilinguae Suedicum Germanicum Latinum Graecum*. Med en inledning och ett alfabetiskt register över de svenska orden - -. Utgivet av Bengt Hesselman. Uppsala: Almqvist & Wiksell.
- SSA (1992–2000) = Erkki Itkonen & Ulla-Maija Kulonen m.fl. (red.): *Suomen sanojen alkuperä. Etymologinen sanakirja*. Helsinki: SKS & Kotimaisten kielten tutkimuskeskus.
- VKS (1985-) = Esko Koivusalo m.fl. (red.): *Vanhan kirjasuomen sanakirja*. Helsinki: Kotimaisten kielten tutkimuskeskus.

### Andra källor och litteratur

- Bureus, Anders (1626/1985): *Orbis Arctoi, Imprimisque Regni Sveciae Nova et accurata descriptio*. I: *Pohjoismaiden kuvaus v. 1626*. Suomennos, johdanto ja selitykset Tuomo Pekkanen. Helsinki: SKS.
- Fieandt, Nanna von (1907): *Kaksi konsonanttia sanan alussa suomen kielessä*. Suomi IV:5. Helsinki: SKS.
- Forsman, Juho Rud. (1900): *Kaarle Fredrik Mennander ja Hänen Aikansa*. Turku.

- Hakulinen, Lauri (1960): *Handbuch der finnischen Sprache II*. Wien: Otto Harrassowitz.
- Hesselman, Bengt (1929): Inledning. I: Schroderus: *Dictionarium Quadrilinguae Suedicum Germanicum Latinum Graecum*, V–XXVI.
- Hesselman, Bengt (1941): Utgivarens förord. I: Schroderus: *Lexicon Latino-Scondicum*, III–XII.
- Hovdhaugen, Even, Fred Karlsson, Carol Henriksen & Bengt Sigurd (2000): *The History of Linguistics in the Nordic Countries*. Helsinki: Societatis Scientiarum Fennica.
- Häkkinen, Kaisa (1987): Ordböcker med finska och svenska från Schroderus till Karlsson. I: *Fennistica festiva in honorem Göran Karlsson septuagenarii*. Åbo: Åbo Akademis förlag, 21–36.
- Jokipii, Mauno 1954 (red.): *Satakuntalaisen osakunnan matrikeli*. Helsinki: Otava.
- Jussila, Raimo (1998): *Vanhat sanat. Vanhan kirjasuomen ensiesiintymiä*. Helsinki: SKS.
- Kettunen, Lauri (1969/1940): *Suomen murteet III Murrekartasto* (3. uppl.). Helsinki: SKS.
- Lauerma, Petri (red.) (2012): *Rudimenta linguae finnicae breviter delineata. Suomen kielen varhaiskielioppi ja sen tausta*. [English Summary]. Helsinki: SKS.
- Lehtivaara, Rauli (1964): *Lexicon Latino-Scondicum in suomenkielisen sanaston murremaantieteellistä tarkastelua*. Pro gradu -työ. Jyväskylän yliopisto, suomen kieli.
- Melander, Toini (1941): *Suomalaista tilapäärinoutta Ruotsin vallan ajalta*. Helsinki: SKS.
- Nummila, Kirsi-Maria (2010): 1600-luvun sanakirjojemme suomalaissanasto ja sen alkuperä. I: *Sanoista kirjakieliin*. Helsinki: Suomalais-Ugrilainen Seura, 189–202.
- Ojansuu, Heikki (1907): Pieniä lisiä purismin historiaan. I:



- Virittäjä*, 46–52.
- Palkki, Riitta (2012): Rudimentan vanhaa kirjasuomea ja murteellisuuksia. I: Lauerma (red.), *Rudimenta*, 38–47.
- Palola, Ari-Pekka (1999): Markus Sadelerus – Oulun ensimmäinen historioitsija ja epäonnistunut kirkkoherra. *Historian viesti*. Oulu: Oulun yliopisto, Suomen ja Skandinavian historia, 27–33.
- Rapola, Martti (1933): *Suomen kirjakielen historia pääpiirteittäin I*. Helsinki: SKS.
- Rapola, Martti (1960): *Sanojemme ensiesiintymiä Agricolasta Yrjö Koskiseen*. Helsinki: SKS.
- Rapola, Martti (1966): *Suomen kielen äännehistorian luennot*. Helsinki: SKS.
- Riksarkivet (Marieberg, Stockholm). *Skoklostersamlingen II*.
- Romppanen, Birgitta (2001): *Från målspråk till källspråk. Utvecklingen av den finsk-svenska och den enspråkiga finska ordboken*. Vasa: Vasa universitet.
- Sainio, Kauko (1959): Suomenkielisten oppikirjojen varhaishistoriaa. I: *Aarni Penttilä juhlakirja 5.8.1959*. Jyväskylä: Jyväskylän yliopisto, 151–166.
- Sarajas, Annamari (1956): *Suomen kansanrunouden tuntemus 1500–1700-lukujen kirjallisuudessa*. Helsinki: Otava.
- Setälä, E. N. (1898): Ensimmäinen suomen kielen sanakirja. I: *Kirjailija-Albumi*. Helsinki: WSOY, 309–317.
- SKVR = *Suomen Kansan Vanhat Runot*.  
<<http://www.finlit.fi/skvr>>
- Strandberg, Carl Henric (samladt och i ordning bragt) (1834): *Åbo Stifts Herdaminne ifrån Reformationens början till närvarande tid*. Sdnare Delen. Åbo: Christ. Ludv. Hjelt.
- Toivonen, Y. H. (1944): *Sanat puhuvat*. Helsinki: WSOY.
- Väänänen, Kyösti (2011): Sadelerus, Marcus Pauli. I: *Turun hiippakunnan paimenmuisto 1544–1721*. Helsinki: Biogra-

LAUERMA

fiakeskus, SKS & Suomen kirkkohistoriallinen seura.

<[www.kansallisbiografia.fi/paimenmuisto/?eid=2179](http://www.kansallisbiografia.fi/paimenmuisto/?eid=2179)>

Warmholtz, Carl-Gust. (1782/1966–1967): *Bibliotheca historica Sveo-Gothica*. København: Rosenkilde og Bagger.

Petri Lauerma

äldre forskare, FD, docent

Institutet för de inhemska språken

Berggatan 24

00100 Helsingfors

[plauerma@kotus.fi](mailto:plauerma@kotus.fi)

## **Petter Dass' ordforråd – hans kilder og hans forfatterskap som kilde for ordbøger**

*Hanne Lauvstad*

The vocabulary of the first great Norwegian author, Petter Dass (1647–1707), is presented and discussed based on his major secular work, *Nordlands Trompet* (The Trumpet of Nordland). This work is a learned, versified description, an *encomium*, modeled on the classics, to a region, i.e. what at that time was known as Nordlands Len (the county of Nordland). Petter Dass' wording reflects the rhetorical norms of his time and the language purism of the late 17<sup>th</sup> century. His vocabulary is related to that of Danish literature of the time, but is supplemented by Norwegianisms and dialectal words and phrases from Nordland, Petter Dass' home region, as well as loans from Latin, Low German, Dutch and Sami. The role of the Norwegianisms, the dialectal elements and the foreign parts of the vocabulary in Petter Dass' poetry is discussed, as well as the representativeness of Petter Dass' Norwegian vocabulary in general and of the Nordland dialect specifically. Petter Dass' choice of words, his rhetorical intention and his most important literary sources and models are presented. Furthermore, the references to Petter Dass' literary works in the most prominent Norwegian and Danish literary dictionaries are accounted for. Finally, the lexicographic value of early-modern literature is discussed. A more thorough lexicographic description of the language of Petter Dass has yet to be put forward, and alternative lexicographic presentations of Petter Dass' literary vocabulary are suggested.

## Innledning

Petter Dass (1647–1707) regnes som den første store forfatteren i norsk nasjonallitteratur. Han var prest, og hans diktning består først og fremst av salmer og åndelige sanger, men også bl.a. av leilighetsdikt, som var en prestisjetung sjanger før romantikken. *Nordlands Trompet* er en større beskrivelse på vers av den landsdelen hvor forfatteren selv var født og levde mesteparten av sitt liv. Det oppfattes i dag gjerne som hans hovedverk.

Petter Dass' diktning har dels vært tolket som ledd i en norsk nasjonsbyggende prosess. Den ble «gjenoppdaget» av forfatteren Johan Sebastian Welhaven i 1856 (Hauge (utg.) 1992:365–390). I en berømt artikkel plasserte Welhaven Dass' forfatterskap innenfor rammene av den norske nasjonale litterære kanon som ble etablert i forbindelse med nasjonsbyggingen utover på 1800-tallet og på begynnelsen av 1900-tallet. Senere, fra 1970-tallet, har Petter Dass' diktning også blitt oppfattet som uttrykk for nordnorsk regional identitet, dvs. den er blitt brukt i det man kunne kalle regionsbyggende øyemed.

Nyere forskning har også lagt vekt på de retoriske sidene av forfatterskapet, og i Lauvstad 2006 har jeg arbeidet med å plassere Dass' forfatterskap i sin historiske kontekst, både språklig, retorisk og litterært. Her vil jeg løfte frem materialets leksikografiske aspekter.

Et verk fra slutten av 1600-tallet bør leses ut fra de forutsetninger det er blitt til under, og tolkningen bør ta hensyn til hvilke litterære og språklige normer som har preget det. Petter Dass' ordforråd skal her beskrives med utgangspunkt i hans største verdslige verk, *Nordlands Trompet*. Litteraturen fra før slutten på 1700-tallet hadde andre normer og målsetninger enn

litteraturen fra romantikken og senere, for romantikken medførte et litterært paradigmeskifte, som førte til at man bl.a. la vekt på individualitet og originalitet. Gjennom 1600-tallets utdannelsessystem, fra latinskole til universitet, ble man som lærd skolert inn i en litterær tradisjon som bygde på den klassiske greske og latinske litteraturen. Målet var å tilegne seg tekstene og kunne imitere dem senere, både i form av emner, motiver og uttrykksmåter. Litteraturen utgjorde et repertoar, et forråds-kammer, hvor forfatteren kunne låne det han trengte av motiver, topoi, troper, figurer m.m. Det gjaldt likeledes å imitere klassikerne, men på en måte som fikk imitasjonene til å bli en organisk del av den teksten man selv skrev.

Skriveferdighetene ble innlært etter mønster av den klassiske retorikkens regler, hvor konteksten var viktig for hvordan man utformet budskapet, og hvor forskjellige sjangre krevde forskjellige fremgangsmåter. Idealet var altså ikke originalitet, det viktige var å kunne vise at man var lærd, *doctus*, og hadde den kunstferdighet, *ars*, som var forventet av en forfatter. Språk og sjanger skulle tilpasses fremstillingens tema, motiv, kontekst og mottager i henhold til reglene for det som passet seg, *decorum*. Man opererte med høystil, mellomstil og lavstil, og hver stil hørte sammen med visse motiver, språklige virkemidler og mottagere.

*Nordlands Trompet* er en beskrivelse på vers, et hyllingsdikt (*encomium*) etter klassisk forbilde til et geografisk område, nemlig daværende Nordlands len. Verket henvender seg til et lærd publikum og skildrer bl.a. geografiske forhold, klima, dyreliv, næringsveier og innbyggernes levesett, motivet er altså delvis «bondsk». *Nordlands Trompet* er derfor gjennomgående holdt i mellomstil, med visse innslag av lavstil.

På denne bakgrunn er det interessant å studere Petter Dass' ordforråd. La oss deretter se nærmere på hvordan dette er

representert i de historiske og litterære ordbøkene som finnes, før vi til slutt drøfter andre mulige leksikografiske presentasjonsmåter.

## Petter Dass' ordforråd

Petter Dass' ordforråd byr på en variert blanding av norvagismer og dialektord, lånord og importord. Den første som studerte ordforrådet systematisk, var språkforskeren Alfred Jakobsen, som publiserte sin hovedfagsavhandling *Norskhet i språket hos Petter Dass* i 1952–53. Det er nærliggende å tro at han har valgt emne og tittel påvirket av veilederen Didrik Arup Seips nasjonsbyggende ideologi (jf. Seips egne vitenskapelige publikasjoner med de nesten likelydende titlene *Norskhet i Sproget hos Wergeland og hans samtid* fra 1914 og *Om norskhet i språket hos Ludvig Holberg* fra 1954). Men av Jakobsens fremstilling ser man at han underveis har oppdaget at språket til Petter Dass var preget også av andre egenskaper enn «norskhet», ikke minst av en rekke fremmedord. Jakobsen har overskredet det nasjonsbyggende utgangspunktet som tittelen tilsier og utvidet perspektivet underveis. Riktignok har hans vært særlig opptatt av hvordan Dass med sine norvagismer og dialektord gjenga det språket han selv kjente. Som nordlending kunne Jakobsen vurdere autentisiteten i nordlendingen Petter Dass' dialektbruk. Men Petter Dass' ordforråd skifter etter hva som er dikterens retoriske hensikt og kravet om det passende, *decorum* (jf. Lauvstad 2006:99–103) og viser slik langt mer enn «norskhet».

Dette ordforrådet er beslektet med det man ellers finner i tidens danske litteratur, selv om det samtidig har innslag av norske og nordlandske ord og -uttrykk. Mange av disse dreier seg om lokale nordnorske forhold, f.eks. *værforhold* [jf. betegnelsene på forskjellige typer vind: «Landsønning», 'sørøstlig vind', «Skottung», 'vind som blåser fra en stor skybanke i det fjerne, vind fra sør-sørvest' (begge *a*:25), *dyreliv*: «Qveite» (*a*:42), 'hellefisk', *klær*: «Skind-stack» (*a*:20), 'lang (fisker)trøye av barket skinn til å dra over hodet' eller *matretter*: «Gomme» (*a*:12), 'en type osterett av innkokt, brestet melk'. Det sterke innslaget av dialektord og norvagismer som vi ser i *Nordlands Trompet*, og i enkelte av hans verdslige viser og leilighetsdikt, mangler i Dass' åndelige sanger (jf. Jakobsen 1952–53:34).

Det er først er med et senere, nasjonsbyggende blick at man kan skjelne mellom norvagismer og dialektord i Petter Dass' forfatterskap. Sett fra datidens dansk-norske hovedstad København, var norvagismene en av flere typer dialekter, jf. Søren Poulsen Gotlænders ord i *Prosodia Danica eller Danske Riimkunst* (skrevet 1634–40, publisert posthumt 1671):

derfor skal man ikke kalde en Dræng en Gut paa Norsk, eller Paag paa Skaansk, eller Sorck paa Gullandsk [dvs. gotlandsk], eller Pois paa Øselsk, eller en Baan, oc Bar, paa Jydsck (...) det vilde falde heel Ureent i vort Spraag (sitert etter Skautrup 1968 [1947]:317).

At de trykte utgavene av *Nordlands Trompet* var beregnet på et lærd publikum utenfor forfatterens landsdel, ser man bl.a. av at visse norvagismer og dialektord har måttet forklares i de trykte utgavene. I utgaven fra Bergen 1739 finner vi f.eks. ordforklaringer (verbalkommentarer) til betegnelser for seler (*a*:29–30): «Kaabbe», «Amøring» og «Ærtne».

Uttrykksmåter som «den Nordlandske Torsk, Som Fiskeren kalde mon Skrejen paa Norsk» (a:33)<sup>1</sup>, og «Havgulle<sup>2</sup> paa Norsk til at sige» (a:25) viser at Petter Dass har vært bevisst at han brukte norvagismer. Petter Dass' språk er interessant pga. sin allsidighet. Kombinasjonen av lærd språk med dansk og fremmedspråklige innslag og lokalt, dialektalt ordforråd er interessant og skyldes altså datidens retoriske språksyn. Derfor gjenfinnes blandingen også hos andre forfattere i perioden før romantikken.

I beskrivelsen av et tingsted finner vi referat av en dialog mellom to bønder som krangler om en merr («Skiut») og en kvern. Opptrinnet begynner med at en bonde legger frem en klage på naboen:

Hand sagde goe Skrivar, Æg haver ey Sag,  
Angaaende mæg og min Grande;  
I giære vel at I her grey os den ud,  
Æg købte ey Miølgvein, og Granden eit Skiut,  
I Fælles-Lag skulle de stande. (a:53)

Dette utgangspunktet leder til forviklinger, og situasjonen har utartet etter at naboen har byttelånt merr og kvern:

Saa kom æg for skade, æg seia da maa,  
Gu naa den som Kaars og Modgangjen fæll paa,

- 
1. Emendasjon etter *c*, *a* har her betegnende nok forvanskningen «kalder Manskrejen», som tyder på at utgiveren ikke har forstått stedet.
  2. Dialektal form, muligens en gjengivelse av *havgulu* (Seip 1974:164), jamvektsform av *havgule* 'kjølig, fuktig vind som blåser fra sjøen om sommeren' (*Bokmålsordboka*).



At æg kiøura mærra fordæva.  
 Som Granden nu hour, æg skandera si mær<sup>3</sup>  
 Og dermed saa blev hand ildhærskens og tvær,  
 Aa sloe saa mi qvæn udi<sup>4</sup> stykia. (ibid.)

Bonden presenteres med bondsk språkbruk og krangleveren oppførsel. Dette skjer på retorisk måte ved hjelp av en monolog lagt i hans munn, en klassisk retorisk *ethopoeia*. Monologen gjengir krangelen mellom ham og naboen som referat av en dialog. Dialekten kan virke tilforlatelig, men inntrykket av at Petter Dass her har gjengitt en episode fra virkeligheten, fortar seg når man oppdager et lignende avsnitt i Christen Jensøns *Norske Dictionarium*, en ordsamling bygd på sunnfjordsmålet publisert 1646 (Jensøn 1946 [1646]:112f.). Her finner man under oppslagsordet «Spaancke» (dvs. 'spankulere') en historie om en bonde som ber en «Hermand» ('herremann') om råd for hvordan han og broren bør dele sin kvern og si «Skiut» ('merr'). Bondens klage er et typisk element fra et folkelig repertoar som Dass har hentet inn for å gi en lavstilsbeskrivelse av en stridbar bonde. Samtidig har Petter Dass gjengitt den på den dialekten han kjente (Jakobsen 1952–53:43–46). Det er vanskelig å vite sikkert om dette er et direkte lån fra Christen Jensøns ordbok, men det kan tenkes at Petter Dass har hatt den som litterært forelegg, og at teksten derfor har bevart visse trekk derfra (Jakobsen 1952–53:105).

I tillegg til norvagismene og dialektordene finner vi lån fra latin, nedertysk, nederlandsk og samisk i *Nordlands Trompet*. Disse innslagene av fremmedspråk har bl.a. til hensikt å demonstrere at forfatteren er lærd, samtidig som de er en

---

3. Emendasjon etter *b*, *a* har en åpenbar trykkfeil («mæe»).

4. Åpenbar trykkfeil i *a* («adi») emendert etter *b* og *c*.

retorisk tilpasning av stilen til tekstens motiver. En rekke steder brukes latinske og latiniserte ord, uttrykksmåter og egennavn. Geografisk beliggenhet kan beskrives med ord som «circulum arcticum» (polsirkelen) og «globo» (ablativ av *globus*) (a:81). Det kan være personifisert framstilling av naturkreftene som romerske guder: Aurora, Eolus, Flora, Jupiter, Neptun og Saturn, eller spøkefulle «lærde» omdanninger, ofte personifikasjoner etter latinsk mønster, som «Marchus» (avledet av (dyret) *mark*), «Garpus» (avledet av *garp*, altså 'hanseat'). Bonden omtales om «Rusticus», altså den rustikke, bondske». Det latiniserte navnet er for øvrig lånt direkte fra den burleske folkeboken om Markolfus (Jacobsen et al. (utg.) 1936).

Er det handel med utlandet eller nordlendingenes næringsveier som omtales, kan forfatteren bruke hhv. nedertyske eller nederlandske fraser. Et eksempel er uttrykksmåten «Got slaper die Bauren» i omtalen av eggsanking på en holme (a:28). Dette er antagelig en forvanskning oppstått i avskriftene av nedertysk «Göt slapen die Bauren/ Büren», dvs. 'godt sover bøndene', underforstått: 'fordi de har rikelig tilgang på fugler og egg, med andre ord: god mat. Et annet eksempel er «Dat is na de Dyvel to loopen» (a:125), nederlandsk med betydningen 'det er å løpe etter djevelen'. I omtalen av samene (med 1600-tallets betegnelser «Lapperne og Finderne» (a:61), kan forfatteren bruke samiske ord, f.eks. ordet for tallordet ti, *logi* (med moderne ortografi, jf. SaN), «Lach» (i andre versjoner av teksten «Lock» el. «loch») (a:66). Det presiseres at samene «slider Koemager i stæden for Skoe» (a:65), altså en fremheving av det samiske ordet som en eksotisk ekvivalent til den alminnelige, norske betegnelsen.<sup>5</sup>

---

5. Transkripsjonen *kommag* er den norske gjengivelsen av samisk *gama*, pl. *gabmanak*, jf. SaN.

Petter Dass' språk har dessuten vært påvirket av den sene 1600-tallets språkpurisme, slik det kommer til uttrykk i Peder Syvs *Nogle Betenkninger om det cimbriske Sprog* (1663) og Henrik Gerners *Orthographia Danica* (1678). Idealet var å erstatte fremmedord med danske ekvivalenter (jf. Skautrup 1968 [1947]: 388–389). Språkpuristene ville vise at morsmålet var vel så rikt på ord og uttrykksmuligheter som gresk og latin (jf. Skautrup 1968 [1947]:307–309). Man aner av forskjellen mellom en del av de bevarte (riktignok ikkeautografiske) manuskriptene og Bergens-utgaven fra 1739 at Petter Dass i verkets tilblivelsesprosess har erstattet franske eller latinske fremmedord og uttrykksmåter med hjemlige danske – altså ikke norvagismer, så langt strakte ikke purismen seg. F.eks. er uttrykket «med høflig Convent» er erstattet med «som Giæster i Rad» (a:10): «Compagni» er byttet ut med «Selskab (der)» (a:30), «Confrater» er endret til «Stalbroder» (a:49). Teorien om at disse endringene er gjort av forfatteren selv og ikke av avskrivere eller utgivere, bestyrkes av at en lignende revisjon har vært gjort i Petter Dass' Katekismesanger.

## Petter Dass' kilder

Som nevnt er ikke beskrivelsen av de to bøndene på tingstedet som krangler om en hest og en kvern, primært et folkelivsbilde fra Helgeland, men snarere et lån fra det litterære repertoaret med en retorisk hensikt og utformet i henhold til læren om tilpasning av stilen i tråd med emne, motiv, kontekst og

mottager. Her kan det igjen passe å sitere Søren Poulsen Gotlænder, som i den allerede omtalte *Prosodia Danica eller Danske Riimkunst* (Skautrup 1968 [1947]:317) sier at man i oversettelser bør unngå dialektord, herunder norvagismer, «med mindre saaadant i Comœdier skeede, at saadanne slags gemeene Folck, selfftalendis blefve indførte, for Lyst skyld, deris egne Ord at bruge i deris Maal».

Her gis det altså en anbefaling om hvordan man bør utforme en oversettelse, men reglene er de samme for dikterisk språkbruk generelt og håndheves av alle som vil regnes som lærde skribenter. Bøndene på tingstedet er et motiv som tilhører lavstilen, som også er komediens nivå, mens den lærde leserkretsen forfatteren henvender seg til, krever en fremstilling i mellomstil.

De nedertyske og nederlandske innslag i ordforrådet i *Nordlands Trompet* er tilpasset emnene handel og næringsveier, mens latinske elementer er med på å gjøre fremstillingen lærd og heve teksten til det forventede mellomstilsnivået.

I *Nordlands Trompet* har Petter Dass imitert og lånt fra klassiske forfattere som Ovid, Vergil og Horats i tråd med repertoartankegangen. Mye av det som ved første øyekast virker liketil og virkelighetsnært i verket, viser seg å være imitasjoner og lån – og målet er å skaffe forfatteren anerkjennelse som en lærd og kunstferdig poet. I visse tilfeller kan riktignok norske eller dialektale betegnelser på natur, værforhold, klær eller mat nærmest ha karakter av termer, de betegner forhold eller foreteelser som er stedeagne, og som det derfor kan være vanskelig å finne noe annet vokabular for. Men ordforrådet fremste hensikt er å gi fremstillingen det rette rustikke preg, noe også dikterens metaspråklige kommentarer illustrerer. Riktignok gjengis bruddstykker av forskjellige språklige registre i *Nordlands Trompet*, men verket har en retorisk, snarere enn en

mimetisk, dvs. virkelighetsgjengivende funksjon. Dette forholdet gjør det problematisk å bruke verket som en historisk kilde til virkelighetens språkbruk og ta det til inntekt for nasjonsbygging eller regionsbygging. Det som dokumenteres, er hvordan en lærd forfatter var i stand til å forene sin samtids retoriske og litterære normer med sitt samlede og svært varierte ordforråd, som ikke bare stammet fra hjemstedet, men også fra lesning og lærde studier.

### **Petter Dass' ordforråd som kilde for ordbøker**

Petter Dass døde i 1707, og med unntak av Dalevisen ble ingen av hans verk publisert i trykt form mens han levde. Derimot sirkulerte de i atskillige avskrifter blant interesserte lesere. Situasjonen er typisk for perioden, hvor publisering i bokform var kostbart. Heller ikke Matthias Moths ordbok (*Moth*) ble trykt. Den er imidlertid nylig (høsten 2013) blitt publisert digitalt av Det Danske Sprog- og Litteraturselskab. Petter Dass' ordforråd er i liten grad behandlet i de eksisterende danske og norske ordbøker som bygger på litterære ekserpter fra 1600- og 1700-tallet.

*Moth* har med et par unntak først og fremst latinske litterære sitater. Som direkte informant kan imidlertid Petter Dass ha bidratt til Moths ordbok, selv om dette ikke kan dokumenteres. Med Petter Dass' interesse for språket ville det ikke være utenkelig at han skulle ha ønsket å bidra til den innsamlingen av ordbokmateriale som *Moth* via biskopene oppfordret prester,

lærere og andre til i 1697. Mange av Dass' norske embetsbrødre sendte ordsamlinger. Fra et bevart brev vet man at også Petter Dass' overordnede, biskop Peder Krog i Trondheim, hadde sendt eller hadde villet sende en samling ord, et materiale som nå dessverre er gått tapt, visstnok enten da Krogs bispegård brant i 1699 eller i Københavns brann i 1728 (Jakobsen 1952–53:8). Det er sannsynlig at Dass mottok oppfordringen om ordinnsamling fra Krog, og det kan tenkes at også et materiale fra Dass senere er forsvunnet.

I *Dansk ordbog udgiven under Videnskabernes Selskabs Bestyrelse* (VSO), som ellers inneholder kildehenvisninger og iblant også sitater fra ekserpert litteratur, har jeg bare funnet én henvisning til Petter Dass, under oppslagsordet *trind* ('rund') i bind 7 fra 1853: «'Hvo gjør at Sneen bliver hvid, og Hagl er alle *trinde*?' (P. Dass)» (VSO 7:259). Henvisning til verk mangler i VSO, men sitatet er fra «Jobs Historie», fra *Aandelige Tids-For-driv, eller Bibelske Vise-Bog* (jf. Erichsen 1875, s. 110–111).

*Kalkars* litterære 1600-tallskilder er utelukkende danske, f.eks. Thomas Kingos og Anders Bordings diktning.

*Ordbog over det danske Sprog* (ODS) dekker med hovedbindene tidsrommet 1700–1950, med supplementsbindene 1700–1955, i tillegg har man tatt inn sitater fra Christian Vs danske lov (1683) og Kingos salmer. I ODS er ekserperingen avgrenset til litteratur på riksdansk og unngått posthume utgaver. Petter Dass er ikke ekserpert i ODS, visstnok både fordi hans språk ble oppfattet som for norsk, fordi hans forfattervirksomhet foregikk stort sett før år 1700 og/eller fordi det som nevnt stort sett foreligger i posthume utgaver.

Ivar Aasen har ekserpert Petter Dass' forfatterskap med vekt på norvagismer og dialektord, og noen av de ekserperte ordene

---

6. VSOs utheving.

er tatt inn i *Norsk Ordbog* (Aasen 1873), bl.a. *dande*, *Fjellfross*, *Hovmann*, *illherdsken* og *Skjelk*.<sup>7</sup> Imidlertid var Aasen i forordet til 1873-ordboken tilbakeholden med å bruke litterære kilder som Petter Dass' forfatterskap, fordi han mente at valg av tekstgrunnlag kunne være vanskelig, og at kildene ofte hadde lange titler:

Det kunde ofte være oplysende at henvise til visse Forfattere, som have behandlet norske Ord eller brugt dem i sine Skrifter (f. Ex. P. Dass); men dette vilde ogsaa have sin Uleilighed, da disse Forfatteres Værker deels forefindes i flere forskjellige Udgaver og deels ere indførte i Tidsskrifter med meget lange og ubekvemme Titler, saa at en Henviisning til dem vilde falde noget besværlig (Aasen 1873:XIII).

Dette er overraskende defensivt. På dette tidspunktet var riktignok ikke Erichsens tekstritiske utgave av Petter Dass' samlede skrifter (Erichsen 1874, 1875 og 1877) utkommet, men problemene Aasen nevner, var ikke uløselige.

*Norsk Ordbok* (NO) bygger på Ivar Aasens og Hans Ross' ordbøker og dekker for øvrig talemålsoppskrifter og litteratur på nynorsk og dialekter fra 1840 til i dag, samt ordsamlinger fra ca. 1600–1895 (altså frem til Ross 1895). Bare et par av Petter Dass-ekserptene som er brukt i Aasen (1873) har kommet inn i *Norsk Ordbok*, f.eks. *fjellfross* ('jerv', bare med forfatterhenviisning), *illherdsken* ('streng, morsk', dessverre med Dass-sitat via Jensøn 1946 [1646], med ordformer som synes påvirket av sunnfjordsdialekten).

---

7. Takk til Dagfinn Worren for opplysninger om Aasens Petter Dass-ekserpter og referanser til Petter Dass i Aasen 1873 og *Norsk Ordbok*.

Norsk riksmålsordbok (NRO) dekker hovedsakelig perioden 1814–1957, supplementsbindene frem til 1995. (Det er gjort noen spredte unntak for eldre litteratur.) Startpunktet er definert historisk-politisk, men avgrensningen bakover i tid er også fornuftig pga. av det normskiftet som skjer i litteraturen på denne tiden. NRO har tre sitater fra Petter Dass' *Nordlands Trompet: utligger, vargtoll* (bind IV) og *bondegods* (bind V). Ingen utgave av *Nordlands Trompet* er oppført på litteraturlisten, og *vargtoll* har indirekte sitat fra *Nordlands Trompet* via et verk av kulturhistorikeren Carl Schøyen.

*Det Norske Akademis Store Ordbok* (NAOB), som er under arbeid, er en modernisering og utvidelse av NRO og dekker samme periode, ført opp til i dag. I NAOB har vi foreløpig tatt inn noen få ekstra sitateksempler fra Petter Dass' forfatterskap. Petter Dass-eksemplene er tatt inn når vi har manglet belegg på foreledede eller dialektale oppslagsord eller underbetydninger. Sitatene fra Petter Dass' forfatterskap vil imidlertid forbli få, siden det ligger utenfor den perioden ordboken hovedsakelig dekker.

## Diskusjon

Petter Dass' forfatterskap er preget av et rikt ordforråd, slik vi har sett med eksemplene fra *Nordlands Trompet*. Dette materialet fortjener en grundig leksikografisk behandling. Som vi har sett, er dette ordforrådet dårlig dekket i eksisterende ord-bøker. Å bruke Petter Dass' diktning som råmateriale for en ordbok byr imidlertid på flere metodologiske problemer.



Siden språket er brukt i en retorisk og litterær hensikt, kan man ikke uten videre bruke det som kilde til f.eks. det helgelandske talemålet på Petter Dass' tid, i så fall trengs kontrollkilder. På denne måten er *Nordlands Trompet* som kilde av en helt annen karakter enn ordsamlinger som Christen Jensøns. Poetiske kilder er generelt problematiske i ordboksarbeidet, fordi tekstenes poetiske funksjon medfører at språket ikke er typisk eller representativt for det man kunne kalle normal-språket. Vanskelighetene blir ikke mindre når de litterære kildene tilhører en førromantisk litteratur med dens særegne forutsetninger.

Nok et problem når det gjelder å bruke Petter Dass' forfatterskap som språklig og leksikografisk kilde, ligger i forfatterskapets brokete tilblivelses- og publiserings-situasjon. Både manuskriptene og primærutgaver av forfatterskapet inneholder dessuten en rekke forvanskninger av språket og har ortografi bestemt av de enkelte boktrykkerne. Det var dette problemet Aasen (1873) omtalte, og som førte til at han helst unngikk denne typen kilder.

Allikevel bør man ikke la seg skremme fra å bruke materialet leksikografisk. Norsk språk fra perioden 1537–1814 er som nevnt svært dårlig dekket i ordbøkene. Dette er et hull i den leksikografiske dokumentasjonen som bare kan forklares ut fra nasjonsbyggingstankegangen, hvor tidlig-moderne litteratur enten er blitt nedvurdert eller tilpasset senere tiders forståelses-horisont og hvor språket først og fremst er blitt studert ut fra sin «norskhet». Språket i de litterære kildene fra perioden er interessant i sin mangslungenhet, også fremmedordene og de mange orddannelsesmåtene, og det er denne allsidigheten – og ikke bare f.eks. dialektordene – som bør dekkes i en litterær, historisk ordbok.

En mulighet kunne være å utvide NAOB bakover til også å

omfatte perioden 1537–1814. Men pga. det store litterære paradigmeskiftet som kom med romantikken, ville det være mer hensiktsmessig å lage et eget litterært ordboksverk for perioden fra reformasjonen til 1814. Imidlertid er kildematerialet sparsomt, og redigeringen ville måtte gjøres av noen med litterær og historisk kunnskap om kildene. En tredje løsning for å beskrive Petter Dass' forfatterskap leksikografisk kunne være en egen forfatterskapsordbok, noe som krever samme type spesialkompetanse. Ved å inkludere norvagismer og dialektord og alle typer av fremmedord og orddannelser ville man med en slik ordbok kunne tydeliggjøre spennvidden i forfatterens ordforråd. Som en fjerde mulighet kunne man lage en moderne tekstkritisk utgave av hele Petter Dass' forfatterskap og deretter snu materialet digitalt for å lage en ordbok basert på kommentarene, især verbalkommentarene. Kommentarene er i sin natur knyttet til den litterære konteksten, men dette ville være et mindre problem i en forfatterskapsordbok enn i en allmennordbok. I alle tilfeller har vi med Petter Dass' forfatterskap et interessant ordtilfang som venter på en systematisk leksikografisk behandling.

## Litteratur

### Ordbøker

*Bokmålsordboka* (2005) = Boye Wangensteen (red.): *Bokmålsordboka. Definisjons- og rettskrivningsordbok* (3. utg).

Oslo: Kunnskapsforlaget.

*Kalkar* = Otto Kalkar (1881–1925): *Ordbog til det ældre danske*

- sprog (1300-1700)*. København: Universitets-jubilæets danske Samfund.
- Moth* = Matthias Moth: manuskript til dansk ordbok fra omkring år 1700, publisert digitalt av Det danske Sprog- og litteraturselskab: <<http://mothsordbog.dk/>> (oktober 2013).
- NAOB = *Det Norske Akademis Store Ordbok*. Digital ordbok under utarbeidelse. Kunnskapsforlaget.
- NO = *Norsk Ordbok* (1966–) = *Norsk ordbok: ordbok over det norske folkemålet og det nynorske skriftmålet*. Oslo: Samlaget.
- NRO (1937–1995) = *Norsk riksmålsordbok*. B. I–IV 1937–57, b. V–VI 1995. Oslo: Kunnskapsforlaget.
- ODS (1918–2005) = *Ordbog over det danske Sprog*. [B. 1–28 1918–56, supplementbind 1–5 1992–2005.] København: Det danske Sprog- og litteraturselskab og Gyldendal.
- Ross, Hans (1895): *Norsk Ordbog: Tillæg til "Norsk Ordbog" af Ivar Aasen*. Christiania: Cammermeyer.
- SaN = Brita Kåven et al. (1995): *Sámi – daru sátnegirji: Samisk – norsk ordbok*. Kárášjohka: Davvi Girji.
- VSO = *Dansk ordbog: udgiven under Videnskabernes Selskabs Bestyrelse*. B. 1–8 1793–1905. København.
- Aasen, Ivar (1850): *Ordbog over det norske Folkesprog*. Kristiania: Carl C. Werner.
- Aasen, Ivar (1873): *Norsk Ordbog: med dansk Forklaring*. Christiania: Mallings Boghandel.

## Annen litteratur

- a* = Dass, Petter (1739): *Nordlands : Beskrivelse/ : Som er : Helgeland/ Saltens/ Lofodens : Westeraalens/ Senniens og Tromsens Fogderier/ med dets Beliggende/ oc hvorudi*

*enhvers Næring og Brug bestaar. : saa og : Hvad slags Fugle i Luften og svømmende Dyr i Havet sig der opholder. : Samt om dets : Horizont/ Elementer og Veyrlig/ : item : Om Finderne og Lapperne. : Forfattet af : Hr: Peder Dass : Sogne-Præst til Alstahaug i Helgelands Fogderie. Bergen: Kongl. Majests privil: Bogtryckerie.*

*b = Dass, Petter (1739): Nordlands Trompet : Forfattet : af : Herr PEDER DASS : Sogne-Præst til Alstahaug : Udi : Nordlandene. Kiøbenhavn. Kiøbenhavn: Friderich Jacobsen Bruun.*

*c = Dass, Petter (1763): Beskrivelse : over : Nordlands Amt : i : Tronhiems Stift. : Først udgivet under Titel : Nordlands Trompet, : i Vers forfattet : af : Peter Dass, : Fordum Sogne-Præst til Alstahoug Menigheder i Nordlandene./ Oplagt paa nye, : /og med Anmerkninger : oplyset og forøget : af : Albert Christian Dass. Kjøbenhavn: Andreas Hartvig Godiche.*

*Dass, Petter (1715): Aandelige Tids-Fordriv, eller Bibelske Vise-Bog, det er adskillige Bibelske Historier uddragen af den Hellige Skriftes Bøger, alle Lysthavende til Fornøye[l]se udi Riim forfattet. Kjøbenhavn: Pet. Nørwig.*

*Erichsen, Andreas Emil (1874): Petter Dass's Samlede Skrifter : Udgivne af A. E. Eriksen : Første Del. Kristiania: Den norske historiske Forenings Forlag.*

*Erichsen, Andreas Emil (1875): Petter Dass's Samlede Skrifter : Udgivne af A. E. Eriksen : Anden Del. Kristiania: Cappelen.*

*Erichsen, Andreas Emil (1877): Petter Dass's Samlede Skrifter : Udgivne af A. E. Eriksen : Tredje Del. Kristiania: Cappelen.*

*Hauge, Ingard (utg.) (1992): Johan Sebastian Welhaven : Samlede verker : Utgitt med innledning og kommentarer av Ingard Hauge. B. IV: "Prosa", 365–390. Oslo: Det norske språk- og litteraturselskap og Universitetsforlaget. [Artik-*

- kelen ”Digteren fra Alstadhaug, Peder Dass” ble først trykt i *Nordisk Universitets-Tidskrift*, ”Anden Aargang, Tredie Hefte”, Christiania 1856.]
- Jacobsen, J.P. et al. (utg.) (1932): Marcolfus/ Det er:/ En lystig Samtale/ imellem Kong Salomon/ og Marcolfum/ Saare kortvillig at læse. I: *Danske Folkebøger : fra 16. og 17. Aarhundrede. Trettende bind*. København: Det danske Sprog- og litteraturselskab. [Faksimileutgave av den eldste bevarte utgaven, forsynt med innledning og kommentarer.]
- Jakobsen, Alfred (1952–53): *Norskhet i språket hos Petter Dass*. Svorkmo: *Håløygminne og Maal og Minne*. [B. 1 tilleggsbok til *Håløygminne* 1952, B. 2 til *Håløygminne og Maal og Minne* 1953.]
- Jensøn, Christen (1946 [1646]): *Den Norske DICTIONARIUM eller Glosebog/ : Vdi huilcken indeholdis mange Norske Gloser/ oc daglige Tale/ atskillige Redskaber/ Fuglis oc Fiskes/ saa oc Diurs Naffne/ som i Norge findis oc kaldis. Item Bondens Seyerserck oc Allmanach/ som kaldis Primstafven med sine Characteribus oc forklaring. : Alle dem udi Danmarck oc Norge/ som haffue lyst at vide det Norske Lands Vjs oc Maal/ til en løstlig tjds fordrift oc Kundskab/ Colligerit oc forfattet. : Ved : C. I. S. P. A. Bergen: J. W. Eides forlag* [Faksimileutgave med forord av Per Thorson, 1. utg. ”Prentet i Kiøbenhaffn/ Aar 1646 : Paa Peder Anderssons bekostning/ oc findis : hos hannem til Kiøbs”.
- Lauvstad, Hanne (2006): *Helicons Bierge og Helgelands schiær: Nordlands Trompets tekst, repertoar og retorikk*. Oslo: Unipub [avhandling for graden dr.art.].
- Seip, Didrik Arup (utg.) (1974): *Nordlands Trompet: av Petter Dass: Utgitt ved Didrik Arup Seip*. Oslo: Aschehoug. [”Femte utgave”, eg. opptrykk av 3. utg.]
- Skautrup, Peter (1968 [1947]): *Det danske Sprogs Historie*. B.

LAUVSTAD

2: *Fra Unionsbrevet til Danske Lov*. København: Gyldendal.

Hanne Lauvstad  
vitenskapelig leksikograf, dr.art.  
Det Norske Akademis Store Ordbok  
Kunnskapsforlaget  
Postboks 4432 Nydalen  
NO-0403 Oslo  
hanne.lauvstad@kunnskapsforlaget.no

# Vinsmagningsordbogen *Enolex Bourgogne*. En milepæl i pragmatisk fagleksikografi

*Patrick Leroyer & Asta Høy*

This paper presents the design of a new online lexicographic information tool, *Enolex Burgundy* (= OB). Initiated in the autumn 2012, OB is an ongoing interdisciplinary, international research project between specialised lexicographers, linguists, and wine experts. The project is co-financed by the Burgundy Wine Board (BIVB/VDB 2013) and by the French region Burgundy. It is aimed at the development of new functions and multimodal usages for its intended users in specific, wine-related user situations. In the first phase of its development, OB will address the information needs of trainees engaged in wine tasting training courses at the Burgundy Wine School. We will stress some of the new lexicographic and technological features of OB, including lemma selection, multimodal data types, and alternative access routes, all of these contributing to the satisfaction of evaluative information needs in the foreseen user situations.

## 1. Indledning

Den (meta)leksikografi, der knytter sig til emnefeltet vin, omfatter forskning og udarbejdelse af vinartikler i almensprog-

lige ordbøger eller fagordbøger, encyklopædier og termbanker. Skønt mere sjældent omfatter det også forskning i og udarbejdelse af specielle ét-fagsordbøger om vinens særlige sprog (DLV 2007) eller om faglige emner om vin generelt (Van der Merwe 2010; WAS 2012; WIKI 2013). Under vinens (meta)leksikografi bør også indregnes leksikografisk strukturerede informationsværktøjer, såsom vinguides, vinhosider, vinatlasser, vinapplikationer til smartphones m.fl. Eksempler på anvendelse af disse informationsværktøjer er operative situationer, dvs. behov for at få vejledning i form af instruktioner, anbefalinger og tip til, fx hvilken vin passer til maden (Leroyer & Gouws 2009), og evaluative situationer, fx i forbindelse med en blindsmagning, hvor opgaven er med præcise ord og udtryk at beskrive vinen for at kunne identificere den. Der vil ofte her være tale om et nyt ordforråd, som skal indlæres, da vinforbrugerne normalt kun benytter sig af et noget begrænset og upræcist ordforråd med udsagn som «Denne vin har en flot mørkerød farve, den dufter af røde bær og krydderier, den er blød, rund og kraftig, den er rigtig god». Der skal med andre ord sættes mere præcise ord på vin.

Vi vil her introducere *CENOLEX Bourgogne* (herefter OB), et nyt, internationalt (meta)leksikografisk forskningsprojekt, som er initieret af den franske vinbrancheforening i Bourgogne BIVB (Bureau Interprofessionnel des Vins de Bourgogne) og gennemføres i et samarbejde mellem Center for leksikografi ved Aarhus Universitet og Université de Bourgogne. Formålet er at udvikle et online, multimodalt, leksikografisk informationsværktøj – dvs. indeholdende forskellige dataformater, såsom tekst, billeder, videoer, illustrationer m.fl. Der er hermed tale om en ny type brancheordbog, der i første omgang skal omfatte data til beskrivelse og vurdering af vine fra Bourgogne i én specifik brugersituation, nemlig indlæring af vinsmagning på vin-

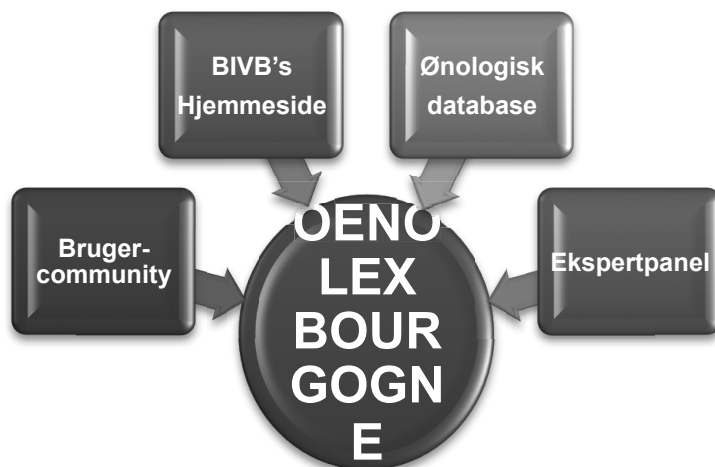


skolen. Ordbogen retter sig mod såvel eksperter som lægfolk og dermed mod medlemmer såvel som ikke-medlemmer af BIVB. På længere sigt skal OB desuden kunne tilfredsstille de særlige kommunikationsbehov, som branchen har i forbindelse med sin interne og sin eksterne kommunikation og med uddannelserne på vinskolen. Vi vil her redegøre for nogle af de leksikografiske principper, vi har udarbejdet for konceptet, herunder dataselektion, datastrukturering og datapræsentation, men først og fremmest alternative tilgangsveje, der muliggør forskellige typer søgninger.

## 2. OB's arkitektur

OB er en brancheordbog, og det leksikografiske design af OB er følgende udviklet til at opfylde branchens informations- og kommunikationsbehov samt de potentielle brugeres behov, jf. Bergenholtz og Tarp 1994. OB benytter sig af en terminalliggende arkitektur (se figur 1) og integrerer fire platforme: en bruger-community, BIVB's hjemmeside, selve den ønologiske database samt et ekspertpanel. OB er derved i tråd med de tanker, som Béjoint 2010 formulerer:

Probably the dictionary as we know it is on its way out, and we will see the emergence of new kinds of tools, reference tools encompassing more than the dictionary, containing other kinds of information and providing a better treatment of the more traditional presentations (Béjoint 2010:386).



Figur 1: OB's arkitektur.

I det følgende vil vi primært kommentere principperne for opbygningen af databasen, som viderefører og udbygger nogle af de situationelle principper skitseret i Leroyer 2011.

### 3. Situationsbaserede datapræsentationer

#### 3.1. Kommunikative situationer: fra *appellation* til indtryk

Kommunikative situationer er i OB defineret som situationer, hvor potentielle vinsmagere har brug for hjælp til at løse et

receptions- eller produktionsproblem. Det kan være hjælp til forståelse af specielle fagtermer eller -udtryk under en vinpræsentation forud for vinsmagningen eller hjælp til at udtrykke sig præcist og professionelt under selve vinsmagningen. I OB lemmatiseres alle termer, der er direkte relateret til vinsmagningen. Men da brugeren normalt ikke kender de ord og udtryk, der skal bruges til at beskrive den enkelte vin, vil en søgning på den smagte vins *appellation* kunne returnere alle relevante termer og udtryk. Der linkes desuden til vinsmagningsvideoer.

<b>Chablis grand cru</b>
<b>Visuel</b> Or-vert pur, une robe haute couture. Avec l'âge, elle évolue vers le jaune clair. Car ce vin est de belle garde (10 à 15 ans, parfois plus).
<b>Olfactif</b> Arômes minéraux (silex, pierre à fusil) intenses. Ils offrent toutefois leur place au tilleul, aux fruits secs, au miel discret, à l'amande.
<b>Gustatif</b> Equilibre parfait entre acidité et gras, vif et sec : tout le charme d'un vin inimitable et de source authentique. Le plus beau joyau de la couronne des Chablis, riche en nuances selon les climats.
<a href="#">Link til vinsmagningsvideo</a>

Tabel 1: Datapræsentation i kommunikative situationer.

### 3.2. Kognitive situationer: fra *appellation* til fakta om vinen

Præcis viden om den smagte vin er medvirkende til at kvalificere og højne vinsmagningsoplevelsen. Ved vinsmagning af Chablis-vine vil spørgsmål som «Hvor kommer Chablis fra?» og «Hvad kendetegner dette vindistrikt?» melde sig. Alle *appellations* og vindistrikter lemmatiseres, og en søgning på den enkelte *appellation* eller det enkelte distrikt vil returnere en faktaboks, der rummer svar på disse spørgsmål og linker til en præsentationsvideo samt et kort over distriktet.

<b>Chablis</b>
Les vins de Chablis sont des vins blancs secs qui se distinguent par leur pureté, leur fraîcheur, leur finesse, leur minéralité. Le cépage Chardonnay s'exprime à Chablis comme nulle part ailleurs. Il puise dans un sous-sol vieux de 150 millions d'années sa personnalité, son caractère, et mûrit dans des conditions idéales, sous un climat semi-continental, lui permettant d'atteindre un bel équilibre entre sa richesse en sucre et sa fraîcheur.
<a href="#">Link til præsentationsvideo</a>
<a href="#">Link til kort over distriktet</a>

Tabel 2: Datapræsentation i kognitive situationer.

### 3.3 Interpretative situationer: fra etiket til fakta om vinen

Information om vin på Bourgogne-vinetiketter er stærkt kodificeret, og OB konstrueres derfor til også at hjælpe brugeren med at afkode den information, som vinetiketten rummer. Dette opnås ved indeksering af billederne af de enkelte producenters vinetiketter, og ved adressering af de kommunikative og kognitive data til disse (se 3.1 og 3.2 ovenover). Der linkes desuden til tilgængelige bedømmelser og anmeldelser af de enkelte vine.

### 3.4 Operative situationer: fra FAQ til instruktioner

Endelig er OB konstrueret til at yde hjælp i operative situationer, hvor brugeren gerne vil vide – eller få demonstreret – hvordan man bærer sig ad for at smage vin. Tilgang til data sikres via en indholdsfortegnelse, hvori samles hyppige spørgsmål (FAQ) med dertil knyttede instruktioner til de enkelte elementer af vinsmagningsproceduren, eller via indeksord over hyppige, relevante termer (temperatur, glas, kigge, slynge rundt m.fl.).

spørgsmål	index
Hvilken slags glas?	glas form rumindhold kvalitet
Hvor meget skal der skænkes i glasset?	skænke
Hvilken temperatur?	temperatur
Hvordan holder man om glasset?	holde
Hvordan slynger man vinen rundt i glasset?	slynge rundt
Hvor kigger man hen i glasset?	kigge disken sider

Tabel 3: Datatilgang i operative situationer.

#### 4. OB's brugergrænseflade

Det traditionelle polyfunktionelle princip, hvor et leksikografisk informationsværktøj skal opfylde forskellige behov på én gang, risikerer enten at føre til *information overload* eller til irrelevante søgeresultater. Ved at benytte et situationsbestemt, monofunktionelt princip, hvor de data, der returneres ved søgning, er styret af brugersituationen, sikres hurtig og nem adgang til relevant information.

Monofunktionaliteten sikres ved, at brugeren aktivt vælger søgning ud fra sine informationsbehov i den pågældende situation. Dette sker ved brug af søgeknapper inden selve søgningen, hvorved data filtreres og der opnås et specifikt søgeresultat.

Søgeknapp	Brugersituation
Sæt ord på vinen	kommunikativ
Indhent viden om vinen	kognitiv
Aflæs vinetiketten	interpretativ
Få instruktioner til smagningen	operativ
Find frem til vin ved blindsmagning	evaluativ

Tabel 4: OB's søgeknapper.

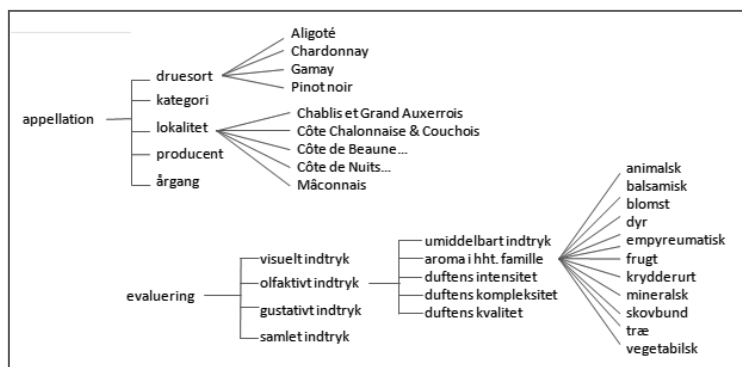
## 5. Organisation af begreberne i OB

Alt efter søgning indgår der et større eller mindre antal dataenheder i søgeresultatet. Hver af disse enheder udgør et begreb, en «unit of knowledge created by a unique combination of characteristics» (ISO 1087), der i OB tilordnes en semantisk kategori, fx:

- lokalitet
- årgang
- druesort
- smagsindtryk
- vinsmagningens procedure

Begreberne opdeles desuden ud fra en skelnen mellem faktuelle data og vinsmagningsindtryk. Da svaret på en søgning typisk vil involvere begge typer af data, skal begreberne kædes sammen ved hjælp af relationer, hvorved der etableres en ontologi, som er med til at sikre en udførlig identifikation af den enkelte vins *appellation*.

I OB er de hierarkiske relationer mellem begreber enten partitive (underbegreberne er dele af det umiddelbare overbegreb) eller generiske (underbegreberne er typer af det umiddelbare overbegreb). Figur 2 herunder illustrerer disse relationer, hvor der skelnes mellem *appellation*, dvs. den betegnelse, der knyttes til den enkelte vin, og *evaluering* af den enkelte vin. I henhold til gængs praksis (Madsen 1999) er generiske relationer vist vha. skrå forbindelsesstreger og partitive relationer vha. horisontale og vertikale forbindelsesstreger.



Figur 2: Generiske og partitive relationer i OB.

De elementer, der indgår i en *appellation*, er underbegreber i hver sit semantiske hierarki og registreres i følgende datafelter:

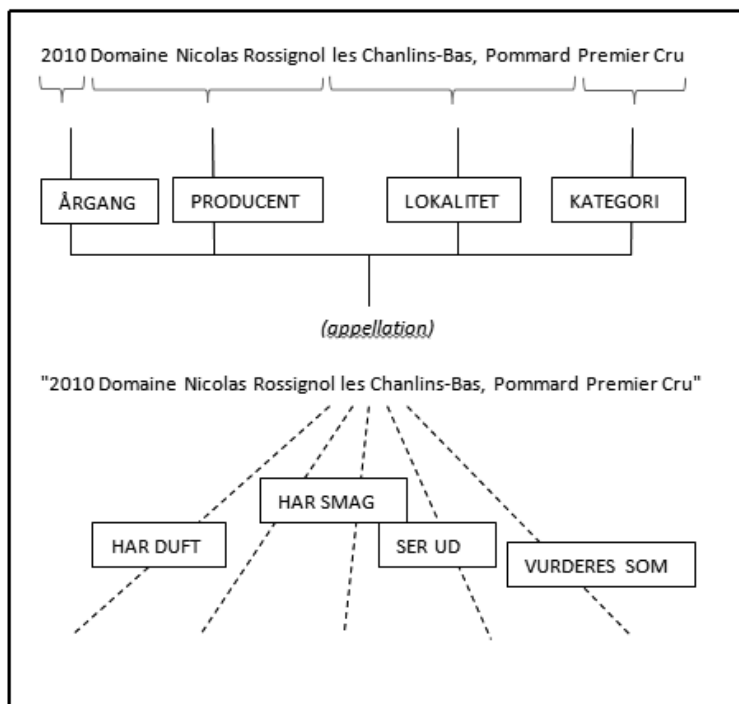


druesort, kategori, lokalitet, producent og årgang. Hver *appellation* tildeles derved sin helt egen datakombination. Aromabetegnelser stammer fra BIVB's egne data. Vi bruger BIVB's egen terminologi, da formålet ikke er at udarbejde et præskriptivt leksikografisk værktøj, men et deskriptivt værktøj. Dette princip – som er forankret i branchens ejerskab af de leksikografiske data – er i øvrigt helt centralt for udarbejdelse af brancheordbøger.

De associative relationer (ISO 74:2000, Madsen 2005) i OB (defineret som tematiske forbindelser mellem to begreber, der hører til i forskellige underordnede begrebsgrupper inden for samme semantiske kategori eller hører til i hver sin semantiske kategori) anvendes til at knytte specifikke træk til en given vin. Der kan indvendes, at der indgår en vis subjektivitet i beskrivelsen af en vins «karakteristiske træk». Men eftersom formålet med OB er at videreformidle gennem læring den ekspertise og den viden, professionelle vinkypere og ønologer ligger inde med, forudsætter vi i OB, at der blandt branchefolk er nogenlunde konsensus om de enkelte vines karakteristiske træk.

Selve evalueringen af vinen består, som det fremgår af figur 3 herunder, af visuelle, olfaktive og gustative indtryk og af det samlede indtryk (kvalitetsevalueringen). Tilknytningen af trækene til den enkelte *appellation* anføres ved hjælp af trækspicifikationerne SER UD, HAR DUFT, HAR SMAG og VURDERES SOM iht. BIVB's egne bedømmelser.

Omdrejningspunktet i OB er og bliver den enkelte *appellation*, som er forsynet med et antal karakteristiske træk. En fordel ved denne dataorganisation er, at brugeren kan få hjælp til identifikation af en given vin ved blindsmagning: Ved en kombineret søgning på de karakteristiske træk, som identificeres under smagningen, kan brugeren få returneret *appellations* af et antal



Figur 3: Vinsmagningsrelationer i OB.

potentielle kandidater til den blindsmagte vin, og OB fungerer derved som et aktivt e-indlæringsværktøj (= evaluativ situation). Der kan selvfølgelig også foretages en række andre kombinerede søgninger; fx kan man finde frem til de Bourgogne-vine, der deler egenskaben *trøffel*, eller ud fra en kombineret søgning på *Grand cru* og *Côtes de Nuit* få returneret listen over samtlige producenter, der matcher denne kombination.

## 6. Konklusion

OB er en pragmatisk, situationsbestemt vinsmagningsordbog, der er specielt udviklet til en brancheorganisation. Den er i første omgang beregnet til at yde hjælp til de kursister, der følger vinsmagningsundervisningen på vinskolen. OB rummer en række innovative tiltag, inklusive en monofunktionel datatilgang samt mulighed for at foretage kombinerede søgninger takket være organisering af vinsmagningsbegreberne i en underliggende ad hoc-ontologi omkring *appellations*. Dataene i OB er udelukkende selekteret ud fra informationsbehovene i de enkelte brugersituationer. Dataene er multimodale og stammer udelukkende fra branchens egne data. OB er således et milepæl i udviklingen af meget specifikke fagleksikografiske informationsværktøjer, der benytter både ny teknologi i overensstemmelse med anbefalingerne i Tarp 2010 og Fuertes-Olivera og Tarp 2014, en avanceret funktionel arkitektur og en pragmatisk tilgang. I OB er den konventionelle fagleksikografiske ordbogskoncipering, som styres af fagets vidensrepræsentation som fx i Fata 2010, Lerat 2011 og Van der Merwe 2008, erstattet af en koncipering, der udelukkende styres af de specifikke, faglige brugssituationer, som brugerne deltager i med de opgaver, de skal løse, og af de højrelevante informationsbehov, som disse brugssituationer og opgaver netop afføder.

## Litteratur

Ordbøger og online informationsressourcer

BIVB = *Bureau Interprofessionnel des vins de Bourgogne*.

<[www.vins-bourgogne.fr](http://www.vins-bourgogne.fr)> (juni 2013).

DLV (2007) = Martine Coutier: *Le dictionnaire de la langue du vin*. Paris: Éditions du CNRS.

WIKI = *Wikipedia. L'encyclopédie libre*. Le portail de la vigne et du vin.

<[http://fr.wikipedia.org/wiki/Portail:Vigne\\_et\\_vin](http://fr.wikipedia.org/wiki/Portail:Vigne_et_vin)> (juni 2013)

WAS = WINETECH AND SAWIS = *South African Trilingual Wine Industry Dictionary*.

<[www.sawis.co.za/dictionary/index.php](http://www.sawis.co.za/dictionary/index.php)> (juni 2013).

VDB = *Vins de Bourgogne*. Beaune Cedex: BIVB.

<[www.vins-bourgogne.fr](http://www.vins-bourgogne.fr)> (juni 2013).

## Anden litteratur

Béjoint, Henri (2010): *The Lexicography of English*. Oxford: Oxford University Press.

Bergenholtz, Henning & Sven Tarp (1994): *Manual i fagleksikografi*. Herning: Systime.

Fata, Ildikó (2010): *The Bilingual Specialised Translation Dictionary for Learners*. I: Pedro A. Fuertes-Olivera & Henning Bergenholtz (red.): *Specialised Dictionaries for Learners*. Berlin/New York: De Gruyter, Lexicographica Series Maior, 83–103.

- Fuertes-Olivera, Pedro A. & Sven Tarp (2014): *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin, New York: De Gruyter (i trykken).
- ISO 1087:2000 = *Terminology work - Vocabulary- Part 1: Theory and application*. International Standards Organisation.
- ISO 74:2000 = *Terminology work - Principles and methods*. International Standards Organisation.
- Lerat, Pierre (2011): Dénominations spécialisées, connaissances professionnelles et connaissances linguistiques en terminologie. Le cas du vin. I: Miguel Ibáñez Rodríguez & Maria Teresa Sánchez Nieto: *El lenguaje de la vid y el vino y su traducción*. Valladolid: Universidad de Valladolid, 201–224.
- Leroyer, Patrick & Rufus Gouws (2009): En termes de vin – Modularisation lexicographique du guide oenotouristique en ligne. I: *Revue Française de Linguistique Appliquée, Vol. 14, Nr. 2, 2009*, 99–116.
- Leroyer, Patrick (2011): Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographical Information Tools. I: Pedro A. Fuertes-Olivera & Henning Bergenholtz (red.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London, New York: Continuum, 121–140.
- Madsen, Bodil Nistrup (1999): *Terminologi. Principper og metoder*. København: Gads Forlag.
- Madsen, Bodil Nistrup (2005): *Håndbog i begrebsarbejde, Principper og metoder for arbejdsgrupperne nedsat af Det Nationale Begrebsråd for Sundhedsvæsenet, Del 1: Principper*. København: Sundhedsstyrelsen.
- Tarp, Sven (2010): Lexicography or Terminography? The Lexicographer’s Point of View. I: Pedro A. Fuertes Olivera (red.): *Specialised Dictionaries for Learners*. Berlin/New

LEROYER & HØY

York: De Gruyter, 27–36.

Van der Merwe, Michele (2008): Wine and Words: A Trilingual Wine Dictionary for South Africa. I: *Lexikos Vol. 18*, 337–348.

Patrick Leroyer

lektor, ph.d.

Aarhus Universitet, Center for leksikografi

Jens Christians Skous Vej 4

DK-8000 Aarhus C

pl@asb.dk

Asta Høy

videnskabelig assistent, ph.d.

Aarhus Universitet, Center for leksikografi

Jens Christians Skous Vej 4

DK-8000 Aarhus C

astah@asb.dk

## **At have begreb skabt om noget – om Den Danske Begrebsordbog**

*Henrik Lorentzen, Sanni Nimb & Thomas Troelsgård*

This paper describes the process of compiling a comprehensive Danish thesaurus at the Society for Danish Language and Literature (DSL) to be published in 2014. The thesaurus is mainly based on the sense descriptions and the lemma selection in our monolingual dictionary, Den Danske Ordbog (DDO). We will focus on the editing process and show how we use digital methods to establish annotated semantic groups which are afterwards automatically transferred to word class groups in the printed version of the thesaurus. The index constitutes an important part of the dictionary and will also be discussed, as will the future plans of making use of the thesaurus data, in particular in other dictionaries published by DSL.

En begrebsordbog over moderne dansk har været savnet i mange år. I denne artikel vil vi fortælle om baggrunden for og arbejdet med en ny dansk begrebsordbog der udkommer i begyndelsen af 2014. Artiklen er disponeret som følger. Først omtaler vi genren begrebsordbøger generelt, hvorefter vi går i detaljer med Den Danske Begrebsordbog (DDB). Fokus vil ligge på det praktiske arbejde med udtræk af ordstof fra andre baser og på sortering og opmærkning af materialet. Vi vil også komme ind på overvejelser vedrørende brugssituationer og de konsekvenser de har for præsentationen i den trykte bog, for dens layout og typografi og for indekset der udgør en væsentlig

del af bogen og dens tilgangsstruktur. Endelig afslutter vi med fremtidsperspektiverne, herunder især hvordan vi forventer at udnytte begrebsordbogens data på ordbogssitet <ordnet.dk> under Det Danske Sprog- og Litteraturselskab (DSL).<sup>1</sup>

## 1. Begrebsordbøger

Slår man op i *Den Danske Ordbog* (DDO) under *begrebsordbog*, får man følgende definition: ‘særlig ordbog eller ordliste hvor ordene er ordnet efter betydning (og altså ikke alfabetisk)’, idet man oplyses om synonymet *thesaurus*. Man får altså det indtryk at den prototypiske ordbog er alfabetisk, og at begrebsordbogen således udgør en undtagelse. Det stemmer givetvis også med den moderne opfattelse (der dog er under forandring med de digitale ordbøger, jf. Trap-Jensen 2014), men i historisk perspektiv er tematisk opstillede ordlister ikke noget særsyn. Blandt andet af indlæringsmæssige hensyn er denne ordning mere velegnet og kendes tilbage fra oldtid, middelalder og renaissance (se hertil bl.a. Boeck 2009).

I 1800-tallet dukker en ny type tematisk ordbog op hvor ord med både konkrete og abstrakte betydninger ikke blot listes under et emne, men yderligere grupperes ud fra betydning. Den første og internationalt stadig mest kendte er for engelsk og udkom i 1852: *Roget's Thesaurus* med 15.000 ord og 1.000

---

1. Vi vil gerne takke vores kolleger Liisa Theilgaard og Lars Trap-Jensen for nyttige kommentarer til denne artikel.



emnegrupper i førsteudgaven. Først langt senere fik tysk sin begrebsordbog da Dornseiffs *Der deutsche Wortschatz nach Sachgruppen* udkom i 1934. Den bruger en lidt anden emneinddeling og lander på 906 grupper. Cirka 10 år senere kom så også en begrebsordbog for dansk: Harry Andersens *Dansk Begrebsordbog* (1945). Andersen inddeler ordstoffet i 892 grupper, idet han lægger sig tæt op ad Dornseiff, »hvis Inddeling under nummererede Grupper næsten helt er fulgt« (Andersen 1945:IX).

*Roget's Thesaurus*, der altså var pioner og forbillede på området, kan roligt regnes for en af ordbogshistoriens største salgssucceser. Siden førsteudgaven er den udkommet i 32 mio. eksemplarer og er mere end tyvedoblet i omfang: Den nyeste udgave (2002) indeholder ifølge bagsideteksten mere end 300.000 ord og udtryk. Denne sproglige rigdom, som også betegnelsen *thesaurus* (oldgræsk *thesauros* 'forråd', 'skatkammer') antyder, stilles til rådighed for den bruger der søger sproglig inspiration, ønsker at finde det helt rigtige ord. Skulle man have glemt det, kan man med fordel slå op i Roget under fx *forgotten* og der finde en lang række forslag til ord og udtryk der betyder noget af det samme:

**Adj.** *forgotten*, clean f., beyond recall; well forgotten, not missed; unremembered, left in limbo 458 *neglected*; disremembered, misremembered etc. vb. almost remembered, on the tip of one's tongue; in the recesses of one's mind, gone out of one's head, passed out of recollection; burried, suppressed ...

Ikke mindst udtrykket *on the tip of one's tongue* inkarnerer en væsentlig funktion for begrebsordbøger, nemlig at lokke det frem der venter på at blive brugt, at hjælpe brugeren med at komme i tanke om synonymmer og beslægtede ord og begreber.

## 2. Den Danske Begrebsordbog (DDB)

Nu er der så en ny dansk begrebsordbog på vej. Projektet og udgivelsen er muliggjort takket være en bevilling fra Carlsbergfondet og gives eksplicit til en trykt ordbog. Projektet løber over 3 år, fra 2010 til 2013, manuskriptudarbejdelsen nærmer sig i skrivende stund sin afslutning, og bogen forventes at udkomme i efteråret 2014. Redaktionen består af Sanni Nimb, Thomas Troelsgård, Liisa Theilgaard, Lars Trap-Jensen og Henrik Lorentzen. Titlen på den nye ordbog bliver *Den Danske Begrebsordbog*, DDB, og lægger sig derved i slipstrømmen af DDO.

### 2.1. Ønsker og omfang

Udover at der skulle komme en trykt ordbog på gaden, havde vi en række andre ønsker til projektet. Først og fremmest vil vi gerne fremtidssikre vores data i den forstand at den trykte ordbog blot er ét udtræk blandt flere mulige, og at det bliver muligt at anvende begrebsordbogens data i vores andre ordbøger. Det kan dels være i form af integration med de eksisterende ordbøger, dels i form af en selvstændig onlineversion på <ordnet.dk>.

Med hensyn til omfanget har vi haft som ideal at medtage alle ord og udtryk i DDO suppleret med nyt materiale. På nuværende tidspunkt estimeres det at vi når 80 % af DDO's materiale, dvs.

- 180.000 ord og udtryk (inkl. gengangere)  $\approx$
- 100.000 forskellige ord og udtryk
- 5.000 nye ord og udtryk

Det store antal gengangere dækker over at mange ord og udtryk helt naturligt optræder i flere emnegrupper, fx findes ordet *ordbog* i øjeblikket i tre afsnit: *11.30 Kundskab, viden, erfaring;* *12.16 Betegnelse, ord;* *12.58 Bog, hæfte.*

## 2.2. Forudsætninger

Der er nogle gode forudsætninger for at dette projekt kan lade sig gennemføre på så kort tid. Begrebsordbogen ligger nemlig i klar forlængelse af DSL's seneste ordbogsarbejde for moderne dansk: DDO og det danske wordnet DanNet (se Pedersen mfl. 2009), og det ligger lige for at udnytte data fra disse to resurser.

DDO indeholder ca. 117.000 betydninger der er fordelt på 110.000 ord og faste udtryk og i mange tilfælde er forsynet med emnetilhørsforhold (som fx *militær* eller *håndarbejde*). Dertil kommer ca. 18.000 afledninger og sammensætninger uden betydningsbeskrivelse. I DanNet finder man cirka halvdelen af dette materiale organiseret i begrebshierarkier, hvor fx spil som 'håndbold' og 'volleyball' er placeret som underbegreber til 'boldspil', der igen er underbegreb til 'sportsgren'. Onlineudgaven af DDO giver allerede i dag brugeren mulighed for at se en del oplysninger om synonymer og andre beslægtede ord. Den trykte bog var forsynet med en del synonymer, antonymer, ord fra samme ordfelt osv. De fungerer som klikbare links i netversionen og kan sende brugeren på opdagelse i ordforrådet.

Begrebshierarkiet fra DanNet er også udnyttet i funktionen »Beslægtede ord« (se Sørensen & Trap-Jensen 2010), hvor man får oplysninger om over- og underbegreber og kan få dem visualiseret i klikbare figurer (stillet til rådighed på <www.andreord.dk>). I figur 1 ser vi således eksempler på synonymer og betydningsbeslægtede ord i artiklen **vigtig** i DDO:

The screenshot shows the DDO entry for the word "vigtig". At the top, the word "vigtig" is followed by "adjektiv" and "BØJNING -t, -e || -ere, -st/(-est)". Below this is a section titled "Betydninger" with a minus sign icon to its right. The first meaning listed is "1. som har stor betydning eller relevans". Underneath this meaning are several clickable tags: "SYNONYMER", "betydningsfuld", "væsentlig", "SE OGSÅ", and "magtpåliggende". A grey text box contains the sentence: "For os er det meget vigtigt at understrege, at pasningsordningen ikke er en opbevaringsordning men skal indeholde stimulation for børnene DetGrøOmr86".

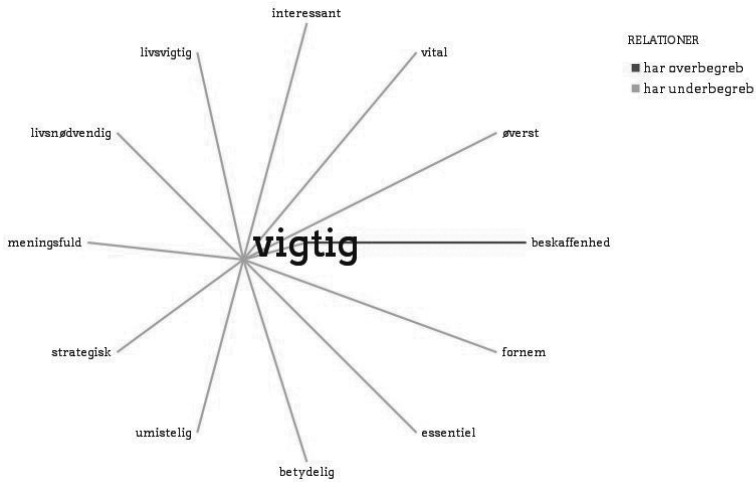
Figur 1: Uddrag af artiklen **vigtig** i DDO.

Tilsvarende ser vi i figur 2 en grafisk fremstilling af *vigtig*'s beslægtede begreber. Bemærk at man får serveret flere ord end i DDO-visningen ovenfor, fx *vital*, *interessant*, *betydelig*, *umistelig*, *essentiell* osv.

**vigtig**

ADJ

som har stor betydning el. relevans



Figur 2: Ordet *vigtig* med over- og underbegreber på <[www.andreord.dk](http://www.andreord.dk)>.

Ingen af disse visninger er dog så detaljeret og omfattende som den kommende begrebsordbog, idet vi ikke bare medtager flere synonyme, men også semantisk relaterede ord og udtryk fra andre ordklasser. Vi kan her sammenligne med afsnittet *Væsentlighed* i DDB, hvor *vigtig* optræder sammen med ca. 250 andre ord og udtryk, her kun et mindre udsnit (fra redigeringsværktøjet, se figur 3).

·væsentlig, ·relevant, ·essentiell, substantiel, af interesse, ·bemærkelsesværdig, ·interessant, ·ikke uinteressant, ·aktuel, mere aktuel end nogensinde, ·tung, ·vægtig, ·tungtvejende, ·betydningsfuld, ·af stor betydning, ·vigtig, ·værdifuld, ·betydelig, ·toneangivende, ·fremherskende, ·betydende, ·vedkommende, ·meningsfuld, ·bærende, ·fundamental, ·grundlæggende, ·central, ·nøgle-, ·egentlig, ·strategisk, ·uundværlig, ·afgørende, ·skelsættende, ·udslagsgivende, ·vital;

Figur 3: Afsnittet *Væsentlighed* i DDB (uddrag).

I den del af afsnittet vi ikke ser her, findes også andre ordklasser end adjektiver. Her nævnes fx substantiver og substantiviske udtryk som *essens*, *sagens kerne* og verbaludtryk som *spille en rolle*, *gøre sig gældende*. Med lidt andre nuancer følger senere bl.a. *hovedperson*, *nøglefigur*; *vægte*, *prioritete* osv., forhåbentlig relevante forslag for den bruger søger hjælp til at udtrykke sig, jf. udvalget fra Roget i afsnit 1.

### 2.3. DDB overfor andre begrebsordbøger

Inden vi går over til at beskrive detaljerne i redigeringsarbejdet, vil vi kort sammenligne DDB med eksisterende begrebsordbøger. Hvad angår omfang, kommer DDB til at placere sig i den tungere ende. Med sine ca. 100.000 unikke ord bliver den ca. tre gange større end forgængeren *Dansk Begrebsordbog* (Andersen 1945, ca. 33.500 ord), den bliver væsentligt større end den nyeste større norske begrebsordbog *Ord og begreper* (Rosbach 2001, ca. 7.000 ord). DDB bliver også noget større end den

tyske Dornseiff (2004, ca. 90.000 ord inkl. gengangere), men til gengæld bliver den mindre end *Roget's Thesaurus* (2002, 300.000 ord inkl. gengangere).

Pladsen tillader ikke en grundig komparativ analyse, men det er oplagt at sammenligne lidt nærmere med den danske for-gænger Andersen (1945). Vores ordbog er for det første korpus-baseret, noget der ikke var muligt for næsten 70 år siden, hvor introspektion var det bedste man kunne gøre. Således indeholder den flere specifikke begreber, bl.a. i form af sammen-satte ord fra korpus. Disse er registreret i DDO og lette at få fat i, som vi ser i figur 4 med ordet *problem*:

▷·**problem**, ·**problematik**, ·**problemstilling**, ·**dilemma**,  
·**vanskelighed**, ·**kattepine**, ·**knibe**, ·**klemme**, ·**suppedas**<sup>ufarm.</sup>  
...  
samspilsproblem, problembarn; ▷**samfundsproblem**,  
verdensproblem, beskæftigelsesproblem,  
·**arbejdsløshedsproblem**, flaskehalsproblem, ·**miljøproblem**  
, ·**affaldsproblem**, resurseproblem◄; ▷·**narkoproblem**,  
·**narkotikaproblem**◄; ▷·**støjproblem**, ·**trafikproblem**◄;  
strukturproblem; ▷**helbredsproblem**, vægtproblem,  
spiseproblem, søvnproblem, ·**vejtrækningsproblem**,  
hjerteproblem, ·**rygproblem**, potensproblem,  
·**orgasme**problem◄; ▷·**misbrugsproblem**,  
spiritusproblem, alkoholproblem◄; ·**ensomhedsproblem**;  
▷·**pengeproblem**, likviditetsproblem◄; ·**transportproblem**

Figur 4: Uddrag af afsnittet *Problem* i DDB.

Foruden synonymer som *dilemma*, *vanskelighed*, *kattepine* osv. er der også nævnt mange specifikke problemer, fx samfundsproblemer som *beskæftigelsesproblem*, *miljøproblem*, *narkoproblem*. Andersen udelod med vilje sammensætninger ligesom han bevidst ikke gjorde meget ud af områder som dyr og planter

(Andersen 1945:IX); her tager vi meget mere med, igen i tråd med DDO's lemnaudvalg.

Der er heller ingen områder der på forhånd er erklæret tabu, så vi medtager også mange begreber indenfor områder som sex, racisme osv., felter der af forskellige grunde kunne være vanskelige at berøre i 1940'erne. Strukturelt gør vi også det at vi fortsætter den semantiske gruppering så langt ned vi kan komme, til laveste niveau (se Nimb 2013 for en nærmere beskrivelse), hvor Andersen på dette niveau lister begreberne alfabetisk, jf. figur 5 og 6:

**A. aaben, fremkommelig, fri, gennemtrængelig, gisten, hul, hullet, læk, porøs, svampet, tilgængelig, utillukket.**

Figur 5: Adjektivafsnittet under *Aabne* hos Andersen (1945).

åben, åben, ·åben, helt åben, vidtåben, pivåben<sup>uform.</sup>, ·åbentstående, utillukket, oplukket;  
 ▷gennemtrængelig, ·fremkommelig, ·fri, tilgængelig◁;  
 ·opknappet<sup>uform.</sup>; ▷halvt åben, halvåben, på klem◁;  
 stå åben, stå på klem, stå på vid gab, ·gabe, ·gabe,  
 ·lække→→→→ Tivoli er åben<sup>spøg.</sup>, oplukkelig  
 hullet, ·gennenhullet, ·gennembrudt, gennemstukket  
 ; ▷læk, gisten<sup>gl.</sup>◁; ·porøs, ·svampet

Figur 6: Grupperne med *åben* og *hullet* i DDB-afsnittet *Åben, åbning*.

På denne måde mener vi at sikre den størst mulige semantiske sammenhæng mellem et begreb og dets naboer i begrebsordbogen.



### 3. Det praktiske arbejde

#### 3.1. Afsnit

DDB's overordnede struktur er baseret på Dornseiff (2004), dvs. med en opdeling i 22 kapitler, med i alt ca. 900 afsnit (ca. 20 i det mindste kapitel, over 80 i det største). Inddelingen begynder med verden omkring os og går derefter via handlinger, følelser, samfundsmæssige fænomener m.m. til kapitel 22: *Religion*. Figur 7 viser et udsnit af DDB's makrostruktur.

1	Natur og omverden	
2	Liv	
3	Rumlig dimension, bel...	1.1
4	Størrelse, mæ...	1.2
5	Væsen, forhold,	1.3
6	Tid	1.4
7	Synlighed, lys, farve,	1.5
8	Sted og stedforandring	1.6
9	Vilje og handling	1.7
10	Følelse, affekt, karakter	1.8
11	Tænkning	1.9
12	Tegn, meddelelse, spro	1.10
13	Videnskab	1.11
14	Kunst og kultur	...
15	Socialt liv	1.25
16	Mad og drikke	
17	Sport og fritid	
...		
22	Religion, oversanselig	

Figur 7: Udsnit af makrostrukturen i DDB med kapitler og afsnit.

Den overordnede struktur i DDB er overtaget fra Dornseiff, idet det dog – udover selve oversættelsen af afsnittenes overskrifter – har været nødvendigt også at foretage en del justeringer i strukturen på grund af sproglige og kulturelle forskelle. Således har det bl.a. vist sig at den »nedarvede« makrostruktur på nogle områder afspejler et ældre samfunds- og menneskesyn der ikke matcher ordforrådet i moderne dansk, ligesom den tekniske udvikling har medført at ord fra nogle afsnit i dagens sprogbrug er blevet perifere.

Som følge af sådanne forskelle har vi fx opdelt afsnittet *Spiele* i Dornseiff i to afsnit i DDB: *Spil* og *Leg*, og vi har nedlagt afsnit som *Vandvej* og *Bilindustri*, og et afsnit som *Arbejdsløshed* er blevet oprettet, hvor det hos Dornseiff blot står under *Untätigkeit*, bl.a. sammen med ord for ‘pause’ og ‘dovenskab’. Også afsnit som fx *Agtelse*, *Lydig*, *Kryberi*, *Servilitet*, *Skabagtighed*, *Soning*, *Ildsted*, *Ondsindet*, *Vogter*, *Barbari*, *Befaling* og *Tvang* er blevet nedlagt, idet de ord der ville stå i disse afsnit, har kunnet placeres i andre afsnit; fx kandiderer ord fra *Soning* til placering under *Dårlig samvittighed* og *Afsoning*, og ord fra *Ondsindet* kan placeres under *Had* eller *Fjendskab*.

### 3.2. Ordstoffet

Indhentning af egnede ord og udtryk sker fra flere kilder:

- DDO – via søgninger på ordformer, faste udtryk (sublemmaer), kollokationer, definitioner, genus proximum, fagmarkeringer m.v.
- DanNet – via søgning på overbegreb

- manuelle opslag, primært i Andersen (1945), men også forskellige synonymordbøger og fremmedordbøger
- introspektion

Tager vi et eksempel som *hat*, vil man i DDO – foruden selve ordet *hat* – kunne finde sammensætninger som *stråhat* og *tyrolerhat*, sublemmaet *høj hat* samt kollokationseksempler som *blød hat* og *bredskygget hat*.

Det er typisk under arbejdet med indhentning af egnet ordmateriale at det har vist sig om der var grundlag for at ændre et afsnits struktur eller overskrift (jf. 3.1. ovenfor).

### 3.3. Opmærkning og sortering

Hvert afsnit består af en række grupper defineret ved en formel relation der beskriver forholdet mellem ordene i gruppen og et eller flere nøglebegreber (se Nimb 2011, Nimb 2013), fx:

- genstande der har et fælles overbegreb
- personer der udfører en bestemt handling
- egenskaber der vedrører et bestemt fænomen
- handlinger der vedrører et bestemt fænomen

Ordene i hver gruppe ordnes med det mest overordnede først, efterfulgt af det mere underordnede og specifikke, ofte undergrupperet, så det i den trykte bog (ved hjælp af separatorer) kan markeres at en mindre gruppe ord er nært knyttet til hinanden semantisk. Endelig har vi givet mulighed for til sidst i en gruppe at placere nogle ord i en løsere tilknyttet »hale«, dvs. en

samling af ord der ikke dækkes af den formelle relation, men alligevel er relevante for den pågældende gruppe (og som ikke passer ind i andre af afsnittets grupper).

Grupperne ordnes så de mest centrale og overordnede udtryk placeres først; på grund af materialets heterogenitet har vi ikke fulgt et fast mønster, men taget udgangspunkt i de grupperinger der naturligt viste sig under redigeringen af det konkrete afsnit, idet semantisk nærhed har været det overordnede princip.

Figur 8 viser afsnittet *Utroskab* i DDB. Efter afsnittets nummer og navn følger tre henvisninger til relaterede afsnit, og herefter selve ordstoffet, ordnet i grupper. Først i hver gruppe er den formelle relation anført; således rummer den første gruppe ord og udtryk der betegner handlingen ‘at være utro’. Næste gruppe rummer adjektiver og adverbier med betydningen ‘som

### 15.008.Utroskab

↗10.052 *Kærlighed, forelskelse* ↗15.006 *Ægteskab, samliv* ↗15.036 *Prostitution, frigjorthed*

- ⊗ {08\_Vb\_SbAflædning/has\_hyperonym: ‘handle involved\_agent: ‘person}
  - ▷·være nogen utro, ·være utro mod nogen, ·bolle udenom<sup>uform.</sup>, ·bolle ved siden af<sup>uform.</sup>, ·snyde◄; ▷·begå ægteskabsbrud, ·bedrage, ·besvige◄; ▷·beligge<sup>gl.</sup>, ·ligge med<sup>gl.</sup>, ·ligge hos<sup>gl.</sup>, ·ligge i med<sup>uform.</sup>◄; ▷·bole<sup>gl.</sup>, ·hore, ·bedrive hor<sup>gl.</sup>◄; ▷·gøre til hanrej<sup>gl.</sup>, sætte/give nogen horn i panden◄;
- ⊗ {01\_Overbegreb/has\_hyperonym: ‘uægteskabelig}
  - ▷·uægteskabelig, ·førægteskabelig, ·udenoms◄; til venstre hånd
- ⊗ {01\_Overbegreb/has\_hyperonym: ‘utroskab}
  - ▷·utroskab, ·ægteskabsbrud, ·sidespring, ·troløshed◄; ▷·bigami, ·flerkoneri◄; ▷·hor<sup>gl.</sup>, ·utugt, ·usædelighed◄; ▷·fri kærlighed, ·partnerbytte, ·åbent ægteskab◄; ▷·trekantsdrama, ·trekantsforhold◄; ægteskabsdrama
- ⊗ {03\_Kerne\_Adled/has\_hyperonym: ‘person}
  - ægteskabsbryder, ·bigamist; ▷·elsker, ·elskerinde◄; ▷·maitresse, frille, ·konkubine◄; ▷·horebuk<sup>eds.</sup>, ·horkarl◄; ·hore<sup>gl.</sup>, ·hanrej<sup>gl.</sup>, ·swinger↔↔↔↔
  - ▷·utro, ·troløs, ·svigefuld<sup>gl.</sup>, ·polygam, ·utugtig◄;

Figur 8: Afsnittet *Utroskab* i DDB’s redigeringsværktøj.

har med utroskab at gøre'. Derefter følger en gruppe af betegnelser for 'utroskab' (i forskellige former), og endelig er der en gruppe af personbetegnelser der både rummer udøvere af og ofre for utroskab, og som også (i en hale) har en undergruppe af adjektiver med betydningen 'som udøver utroskab'.

## 4. Den trykte bog

### 4.1. Ordklassesortering

Efter redigeringsfasen, hvor semantisk nærhed er det primære princip for gruppernes rækkefølge, følger til brug for den trykte bog en omgruppering hvor udtrykkene i hvert afsnit omstruktureres, idet:

- ordene grupperes efter ordklasse
- de oprindelige grænser mellem grupperne gengives ved separatorer
- nøgleord bevares og fremhæves; ord der kommer til at stå først i en gruppe, ændres til nøgleord
- dubletudtryk der kommer til at stå i samme gruppe, fjernes

Når vi har fundet det nødvendigt med denne omgruppering, skyldes det at vi mener at brugeren er bedst tjent med en ordklassevis præsentation af begreberne, især den bruger der

ønsker et paradigmatiske alternativ i form af det samme eller et beslægtet begreb manifesteret i samme ordklasse. Desuden bliver bogpræsentationen homogen da hvert opslag følger samme struktur (jf. diskussionen i Trap-Jensen 2011:194ff.). I figur 9 ses et eksempel på hvordan den bogtrykte version forventes at blive, her med de fire store ordklasser substantiver, verber, adjektiver og adverbier.

## 4.21 Fuld

4.10 Tyk; 4.20 Mange, meget; 4.41 Helhed

SB • *komplet samling*, den totale mængde • *menneskefyldt* • fuldt hus  
• *overfyldte busser* • overflod, overmål • virvar.

VB • **fylde op**, fylde, udfylde, fylde ud, fylde helt op, stoppe (med noget), proppe (med noget), stoppe til, overfylde, belæse • **være fyldt med vand** • være fyldt til randen, være proppet til randen • **fuldende** • løbe fuld • stå som sild i en tønde.

ADJ • **fuld**, helt fuld, fyldt • stopfuld, propfyldt, propfuld, stopfyldt, overfyldt, stuvende fuld/fyldt • tungt lastet • pakket, tæt pakket, fuldt lastet ANT tom • **komplet**, fuldendt, hel, fuldstændig, total, fuldtallig ANT ukomplet • **fyldt**, fyldt med/af mennesker, fuld af mennesker • helt fuld, propfyldt, propfuld, stopfuld, stopfyldt, overfyldt, stuvende fuld, stuvende fyldt • pakket, tæt pakket, fuldt lastet • (alt) besat, fuldt optaget, udsolgt ANT mennesketom, tom • fuldt op af/med noget.

ADV • *som sild i en tønde*, så tæt som sardiner i en dåse

Figur 9: Foreløbig udgave af afsnittet *Utroskab* i bogversionen af DDB.

Efter omgrupperingen følger en slutredigering, hvor de nye afsnit gennemgås med henblik på omplacering eller sletning af udtryk der er kommet til at stå uhensigtsmæssigt efter ændringen.

## 4.2. Indeks

Omkring halvdelen af den trykte bog kommer til at bestå af et alfabetisk indeks der henviser til forekomsterne i selve indholdsdel af ordbogen. I indekset medtages de flerordsudtryk der er sublemmaer i DDO (under deres hovedlemma), samt alle enkeltord (og DDO-flerordslemmaer). For hvert henvisningsmål anføres i indekset:

- afsnitsnummer
- ordklassegruppe
- nærmeste foregående nøgleord

Figur 10 viser et udsnit fra bogstavet B i en foreløbig udgave af det alfabetiske indeks. Hvis vi ser på ordformen *bakke*, fremgår det at den bl.a. findes i afsnittene *1.15 Geologi* ('forhøjning i landskabet'), *3.12 Skrå* ('hældende underlag'), *3.15 Understøtte* ('plade til at bære noget på', sammen med bl.a. *fad* og *bageplade*), *8.8 Trafik* ('køre baglæns'), *8.9 Vej, gade* ('hældende vejstrækning'), *8.29 Tilbage, baglæns* ('køre baglæns'), *19.4 Støtte* ('del af en skruestik'), *19.5 Beholder til flydende ting* ('bæger/holder til madvarer'), *19.6 Beholder til faste ting* ('plade til at bære noget på'). Derefter følger en række henvisninger vedr. flerordsudtryk med *bakke*.

bajonetsav 19.11 SB sav  
bak 8.14 SB *dæk*, 8.29 ADV *tilbage*  
bakelit 1.22 SB *plaststof*  
bakgear 8.29 SB *bakgear*, 19.21 SB *gear*  
bakkanal 17.27 SB *raveparty*  
bakke 1.15 SB *bakke*, 3.12 SB *skrånende grund*, 3.15 SB *bakke*, 8.8 VB *dreje*, 8.9 SB *vejstrækning*, 8.29 VB *gå baglæns*, 19.4 SB *klemme*, 19.5 SB *bakke*, 19.6 SB *bakke*  
bakke op 9.68 VB *støtte*  
bakke på en pipe 16.11 VB *ryge*  
bakke ud 21.5 VB *fortryde*  
det går ned ad bakke 5.261 VB *gå tilbage*  
en bakke.. 4.17 SB *en skål suppe*  
op ad bakke 9.38 ADJ *anstrengende*, 9.53  
ADJ *strabadserende*  
bakkebord 19.3 SB *skænk*  
bakkedrag 1.14 SB *bakkelandskab*, 4.12 SB *højdedrag*  
bakkehæld 1.15 SB *bakke*, 3.12 SB *skrånende grund*  
bakkekam 1.15 SB *bakke*  
bakkeland 1.14 SB *bakkelandskab*, 1.15 SB *bakke*, 4.12 SB *højdedrag*

Figur 10: Udsnit af bogstavet B i DDB's alfabetiske indeks.

## 5. Fremtiden

Data fra DDB vil – udover at kunne publiceres som selvstændig onlineordbog i en eller anden form – kunne udnyttes i flere sammenhænge:

- DDB-data kan vises integreret i DDO eller udnyttes til forbedring af DDO's oplysninger om synonymer og især »Beslægtede ord« (jf. Trap-Jensen 2011).
- DDB-data kan berige DanNet, idet nogle af de registre-



rede relationer vil kunne oprettes i DanNet og linkene overføres dertil. Alene det faktum at alle ord i et afsnit per definition vedrører det område der beskrives af afsnittets overskrift, kan medvirke til en finere underopdeling af nogle af DanNets største »søskendeflokke« (jf. Nimb & Pedersen 2012, Nimb et al. 2013).

- DDB-data vil kunne anvendes ved semantisk opmærkning af korpuser, idet man ud fra ordbestanden i et bestemt tekststykke vil kunne identificere tekstens emneområde i de tilfælde hvor mange af ordene optræder i et bestemt DDB-afsnit. Ved automatisk semantisk analyse af korpustekster vil en sådan opmærkning bl.a. kunne understøtte disambiguering af homografe former samt udnyttes ved betydningsbestemmelse af polysemer.
- Nyt ordstof fundet ved introspektion eller fra andre kilder end DDO er oplagte kandidater til nye ord i DDO (se Lorentzen & Nimb 2011).

Som nævnt ovenfor er det oplagt at udnytte begrebsordbogens data til at forbedre oplysningerne under »Beslægtede ord« i DDO. For tiden er disse oplysninger baseret på DanNet, der har sit hovedfokus på hierarkier af over- og underbegreber. Ved inddragelse af data fra DDB vil man kunne vise flere af de semantiske relationer et begreb indgår i. Figur 11 viser hvordan fx *røg*, udover sin placering i ontologien som 'partikelholdig luftform' (øverste linje i tabellen), med ændret fokus tillige kan tilknyttes en række andre emnegrupper, hver gang med nye ord som søskende.

<i>røg</i>	Luftform, fordampning	<i>em, os, dunst, uddunstning, ..</i>
	Ild	<i>os, røgsøjle, røgmasse, røgslør, aske, sod, ..</i>
	Uigennemsigtig	<i>tåge, damp, em, sky, ..</i>
	Let	<i>støv, luft, æter, ..</i>
	Tobak	<i>røgtåger, røgsky, tobaksrøg, piberøg, røgring, nikotin, tjære, kondensat, ..</i>

Figur 11: Fem mulige placeringer af *røg* med eksempler på søskendeord i den pågældende sammenhæng.

Sådanne oplysninger om et ords tilknytning til forskellige semantiske »klumper« bruges allerede i nogle ordbøger, fx hos Macmillan, hvor man ved de enkelte betydninger kan vælge tesaurus-visning af den pågældende betydning, dvs. at overskriften for den konkrete relation vises, efterfulgt af de første ord i emnegruppen. Macmillan giver endvidere brugeren mulighed for at klikke sig et skridt videre til en ren tesaurus-artikel med overskrift og alle gruppens ord.

Vi håber at kunne udnytte dataene fra DDB til en lignende berigelse af DDO-artiklerne på <ordnet.dk>, men om det bliver på samme måde, må tiden vise.

## Litteratur

Alle webkilder er besøgt i oktober 2013.

### Ordbøger

Andersen, Harry (1945): *Dansk Begrebsordbog*. København: Munksgaard.

*andreord.dk* = <[www.andreord.dk](http://www.andreord.dk)> ved Anders Johannsen.

Viser data fra DanNet.

DanNet = Bolette Sandford Pedersen mfl.: *DanNet*,

<[wordnet.dk](http://wordnet.dk)>.

DDO = Ebba Hjorth, Kjeld Kristensen mfl.: *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab & Gyldendal, online: <[ordnet.dk/ddo](http://ordnet.dk/ddo)>.

Dornseiff, Franz (2004): *Der deutsche Wortschatz nach Sachgruppen* (8. Auflage) (von Uwe Quasthoff). Berlin/New York: Walter de Gruyter.

Macmillan = *Macmillan Online Dictionary*, <[www.macmillan-dictionary.com](http://www.macmillan-dictionary.com)>.

Roget, Peter Mark (2002): *Roget's Thesaurus* (150th anniversary edition edited by George Davidson). London: Penguin.

Rosbach, Johan Hammond (2001): *Ord og begreper. Norsk thesaurus*. Oslo: Pax.

### Anden litteratur

Boeck, Simon Skovgaard (2009): Etablering af dansk som kil-  
desprog. I: *LexicoNordica* 16, 255–270.

- Lorentzen, Henrik & Sanni Nimb (2011): *Fra krydderkage til running sushi – hvordan nye ord kommer ind i Den Danske Ordbog*. I: Margrethe Heidemann Andersen & Jørgen Nørby Jensen (red.): *Nye ord – København 12.-13. april 2011, Sprognævnets konferenceserie 1*. København: Dansk Sprognævn, 69–85.
- Nimb, Sanni (2011): Semantiske relationer i en ny dansk begrebsordbog: genbrug på tværs af ordbøger. I: *LexicoNordica 18*, 135–154.
- Nimb, Sanni & Bolette Sandford Pedersen (2012): Towards a richer wordnet representation of properties – exploiting semantic and thematic information from thesauri. I: Nicoletta Calzolari (red.): *LREC 2012 Proceedings*. Istanbul: European Language Resources Association, 3452–3456.
- Nimb, Sanni (2013): Leksikalsk-semantisk information i en ny dansk begrebsordbog. I: Dorthe Duncker, Anne Mette Hansen & Karen Skovgaard-Petersen (red.): *Betydning & Forståelse. Festskrift til Hanne Ruus*. København: Selskab for Nordisk Filologi, 251–266.
- Nimb, Sanni, Bolette Sandford Pedersen, Anna Braasch, Nicolai Hartvig Sørensen & Thomas Troelsgård (2013): Enriching a wordnet from a thesaurus. I: Lars Borin, Ruth Vatvedt Fjeld, Markus Forsberg, Sanni Nimb, Pierre Nugues & Bolette Sandford Pedersen (red.): *Proceedings of the workshop on lexical semantic resources for NLP. NEALT Proceedings Series 19*. Linköping Electronic Conference Proceedings; Volume 85 (ISSN 1650-3740).
- 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, Volume 43, Number 3*

/ *September*, 269–299.

- Sørensen, Nicolai Hartvig & Lars Trap-Jensen (2010): Den Danske Ordbog som begrebsordbog. I: Harry Lönnroth & Kristina Nikula (red.): *Nordiska studier i lexikografi 10. Rapport från Konferensen om lexikografi i Norden, Tammerfors 3–5 juni 2009*. Tammerfors: Nordiska föreningen för lexikografi, 164–179.
- Trap-Jensen, Lars (2011): Dansk begrebsordbog: baggrund og metode. I: *LexicoNordica 18*, 183–197.
- Trap-Jensen, Lars (2014): Leksikografisk tradition og fornyelse: tre revolutioner på 100 år? I: Ruth E. Vatvedt Fjeld & Marit Hovdenak (red.): *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.–16. august 2013* (dette bind).

Henrik Lorentzen  
seniorredaktør  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
hl@dsl.dk

Sanni Nimb  
seniorredaktør  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
sn@dsl.dk

LORENTZEN, NIMB & TROELSGÅRD

Thomas Troelsgård  
seniorredaktør  
Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K  
tt@dsl.dk

# Om (bristande) bruklighetsstruktur i SAOB

*Anki Mattisson*

The Swedish Academy Dictionary (SAOB) is a historic descriptive dictionary. The period described is 400-500 years, depending on the volume. From the beginning in 1884 plenty of use markers have been in use, as well as a lot of additional qualifications. The use markers can be placed in the beginning or in the later part of the definition, within or without brackets and in full text or condensed (e.g. technical markers). The markers often seem to be a part of the definition. This article deals with the lack of structure and the obsolete system of use markers. This ought to be changed in an updated version of the dictionary.

## 1. Inledning

Om man slår upp ett huvudord i SAOB på nätet och läser den «finstilta» texten längst upp till höger på skärmen får man veta vilket år det aktuella häftet lämnade trycket. Vill man veta samma sak i den tryckta upplagan måste man leta upp tryckarkets början där datum för tryckning respektive slutredigering står. Det är en viktigare uppgift än man tror, i synnerhet om man är intresserad av ett ords eller en betydelses bruklighet och stil.

Ett ord som åsatts markören «starkt arkaiserande, numera knappast br.» år 1932 är sannolikt obrukligt idag. Ett ordboksprojekt som hållit på i 125 år och ännu inte avslutats kan inte ha enhetliga markörer som under hela den långa tiden avser eller har avsett samma sak. Språket är alltför dynamiskt för detta, och både skriftspråket och talspråket har sannolikt utsatts för större förändringar under 1900-talet än någonsin tidigare. Till detta kommer att det är fem generationer redaktörer som redigerat, och de har utan tvekan haft olika bildning, utbildning, geografisk hemvist och inte minst referensramar. För att inte tala om tillgång till olika medier.

När SAOB planerades på 1880-talet var avsikten att göra ordboken till en normerande diakronisk samtidspråklig ordbok. Idag, efter drygt 100 år, är ordboken snarast på väg att bli en diakronisk historisk ordbok. Svenska Akademien har fattat ett principbeslut om en uppdatering efter Ö. Det sannolika är att ordboken efter uppdateringen kommer att täcka de 500 åren mellan 1521 och 2020. Första upplagan kommer alltid att ligga sökbar på nätet i sin ursprungliga form och en andra upplaga uppdateras (med successiv utläggning på webben). Nya ord och nya betydelser ska naturligtvis läggas till, men i vilken utsträckning struktur, definitionsspråk, de diasystematiska markörerna (se Svensén 2004: 378 f.) m.m. ska uppdateras och om nya källor ska tillkomma vet redaktionen inte än. Ständige sekreteraren ska utreda frågan och lägga fram ett förslag inom en inte alltför avlägsen framtid. Naturligtvis funderar vi som dagligdags arbetar i projektet också. Vi är fullt medvetna om att det inte är aktuellt med en ny ordbok som får ta sina 100 år och kosta en miljard (som första upplagan har gjort hittills). Det viktiga är att det blir en historisk ordbok i digital form värd namnet med god läsbarhet och användbarhet, bl. a. genom möjlighet att dölja delar (t.ex. formparentes eller undermoment) för att ge en bättre överblick över långa artiklar.



Med mitt bidrag vill jag visa dels hur föråldrade bruklighets- och stilmarkörerna i SAOB är, dels hur den bristfälliga strukturen ter sig. Inför en kommande uppdatering sneglar vi alla på Oxford English Dictionary (OED) på nätet, och detta med vetskapen om att även den första generationen SAOB-redaktörer hade OED:s föregångare NED som mönster.

## 2. Bruklighet och stil

Att Theodor Wisén och K.F. Söderwall var väl förtrogna med dr Murrays arbete med New English Dictionary (NED; idag: OED) när planen för en Svenska Akademiens ordbok 1883 lades fram för akademiledamöterna är sannolikt. Murray hade kallats till tjänsten som huvudredaktör för det avstannade projektet 1876, och den plan som han gjorde torde snabbt ha spritts till de etablerade europeiska lexikograferna, till vilka ju Söderwall hörde. Wiséns plan för SAOB i det handskrivna skick som den lämnades in till Akademien följer Murrays på i stort sett varje punkt (se Hjelmqvist 1896: 91 ff.).

Det normala är att markeringar för bruklighet och stil kondenseras i tryckta ordböcker. Den vanligaste formen av kondenserad markering är en standardiserad förkortning av den typ som redan Dalin använde i Ordbok öfver svenska språket (1850-53), som t.ex. fam[iljär], bib[lisk stil], prov[insiell]. Och det var en sådan uppsättning kondenserade markörer som valdes till NED. NED har ganska få (tolv) markörer och de är trots det korta formatet tydliga och väl synliga för användaren

genom att de är kursiverade. Några exempel: *arc[haic]* 'ålderdomlig', *erron[eous]* 'felaktig, oriktig', *joc[ular]* 'skämtsam', *obs[olete]* 'obsolet, föråldrad, omodern', *occas[ionally]* 'tillfällig, enstaka'. De åtföljs emellanåt av ett «now», vilket torde ha utgjort mönster för SAOB:s «numera».

SAOB:s upphovsmän valde en helt annan väg, nämligen tolv (12) markörer för bruklighet gällande uttryckssidan, två (2) markörer för bruklighet gällande innehållssidan och 18 markörer för stil inklusive dialekt samt modifieringar (i sht (= i synnerhet), nästan, bl. (= blott) m.fl.). Någon kondensering var det egentligen inte fråga om mer än för några förkortningar som «ä. » för äldre och «br. » för bruklig. Den enda grupp som kondenserades var fackbeteckningar som också – i likhet med NED – kursiverades.

Systemet för markörer i SAOB kan i korthet presenteras så här:

**A:** initialt efter momentbeteckningen parentes om uppgifter om bruklighet och stil i samma typsnitt och storlek som definitionstexten.

Ex.: **SPÖRJA** (S 10637, 1986)

I 1) (†) om hund: spåra upp (villebråd).

I 2) (med ålderdomlig prägel) gm en fråga l. frågor begära upplysning(ar) l. besked, framställa en fråga l. frågor, fråga.

**B:** längre brukligheter i slutet av en definition, mestadels inom parentes men inte alltid. Oftast handlar det om att definitionen börjar med en obruklig betydelse för att glida över i en mer bruklig eller fullt bruklig betydelse.

Ex.: **SKÖR** (S 6163, 1976)

3) [...] otuktig l. liderlig l. okysk o. d.; utom i ssgrna  
**SKÖR-LEVANDE** o. **-LEVNAD** numera bl. ngn gg skämts.  
l. arkaiserande.

**C:** kursiverad fackbeteckning (förkortad och utan eller med parentes).

Ex.: **TUBULERA** (T 3060, 2009)

1) i sht *kem.* förse (kolv l. retort o. d.) med rör.

Ex.: **TVILLING** (T3287, 2009)

2 b) (i fackspr., särsk. *mineral.*) om (l. elliptiskt för) tvillingkristall.

**D:** bruklighetsmarkering inom parentes (oftast †) + kursiverad fackförkortning. Denna typ är slopad sedan slutet av bokstaven S.

Ex.: **KROMATIK** (K 2884, 1938)

1) (†) *konst.* färglära; konsten att blanda färgerna.

I stället för att besluta om sådana traditionella förkortningar som sedan relativt lång tid var etablerade inom lexikografin kom SAOB:s redaktion alltså att blanda ett kondenserat markerings-system med ett fulltextsystem. Olika markörer kunde och kan kombineras tämligen fritt och modifieras. I princip alla bruklighetsuppgifter kunde också sättas ut i huvudet tillsammans med uttal, stavning och böjningsformer. En sökning i nuvarande SAOB på nätet efter diasystematiska markörer visar att det förekommer ca 4 500 olika markörer.<sup>1</sup> Av dessa är 3 200 eller ca 75 % unika

---

1. Tack till Rudolf Rydstedt, Institutionen för svenska språket, Göteborgs universitet, för sökning av bruklighetsmarkörer i OSA åt mig.

MATTISSON

markörer av typen (numera föga br. o. med en viss ålderdomlig anstrykning). Några andra exempel på några markörer av det vidlyftigare slaget:

**HISTRION** (H 972, 1931)

(numera dels arkaiserande, i fråga om ä. förh., dels nedsättande, föga br.).

**REDLIG** (R 592, 1956)

(utom i vissa trakter bygdemålsfärgat numera bl. med ålderdomlig prägel, i sht i vitter l. högre stil) **REDELIG**.

**RÄT-BÄNK** (R 3991, 1961)

(i fackspr., utom i vissa trakter, bygdemålsfärgat, numera föga br.).

Med undantag för fackbeteckningarna och korset för obruklighet kan man säga att SAOB:s markeringssystem är en integrerad del av definitionstexten. Markeringar kan förekomma i stort sett överallt. Eftersom platsen för markören inte är bestämd och eftersom en markör initialt kan ha eller inte ha parentes uppfattas den inte alltid som en markör. En definition kan dessutom börja med en parentes som ringar in sfären eller regionen eller dylikt för ett ords betydelse, vilket lätt kan uppfattas som en bruklighetsmarkör (och delvis är det).

### 3. Varför blev markeringsystemet så här?

Uppgifterna om ordens bruk i SAOB är detaljerade, t.o.m. «i överkant specificerade» (Hast 1993: 74). Beteckningarna i det subtila systemet var inte valda på måfå utan med omsorg, skriver Loman (1986: 116). Upplysningarna är på ett ovanligt sätt integrerade i den semantiska och syntaktiska beskrivningen. Området för ord, former och ordbetydelser, som räknas som fullt brukliga och vanliga i allmänspråket, behandlas snävare i SAOB än i t.ex. OED (Ekbo 1965: 28 f.). SAOB:s bruklighetsuppgifter blir ofta föremål för intern diskussion på grund av både individuella ställningstaganden och svårigheter att tolka reglerna. I de tryckta böckerna är, skriver Ekbo, markörerna ett uttryck för en bedömning inom en arbetskår som ägnat dessa frågor ett omfattande intresse. Det skulle jag kalla ett understatement om något.

Min teori är att markeringsystemet hänger samman med det dubbla syftet med SAOB, dvs. att ordboken skulle vara både deskriptiv och normativ. Definitionen av allmänspråket eller normalspråket i SAOB är ganska allmänt hållen: «sådan [språk] som i högre grad kunnat anses vara gemensamt för ett betydande antal bildade svenskar från olika delar av landet» (Söderwall 1893: 2). Här står inget om vårdat språk. Det kommer i stället fram när markörernas uppgift beskrivs. De avskiljer «genom vissa tecken eller uttryck det föråldrade, det oriktiga, det ovanliga, det vardagliga eller det någon stilart eller yrkesterminologi tillhöriga från det, som nu för tiden kan allmänt brukas i vårdadt språk». Ett av ordbokens syften är att «verka för riktig användning af vårt modersmål i tal och skrift». Det sägs alltså uttryckligen att ordboken inte bara skulle vara deskriptiv utan också normativ och behandla både talat och

skrivet språk (se vidare Larsson 2012: 385 f.). Det normativa kommer till uttryck i äldre band också genom markörer som «i vårdat språk», «i vårdat talspråk» och «klandrat språk».

Det var sannolikt den normerande aspekten som kom att prägla SAOB:s markeringssystem. Det normativa lyftes fram av Söderwall och ännu för 50 år sedan när SAOB-användare och studenter fick sin första vägledning till SAOB var det ett framträdande drag. Sven Ekbo (1965: 27 ff.) menade att genom de detaljerade bruklighetsuppgifterna lämnar SAOB värdefulla bidrag till kännedomen om vad som vid tryckningen av ett visst häfte eller ark anses eller ansetts riktigt eller användbart eller mindre eller alls icke acceptabelt. Därför, menade han, kunde SAOB ses som ett fortlöpande väsentligt bidrag i språkvårdsdebatten.

Vad som mig veterligt aldrig har påpekats är att ett så detaljerat markeringssystem som SAOB:s egentligen kräver ständig uppdatering. All denna information har ett bästföre-datum. Om hela SAOB i likhet med Dalins ordbok getts ut under en begränsad tid, t.ex. 20 år, hade upplysningarna haft stort värde och varit användbara – om än för en avgränsad tid. Nu har vi bl. a. ca 40 000 ord och betydelser som har en bruklighet som börjar med «numera», och detta redan från bokstaven A som trycktes mellan 1893 och 1903.

#### **4. Föråldrade markörer**

Det är naturligt att den diasystematiska informationen åldras, i synnerhet i en diakronisk samtidspråklig ordbok av SAOB:s

typ. Ord och betydelser får ibland nytt liv eller dör, och det bör komma till uttryck vid en uppdatering. Ibland har tidigare redaktioner haft för lite material (språkprov) att göra sina bedömningar på och därför satt en markör på ett ord eller en betydelse som idag uppfattas som helt fel. Några exempel ska nämnas:

**HÄVA** (H2191, 1932)

2) (numera nästan bl. dels vard., ofta bygdemålsfärgat, dels i vitter stil, ålderdomligt) kasta l. vältra l. vräka l. störta (ngn l. ngt ngnstädes).

**PRIVET** (P1910, 1954)

(utom bygdemålsfärgat i vissa trakter numera bl. med ålderdomlig l. skämts. prägel) avträde (se d. o. 6), avträdeshus, dass.

I det första exemplet har vi en lång markör innehållande flera olika stilinskränkningarna med modifieringarna «numera nästan bl.» och «ofta». Om detta moment hos **HÄVA** skulle skrivas idag skulle ordet stå utan markör, dvs. betraktas som fullt brukligt.

1954 betraktades **PRIVET** som dialektalt och skämtsamt. När artikeln skrevs medtogs dock belägg som visade att ordet på 1940-talet förekom i (arkeologiskt och konstvetenskapligt) fackspråk, och troligen skulle markören idag bli «numera mindre br.» eller «i sht i fackspr.».

Flera stilmarkörer som t.ex. «i bibelpåverkat spr.», «i högre stil» och «i vitter stil» lämnar sannolikt många användare i sticket. Bibeln är långt mindre känd än tidigare, och dessutom har vi i Sverige en ny bibelöversättning, som enligt direktiven 1975 inte skulle utföras «i någon mer eller mindre särspråklig

bibelstil av genomgående högtidlig karaktär» (Albrektsson & Åsberg 1996: 55). I SAOB finns ca 11 000 ord och betydelser som markerats «i vitter stil». Idag vet inte många vad som menas med «vitter stil», och dessvärre tror jag inte att alla redaktörer som använt markörerna heller har varit säkra på innebörden. Ordet *vitter* används idag huvudsakligen i tidningarnas dödsrunor för att karakterisera någon lärd professor – och i tal och skrift av ledamöter av Svenska Akademien.

Markören «i vitter stil» är ett exempel på en tidsbunden markör, knuten dels till SAOB-projektets start i slutet av 1800-talet, dels till det då befintliga skönlitterära källmaterialet. Litteraturen dominerades av de välkända 1700- och 1800-talsförfattarna, dvs. de vittra. Med åren har materialet ökat flerfaldiga gånger och den stil som kallas vitter läses knappast längre. Även om ett ord som **PEJLA** i betydelsen 'mäta l. undersöka l. sondera l. utforska' (P 540) första gången förekom i Polyfen från 1812 kan man inte säga att ordet har vitter prägel idag (och kunde det nog knappast heller 1952 när häftet i fråga trycktes).

## 5. Bruklighetsstruktur

Dagens SAOB har en struktur som hänför sig till senare hälften av 1800-talet. Strukturen är inte homogen; de tio första banden skiljer sig mest från de övriga. De omfattande och utspridda diasystematiska markörerna ger intryck av att sakna struktur i dagens mening – med undantag för de kondenserade och kursiverade fackbeteckningarna. I den digitala ordboksvärlden är det



idag självklart med en artikelstruktur som gör att användaren alltid hittar den information som han eller hon söker på samma ställe. All information, inklusive markörerna, förväntas komma i en fast ordning och med så enhetlig utformning som möjligt. Bruklighets- och stilmarkörerna ska helst vara få och lättolkade.

SAOB behöver en ny struktur med kondenserade och sökbara markörer. En invändning som jag har hört ett par gånger är att de nuvarande markörerna är tidstypiska och en del av charmen med SAOB. Denna synpunkt tillgodoses genom att första upplagan kommer att finnas både i tryckt och digital form. Några lexikografer, däribland Bo Svensén (2004: 380), menar att behovet av markörer inte är så stort i enspråkiga ordböcker därför att definitionsspråk och källspråk är desamma, men det gäller knappast diakrona historiska ordböcker. Användaren av en historisk ordbok som täcker 500 år behöver en viss vägledning. Vi torde inte kunna förvänta oss bättre läskunnighet av äldre texter om 50 år än idag. Det språk som beskrivs är inte bara dagens allmänspråk, det är flera hundra års, och det är inte bara döda ord och betydelser som bör markeras. Skillnaden mellan t.ex. (förr), som handlar om innehållssidan, samt (ålderdomligt) och (numera mindre br.), som handlar om uttrycksidan, bör upprätthållas som idag, liksom ett litet antal stilmarkörer.

Sammanfattningsvis: Den nuvarande bruklighets- och stilinformationen i SAOB saknar egenskaper som ett modernt markeringsystem bör ha och utgör närmast en del av betydelsebeskrivningen. Den är detaljerad och tidsbunden, vilket gör det nödvändigt att vid en uppdatering av ordboken också åtgärda denna information. Skulle markörerna utformas som idag, skulle det ta alldeles för mycket tid i anspråk och värdet av sådan information är nog – som framgått – tämligen begränsat. Flera markörer för bruklighet och stil bör bytas ut och konden-

MATTISSON

seras för att användare lätt ska förstå till vilken sfär eller period ordet ska hänföras. Markörerna måste också få ett tydligt innehåll. Och – inte minst – bör de placeras på ett och samma ställe i strukturen. När det gäller ordböcker är placeringen och utformningen av vissa element i artikelstrukturen av största vikt. Både redaktören och användaren ska känna igen sig.

## Litteratur

### Ordböcker

Dalin, A. F. (1850-55): *Ordbok öfver svenska språket*.

Stockholm. (Samlingar utgivna av Svenska fornskriftsällskapet: Svenska skrifter 91.).

NED = James A. H. Murray et al. (red.) (1884-1928): *A New English Dictionary on historical principles*. Oxford.

OED = J. A. Simpson & E. S. C. Weiner (1989): *The Oxford English Dictionary*. Oxford: Clarendon Press.

OSA = Om Svar Anhålles. SAOB på nätet. <<http://g3.spraak-data.gu.se/saob/>>

SAOB = *Ordbok över svenska språket, utgiven av Svenska Akademien* (1893 ff). Lund.

### Annan litteratur

Albrektsson, Bertil & Åsberg, Christer (1996): *Det nya Gamla*

- testamentet: från forntida hebreiska till nutida svenska.*  
Stockholm: Svenska bibelsällskapet.
- Ekbo (1965) = Sven Ekbo & Bengt Loman: *Vägledning till Svenska Akademiens ordbok*. Stockholm: Skrifter utgivna av Nämnden för svensk språkvård 32.
- Hast, Sture (1993): SAOB och dess likar. De fyra stora historiska nyspråkliga germanska ordböckerna. En kort redogörelse och jämförelse. I: *Ord och lexikon. Festskrift till Hans Jonsson 10 juni 1993*. Stockholm: Norstedts.
- Hjelmqvist, Theodor (1896): *Några anteckningar om de historiska ordböckerna i Tyskland, Holland och England*. Lund: C. W. K. Gleerups förlag.
- Larsson, Lennart (2012): Varför inte dub(b)lettformer? Om SAOB som normativ ordbok. I: B. Eaker, L. Larsson & A. Mattisson (red.): *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden. Lund 24-27 maj 2011*. Lund: Svenska Akademiens ordboksredaktion, 385–395.
- Loman, Bengt (1986): En Inrättning, ägnad endast till Språkets förbättring. I: S. Allén, B. Loman & B. Sigurd: *Svenska Akademien och svenska språket: tre studier*. Stockholm: Norstedts, 1–141.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik* (2 omarb. och utökade uppl.). Stockholm: Norstedts akademiska förlag.
- Söderwall, K. F. (1893): [Företal till första häftet av SAOB]  
Lund.

MATTISSON

Anki Mattisson  
ordbokschef, docent  
Svenska Akademiens ordbok  
Dalbyvägen 3  
SE-224 60 Lund  
[anki.mattisson@svenskaakademien.se](mailto:anki.mattisson@svenskaakademien.se)

# Nja... En korpusbaseret undersøgelse af interjektioner i talesprog

*Jessie Leigh Nielsen*

Interjections have often been viewed as unsystematic and unimportant to the description of language. *Ordbog over Dansk Talesprog* challenges this idea of interjections as irregular and insignificant by creating a dictionary of Danish interjections, in this article exemplified by the Danish interjection *nja*. By examining interjections in Danish spoken language, it is found that interjections can in fact be systematically described and are very relevant to language description. Also a sound based approach to language produces new lemmas which are not orthographically distinct and thus not visible in the written language. This way *Ordbog over Dansk Talesprog* can supplement traditional orthography based dictionaries and contribute to a more precise and thorough description of the Danish language.

## 1. Indledning

Interjektioner er en ordklasse som har fået forholdsvis lidt opmærksomhed gennem tiden, og teoretikere er meget uenige om hvordan de bør behandles. De græske grammatikere klassi-

ficerede interjektioner som en slags adverbier fordi de mente de modificerede verber (Ameka 1992). Heller ikke Otto Jespersen mente at interjektioner skulle klassificeres som en særskilt ordklasse, men allokerede dem til en klasse af partikler sammen med adverbier, præpositioner og konjunktioner (Jespersen 1924).

En af grundene til at interjektioner i tidens løb er blevet til-sidesat, er at interjektioner umiddelbart ikke lader sig beskrive lige så let og nøjagtigt som andre ordklasser. De har ingen fast plads i syntaksen, men siges at udgøre deres egen ytring (Ameka 1992), og de overholder tit ikke sprogets fonologiske regler (Wilkins 1992), for eksempel følger interjektionen «sh» (ti stille!) ikke de danske fonologiske regler da vi som regel ikke tillader stavelser uden vokaler. Fonetisk findes der selvfølgelig stavelser uden vokaler i talesprog, men fonologisk er der altid en underliggende vokal. Da interjektioner udgør deres egen ytring, er de ikke underlagt den almindelige sætningsintonation og er altså også fonetisk afvigende. I det hele taget er interjektioner ved første øjekast uforudsigelige og regelbrydende fra fonologi til syntaks.

Ordbog over Dansk Talesprogs (ODT) leksikografiske mål er at udarbejde en netbaseret talesprogsordbog som er tilgængelig, forståelig og interessant for både den generelt sprogligt interesserede dansker og for sprogforskeren som ønsker detaljeret faglig viden om talesprog. Det er ikke ODT's hensigt at udarbejde en ordbog helt fra bunden, men snarere at virke supplerende i forhold til traditionelle skriftsprogsbaserede ordbøger.

I den forbindelse ønsker ODT at gøre op med idéen om interjektioner som en undefinerbar rodekasse ved at undersøge interjektioners funktion i talesprog og udarbejde en selvstændig ordbog over danske interjektioner som en underdel af den mere

omfattende ordbog over talesprog. Da interjektioner grundlæggende er talesprogsforankrede, er det oplagt at undersøge interjektioners funktion i talesprog.

### 1.1. Ordbog over Dansk Talesprog

Ordbog over Dansk Talesprogs arbejde bygger på LANCHART-korpusset, som består af cirka 6,5 millioner transskriberede tokens af 1.600 samtaler. Korpusset består primært af sociolingvistiske interviews fra 1978 til i dag fra syv forskellige områder i Danmark. Korpusset omfatter ikke kun traditionelle sociolingvistiske interviews med én informant og én interviewer, men også gruppesamtaler hvor flere informanter taler sammen med eller uden interviewer.

Da ODT's primære fokus er hvordan talesprog afviger fra skriftsprog, er det oplagt at beskæftige sig med ord og partikler som primært er interaktionelt forankret, som for eksempel interjektioner. I denne artikel præsenteres arbejdet med *nja* som et eksempel på ODT's generelle arbejde med interjektioner. Da der ikke allerede findes nogen teorier for leksikografisk arbejde med talesprog, må ODT i høj grad selv udvikle metodiske og teoretiske praksisser. På den måde eksemplificerer denne artikel om *nja* ikke kun arbejdet med interjektioner og andre partikler, men også hele ODT's metodiske praksis, hvilke problemer det medfører at arbejde med talesprog, og hvordan de kan løses.

## 2. ODT's metode eksemplificeret med *nja*

### 2.1. Søgning i korpus og stikprøve

Et talesprogslemma må nødvendigvis være lydbaseret, og derfor ønsker ODT at tage udgangspunkt i udtale. LANCHART-korpusset er dog udskrevet i ortografi, og man må altså tage udgangspunkt i en ortografisk form, for eksempel *nja*, når der søges i korpusset. Korpusset er udskrevet efter både semantiske og fonetiske kriterier, hvilket betyder at transskriptionen ikke altid stemmer overens med udtalen. For eksempel bliver alle tøvefænomener udskrevet *øh* uanset udtalen. Transskriptionen udgør altså et omfattende problem: Den er som nævnt udskrevet i ortografi, primært i overensstemmelse med Retskrivningsordbogen og Den Danske Ordbog. Ved ord som ikke fremgår i ordbøgerne, må udskriverne selv vurdere hvilket ordbogslemma ordet minder mest om og udskrive det sådan. Det betyder at der ikke altid tages hensyn til udtalen, og selv ved de ord hvor stavemåden i stor grad afspejler udtalen, for eksempel interjektioner, spiller semantiske kriterier en stor rolle for hvordan et givet ord bliver udskrevet. Man kan altså sige at det sproglige materiale på forhånd er blevet analyseret gennem transskriptionen, hvilket hindrer ODT i at opretholde en lydbaseret tilgang til data. I fremtiden ønsker ODT primært at basere sig på lyd og altså komme uden om udskriverens fortolkning af materialet, men indtil videre må der først tages udgangspunkt i transskriptionen, og derefter må de afvigende udtaler sorteres fra.

Da højfrekvente ord kan optræde i korpusset flere tusinde gange, er det ofte nødvendigt at lave en stikprøve. Stikprøven



udarbejdes sådan at den er geografisk repræsentativ i forhold til hele søgningen. I fremtiden ønsker ODT en automatiseret måde at sikre repræsentativiteten på som tager højde for de øvrige sociolingvistiske variable som køn, alder, socialklasse og optagelsestidspunkt.

Søgning på *nja* i korpusset giver 965 resultater. Det er derfor nødvendigt at lave en stikprøve. I første omgang er stikprøven på 10 %, altså 97 excerpter med *nja*. Ved systematisk gennemlytning viser det sig dog at mange af eksemplerne på *nja* er fejltransskriberede i forhold til den faktiske udtale, og at det derfor er nødvendigt at udvide stikprøven for at få nok data. Stikprøven udvides derfor til 20 %, altså 193 excerpter hvoraf kun 69 udtales [nja] og kan indgå i undersøgelsen.

For at lave den bedste stikprøve bør alle eksemplerne i princippet gennemlyttes først og de forkerte frasorteres således at stikprøven kun laves på de korrekt transskriberede eksempler. Det er dog ikke praktisk muligt at gennemlytte så mange excerpter, og ODT må derfor følge ovenstående praksis indtil det er muligt at søge på lyd og undgå transskriptionens bias.

## 2.2. Trækopmærkning

I arbejdet med interjektioner udarbejder ODT et trækinventar til beskrivelse af lemmaer. På denne måde kan forskellige lemmaer eller betydningsvarianter let sammenlignes. Trækinventaret dækker tre områder: pragmatik, interaktion og udtale. I denne undersøgelse af *nja* fokuseres der især på de pragmatiske træk. De pragmatiske træk er udvalgt ud fra ODT's egne empiriske erfaringer og ud fra teoretiske klassifikationer af interjektioner,

specielt Ameka (1992), som drager inspiration fra blandt andre Karl Bühler og Roman Jakobson. Foreløbigt ser ODT's pragmatiske sådan ud (se tabel 1, hvor trækkene står i øverste række, og de mulige trækverdier står i kolonnerne under. «+» betegner at trækket er en del af målordets funktion. «Ø» betegner at trækket ikke er en del af målordets funktion.)

<b>Attitude</b>	<b>Konativ</b>	<b>Fatisk</b>	<b>Ekspressiv</b>
Positiv	+	+	Emotiv
Negativ	Ø	Ø	Kognitiv
Absolut negation			Ø

Tabel 1: Oversigt over ODT's pragmatiske træk og deres værdier.

Attitudetrækket beskriver talerens holdning til egen eller andens ytring. *Positiv attitude* defineres som at taleren forholder sig positivt til indholdet i egen eller andens ytring, for eksempel at taleren udtrykker emfase eller entusiasme. *Negativ attitude* defineres som at taleren forholder sig negativt til indholdet i egen eller andens ytring, for eksempel udtrykker forbehold eller ærgrelse. Derudover kan interjektionerne være *konative*, *fatiske* og *ekspressive*. Et konativt målord har emfase på talerens ønsker, for eksempel *sh*, som kræver en handling fra en anden part, nemlig at vedkommende holder op med at tale. Fatiske målord etablerer og vedligeholder kontakt såsom minimalrespons og hilsner. Ekspressive målord udtrykker talerens mentale tilstand og kan deles op i emotive, som udtrykker talerens følelser, og kognitive, som udtrykker talerens viden og tanker. De ekspressive, fatiske og konative træk udelukker ikke hinanden. Et ord kan altså for eksempel være både fatisk og eks-

pressivt. Desuden er de konative, fatiske og ekspressive træk kompatible med alle værdierne i attitudetrækket.

ODT ser trækkene som de mindste betydningsenheder, der sammen danner funktionsgrupper. De kunne eksempelvis kaldes *begejstring* eller *overraskelse*. I den færdige ordbog skal det være muligt også at søge på funktionsgrupper så man for eksempel kan undersøge forskellige måder at udtrykke *begejstring* på. Det vil også være muligt at søge på både funktionsgrupper og sociolingvistiske variable så man kan søge efter hvordan unge udtrykker *begejstring* eller om det gennem tiden har ændret sig hvordan man udtrykker *begejstring*.

### 2.3. Opdeling i betydninger

Ud fra trækopmærkningen opdeles målordet om nødvendigt i forskellige betydninger. Det viser sig at *nja* har to betydninger som primært adskiller sig fra hinanden ved de pragmatiske træk.

Nja 1: Udtrykker **positiv attitude** i form af bekræftelse af andens ytring eller genbekræftelse af egen ytring. Er **fatisk** ved at det vedligeholder kontakten, og minder om en minimalrespons. (53 excerpter)

Eksempel:

Taler 1: *Men jeg har ikke lavet så mange slåskampe*

Taler 2: *Mm*

NIELSEN

Taler 1: *Nja* (...) *jeg slås ikke meget jeg er blevet mere moden*

*Nja 2: Udtrykker negativ attitude, som tvivl og forbehold, og er kognitivt ved at det siger noget om talerens viden. Det kan også siges at være fatisk da det vedligeholder kontakten. (16 excerpter)*

Eksempel:

Taler 1: *Ja men det er jo da trods alt et temmelig stort hospital*

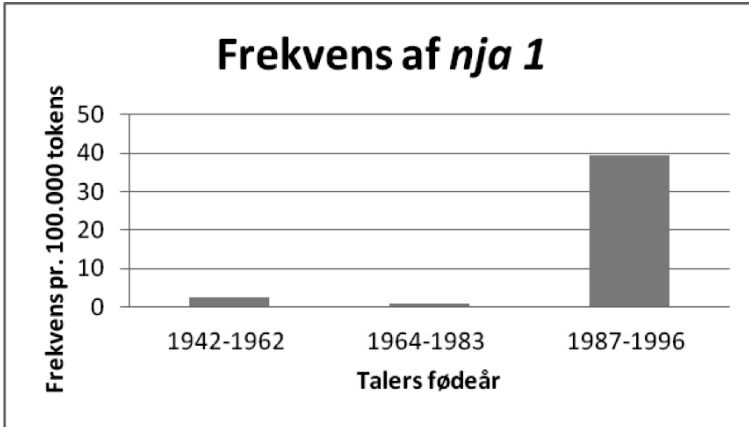
Taler 2: *Nja ja halvanden hundrede patienter*

*Nja 1* beskrives altså med trækkene: positiv attitude og fatisk. *Nja 2* beskrives med trækkene: negativ attitude, kognitiv og fatisk. *Nja 1* og *nja 2* har altså kun ét pragmatisk træk tilfælles: det fatiske.

*Nja 1* og *nja 2* opfører sig også forskelligt interaktionelt. *Nja 1* optræder både som respons på en andens ytring og på egen ytring, mens *nja 2* kun optræder som respons på en andens ytring. *Nja 1* synes også ofte at findes i en speciel konstruktion hvor taleren genbekræfter det vedkommende har sagt tidligere, som det også fremgår i eksemplet på *nja 1*.

Det viser sig også at *nja 1* er betydeligt kortere end *nja 2*. *Nja 1* varer gennemsnitligt 0,27 sekunder, og *nja 2* varer gennemsnitligt 0,6 sekunder. Det stemmer overens med sproglig intuition at det tøvende *nja* (*nja 2*) skulle være langt. Det stemmer også overens med den transskriberede udtale i Den Danske Ordbog hvor *nja 2* gengives [nja:a]. *Nja 1* synes også ofte at blive udtalt med diftong og stød som [njae̥ʔ].

## 2.4. Sociolingvistiske parametre



Figur 1: Diagram over frekvensen af *nja 1* fordelt på taleres fødeår.

Det er muligt at lave statistiske beregninger over de sociolingvistiske parametre. Beregningerne tager højde for korpussets opbygning og udligner eventuelle uligheder for at sikre undersøgelsens validitet. På den måde er det muligt at tegne en profil af den typiske taler af et målord med hensyn til geografisk oprindelse, alder, socialklasse og køn. Det er også muligt at undersøge i hvilke typer samtaler målordet er mest frekvent (enkelt- eller gruppesamtaler med kendt eller ukendt interviewer). Dette siger meget om i hvilken kontekst målordet optræder og bidrager altså i høj grad til beskrivelsen af et givet lemma.

Da korpusset er skævt fordelt i forhold til sociolingvistiske parametre, er de statistiske beregninger baseret på relative frekvenser så det er muligt at undersøge eventuelle sammenhænge

mellem et bestemt målord og de sociolingvistiske parametre.

Der er et særligt stort udsving på et af parametrene, nemlig talers fødeår. *Nja 1* er generelt meget mere frekvent end *nja 2*. *Nja 1* forekommer gennemsnitligt 11,1 gange pr. 100.000 tokens, og hvor det er mest frekvent, ved talere født efter 1987, forekommer det 39,2 gange pr. 100.000 tokens. Til sammenligning forekommer *nja 2* gennemsnitligt kun 1,6 gange pr. 100.000 tokens. Der er også en klar tendens til at *nja 1* oftest bruges af unge. Dette illustreres i figur 1.

### 3. Interjektioner og talesprog

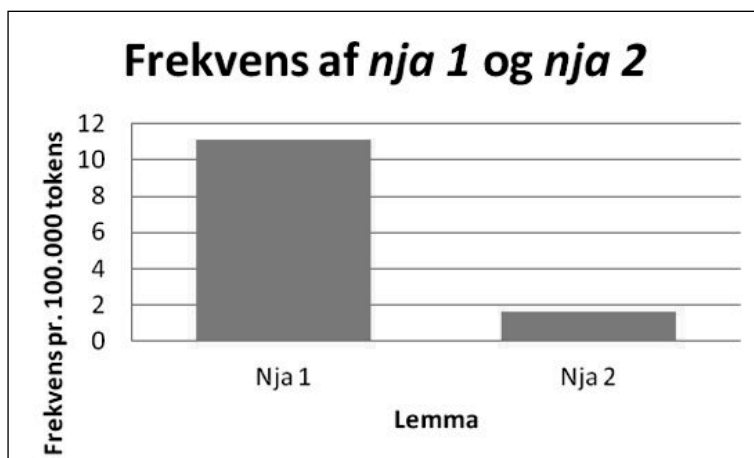
Arbejdet med *nja* viser at den traditionelle håndtering af interjektioner ikke er tilstrækkelig. Ved at undersøge interjektioner på deres egne præmisser, nemlig i talesprog og med udgangspunkt i lyd, viser det sig at *nja*, og måske også andre interjektioner, kan beskrives præcist og uddybende både semantisk, distributionelt og udtalemæssigt. *Nja* er altså bestemt ikke usystematisk eller uvigtig for sprogbeskrivelse som ældre teorier om interjektioner ellers foreslår, og undersøgelser som denne om *nja* kan bidrage til en mere nøjagtig beskrivelse af ordklassen interjektioner.

Ud over den generelle beskrivelse af interjektioner har ODT også vist at man med en lydbaseret tilgang til talesprog kan afdække nye lemmaer. Tabel 2 viser forskellen på ODT's og DDO's behandling af *nja*. DDO (Den Danske Ordbog) er primært baseret på skriftsprogsdata.

Lemma	Ordbog over Dansk Talesprog	Den Danske Ordbog
<b>Nja 1</b>	Udtrykker positiv attitude i form af bekræftelse af andens eller genbekræftelse af egen forudgående ytring.	(Fremgår ikke)
<b>Nja 2</b>	Udtrykker negativ attitude som tvivl og forbehold.	Bruges for at udtrykke tøven, tvivl eller forbehold, fx når man ikke mener at kunne svare entydigt ja eller nej på et spørgsmål.

Tabel 2: ODT's og DDO's behandling af *nja*.

ODT har altså ved hjælp af et talesprogs-korpus både fundet det lemma som findes i DDO, og afdækket et nyt lemma eller en ny betydningsvariant som ikke fremgår i traditionelle ordbøger. Det tyder på at talesproget gemmer på mange, og endda meget frekvente, lemmaer eller betydningsvarianter som ikke optræder eller er synlige i skriftsprog. I figur 2 ses det at *nja 1*, den tidligere ukendte betydningsvariant, faktisk er meget mere frekvent end det kendte *nja 2*. Med en lyd-baseret tilgang til sproget og med udgangspunkt i samtalesprog er det altså muligt at supplere traditionelle ordbøger og bidrage til en mere præcis beskrivelse af dansk sprog.



Figur 2: Diagram over frekvensen af *nja 1* og *nja 2*.

## Litteratur

- Ameka, Felix (1992): Interjections: The universal yet neglected part of speech. I: *Journal of Pragmatics* 18, 2-3, 101–118.
- Jespersen, Otto (1924): *The philosophy of grammar*. Great Britain: Unwin Brothers LTD.
- Wilkins, David P. (1992): Interjections as deictics. I: *Journal of Pragmatics* 18, 2-3, 119–158.

Jessie Leigh Nielsen  
Ordbog over Dansk Talesprog  
Københavns Universitet  
mgk690@alumni.ku.dk



# **Att läsa ordboken baklänges. Om betydelse- utveckling hos partikelinledda fasta samman- sättningar och presentationen i SAOB**

*Pär Nilsson*

Volume 36 of the Swedish Academy's dictionary (SAOB) was finished in december 2012. The main part of this volume consists of the particles *under*, *upp*, *ur* and *ut*, and compounds with these words. Development of word meaning of particles differs in some respects from that of other words. This paper deals with development of word meaning for compounds with a particle as the first element, and different interpretations of SAOB:s system of description of such development.

## **1. Inledning och frågeställning**

SAOB:s band 36, TYNGA – UTSUDDA, färdigställdes under slutet av 2012. Vid arbetet med detta band har alla på redaktionen på ett eller annat sätt varit sysselsatta med partiklarna UNDER, UPP, UR och UT och sammansättningar till dem. Ord som upptar mer än hälften av utrymmet i bandets spalter. En vanlig kommentar i mötet med läsaren är att SAOB-texten är svårforcerad och kompakt och besvärlig att tolka. Många kryptiska förkortningar och svårtolkade siffersystem ska tas hänsyn till och uttydas.

I mitt bidrag kommer jag att diskutera lite av snårigheten i SAOB, och förhoppningsvis tydliggöra innehållet lite grand. Jag kommer att analysera SAOB:s system att hos sammansatta ord hänvisa läsaren tillbaka till betydelsen hos sammansättningens förled såsom simplex. Härigenom vill jag förklara en ibland rätt så oklar grafisk struktur i ordboken. Samtidigt vill jag uppmärksamma hur partiklars betydelser utvecklas dels såsom simplex, dels som förled i sammansättning, och också diskutera hur dessa utvecklingsmönster representeras i SAOB.

## 2. Teoretisk bakgrund

Vid beskrivningen av partiklar och partikelverb har man traditionellt sett ofta fokuserat på grammatiken och de syntaktiska funktionerna. Partikelverbens semantiska utveckling från en betydelse till en annan har inte ägnats stor uppmärksamhet i litteraturen. Snarare har fokus riktats mot utvecklingen från simplexverbets betydelse till det sammansatta verbets betydelse: på vilket sätt skiljer sig betydelserna åt i sådana fall som *hämta* respektive *hämta ut* och *skriva* respektive *skriva upp*?

Vid jämförelser av det här slaget är den semantiska divergensen ofta marginell, och istället har olikheter i fråga om aktionsart och transitivitet mellan partikelkonstruktioner och simplexverb hamnat i förgrunden. Denna tradition kan spåras tillbaka till Adolf Noreens «Vårt språk» (1904), och bevaras bl.a. i en av den senaste tiden mest grundläggande analyser av

svenska partikelverb, Kerstin Noréns «Svenska partikelverbs semantik» (1996).

Svenska Akademiens Grammatik (SAG) lägger inte oväntat också stort fokus på funktionen vid beskrivningen av verbpartiklar. I avsnittet om partikelförbindelsernas betydelse (3:428–9) poängteras förvisso inledningsvis att «helhetens betydelse (*ofta*) ... något så när (*kan*) förstås på grundval av delarnas», men störst uppmärksamhet riktas ändå just mot aktionsart och det sammansatta verbets förhållande till simplex verbet i sådana jämförelser som exempelvis *brinna upp/ner* respektive *brinna* (s. 428).

Med målsättningen att förklara det som är generellt och absolut för hela gruppen partikelverb är det också förståeligt att en sådan svagt semantisk företeelse som aktionsart uppmärksammas i så hög grad – det är något som kan undersökas och beskrivas ord för ord. Men att detta fokus har större relevans för förklaringen av den grammatiska funktionen än för partikelverbens och de enskilda partiklarnas betydelse är inget djärvt påstående. Partiklar och partikelverb befinner sig i praktiken i gråzonen mellan grammatik och lexikon och detta kan ses som en möjlig förklaring till att de historiskt sett fått förhållandevis litet utrymme i litteraturen. Lexikonets status har också historiskt sett varit lägre än grammatikens, vilket även kan antas ha bidragit till det faktum att då partiklar och partikelförbindelser väl undersökts semantiskt, så har det grammatiska perspektivet ändå fått det sista ordet.

En mer betydelsefokuserad analys utförs däremot av Strzelecka (2003), där partiklarna och dess förbindelser undersöks utifrån ett kognitivt perspektiv. Analysen utgår ifrån människans sätt att tänka och conceptualisera, och betydelser och betydelseutveckling förklaras i termer av *perspektiv*, *metaforer* och *metonymer* (s. 15 f.). Istället för att söka efter de

gemensamma dragen för de partiklar och förbindelser som undersöks, analyseras begreppen utifrån sina egna förutsättningar, sina unika kontexter och också i förhållande till sina potentiella motsatser (*upp – ner, in – ut*)<sup>1</sup>.

Vid beskrivningen av partiklar och partikelförbindelser, deras olika betydelser och betydelseutvecklingar – utifrån SAOB:s metod – är detta synsätt enligt min mening tilltalande och också förenligt med ordbokens grundsyn. Det är denna tolkning av begreppen som jag i det följande ansluter mig till vid analysen av partiklar och partikelkonstruktioner i SAOB.

### **3. SAOB:s hänvisningssystem och partiklars betydelseutveckling**

Partiklar befinner sig alltså i skärningspunkten mellan grammatik och lexikon, och att de är semantiskt kontextberoende och osjälvständiga är ett välkänt faktum (jfr t.ex. SAG 3:417f. och Wessén 1965:109f.). Tillsammans med olika (typer av) verb antar och utvecklar de sina olika betydelser. SAOB:s utgångspunkt vid definitionen är följaktligen just partikeln i förbin-

---

1. Betydelseskillingen mellan *in* i ”gå in i ett rum”, och i ”borra in sina fingrar i någons arm”, kan i dessa termer förklaras av att man i det förra fallet har att göra med en prototypisk behållare (ett rum), och i det senare något som genom sin förändrade, något inbuktande form (en arm) i tanken fungerar som en (således mindre prototypisk) behållare (Strzeleckas 2003:159).

delse. I relation till (i första hand) verb beskrivs den enskilda partikeln med hjälp av flera olika exempel och talande belägg i definitionstexten såväl som genom språkproven.

En partikel kan, som andra ord, givetvis förekomma i en mängd olika kontexter och förbindelser och därmed generera och etablera en mängd olika betydelser. En etablerad och mer eller mindre klart avskiljbar betydelse presenteras i SAOB som ett eget *betydelsemoment*, med arabisk siffra. För adverbet UT exempelvis, den sista av de ovan nämnda partiklarna i bd 36, presenteras betydelserna i SAOB i 13 olika betydelsemoment (se vidare nedan).

Sammansättningar grupperas och presenteras i SAOB normalt utifrån den aktuella förleden i en särskild underavdelning inom artikeln, den s. k. *sammansättningsramsan*. Inom denna ramsa ansluts varje sammansättning i regel till aktuell betydelse hos förleden såsom simplex (då detta ord är polysemt) för att leda läsaren rätt. Med en siffra hänvisas läsaren till aktuellt betydelsemoment hos förleden. Rent praktiskt kan sådana hänvisningar utformas på två olika sätt. Antingen placeras siffran precis före sammansättningens efterled, eller också, vid flera olika betydelser hos sammansättningen, och då de kan hänföras till olika moment av förleden, i anslutning till varje enskild betydelse hos sammansättningen.

Exakt hur detta system fungerar är inte alldeles genomskingligt och lätt att förstå, och det har av den anledningen även uppmärksammats i tidigare sammanhang, bl. a. av Lundbladh (1992:55f) och Eaker (2006:50). En rimlig slutsats av vad momenthänvisningen faktiskt innebär, utifrån SAOB:s kronologiska och logiska tillvägagångssätt, där man tar sig från då till nu och från det enkla till det sammansatta, är att den förklarar sammansättningens bildningssätt, dvs. hur den aktuella betydelsen har uppstått. Denna tolkning kan exemplifieras med

hjälp av verbet TRYCKA som i SAOB presentas med bl. a. följande fyra betydelser, och med en utvecklingsgång för det enkla ordets betydelser som i exemplet illustreras av pilarna:

### TRYCKA

↓ **1)** gm direkt kontakt utsätta (ngt l. ngn ) för mer l. mindre  
↓ stor kraft (l. tyngd l. vikt) ...

↓ **2)** förskjuta l. förflytta l. tvinga l. tränga (ngt l. ngn ngnstans  
↓ (l. så l. så l. med visst resultat)) gm tryckande (i bet. **1**) ...

↓ **3)** trycka (i bet. **1** l. **2**) (ngt) mot ett underlag o. d. så att det  
↓ uppstår ett avtryck l. märke o. d. ...

**4)** [eg. specialanv. av **3**, jfr motsv. anv. i mlt. o. t.] medelst tryckpress o. d. (låta) mångfaldiga (skrivet alster (ss. bok l. tidning o. d.) l. bild(verk) o. d. ...

Ordboksexempel (i)

Vid sammansättningar med TRYCKA som förled, t.ex. SAOB:s «TRYCK-(2)-PUMP» och «TRYCK-PRESS. till 4», förklarar hänvisningssystemet att betydelse 2 respektive 4 av uppslagsordet TRYCKA aktualiseras. Systemet verkar här bygga på en additiv princip och förutsätta en kronologisk tolkning: en viss betydelse av förleden + (en viss) betydelse av efterledet = sammansättningens betydelse.

Vid fall med partikel såsom förled i sammansättning fungerar emellertid inte SAOB:s modell för momenthänvisning lika bra, eller åtminstone inte den ovan angivna tolkningen av den. Beskrivningen av UT kan återigen tjäna som exempel, i ordboksexempel (ii) nedan återgiven i form av SAOB:s sammanfattande översikt.

**UT**, adv. *Översikt av betydelserna*

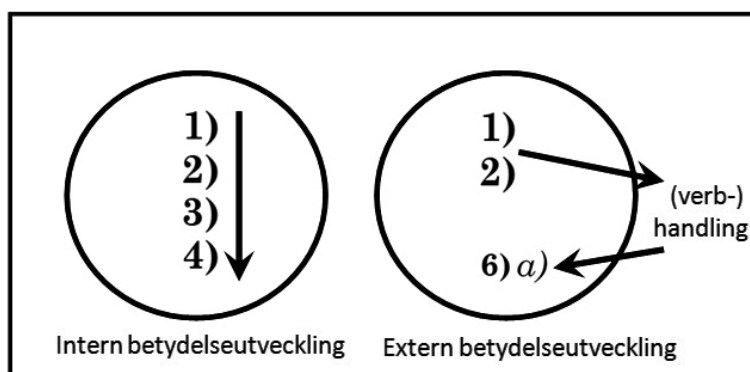
**1**) i fråga om riktning från ett (tänkt) centrum. **2**) i fråga om riktning från det inre av ett avgränsat område. **3**) i fråga om förflyttning av ngt som består av olika delar så att större avstånd mellan de olika delarna åstadkoms. **4**) i fråga om förflyttning som innebär att en del av ngt avlägsnas från detta. **5**) i fråga om handling l. skeende som har en startpunkt l. riktar sig från ngt l. mot ngt. **6**) i fråga om att ngt lämnar l. kommer från l. uppstår ur ngt annat. **7**) i fråga om (ekonomiska) transaktioner. **8**) i fråga om att ngn drabbas av ngt. **9**) i fråga om handling som berör skicket l. ordningen l. fullbordandet av ngt. **10**) i fråga om tid. **11**) i fråga om att ngt försvinner l. förbrukas l. dör bort; äv. i rent förstärkande anv. **12**) i rumslig bet., särsk. i fråga om befintlighet på en plats som är längre ut l. utanför en annan plats. **13**) i fråga om tillstånd l. belägenhet.

*Ordboksexempel (ii)*

SAOB:s beskrivning av **UT** tar sin utgångspunkt i två grundbetydelser, eller två prototypiska användningar, i moment **1** och **2**. Det är betydelsen som rör centralitetsbegreppet: ut från centrum (mot periferin), och betydelsen som behandlar behållarbegreppet: ut ur behållare. Härur kan sedan ett stort antal användningar, eller andra betydelser härledas och förklaras. Som exempel kan nämnas betydelser som: ut  $\approx$  i ljuset eller offentligheten (**UT 6 a**), som exempelvis i förbindelsen *ropa ut*, ut  $\approx$  transaktion (**UT 7**), som i *hyra ut*, eller ut  $\approx$  bort (**UT 11**), som i *radera ut*.

Liksom vid verbet **TRYCKA** har betydelserna här utvecklas ur en mer eller mindre ursprunglig betydelse till en annan eller flera andra. Skillnaden här är dock att när betydelseutvecklingen hos **TRYCKA** har skett internt, inom ramen för begreppet självt så är motsvarande utveckling hos partikeln (i

detta fall UT) extern i förhållande till begreppet självt. Dvs. TRYCKA:s betydelse 1 i SAOB innehåller i sig så många semantiska komponenter att den i en metaforisk användning själv kan ge upphov till en ny betydelse. Riktningen ut ur behållare hos UT är däremot en semantiskt tunnare betydelse och kräver för att utvecklas till en ny betydelse tillskott av annat innehåll. För att ta sig från UT:s betydelse 2 till 6 a i SAOB (vilka uppges samhöriga) måste man således ta hjälp utifrån från en (verb)handling som inte rymms i begreppet självt. Och som nämdes tidigare tas då i själva definitionen användningar av begreppet i olika uttryck tillvara, som betonad partikel i de s. k. särskilda förbindelserna, eller såsom förled i sammansättningar.



Figur 1: Intern och extern betydelseutveckling.

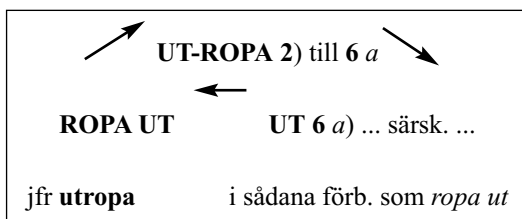
Tillbaka till sammansättningsramarna kan betydelsemoment 2 av UT-ROPA få exemplifiera SAOB:s hänvisningssystem:



**UT-ROPA**

2) till 6 a: (ropande l. ss. ett utrop) yttra (ngt); äv. ...  
 offentligt meddela l. bekantgöra l. tillkännage (ngt) ...  
 Ordboksexempel (iii)

Utifrån hänvisningsmodellen och de ovan anförda exemplena borde tolkningen således ge: UT-ROPA 2 är summan av UT 6 a + ROPA, vilket visst kan sägas förklara ordets betydelse, eller åtminstone hjälpa till att ringa in sammanhanget för läsaren. Men hänvisningarna förklarar knappast hur ordet bildats eller hur sammansättningens aktuella betydelse utvecklats – betydelsen som i SAOB anges som UT 6 a, beskrivs nämligen just utifrån «sådana förb(indelser) som ropa ut». Ur ett kronologiskt ordbildningsperspektiv har modellen i dessa fall ett lågt förklaringsvärde, och man kan också tala om cirkelresonemang i det här avseendet.



Figur 2: Exempel på cirkulärt hänvisningssystem i SAOB.

Rent grafiskt ser hänvisningarna i ramsorna likadana ut i sådana fall som TRYCK-PRESS (med intern utveckling) och i sådana som UT-ROPA (extern). I praktiken föreligger emellertid en principiell skillnad vad gäller tolkningen av dem. SAOB:s modell för hänvisning till betydelsemoment hos förleden ska alltså uppfattas delvis annorlunda i dessa båda fall.

I det förra fallet (den interna) anges att ordets betydelse är sammansatt av förleden i angiven betydelse + efterledens betydelse. Hänvisningen kan således i princip översättas med frasen «bildat till». I det senare fallet (den externa) får man göra en mer allmän tolkning. Sammansättningsens förled har här en viss anslutning (som inte nödvändigtvis har att göra med bildnings-sättet) till den angivna betydelsen. Hänvisningen meddelar i sådana fall snarast att man för att få utökad hjälp med förståelsen av sammansättningsens betydelse, också bör studera den betydelse som hänvisas till. I denna tolkning är hänvisningen i praktiken inne och tassar på definitionens domäner. Vid en sammansättning med något slag av definition borde en sådan upplysning dock te sig tämligen överflödig. Av definitionen som sådan kan man dra tillräckliga slutsatser för betydelsen hos sammansättningen och dess ingående leder. Ytterligare ett exempel får belysa detta faktum. Till betydelsemoment **6 a** hos UT hänförs även UT-BLÅSA moment **3**.

### UT-BLÅSA

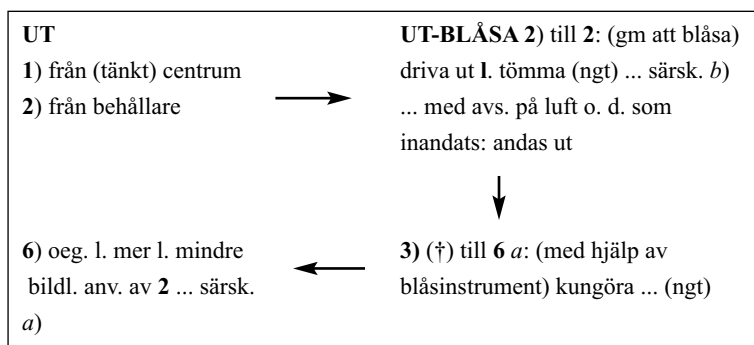
**3** (†) till **6 a**: (med hjälp av blåsinstrument) kungöra l. tillkännage början av (ngt); äv. basunera ut; jfr **blåsa ut 7**.  
Ordboksexempel (iv)

Vad sammansättningen betyder framgår med tydlighet av definitionen. Hur lederna förhåller sig till varandra, och var för sig betyder är också klart. Bilden eller metaforen av att med instrument blåsa ut en nyhet behöver nog inte förklaras ytterligare. Så vad tillför hänvisningen «till **6 a**» förståelsen av betydelsen? Vid exempel av den här typen, menar jag, ökar inte hänvisningen förståelsen för sammansättningen och dess betydelseutveckling och bildning. Däremot kan den sägas öka förståelsen för betydelse(utveckling)en hos det enskilda, simplexa

ordet, alltså hos förleden. Genom att hänvisa läsaren vidare till aktuellt betydelsemoment hos förleden får man exempel på användningar av betydelsen, och ökad förståelse för hur denna betydelse uppstått. Låt vara att det knappast är den information läsaren är ute efter när hon eller han slår upp en sammansättning i ordboken. Genom att följa flera sådana sammansättningar tillbaka till betydelsen hos förleden, får man en större bild av sammanhanget. Det är ett systematiskt sätt att öka förståelsen för betydelsen, men det sker så att säga baklänges. Likartade sammansättningar som UT-BASUNERA, UT-BLÅSA och UT-DUNDRÄ målar tillsammans med hjälp av hänvisningarna upp bilden av den användning och betydelse av UT som i SAOB återges i moment **6 a**, och hur den har utvecklats.

Momenthänvisningarna kan förklara hur ett utvecklingsmönster som **1, 2 → 6** hos UT i SAOB i praktiken äger rum; dvs. hur man via verbförbindelserna har tagit sig från **1** till **6**. En sådan extern betydelseförändring uppvisar alltså inte samma linjära utveckling som den interna (jfr ordboksexempel (i) ovan), och tolkad så här kan momenthänvisningen ses som en komplementär uppgift, den felande länken mellan två betydelser.

Om man går steget längre och undersöker utvecklingen hos den enskilda sammansättningens betydelser framträder en extra dimension. Med fortsatt fokus på sammansättningen UT-BLÅSA och dess redovisade betydelser i SAOB, visar det sig att det är på denna nivå som den bildliga eller metaforiska betydelseförändringen egentligen äger rum. Den förändring som nämns under simplexet UT, men som helt enkelt blir för komplicerad att förklara innebörden av där. Sammansättningarnas momenthänvisningar kan, tolkade som tillsammans utgörande ett system, hjälpa till att åskådliggöra detta förhållande.



Figur 3: Betydelseutveckling hos simplex partikel via användning i sammansättning.

Problematiskt vid fall som UT-BLÅSA är, som nämndes ovan, att läsaren som slår upp ordet i ordboken sannolikt inte är ute efter denna information. Är man nyfiken på utvecklingsförhållandena mellan sammansättningens olika betydelser skymmer snarast en hänvisning som (**6 a**) förhållandet att betydelse **2 b** hos UT-BLÅSA gett upphov till **3**, och man får istället intrycket att denna betydelse är bildad direkt till UT:s betydelse **6 a** utan påverkan från sammansättningens andra betydelser.

Att det finns en konflikt mellan en sammansättnings betydelseutveckling å ena sidan och anslutning till betydelsemoment hos förleden å den andra, är ett faktum och en prövning som SAOB-redaktörerna har fått ställas inför vid en mängd tillfällen. Det har därvid utvecklats ett antal metoder att ta sig runt problemet (momenthänvisningssystemet är alltså inte en tvingande regel för redaktören). Även om det normala för en redaktör är att följa hänvisningsmodellen, så kan man också i sammansättningsramorna stöta på exempel där man utgått ifrån sammansättningens egna betydelseutveckling på bekostnad av sådana hänvisningar.

### UT-FLYTA

**1)** om vatten l. annan vätska (l. lättflytande massa): flyta ut  
l. fram, strömma l. flöda ut l. ur l. från ngt.

**2)** i utvidgad l. mer l. mindre bildl. anv. av **1**; särsk.: utgå  
(o. sprida sig), strömma ut; särsk. om pengar l. andra  
tillgångar.

Ordboksexempel (v)

UT-FLYTA moment **2** i exemplet ovan visar hur man avstått från hänvisning till moment hos förleden och istället tagit utgångspunkt i sammansättningens ursprungliga eller egentliga betydelse. Ännu tydligare blir ett fokus på sammansättningens egna betydelseutveckling när ordet så att säga lyfts upp en nivå ur sammansättningsramsan och får bilda eget uppslagsord. Vid sådana egna huvuduppslagsord, *utbrutna sammansättningar* som de kallas på redaktionen, tillämpas inte heller hänvisningsmodellen. Det mest extrema tilltaget för att undgå konflikten mellan sammansättningens betydelseutveckling och momentanlutning till sammansättningens förled hittar man annars i sammansättningsramsan till adverbet IN, där modellen inte överhuvudtaget tillämpas.

För att göra bilden av partikelsammansättningarnas betydelseutveckling än mer komplex, och för att lyfta fram vad som trots allt kan motivera systemet med hänvisningar till förledsbetydelse i SAOB, bör också det faktum uppmärksammas, att de enskilda partiklarnas betydelser, vilka alltså utvecklats externt, också i högre eller lägre grad kan vara lexikaliserade (såsom enskilda element och inte enbart i förbindelser). Genom användning i de olika uttrycken, i sammansättningar och särskilda förbindelser, stelnar betydelsen hos partikeln och kan sedan sättas samman direkt med annat led (vanligen ett verb) efter mönster av liknande ord.

**UT-RUTTEN**

(11 (c)) (†) helt rутten l. förruttnad; förstörd av röta.

Ordboksexempel (vi)

Här fungerar hänvisningsmodellen utan konflikt, och tydligast blir denna företeelse vid sammansättningar som bara uppvisar en enda betydelse. Det är svårt att tänka sig en egentlig eller ursprunglig betydelse hos sammansättningen UT-RUTTEN här, som den aktuella betydelsen sedan utvecklats ifrån. Snarare har den uppkommit efter påverkan av liknande bildningar (t.ex. *utplåna*, *radera ut* osv.). Betydelsen kan således sägas utgöra en regelrätt sammansättning av förledens aktuella betydelse + efterledens dito.

Med dessa två möjligheter: att en partikelinledd sammansättnings betydelse antingen har utvecklats inom sammansättnings egen ram, eller också uppstått genom sammansättning av två redan befintliga betydelser till en ny, följer även en potential för betydelsen att höra hemma inom båda dessa kategorier samtidigt. Betydelsen kan ursprungligen ha utvecklats ur en annan och ha tydliga kopplingar till dit, och samtidigt genom lexikaliserad betydelse hos förleden (som ursprungligen uppstått bl. a. ur denna sammansättnings utvecklade betydelse) bildas av andra språkbrukare på nytt. Detta förhållande kan också anges i SAOB genom ett slags kompromiss. I fall som UT-PUMPA 2, i ordboksexempel (vii) nedan, ansluts betydelsen samtidigt till både förleden (UT) och till en egentlig eller ursprunglig, egen betydelse hos sammansättningen (UT-PUMPA 1). UT-PUMPA 2 antas alltså ha både sammansatt och internt utvecklad betydelse.

## UT-PUMPA

1) gm pumpning avlägsna (vätska l. gas o. d. ur behållare l. utrymme o. d.) (jfr **pumpa ut 1**)

2) i mer l. mindre bildl. anv. av 1. särsk. [delvis med anslutning till **ut**, adv. **11 a**] i p. pf. i adjektivisk anv.: utmattad l. uttröttad l. slutkörd.

Ordboksexempel (vii)

Slutligen finns för den som analyserar partikelinledda sammansättningar fall där det är svårt eller omöjligt att avgöra om ursprunget är det ena eller det andra. SAOB bygger på ett citerat material, och att bedöma författarens avsikt i varje fall av ordanvändning är inte möjligt. I sådana fall finns även möjligheten för redaktören att redovisa tvetydighet genom en kompromiss av ovan angivet slag.

## 4. Sammanfattning

SAOB beskriver ord och vad de betyder ur ett historiskt perspektiv. Principen bygger på tanken om det självförklarande i att presentera information kronologiskt (med antydande uppgifter om ursprung, men) utan explicita kommentarer i övrigt. Tillvägagångssättet är präglat av de teoretiska ideal som rådde vid projektets början, men har också praktiskt, utrymmesbesparande syfte. Idag blir ordbokstexten ibland svårtolkad med

NILSSON

de olika implicita systemen och kryptiska formlerna för hänvisningar och olika samband. SAOB:s modell att beskriva en sammansättnings förhållande till aktuell betydelse hos förleden är ett sådant implicit system som inte alla gånger är helt lätt att tolka och som ibland även är missvisande.

Vi har sett att partikelinledda sammansättningar betar sig delvis annorlunda i förhållande till andra sammansättningar (inledda av verb eller substantiv) i fråga om bildningssätt och betydelseutveckling. SAOB:s hänvisningssystem för momentanslutning är emellertid detsamma i båda fallen. För partiklar genereras betydelseutveckling hos simplex i relationen med andra element, t. ex. efterleden i en fast sammansättning. Principen för denna utveckling har inte uppmärksamats i vidare utsträckning i litteraturen. Med lite god vilja är det möjligt att nå också denna utvecklingsgång i SAOB, men då genom att följa hänvisningssystemet baklänges.

## Litteratur

### Ordböcker

SAOB (1898–) = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1–. Lund.



## Övrig litteratur

- Eaker, Birgit (2006): Historisk lexikografi i praktiken. Om momentordningen i SAOB. I: LexicoNordica 13, 41–54.*
- Lundbladh, Carl-Erik (1992): Handledning till Svenska Akademiens ordbok. Lund: Svenska Akademien & Norstedts.*
- Noreen, Adolf (1904): Vårt språk: nysvensk grammatik i utförlig framställning. Lund.*
- Norén, Kerstin (1996): Svenska partikelverbs semantik. (Nordistica Gothoburgensia 17. Acta Universitatis Gothoburgensis.) Göteborg.*
- SAG = U. Teleman, S. Hallberg & E. Andersson (1999): Svenska Akademiens Grammatik. Stockholm.*
- Strzelecka, Elżbieta (2003): Svenska partikelverb med in, ut, upp och ner. En semantisk studie ur kognitivt perspektiv. (Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet 62) Uppsala.*
- Wessén, Elias (1965): Svensk språkhistoria 2. Ordbildningslära. Stockholm: Almqvist & Wiksell.*

Pär Nilsson  
redaktör SAOB  
Redaktionen för Svenska Akademiens ordbok  
Dalbyvägen 2, 224 60 Lund  
Par.Nilsson@svenskaakademien.se

# Från *aspekt till övergripande* – en ordlista över svensk akademisk vokabulär

*Judy Ribeck, Håkan Jansson & Emma Sköldberg*

This report describes a project to develop an academic word list for Swedish. The resulting word list is published at <http://spraakbanken.gu.se/ao/>. It comprises 655 headwords, extracted from a 25 million word corpus of Swedish academic texts. Both the word list and the corpus are openly accessible through Språkbanken's lexical and corpus infrastructures.

## 1. Inledning

Betydelsen av att tillägna sig ett adekvat ordförråd för akademisk verksamhet har sedan en tid fått ökad uppmärksamhet. Till en början gällde intresset i huvudsak akademisk engelska, men nu växer insikten om behoven även för andra språk. Mot denna bakgrund har forskare sedan 1970-talet tagit fram olika *akademiska ordlistor*.

Vi presenterar här vårt arbete med att framställa en *svensk akademisk ordlista*. Syftet med denna är att stödja, i första hand, andraspråksinlärare på mer avancerad nivå, men även modersmålsstudenter som är ovana vid det akademiska språket. Listan riktar sig också till dem som väljer att skriva på svenska i stället för på engelska inom ramen för sina högre studier.

Ordlistan är främst tänkt att användas vid produktion av akademisk text, men tack vare att uppslagsorden försetts med svenska betydelseangivelser och engelska översättnings-ekvivalenter, kan den även vara till hjälp vid reception. Dessutom kan vårt arbete tjäna som underlag för utveckling av framtida ordtester och undervisningsmaterial samt bidra till att dokumentera det svenska akademiska ordförrådet.

Ända sedan projektets början har vi kontinuerligt redogjort för dess fortgång i olika sammanhang. Vi hänvisar därför den som vill följa våra metodologiska överväganden till tidigare publikationer.<sup>1</sup>

Denna artikel inleds med en kort forskningsbakgrund, varefter vi beskriver vårt korpusmaterial, vår metod för urval av uppslagsord samt vår resulterande ordlistas innehåll och utformning. Texten avslutas med några ord om framtida utvecklingsbehov.

## 2. Tidigare akademiska ordlistor

I engelskspråkig litteratur står den övergripande termen *aca-*

---

1. Se Sköldberg & Johansson Kokkinakis (2012) för en allmän projektpresentation, Jansson et al. (2012) om insamling av akademiska texter, Johansson Kokkinakis et al. (2012) om nordiskt samarbete och Carlund et al. (2012) om CALL. Vi vill dock understryka att dessa arbeten behandlar en tidigare version av ordlistan, som byggde på ett mindre omfattande korpusmaterial.

*demic vocabulary* för en typ av ord som förekommer frekvent i löpande akademisk text från flera olika discipliner (se Paquot 2010:17–21 för en termutredning). Med början på tidigt 1970-tal har en rad olika engelska akademiska ordlistor framställts, för att möta behoven hos studenter på avancerad nivå. Den som har haft störst genomslag är *The Academic Word List* (AWL, Coxhead 2000 m.fl.).

AWL omfattar sammanlagt 570 *ordfamiljer*<sup>2</sup>, fördelade över 10 frekvensbaserade dellistor. Listan är baserad på en korpus på 3,5 miljoner ord som utgörs av akademiska artiklar och kurslitteratur. Korpusen är indelad i fyra discipliner, vilka, var och en, innehåller sju ämnesområden. Förutom de allmänna, ovan nämnda kriterierna för akademiska ord, gäller för ordfamiljerna i AWL att dess medlemmar inte får tillhöra de 2000 vanligaste orden i språket.

Under det senaste decenniet har AWL använts flitigt i språkundervisning, ordkunskapstest och läroböcker, samt som forskningsunderlag (Coxhead 2011). Listan har dock inte undgått kritik. I huvudsak går kritiken ut på följande:

1) Användningen av ordfamiljer snedvrider ordurvalet. Ord, som inte av egen kraft uppfyller kriterierna, «räddas» genom avledningar med annan betydelse. Vidare tas ingen hänsyn till homografi och polysemi; gemensamt etymologiskt ursprung behöver inte innebära att ordstammen har samma betydelse i besläktade ord (Wang Ming-Tzu & Nation 2004).

2) Indelningen och urvalet av korpustexter är godtycklig (Hyland & Tse 2007).

---

2. Idén att använda s.k. *ordfamiljer* vid ordinläring presenterades i Bauer & Nation (1993). Med en ordfamilj avses en ordstam och alla dess vanliga böjningar och avledningar, t.ex. *react*, *reacting*, *reaction*, *reactionary*, *reactive*, *unreactive*, *reactivate*, *reactor*.

3) Föresatsen att exkludera vardagliga ord har ifrågasatts, då vissa ord ur basordförrådet anses ha speciella akademiska funktioner (jfr Paquot 2010, Gardner & Davies 2013). Här utmanas följaktligen hela idén om att ordförrådet kan delas in i diskreta stycken, som kan studeras var för sig (jfr Nation 2001). Det är dessutom problematiskt att, som Coxhead, grunda exkluderingen av ord på en jämförelse med en föråldrad ordlista som *General Service List* från 1953.

Olika forskare är alltså inte överens om hur man bäst beskriver och identifierar akademiskt ordförråd. Detta har, på senare år, resulterat i åtminstone två seriösa utmanare till AWL: *the Academic Keyword List* (AKL, Paquot 2010) och *the Academic Vocabulary List* (AVL, Gardner & Davies 2013). Båda dessa ordlistor räknar med lemman, i stället för ordfamiljer. Båda kräver också att orden är jämnt fördelade över akademiska texter, samt utmärkande för just denna texttyp (s.k. *nyckelord*). Det senare kravet uppfylls genom jämförelser med referenskorpusar.

AKL innehåller 930 akademiska nyckelord, som extraherats ur en korpus på 3 miljoner ord, fördelade över akademisk prosa och studentuppsatser. Texterna kommer från fem olika discipliner. I materialet ingår även en skönlitterär referenskorpus.

AVL består av 3000 ord, framtagna ur ett korpusmaterial på över 120 miljoner ord. Texterna utgörs av tidskriftsartiklar från nio olika discipliner. Som jämförelsematerial används referenskorpusar med nyhetstexter och skönlitteratur. Förutom det stora empiriska underlaget och listans omfattning, är AVL även unik i det att orden inte får förekomma «oväntat» mycket i någon eller några få discipliner.

### 3. Metod och material

Den metod som vi använt för att ta fram kandidater till den svenska akademiska ordlistan är tydligt inspirerad av tidigare försök att extrahera akademiska ord ur en korpus med akademiska texter. Det första steget består således i att, på bästa sätt, sätta samman en korpus som är representativ för svenskt akademiskt skriftspråk (se 3.1). Nästa steg är att i denna korpus identifiera akademiska ord (se 3.2).

#### 3.1. SveAk

*SveAk* – *Svensk akademisk korpus* – består av sammanlagt 25,4 miljoner ord från avhandlingar och tidskriftsartiklar som publicerats 1997–2012. Korpusen är fritt tillgänglig via korpusinfrastrukturen, *Korp*, i Språkbanken (se Borin et al. 2012b för närmare beskrivning). Vid textinsamlingen har vi utgått från den nationella databasen *SwePub*, som listar alla publikationer från svenska universitet och högskolor enligt en internationell standard.<sup>3</sup>

---

3. Se Jansson et al. (2012:958) och där angivna referenser för närmare detaljer.

Humaniora	Ord <sup>4</sup>	Samhällsvetenskap	Ord
Etnologi	1 669	Ekonomi/näringsliv	1 886
Filosofi	853	Juridik	683
Historia	2 704	Medie-/kommunikationsvetenskap	1 131
Konst	1 650	Psykologi	340
Litteraturvetenskap	2 359	Social/ekonomisk geografi	1 621
Religion	2 957	Sociologi	1 838
Språkvetenskap	2 287	Statsvetenskap	1 557
		Utbildningsvetenskap	1 827
<b>Totalt</b>	<b>14 479</b>		<b>10 883</b>

Tabell 1: SveAk:s sammansättning.

För att summera de tankar som legat till grund för SveAk:s sammansättning (se tabell 1) definierar vi det svenska akademiska skriftspråket som bestående av texter skrivna av och för akademiker, på svenska. Vi har således låtit det representeras av texter från disciplinerna humaniora och samhällsvetenskap, där den svenska akademiska produktionen är tillräckligt hög. Från dessa discipliner har vi sedan valt ämnesområden där förhållandevis många publikationer har funnits tillgängliga som fulltext-pdf:er genom SwePub. De humanistiska ämnesområdena är: *etnologi*, *filosofi*, *historia*, *konst*, *litteraturvetenskap*, *religion* och *språkvetenskap*.

---

4. Alla ordantal i tabellen är angivna i tusental.

*tenskap*, och de samhällsvetenskapliga: *ekonomi och näringsliv, juridik, medie- och kommunikationsvetenskap, psykologi, social och ekonomisk geografi, sociologi, statsvetenskap* och *utbildningsvetenskap*. Sammanlagt består korpusen av drygt 500 texter skrivna av fler än 450 olika författarkonstellationer.

### 3.2. Extraktion av akademiska ord

Till att börja med har allt korpusmaterial automatiskt annoterats med den teknik som används av Språkbanken, vilken bl.a. innefattar tokenisering, ordklasstagning och lemmatisering (Borin et al. 2012b). Därtill har vi valt att utgå från lemmatiserade lexikala enheter i våra beräkningar; med *ord* menar vi alltså grundform inklusive samtliga böjningsformer. Vidare definierar vi *akademiska ord* som typiska för akademiska texter (**nyckelord**), där de är vanligt förekommande (**frekvens**) och jämnt spridda (**dispersion**) oberoende av ämnesområde (**utbredning**). Dessutom ingår de **inte i basordförrådet**.

För att automatiskt kunna extrahera dylika ord ur en akademisk korpus, måste definitionens alla kriterier operationaliseras, dvs. formuleras som regler vilka kan appliceras av ett datorprogram.<sup>5</sup>

---

5. Såväl vår definition av akademiska ord som de tekniska parametrarna i den automatiska extraktionsmodellen har bestämts genom att kombinationer av alla krav som tidigare använts för att extrahera akademiska ord testats och de resulterande listorna manuellt utvärderats. De slutliga kriterierna och tröskelvärdena är alltså de med vilka vi erhöll bäst precision (se vidare fotnot 6 och 7).



För att försäkra oss om att de akademiska orden inte ingår i det svenska basordförrådet har orden i SveAk filterats mot de 1000 mest frekventa orden i en korpus med lättlästa texter – *LäsBarT* (1,1 milj ord, Mühlenbock 2009). Några exempel på ord ur detta basordförråd är: *som, då, exempel* och *språk*.

För att objektivt skatta hur «vanligt» ett visst ord är i SveAk, använder vi ett frekvensmått som tar hänsyn till dispersion. Denna, s.k. *reducerade frekvens* (Savický & Hlaváčová 2002) ligger, enkelt uttryckt, närmare det absoluta frekvensvärdet om ordet är jämnt spritt i korpusen. Vidare räknar vi strikt med relativa frekvenser, för att kompensera för de olika ämneskorpusarnas varierande storlek.

De ord som uppvisar en reducerad frekvens på minst 15 förekomster<sup>6</sup> per miljon ord inom alla ämnesområden räknas som ämnesneutrala, och kvalificerar sig för den sista kontrollen. Denna urskiljer texttypiska nyckelord (Scott 1997) genom att ställa ordens (reducerade) frekvenser i SveAk mot motsvarande värden i en korpus med skönlitterära texter (2,5 milj. ord, Norstedtsromaner från 1999). Det akademiska nyckelordsförhållandet baserat på reducerade frekvenser, kallar vi för *akademiskt index*; ju högre detta värde är, desto mer akademiskt är ordet enligt vår definition. För att kandidera till den slutliga ordlistan måste det akademiska indexet uppgå till minst 1,1<sup>7</sup>.

Slutligen har kandidatlistan rensats manuellt på oönskat brus, såsom förkortningar, textstrukturerande element (som *ii.*) och en del engelska ord, som taggaren inte lyckats identifiera som utländska.

---

6. Valet att sätta tröskelvärdet till 15 förekomster är heuristiskt baserat. 20 förekomster skulle resultera i ett alltför litet antal ord, medan 10 förekomster skulle riskera att öppna för alltför ovanliga eller ämnesspecifika ord.

7. Även detta tröskelvärde är grundat på heuristiska överväganden.

#### 4. Presentation av ordlistan

Det arbete som beskrivs i avsnitt 3 ovan har resulterat i en samling med totalt 655 lexikala enheter. När dessa ordnas enligt fallande akademiskt index hamnar följande ord i topp:

*dock, studie, beskriva, social, enligt, innebära, samt, form, betydelse, fall, begrepp, relation, möjlighet, bild, utifrån skapa, analys, skillnad, utgöra, perspektiv*

Ordklassfördelningen i vår samling liknar i hög grad den i AKL (Paquot 2010). Huvuddelen (42 %) av listan består av substantiv. Vidare utgör verben 26 % och adjektiven 14 % av det totala antalet ord. Listan innehåller också många adverb, hela 8 %. (Jfr t.ex. den allmänspråkliga ordboken *Svensk ordbok utgiven av Svenska Akademien 2009*, som innehåller 68 % substantiv och 2 % adverb.)

Det faktum att vi i vårt arbete tagit fasta på lemman istället för ordfamiljer har tydliga konsekvenser för innehållet i listan. Formellt besläktade ord som *bedöma* och *bedömning*, *diskutera* och *diskussion* samt *omfatta*, *omfattande* och *omfattning* bildar egna uppslagsord. Genom vår metod synliggörs sålunda alla uppslagsord mer och – inte minst – de beskrivs på sina egna premisser. Vårt ställningstagande kan också kopplas till tanken att ordlistan i första hand ska användas vid produktion. Coxhead (2000) menar att bruket av ordfamiljer är befogat med tanke på att psykolingvistiska studier visat att morfologiska relationer mellan ord troligen finns lagrade i det mentala lexikonet. Paquot (2010) konstaterar att Coxheads resonemang håller för att presentera ordfamiljer för receptiva syften.

Däremot är presentationssättet föga meningsfullt vid produktion, då inte alla medlemmar i ordfamiljerna är lika användbara (jfr Gardner & Davies 2013:3f. som förespråkar användning av lemman i pedagogiska lexikala resurser).

Liksom SveAk är den svenska akademiska ordlistan fritt tillgänglig och nedladdningsbar via Språkbanken. Vidare är ordlistan införlivad i Språkbankens lexikala infrastruktur *Karp* (se vidare Borin et al. 2012a). Listans användargränssnitt framgår av figur 1.

Ord	Böjning	Betydelse	Språkprov	Engelska
11 begrepp substantiv	begreppet begrepp begreppen	föreställning, uppfattning	I sin bok från 1994 genomför Derrida en analys av begreppet "demokrati" korpus	concept, conception, idea, notion
12 relation substantiv	relationen relationer relationerna	1. förhållande 2. känslomässigt (ofta sexuellt) förhållande; <även> (formell) förbindelse	I nyare forskning betonas man författarens relation till modernismen korpus	relation, relationship
13 möjlighet substantiv	möjligheten möjligheter möjligheterna	möjlig utväg, tillfälle	Ett flertal kommittéer har tillsatts för att utreda möjligheterna att minska utsläppen korpus	possibility
14 bild substantiv	bilden bilder bilderna	foto, teckning, målning etc.; <även> bildligt> skildring; <även> liknelse	en positiv bild  en heltäckande bild  Mediernas uppgift är att förmedla en rättvisande bild av verkligheten korpus	image, picture

Figur 1. Den akademiska ordlistans användargränssnitt.

I figur 1 återges ett visningsläge där uppslagorden är ordnade efter akademiskt index. Men användarna kan även välja att se orden i alfabetisk ordning.

De 100 översta uppslagorden bildar utgångspunkt för mer traditionella ordboksartiklar. Dessa ord är försedda med upp-

gifter om ordklass, böjning och betydelse, ett eller flera språkexempel samt engelska ekvivalenter. Exempelvis ges följande upplysningar om adverbet *dock*:

**dock** (adverb) (oböjligt), 'i alla fall, ändå, likväl': *Efter ett par rosade novellsamlingar, som dock inte blev några försäljningssuccéer, började författaren att skriva romaner;* however, nevertheless, still, yet

Uppgifterna om ordklass, böjning och betydelse är hämtade från den nyligen uppdaterade *Lexins svenska lexikon* (2011, <<http://lexin.nada.kth.se/lexin/>>). Informationen har tillgängliggjorts via svenska Språkrådet som numera ansvarar för Lexinprojektet. I nuläget återges alla betydelser som anförs i Lexin, även om vissa betydelser torde vara vanligare än andra i akademiska texter. Ett exempel är verbet *uppfatta* som enligt ordboken kan betyda 'förstå, tolka' och 'lyckas höra'. Enligt vår bedömning är det främst den första betydelsen som är aktuell i SveAk. En systematisk granskning av vilken eller vilka betydelser som är vanligare i materialet hade givetvis bidragit till en bättre ordlista, men tyvärr saknades utrymme för ett sådant arbete inom projektets ramar.

Vidare är uppslagsorden försedda med ett eller flera redaktionella språkprov. Dessa är baserade på bruket i SveAk. Användarna kan också (via direktlänkar) klicka sig vidare från artiklarna till korpusen och på så sätt har de tillgång till fler – och autentiska – exempel utöver de enklare i artiklarna. Avslutningsvis är de engelska ekvivalenterna automatiskt hämtade från Lexins engelsk-svenska lexikon som tillhandahållits av Språkbanken.

## 5. Summering och framtida perspektiv

Den akademiska ordlista som presenteras här är tänkt som stöd för dem som behöver hjälp på vägen mot att erövra det svenska akademiska språket. I ordlistans förvalda presentationsform står orden i en ordning, där de mest typiska för akademiskt språkbruk står överst. Det innebär att listan kan rekommenderas till instudering i den ordning orden står.

De engelska översättningsekvivalenterna tillsammans med de svenska betydelseangivelserna gör att listan lämpar sig för såväl andraspråksinlärare som för modersmålstalande med liten erfarenhet av akademiskt språk. Länkningen till SveAk-korpusen ger vidare tillgång till en stor mängd autentiska exempel på hur varje ord kan användas.

Som redan antytts har projektets ekonomiska ramar inte tillåtit att alla uppslagsord försetts med utförligare information. Givetvis är det angeläget att åtgärda denna brist i framtiden. Därutöver kan vi se ett intresse för utbyggnad med återkommande akademiska fraser, i linje med vad som antytts i Carlund et al. (2012). Sett till ordlistans praktiska nytta, kan det finnas skäl att samarbeta med andra forskare vid framtagning av underlag för ordtester och undervisningsmaterial.

Användarstudier kan också bidra till kunskap om vidare utvecklingsbehov. Om det t.ex. skulle visa sig att information om ordfamiljer underlättar ordinläringen, bör ordlistan byggas ut med sådana funktioner.

## Litteratur

- Bauer, L. & P. Nation (1993): Word families. I: *International Journal of Lexicography* 6, 253–279.
- Borin, L., M. Forsberg, L.-J. Olsson & J. Uppström (2012a): The open lexical infrastructure of Språkbanken. I: *Proceedings of LREC 2012*. Istanbul: ELRA, 3598–3602.
- Borin, L., M. Forsberg & J. Roxendal (2012b): Korp – the corpus infrastructure of Språkbanken. I: *Proceedings of LREC 2012*. Istanbul: ELRA, 474–478.
- Carlund, C., H. Jansson, S. Johansson Kokkinakis, J. Ribeck, & J. Prentice (2012): An academic word list for Swedish – a support for language learners in higher education. I: *Proceedings of the SLTC 2012 workshop on NLP for CALL*. Linköping Electronic Conference Proceedings 80, 20–27.
- Coxhead, A. (2000): A new academic word list. I: *TESOL Quarterly*, 34:2, 213–238.
- Coxhead, A. (2011): The academic word list 10 years on: Research and teaching implications. I: *TESOL Quarterly* 45:2, 355–362.
- Gardner, D. & M. Davies (2013): A New Academic Vocabulary List. I: *Applied Linguistics* 4, 1–24.
- Hyland, K. & P. Tse (2007): Is there an "academic vocabulary"? I: *TESOL Quarterly* 41:2, 235–253.
- Jansson, H., S. Johansson Kokkinakis, J. Ribeck & E. Sköldberg (2012): A Swedish academic word list: methods and data. I: R. V Fjeld & J. M. Torjusen (red.): *Proceedings of 15th EURALEX International Congress*. Oslo: University of Oslo, 955–960.
- Johansson Kokkinakis, S., E. Sköldberg, B. Henriksen, K. Kinn

- & J. Bondi Johannessen (2012): Developing Academic Word Lists for Swedish, Norwegian and Danish – a joint research project. I: R. V. Fjeld & J. M. Torjusen (red.): *Proceedings of 15th EURALEX International Congress*. Oslo: University of Oslo, 563–569.
- Mühlenbock, K. (2009): Readable, legible or plain words – Presentation of an easy-to-read Swedish corpus. I: *Multilingualism, Proceedings of the 23rd Scandinavian Conference of Linguistics (Studia Linguistica Upsaliensia 8)*. Uppsala: Acta Universitatis Upsaliensis, 325–327.
- Nation, P. (2001): *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Paquot, M. (2010): *Academic vocabulary in learner writing. From extraction to analysis*. London & New York: Continuum.
- Savický, P. & J. Hlaváčová (2002): Measure of word commonness. *Journal of Quantitative Linguistics* 9, 215–231.
- Scott, M. (1997): PC analysis of key words – and key key words. *System* 25/2, 233–245.
- Sköldbberg, E. & S. Johansson Kokkinakis (2012): A och O om akademiska ord. Om framtagning av en svensk akademisk ordlista. I: B Eaker, L. Larsson & A. Mattisson (red.): *Nordiska studier i lexikografi 11*. Lund: Nordiska föreningen för lexikografi, 575–585.
- Språkbanken*. <<http://spraakbanken.gu.se/>>.
- Svensk akademisk ordlista*. <<http://spraakbanken.gu.se/ao/>>.
- Svensk ordbok utgiven av Svenska Akademien* (2009). Stockholm: Norstedts.
- SwePub*. <<http://swepub.kb.se/>>.
- Wang Ming-Tzu, K. & P. Nation (2004): Word meaning in academic English: homography in the academic word list. I: *Applied Linguistics* 25:3, 291–314.

RIBECK, JANSSON & SKÖLDBERG

Judy Ribeck  
doktorand i språkvetenskaplig databehandling  
judy.ribeck@svenska.gu.se

Håkan Jansson  
doktorand i nordiska språk  
hakan.jansson@svenska.gu.se

Emma Sköldberg  
universitetslektor, docent  
emma.skoeldberg@svenska.gu.se

Inst. för svenska språket, Göteborgs universitet  
Box 200, SE-405 30 Göteborg



## **Från dilation till uppskov. Var ordböcker viktiga i etableringen av ett juridiskt ordförråd?**

*Lena Rogström*

In this paper eight juridical words of Latin origin are examined in order to see if they are included in eight Swedish dictionaries published between 1712 and 1850. The purpose is to see to what extent the dictionaries apply to the general tendency of replacing Latin vocabulary with Swedish, and how this shows in the lexicographic structure.

### **1. Att etablera nya ord via språkbruket**

Projektet *Lexisering bakom kulisserna. En undersökning av lagkommissionens språkliga arbete inför 1734 års lag*, finansierat av Riksbankens Jubileumsfond, har till syfte att kartlägga en del av det svenska ordförrådets etablering och processen bakom etableringen. Undersökningarna tar sin utgångspunkt i de diskussioner som förs i förarbetena till 1734 års lag, vilka finns bevarade i form av protokoll från lagkommissionens arbete, utgivna av Wilhelm Sjögren (1900–1909). I protokollen återges ett antal språkliga förslag om ord och uttryck som man anser vara mindre lämpliga i den nya lagboken och man ger också en del förslag till ersättningsord, som skulle kunna fylla de önskade ordens

funktion. Materialet är intressant på så sätt att det ger en inblick i hur lagkommissionen resonerade rörande ordförrådets innehåll och form både ur ett språknormerande perspektiv, men också ur ett lexiserande perspektiv. (Med lexisering avses här den process i en språkplanerande verksamhet som har till syfte att utöka ett språks ordförråd, se Haugen (1987)).

I protokollen finns även exempel på hur man ville hantera problemet att införa ord som ännu inte var helt etablerade i språket. Strategin verkade vara att bruket blir normerande, vilket visas av följande citat rörande förslaget att byta ut *civil* mot *borgerlig*:

[...] till att der med på vårt eget språk förklara ordet *civile*, och fast det synes wara owant, så kan det blifwa med tiden gått (Sjögren 1900–1909:3:132).

Med historien som facit kan vi konstatera att kommissionens förslag i detta fall inte föll så väl ut, vi använder fortfarande ordet *civil* i svenskan, men förslaget i sig är intressant när man studerar språkpolitiska och språknormerande tillvägagångssätt, eftersom det ger en inblick i hur man tänkte ur ett språknormerande perspektiv. Med utgångspunkt i yttrandet får man också ett bra incitament att undersöka hur de föreslagna utbytesorden beskrivs i samtida språkmaterial, inte minst ordböcker. En jämförelse med de föreslagna orden i protokollen och samtida ordböcker kan ge värdefull kunskap om ordförrådet i ordböckerna och om ordböckernas representativitet för sin samtid. Detta antagande ska nu prövas med utgångspunkt i en lexikografisk närstudie av åtta olika ord från förarbetena till 1734 år lag och deras förslag till ersättningsord samt hur de behandlas i åtta olika ordböcker från 1712 till 1850.

## 2. Äldre lagspråk och det problematiska latinet

Under 1600-talet försöker man flera gånger skapa en enhetlig svensk lag giltig för både stad och land, men det är först med 1734 års lag som man lyckas. Lagkommissionen tillsattes 1686 av kung Karl IX och arbetade alltså i nära 50 år innan lagen stadfästes av riksdagen 1736. Lagen blev en viktig språklig förebild med sitt högtidliga, något ålderdomliga men samtidigt enkla och tydliga språk, och den är tillsammans med Gustav Vasas bibel ett av de verk som betytt mest för etableringen av det svenska skriftspråket.

Under 1700-talet ägnades mycken möda åt att stärka svenskans ställning (se Teleman 2002). En viktig del i arbetet var att utveckla det svenska ordförrådet för att kunna använda svenskan inom alla olika domäner i samhället, inte minst den vetenskapliga, där latinet var *lingua franca*.

För juridikens vidkommande användes svenska och latin för olika syften. Lagarna har alltid varit skrivna på svenska men det juridiska fackspråket var snarast latin, och man ser i protokollen till 1734 års lag att de inblandade personerna växlar obehindrat mellan svenska och latin, ibland i en och samma mening.

Ett av lagkommissionens uppdrag var att försöka ersätta vissa ord av latinskt ursprung med mer inhemskt klingande synonymer. I Rogström (2009) visas hur 50 ord av latinskt ursprung hanteras i protokollen och hur deras ersättningsord förekommer i tre samtida ordböcker: Swedberg (utgiven 2009, dessförinnan endast i manuskript från 1720–30-tal), Serenius (1741) och Sahlstedt (1773). Undersökningen visar att ordböckerna innehåller flertalet av de ersättningsord som lagkommissionen föreslår och att flera av de ursprungliga, latinska, orden

antingen inte finns alls i ordböckerna eller också endast förekommer som synonymer. I Dalins ordbok (1850–55) har förekomsten av de latinska orden ökat och de har en ganska framträdande plats i ordböckerna, ofta som lemman. Detta föranleder en fundering över vad som hänt med orden under mellantiden. Varför har de latinska orden en mindre framskjuten plats i 1700-talsordböckerna när de sedan tydligare framträder i Dalins ordbok? Rimligen har de funnits parallellt hela tiden. Har 1700-talsordböckerna tagit intryck av tidens tankar om ett svenskt ordförråd och helt enkelt undertryckt latinet till förmån för svenskan? Och vad säger oss detta i så fall om ordböcker som representanter för ett visst ordförråd? Min avsikt är att i det följande försöka besvara några av dessa frågor.

### 3. Undersökningens uppläggning och material

De åtta orden valdes ut bland de 50 ord som hade utländskt, för enkelhetens skull kallat latinskt, ursprung. Alla orden förekommer i Dalins ordbok (1850–55) och de flesta känner man igen som fullt gångbara ord än i dag (*agera*, *appellera*, *instans*, *original*, *sterbhus*). Tre av orden är mer ovanliga idag (*citation*, *contumax*, *dilation*). I tabell 1 visas ursprungsorden i protokollen, deras förslag till ersättningar samt vilket ord som kom att användas i 1734 års lag. Orden är stavade med nusvensk stavning.

## FRÅN DILATION TILL UPPSKOV

Protokollen	Förslag	1734 års lag
agera	gå med varannan till rätta ställa/stämman/svara till rätta till rätta gå/föra/ stämman	gå/komma/svara till rätta
appellera	vädja	vädja
citation	stämning	stämning
contumax	motvilligt eller för- fallelost uteblivande	tredskas vara förfallelös ute/ borta
dilation	uppskov	uppskov
instans	domstol	domställe
original	huvudbrev grundskrift	huvudskrift
sterbhus	död mans bo död mans arvingar	där hans bo var som ärvd är den, som död är död mans arvingar i oskift bo

Tabell 1: Ursprungsorden med förslag till ändringar.

Alla orden utom två (*dilation* och *contumax*) finns i Svensk ordbok (2009), men har ibland annan betydelse än i de äldre ordböckerna. En kontroll gentemot tre aktuella juridiska ordlistor (Juridisk ordlista, Melin 2000 och Bergström 2002) visar att orden *agera*, *citation*, *contumax*, *dilation* och *sterbhus* inte finns med där. Några av orden har således ändrat betydelse under årens lopp, från en juridisk till en mer allmän betydelse.

De äldre ordböcker som används i undersökningen av de åtta orden valdes ut för att de är väl undersökta ur ett lexikografihistoriskt perspektiv och på något sätt anses vara viktiga i utvecklingen av svensk lexikografi: Spegel (1712); Swedberg (utg. 2009); Serenius (1741); Lind (1749); Sahlstedt (1773); Möller (1790); Weste (1807); Dalin (1850–55). Som en extra kontroll har också databasen SAOLHist använts. Den består av innehållet i följande upplagor av SAOL: 1874, 1889, 1900, 1923, 1950, 1973, 1986, 1998. Eftersom SAOL i första hand är en ordlista anges oftast bara ordens form och inte alltid betydelsen.

Jag utgår i undersökningen från de åtta orden i protokollen och deras förslag till utbyten, och kontrollerar vilka av orden som används i 1734 år lag, för att klargöra i vilken utsträckning lagkommissionen följde sina egna förslag. Jag har därefter kontrollerat om orden finns i ordböckerna och var i artikelstrukturen de i så fall placerats. I undersökningen noteras i första hand ordens status som lemman, men också om orden förekommer någon annanstans i artikelstrukturen. Jag utgår här från antagandet att ett ord är som mest etablerat om det har lemmastatus.

#### **4. Spelade ordböckerna någon roll?**

Resultaten redovisar respektive ords förekomst i ordböckerna samt ordens placering i respektive artikel.

#### 4.1. Agera / Gå till rätta

Lagkommissionen föreslår att *agera* ersätts av en fras uppbyggd kring huvudordet *rätt*. Typfallet är *gå till rätta*, men även andra förslag ges (se tabell 1). I 1734 års lag används företrädesvis fraserna *gå/komma/svara till rätta*.

I ordböckerna förekommer *agera* som lemma, men har vanligen en annan betydelse än den juridiska som första betydelse. Frasen *gå till rätta* återfinns i sju av de åtta ordböckerna; den saknas bara i den äldsta och minsta ordboken (Spegel). I övriga ordböcker hittas frasen under lemmana *rätt* eller *rätta*.

Lagkommissionen har i detta fall valt en fras som ersättning för *agera*. Ur ett lexikografiskt perspektiv kunde detta ha inneburit problem eftersom det inte alltid är givet under vilket lemma man ska placera fraser. Lagkommissionen har dock valt en fras som var etablerad i äldre lagspråk, *ganga til rätta* (SAOB R4117), vilket nog är förklaringen till den enhetliga placeringen i alla ordböckerna.

#### 4.2. Appellera / Vädja

Förhållandet mellan *appellera* och *vädja* är sådant att *appellera* förekommer som lemma i fyra ordböcker (Lind, Möller, Weste och Dalin) och *vädja* som lemma i sju av ordböckerna och som sublemma i en (Lind). Det verkar alltså som att det föreslagna utbytesordet är vanligare än det latinska. Intressant är att *appellera* förekommer i något högre utsträckning i de senare ordböckerna, och att det idag används igen, både som verb men kanske

oftare i sammansättningen *appellationsdomstol*. I Almqvists ordboksfragment från 1840-talet, utgivet 2013, noterar han i artikeln *appellera* att ”Ordet tillhör den äldre lagstilen”. Uppenbarligen har ordet överlevt i juridiska sammanhang bl.a. i sammansättningen *appellationsdomstol*.

*Vädja* har däremot i högre utsträckning tappat den juridiska betydelsen. Man kan fortfarande *vädja till högre instans*, men annars är det vanligaste ordet i dag att *överklaga*. Här kan man alltså se att det ord som föreslogs bytas ut (*appellera*) till stor del ersattes av *vädja*, som i sin tur försvunnit till förmån för *överklaga*, samtidigt som *appellation* i viss mån fortlevat under hela perioden och fortfarande används.

### 4.3. Citation / Stämning

Substantivet *citation* i betydelsen ’stämning’ finns som lemmat i tre av de äldsta ordböckerna: Spegel, Swedberg och Lind. Hos Spegel finns ordet i «Förteckning på några fransösk ord [...]» och Swedberg hänvisar till Spegel. Detta tyder på att man uppfattar ordet som utländskt. *Stämning* förekommer i ordböckerna tre gånger som lemma och två gånger som sublemma.

*Citation/Stämning* förekommer också som verb i formerna *citera/stämna*. *Citera* förekommer som sublemma två gånger medan *stämna* förekommer som lemma två gånger. I de senare ordböckerna (Weste och Dalin) förekommer ordformerna *stämning* som verbalsubstantiviskt sublemma under *stämna*, vilket visar på en lexikografisk utveckling där avledda ordformer förstås utifrån lemmat.



Detta är det ordpar som ganska tydligt visar en successiv övergång, både från latinet till svenska (*citation* > *stämning*) men också i den lexikografiska utvecklingen där man i allt högre grad anger verbalsubstantivet som ett sublemma under verbet.

#### 4.4. Contumax / Tredska

Detta ordpar är intressant såtillvida att det förslag som lagkommissionen ger (*motvilligt eller förfallelost uteblivande*) inte kommer till så stor användning i lagen. Däremot förekommer *tredska* mycket oftare. *Tredska* är det ord som används i de medeltida landskapslagarna och man undrar varför lagkommissionen inte föreslog det från början i stället för den otympliga frasen.

*Contumax* finns som lemma en gång, i Sahlstedts latinska index, och som lemma två gånger i formen *contumacia* (Serenius och Dalin). Det är däremot vanligare som latinsk ekvivalent, och som förled i sammansättningen *contumaciæ-dom*. En liten lexikografihistorisk utvikning kan ge en förklaring till varför ordet användes i svenska ordböcker, trots att det inte verkar användas i lagen. I Serenius engelsk-svensk-latinska ordbok från 1734 står följande att läsa i artikeln *non-suite*:

NON-SUIT or non-sute [...] V.A. döma, då swaranden blifwer ute (in contumaciam) Item förmå en at öfwer gifwa sin sak. Non-suited or non-suted, Adj. dömd in contumaciam.

Den latinska glosan *in contumaciam* fungerar alltså som komplement till den svenska ordförklaringen *döma, då swaranden blifwer ute* i den engelsk-svenska ordboken. När Serenius sedan skriver sin svensk-engelska ordbok verkar han utnyttja den engelsk-svenska upplagan systematiskt, genom att helt enkelt vända materialet så att det som varit ekvivalenter i den engelsk-svenska ordboken blir lemman i den svensk-engelska (för en utförligare beskrivning, se Rogström 1998). I den svensk-engelska ordboken kan man därför se följande:

Contumacia, s. Döma in contumaciam, To non-suite one. Dömd in contumaciam, Non-suited.

Att Serenius i den svensk-engelska ordboken verkar föredra det latinska ordet som lemma tyder på att det ändå kunde förstås av läsarna, vilket i sin tur torde bero på att latinkännenheten var hög hos dessa. Ordet var troligen levande i en juridisk kontext, om än inte i den svenska lagboken. Ytterligare ett tecken på detta är att ordet lexikaliserats i sammansättningen *contumaciæ-dom*.

Det äldre svenska ordet *tredska* finns som lemma i alla de undersökta ordböckerna och det är också väl företrätt som sammansättningsled (*tredsko-dom*). Ordparet är ett tydligt exempel på att man kunde använda olika ord i olika juridiska kontexter.

#### 4.5. Dilation / Uppskov

*Dilation* verkar inte ha haft så hög frekvens i svenskan överhuvudtaget vid denna tid, eftersom det bara återfinns som lemma i

två av de undersökta ordböckerna: Sahlstedt och Dalin. Hos Sahlstedt står det i hans index över latinska ord. Varför Dalin införlivar det i sin ordbok är mer oklart.

*Uppskov* är däremot etablerat som lemma hos alla lexikograferna och var troligen det normala ord man använde i juridiska sammanhang. Det är fortfarande så i modern svenska.

#### 4.6. Instans / Domstol

Förhållandet mellan orden i detta ordpar verkar vara sådant att *domstol* tas över av *instans* ganska sent, och idag har de olika betydelser. Svensk ordbok (2009) definierar *instans* som: *beslutsorgan på viss nivå inom rättsväsende*, medan *domstol* har följande definition: *auktoriserad offentlig inrättning för rättsskipning*. Betydelseutvecklingen hos orden återspeglar utvecklingen av svenskt rättsväsende i början av 1800-talet med 1809 års regeringsform som ett viktigt exempel (Modéer 2010).

*Instans* förekommer inte i någon av ordböckerna från 1700-talet; däremot förekommer *domstol* hos alla 1700-talslexikograferna. Weste har både *instans* och *domstol* som lemman liksom Dalin.

#### 4.7. Original / Huvudskrift

Förslaget att ersätta *original* med *huvudbrev* eller *grundskrift* vann inte gehör i 1734 års lag. Där använder man i stället *huvudskrift*. Detta ord förekommer dock endast i en av de

undersökta ordböckerna, nämligen Weste. *Original* var troligen så fast etablerat att det var svårt att ersätta, vilket bekräftas av de övriga ordböckerna som har ordet som lemma fyra gånger och som sublemma en gång. Intressant att notera är att SAOL-Hist har ordet *huvudskrift* i alla upplagorna. Det är dock svårt att veta i vilken betydelse ordet användes. Troligen är det i den betydelse som SAOB anger som den första: 'den förnämsta skriften i en författares produktion' (H1568).

#### 4.8. Sterbhus / Dödsbo (död mans bo)

*Sterbhus* förekommer som lemma fyra gånger (Sahlstedt, Mölle, Weste, Dalin) medan ersättningsförslaget *död mans bo* inte återfinns alls, vare sig i lagboken eller ordböckerna. Det idag vanliga ordet *dödsbo* finns inte heller, vilket förklaras av att det enligt SAOB har ett förstabelägg från 1874.

Man kan fråga sig varför *sterbhus* inte förekommer mer i ordböckerna, inte minst med tanke på att det än idag kan förekomma i vissa sammanhang.

### 5. Har ordböckerna spelat någon roll?

De åtta ordparens förekomst och utveckling går att följa i ordböckerna, och man kan ofta koppla ordens etableringsgrad i

språket till artiklarnas lexikografiska struktur. De ord som varit vanligast och som överlevt fram till nu har en mer framträdande plats i ordböckerna genom att de oftare förekommer som lemman och det är också dessa ord, eller former av dem, som finns i juridiskt fackspråk idag.

En sammanställning över ordmaterialet och förhållandet mellan orden i ordparen visar en viss övervikt för ersättningsorden som lemman i ordböckerna (se tabell 2). De klara undantagen är *original* och *sterbhus*. *Dra inför rätta*, som gick att hitta i alla ordböckerna, återfinns inte i Svensk ordbok (2009), och som juridisk term används det bara i muntligt språk idag.

Resultaten av undersökningen visar att de ord som förekommer i de flesta ordböckerna, och därmed kan antas ha haft starkast förankring i språket, också är de ord som används idag (*uppskov*, *tredska*, *dra inför rätta*, *överklaga*). I några fall har orden ändrat eller utvecklats en ny betydelse: *instans/domstol*, *appellera/vädja*, *agera*, *citation/stämning*. Detta indikerar att de undersökta ordböckerna troligen är ganska representativa för sin

Ursprungord	Ersättningsförslag	Förekomst som lemman
agera	dra inför rätta	0/7
appellera	vädja	4/7
citation	stämning	3/3
contumax	tredska	3/8
(contumaciæ-	tredsko-	4/3)
dilation	uppskov	2/8
instans	domstol	2/6
original	huvudskrift	4/1
sterbhus	dödsbo	4/0

Tabell 2: Lemmafördelning mellan ursprungs- och ersättningsord.

tids språk och att ordböcker spelade en roll i etableringen och förankringen av ordförrådet.

I några fall har strävan att ersätta de utländska orden inte lyckats fullt ut (*appellation; akt*) och i ett fall har det egentligen varit det latinska ordet som gällt hela tiden (*original*), även om man i 1734 års lag föredrog ersättningsordet *huvudskrift*. Här kan man konstatera att ordböckerna troligen i högre grad skildrar det samtida bruket.

Undersökningen i fråga är liten men ger ändå vissa uppslag till fortsatt forskning om hur ord kan etableras och utvecklas i ett språk. En fruktbar väg att gå skulle kunna vara att undersöka just ordpar eller kluster av ord med liknande betydelse och hur de utvecklas (jfr Ruthström 2002). Det enskilda ordets utveckling står ju ofta att finna i SAOB men hur ord utvecklas i förhållande till varandra kan ge värdefulla insikter i samspelet mellan enheter i ett ordförråd och lexisering av enskilda ord.

## Litteratur

### Ordböcker

- Almqvist, C.J.L. [1842-44] (2013): *Ordbok öfver svenska språket i dess närvarande skick*. Utg. av Sven-Göran Malmgren. (C.J.L. Almqvist, Samlade Verk 17.) Stockholm: Svenska Vitterhetssamfundet.
- Bergström, Sture (2002): *Juridikens termer*. Stockholm: Almqvist & Wiksell.

- Dalin A.F. (1850–55): *Ordbok öfver svenska språket*. Stockholm. *Juridisk ordlista* <[www.domstol.se](http://www.domstol.se)> (augusti 2013)
- Lind, Olof [1738] (1749): *Teutsch-Schwedisches und Schwedisch-Teutsches Lexicon*. Stockholm.
- Melin, Stefan (2000): *Juridikens begrepp*. Uppsala: Iustus.
- Möller, J.G.P (1790): *Teutsch-Schwedisches und Schwedisch-Teutsches Wörterbuch*. Greifswald.
- Sahlstedt, Abraham (1773): *Swensk Ordbok*. Stockholm.
- SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien* (1898–): Lund: Gleerups.
- SAOLHist.<[http://spraakdata.gu.se/saolhist/saolhist\\_avancerat.phtml](http://spraakdata.gu.se/saolhist/saolhist_avancerat.phtml)> (augusti 2013)
- Serenius, Jacob (1734): *Dictionarium Anglo-Svethico-Latinum*. Hamburg.
- Serenius, Jacob (1741): *Dictionarium Suethico-Anglo-Latinum*. Stockholm.
- Spegel, Haqvin (1712): *Glossarium Sveo-Gothicum Eller Swensk Ordabook*. Lund.
- Swedberg, Jesper (2009): *Swensk Ordabok*. Utgiven efter Uppsala-handskriften, med tillägg och rättelser ur övriga handskrifter, av Lars Holm. Stiftelsen för utgivande av Skarmissalet. Skara.
- Svensk ordbok utgiven av Svenska Akademien* (2009). Stockholm: Norstedts.
- Weste E.W. (1807): *Svenskt och Fransyskt lexicon (= Parallèle des langues Française & Suédoise*. Tome III). Stockholm.

## Annan litteratur

1734 års lag = *Sveriges rikets lag gillad och antagen på Riks-*

- dagen år 1734*. [1984] Skrifter utgivna av Institutet för Rättshistorisk forskning grundat av Gustav och Carin Olin. Rättshistoriskt bibliotek. Lund.
- Haugen, Einar (1987): *Blessings of Babel. Bilingualism and Language Planning*. Berlin: Mouton de Gruyter.
- Modéer, Kjell-Å. (2010): *Historiska rättskällor i konflikt: en introduktion i rättshistoria*. Stockholm: Santérus.
- Rogström, Lena (1998): *Jacob Serenius lexikografiska insats. Meijerbergs arkiv för svensk ordforskning. 22*. Göteborg.
- Rogström Lena (2009): Språkplanering i praktiken? En lexikalisk studie av förarbetena till 1734 års lag. I: *Språk & Stil 19*, 5–34.
- Rogström, Lena (2011): I akt och mening. En jämförelse av ordförrådet i 1734 års lag med samtida ordböcker. I: *Studier i svenska språkets historia 11*. Acta Academiae Regiae Gustavi Adolphi CXIII, 189–199.
- Ruthström, Bo (2002): *Land och få. Strukturellt-rättsfilosofiska studier i fornnordiskt lagspråk över beteckningar för egendom i allmänhet med underkategorier*. (Nordlund 23. Småskrifter från Institutionen för Nordiska språk i Lund.) Lund.
- Sjögren, Wilhelm (1900–1909): *Föarbetena till Sveriges rikes lag: 1686-1736 : efter offentligt uppdrag / utgifna af Wilhelm Sjögren*. Uppsala.
- Teleman, Ulf (2002): *Åra, rikedom och reda. Svensk språkvård och språkpolitik under äldre nyare tid*. Stockholm: Norstedts ordbok.



FRÅN DILATION TILL UPPSKOV

Lena Rogström  
universitetslektor fil. dr.  
Institutionen för svenska språket  
Göteborgs universitet  
Box 200  
SE-405 30 Göteborg  
[lena.rogstroem@svenska.gu.se](mailto:lena.rogstroem@svenska.gu.se)

## **Fypike gjorde selbureis i røytemåned. Om *Norsk-russisk ordbok* av Vladimir Arakin: utvalg, kilder, målgruppe og formål**

*Aud Anna Senje*

*Norwegian-Russian Dictionary* by Vladimir D. Arakin (1963) was the first Norwegian-Russian dictionary of its kind, and for 40 years it was the only one. The dictionary contains approximately 77 000 entries, but it is strikingly inadequate. In addition to insufficient or positively misleading translations, it contains a certain amount of dialectal words, and archaic, obsolete or peculiar words, uncommon in Norwegian everyday speech. In fact, until 2003 (when a more reliable Norwegian-Russian dictionary was published), Norwegian students of Russian were strongly advised not to use Arakin's dictionary. Instead, they had to use Swedish, Danish or English dictionaries. The source language in Arakin's dictionary is outdated, whereas the target language pretends to be in keeping with its time. Thus, the dictionary is unsuitable as a practical tool for learners of Norwegian and Russian. Apart from the curiosities, the author has included highly specialized terminology. In this field he was certainly ahead of his time, as there is still a great need for specialized dictionaries between Norwegian and Russian. For all its deficiencies, Arakin's dictionary has still obtained somewhat of a cult status among Slavists in Norway. Who was Arakin? For whom and for what purpose did he write his dictionary? Why did he include so many peculiar words, and where did he find them? In this paper I have tried to find answers to these questions.

## 1. Ordbok i særklasse

I 2013 er det 50 år siden Vladimir Arakins *Norsk-russisk ordbok* ble utgitt i Moskva. Det gir god anledning til å børste støv av dette usedvanlige verket. Det dreier seg om en allmennordbok, som er så spekket med rariteter at den burde ha vært utstyrt med bruksanvisning.

Dette var den første norsk-russiske ordboken i sitt slag, og i 40 år var det den eneste som fantes. Boken er omfattende, men likevel så mangelfull at studenter ved norske universiteter ble advart mot å bruke den. I stedet måtte de bruke svenske, danske og engelske ordbøker. I tillegg til misvisende oversettelser (eksempel: *høyhus* er oversatt med «høylåve») inneholder Arakins ordbok en mengde ord som sjelden eller aldri opptrer i norsk dagligtale, for eksempel *bygdestripper*, *damebris*, *gråtegauk*, *karnøffel*, *kritikakleri*, *luskemikkel*, *manntimmer*, *oksepannet*, *snilefisk*, *søplenisse*. Med alle sine mangler og merkverdigheter har ordboken nærmest fått kultstatus i slavistmiljøet. Hvem var egentlig Arakin? Hvem var ordbokens målgruppe, og hva var formålet? Hvilke kilder har Arakin brukt? Hadde han noen norske hjelpere i arbeidet? Sist, men ikke minst: Hva kan være grunnen til at han tok med så mange underlige oppslagsord? Disse spørsmålene skal jeg forsøke å besvare i dette innlegget.

## 2. Forfatter, målgruppe og formål

Vladimir Dmitrijevitsj Arakin (1904–1983) var professor i filologi ved det statlige pedagogiske universitetet i Moskva. Han har hatt stor betydning for utviklingen av språkvitenskap og metodikk for undervisning i fremmedspråk i Russland. Hovedfeltet hans var engelsk, men han fordypet seg også i skandinavisk språkhistorie og i afrikanske og asiatiske språk. Han utga blant annet flere ordbøker og et fembinds læreverk i engelsk, som ennå brukes ved flere universiteter og pedagogiske høyskoler i Russland. Mange treff på nettet tyder på at bøkene hans fremdeles er i omløp.

Av kolofonteksten og forordet i *Norsk-russisk ordbok* framgår det at målgruppen er personer som leser og oversetter norsk skjønn- og faglitteratur, tidsskifter og aviser, pluss studenter og alle andre som lærer seg norsk. Formålet er å sette leserne best mulig i stand til å forstå norsk litteratur (i vid forstand). I kolofonteksten står det at ordboken inneholder ca. 77 000 ord «fra moderne norsk». Videre heter det at en viss mengde foreldede ord er tatt med i ordboken fordi de er nødvendige for å kunne lese norsk litteratur fra 1800-tallet.

## 3. Eksempler på oppslagsord

Blant oppslagsordene i Arakins ordbok finner vi både fagtermer, gammeldagse ord og dialektuttrykk. Spesielle termer har sin

naturlige plass i en fagordbok. Men der fagordbøker ikke finnes, vil en ansvarlig allmennordbokredaktør se det som sin oppgave å dekke i det minste noen områder, og det gjorde da også Arakin. Da hans ordbok kom ut, fantes det nemlig ingen fagordbøker mellom norsk og russisk, og i dag er det fremdeles flere viktige fagområder som ikke er dekket. Arakin opplyser at mye plass er viet teknisk terminologi. Størst vekt har han lagt på områder som var sentrale for Norge, for eksempel fiske, hvalfangst, sjøfart og skogbruk. Også særegne tekniske termer er tatt inn, som *pudlejern*, *snøfteventil*, *sporlysprosjekt*, *taljeheis*, *tinnaske* og *treeddik*. Ellers later Arakin til å ha vært svært opptatt av botanikk og zoologi. For eksempel har han tatt med både *grepplyng*, *sprikebygg*, *stankkarse* og *tiggersoleie*, og han kan skilte med en imponerende billesamling. Bare i det første bindet av den søkbare utgaven opptrer minst ni stykker, bl.a. *bløtbille*, *borebille*, *brødbille* og *glansbille*.

Men det er ikke først og fremst fagtermene som gjør Arakins ordbok så spesiell – det er alle de sære ordene og uttrykkene som er forsynt med ufullstendige, tvilsomme eller stilistisk nøytrale oversettelser, noe som kan forlede russiske brukere til å tro at ord som *gaustadkandidat*, *hjerneflue*, *selbureis* og *uværsågod* inngår i et dagligdags norsk ordforråd.

### 3.1. Fypike, røytemåned og selbureis

La oss først se nærmere på ordene i tittelen på dette innlegget. *Fypike* er definert som «lettferdig pike, gatepike» i *Norsk Riksmålsordbok*. Arakins oversettelse (*пазг. девица легкого поведения, уличная девка*) er dekkende, men ordet er neppe mye

brukt i moderne tid. *Røytemåned* er «en svært varm måned» ifølge Arakins knappe oversettelse (*дуал.* жаркий месяц). *Røyte* har i denne sammenhengen ikke noe med hårfelling å gjøre. Forklaringen i *Norsk Riksmålsordbok* er hundedagene (23. juli–23. august, da maten lett råtner på grunn av varmen). En av forklaringene på *røyte* i *Nynorskordboka* er 'tid med rått vær': *ei bløyte gjer lite, ei røyte er verre / tømmeret låg ute i regn og røyte.*

Oppslagsordet *selbureis* forekommer i uttrykket «gjøre selbureis», og den konkrete betydningen er selvfølgelig å reise til Selbu, som er en kommune i Sør-Trøndelag fylke, et stykke fra Trondheim, men her er det den overførte betydningen som gjelder. Arakin har oversatt uttrykket med «føde, sette til verden» (*дуал.* рожать, производить на свет). Det er et dialektuttrykk med begrenset bruksområde, men det er like fullt oppført i *Norsk Riksmålsordbok*:

**selbu|reis**, en, dial., i forb. *gjøre selbureis* (egtl. «reise til Selbu») gjøre barsel (*Egge, Solst.*, 198).

Vi kan finne uttrykket i eldre litteratur, for eksempel i romanen *Hansine Solstad* av Peter Egge (1925): «Går brølløpsreisen til Tordenskjoldsgata eller Enskillingsveita? – Eller kanskje du gjør ei lita Selbureis først?» *Selbureis* er forsynt med fotnote: «Gjøre selbureis (en lokal vittighet) – gjøre barsel.» Uttrykket skal visstnok komme fra Trondheim. Det kan bety å reise til et avsides sted (Selbu) for å føde i dølgsmaal, utenfor ekteskap. Nokså utbredt er også betydningen «abortere», ifølge *Trønderordboka* (1997):

**selbureis** f, *gjera selbureis* få barn; føde (vanleg i større delar av Trl), stundom særleg om fødsel i løynd el utafør

ekteskap. Nokså utbreidd er også tydinga abortere. Ein teori om opphavet er at rikmannsdøtrer frå Trondheim som hadde komme i «ulykka», vart sendt til Selbu for å føde, for at ikkje skandalen skulle bli så synleg. Ein annan teori er at reisa til Selbu var så vanskeleg og strabasios i gamle dagar at ho kunne bli jamført med ei barselferd.

Riktignok har Arakin markert uttrykket som *dialektalt* (*дуал.*), men russiske lesere har neppe forutsetning for å skjønne at det ikke er gangbart over hele Norge.

### 3.2. Utdatert slang og hapax legomena

Et annet eiendommelig oppslagsord er *gaustadkandidat*. Ordet er knyttet til Gaustad sykehus i Oslo, det første statlige sykehus i Norge for behandling av mennesker med alvorlige psykiske lidelser (åpnet i 1855). *Norsk Riksmålsordbok* har denne definisjonen:

**gaustadkandidat**, en, fam., person moden for sinnssykeasylet (Gaustad): *spekulanter; idioter; gaustadkandidater; rævepels* (Heib. III, 188).

Man får flere treff på ordet ved søk i Nasjonalbibliotekets database. Den tidligste forekomsten av ordet er trolig i romanen *Trøtte Mænd* av Arne Garborg (1891): «Jeg [...] vil altså komme, for såvidt ingen anden Gaustadkandidat lægger beslag på mig» (s. 135 i 1970-utgaven).

Som en kuriositet kan nevnes at gaustadkandidaten dukker opp i en norsk oversettelse av en amerikansk roman fra 1918

(pussig nok om en reise i Russland): *The Village : Russian Impressions* av Ernest Cook Poole (1918), utgitt på norsk i 1920 med tittelen *Tarasovs Landsby : fra Rusland 1918*. Den ukjente oversetteren har gjengitt «chatters like a babbling fool» med «skravler som en gaustadkandidat», så ordet må ha vært i bruk. Det er nok et eksempel på slang som har blitt borte. Ordet er oppført ikke bare i Arakins ordbok, men også i eldre norsk-franske ordbøker, der det er oversatt med «charentonesque, fou à mettre à Charenton»<sup>1</sup> og «échappé de Bicêtre»<sup>2</sup>. I en eldre norsk-tysk ordbok er det oversatt med «Tollhäusler»<sup>3</sup>. Arakin har oversatt det med «person som er moden for sinnssykeasylet» (*разг.* кандидат в сумасшедший дом) og tatt med en kort forklaring om Gaustad sykehus, der han omtaler Gaustad som en norsk by.

I tillegg til slangord og dialektuttrykk har Arakin tatt med noen såkalte hapax legomena, dvs. ord med få eller bare ett literært belegg, som *hjerneflue* (Welhaven) og *uværsågod* (Hamsun). *Hjerneflue* er oversatt med «fiks idé» (навязчивая идея, идефикс). *Uværsågod* er oversatt med «uelskverdig, lite forekommende» (нелюбезный, непредупредительный), og det er ikke forsynt med noen annen markør enn *mundlig* (*разг.*). Ordet er laget av Knut Hamsun. Vi finner det i romanen *Markens grøde* (1917), der hovedpersonen undrer seg over at hans kone varsler om middagen på en «forbandet tvær og uværsaagod maate». *Uværsågod* finnes i *Norsk Riksmålsordbok* og har dessuten har funnet veien inn i *Norsk-dansk ordbok* (1955). I *Kryssordexperten* (1997) er det oppført som synonym til *uelskverdig*.

---

1. Skougaard, Løseth (1940): *Norsk-fransk ordbok*.

2. Jacobsen, J. (1941): *Norsk-fransk ordbok*.

3. Brynildsen, J. (1926): *Norsk-tysk ordbog*.



### 3.3. Faste uttrykk med løs konsistens

I mange av ordbokartiklene har Arakin tatt med det han kaller vanlige ordforbindelser og faste uttrykk. Men det han skriver, kan ikke alltid tas for god fisk. Under *laks* har han for eksempel ført opp uttrykket «du skal ha takk for laksen!», med forklaringen «du skal ha takk» (тебя поблагодарят, тебе скажут спасибо). Uttrykket er hentet fra Alexander Kiellands roman *Skipper Worse* (1882), og det står i *Norsk Riksmålsordbok* med forklaringen «nei mange takk, nei vet du hvad!». Vi finner det også i *Norsk-dansk ordbog* (1955).

Ordet *tetuadunk* opptrer angivelig i uttrykket «Hold tetuadunken din!». «Ti stille!» heter på fransk «Tais-toi!». Uttrykket kan bygges ut til «Tais-toi, donc!», altså «Ti stille, da!», så det kan tenkes at *tetuadunk* har fransk opphav. Eventuelt har noen blandet det sammen med *tetut*, som i tillegg til «tut på tekjele» kan bety «trut». Uttrykket står i *Norsk Riksmålsordbok* med referanse til en utgave av Dagbladet i 1939. Arakin har dristet seg til oversettelsen «hold kjeft» (заткни плотку!, замолчи!), med markøren *grovt* (груб.).

## 4. Arakins kilder og hjelpere

Selv skriver Arakin at han i arbeidet med ordboken har brukt klassikere i norsk litteratur, verker fra norsk samtidslitteratur, aviser og tidsskrifter. Kildene hans er ellers hovedsakelig tosp-

SENJE

råklige allmennordbøker (svensk-russisk, dansk-russisk, engelsk-russisk) og fagordbøker. Blant de tospråklige nevnes *Norsk-russisk ordbok* av R. Jankov (1959), som er mer en ordliste enn en ordbok i tradisjonell forstand.

I denne sammenhengen er det de enspråklige norske kildene som har størst interesse. *Norsk Riksmålsordbok* (bb. IIIV, 1937/1959) er åpenbart en viktig kilde. Arakin har også benyttet *Norsk slangordbok* av Ulf Gleditsch (1952). Enkelte oppslagsord finner man bare igjen der, f.eks. *snilefisk* («vemmelig type») og *søplenisse* («lasaron»).

Arakin skriver videre at han har måttet fordype seg i flere fagfelt for å få til en nøyaktig oversettelse av tekniske termer. I tillegg nevner han at han har konsultert norskkyndige russere og navngitte russiske spesialister på en del fagområder, blant annet zoologi og botanikk. Han har også rådspurt spesialister på maritim, teknisk og militær terminologi.

Det ser ut til at ingen nordmenn har vært involvert i arbeidet med ordboken. Arakin har i hvert fall ikke navngitt noen. Han nevner en norsklærer med et ikke-norsk etternavn, det er det hele.

## 5. Bakgrunnen for utvalget av oppslagsord

Vi vil aldri få vite om noen av de mest spesielle oppslagsordene kunne ha blitt luket ut av manuset dersom Arakin hadde hatt norske hjelpere. Kildelisten, kolofonteksten og forordet bringer oss litt nærmere en forklaring på det spesielle utvalget av oppslagsord. I forordet gjør forfatteren rede for de utfordringene

han støtte på under arbeidet. Den viktigste av dem er den norske språksituasjonen med to målformer og den store valgfriheten i rettskrivning og bøyningsformer. Han har basert seg på bokmål, men bemerker at ordtilfanget i bokmål ikke er stabilt, og at det stadig er under påvirkning fra «landsmålet» og talespråket i byene. Dessuten viser han til at mange norske forfattere gjør bruk av dialektord i verkene sine. Han har derfor sett seg nødt til å ta med «en betydelig mengde» slike ord, i tillegg til de fordelde ordene han nevner i kolofonteksten. At enkelte ord har få litterære belegg, synes han ikke å ha reflektert over, og noen søkemulighet i moderne forstand hadde han jo ikke. Dermed har han neppe vært i stand til å undersøke i hvilken grad de utvalgte ordene var i bruk.

De aller fleste av Arakins merkverdigheter er faktisk å finne i *Norsk Riksmålsordbok*, så forfatteren må ha vært i god tro da han valgte dem. Men han har nok tillagt flere av ordene ufortjent stor vekt. Kanskje han bare hadde lest noen få verker (litteraturpensum på sovjetiske universiteter?) og trukket den konklusjonen at ordene han påtraff der, var en naturlig del av norsk dagligspråk. Etter sigende var han aldri i Norge, og det er uvisst om han i det hele tatt hadde noe med nordmenn å gjøre. Hans kjennskap til norsk kan ha vært rent teoretisk, slik at han ikke har vært i stand til å skille mellom det vesentlige og det uvesentlige.

## 6. Arakin i ny drakt

I 1998 ble en tobindsutgave av Arakins ordbok publisert i

Moskva. På omslaget er forfatterens navn utelatt, og boken har fått den nye tittelen *Stor norsk-russisk ordbok*, selv om den neppe er større enn før. Det oppgitte antallet oppslagsord og ordforbindelser (ca. 200 000) er mer enn dobbelt så høyt som i førsteutgaven (77 000), men dét er nok helst et resultat av en annen tellemåte. Opplaget er bare 1550, mot førsteutgavens 12000. Tobindsutgaven har ISBN-nummer, det hadde ikke 1963-utgaven. Utgiveren er et privat forlag, i 1963 ble boken utgitt av «Statsforlaget for ordbøker på fremmede språk og nasjonalitetsspråk».

Kolofonteksten i 1998-utgaven er endret: «Ordboken egner seg for lesning og oversettelse av skjønnlitteratur, avistekster, populærvitenskapelige tekster og fagtekster. [...] Ordboken er beregnet på oversettere, forskere og alle som driver med norsk språk.» Det opprinnelige forordet er litt forkortet og har ellers få endringer. Utgiverne har forsynt boken med et nytt og moderne forord nummer to, der målgruppe og formål er noe tilpasset. Blant annet heter det at ordboken utgis på grunn av økt kontakt mellom Norge og Russland og økende interesse i Russland for norsk språk (noe som vel skulle tilsi et større opplag enn 1550?). Kildelisten er interessant nok utelatt. Det står ikke noe om at ordtilfanget er foreldet, eller om hvilke endringer som angivelig er gjort i denne utgaven. Den gir seg nemlig ut for å være «ny og revidert», men stikkprøver viser at raritetene er intakte, så det dreier seg trolig om et rent opptrykk. At tobindsutgaven ikke er oppdatert med moderne ord, kan man enkelt konstatere ved å lete etter f.eks. sammensetninger med *olje-* eller *data-*.

## 7. Har Arakins ordbok noen verdi i dag?

Arakins *Norsk-russisk ordbok* var nærmest enerådende på markedet i 40 år. Først i 2003 ble den avløst av Kunnskapsforlagets *Stor norsk-russisk ordbok*. Arakins ordbok må brukes med forsiktighet, og den krever gode kunnskaper i både norsk og russisk. Kildespråket er utdatert, mens målspråket pretenderer å være i takt med samtiden. Ordboken er dermed uegnet som praktisk verktøy for nordmenn som vil lære russisk, og for russere som vil lære seg moderne norsk. I dag har den liten eller ingen verdi som oppslagsverk, og den er neppe i salg. Men så sent som i 2007 lå den faktisk på utlånstoppen ved Sør-Varanger bibliotek i Kirkenes. Det er slett ikke sikkert at brukerne er klar over hvor upålitelig den er.

Det er fristende å latterliggjøre Arakin for det sære utvalget av ord, men intensjonen hans var god. For det første tok han ansvar for fagterminologien. Han visste at det ikke fantes fagordbøker mellom norsk og russisk, derfor valgte han å ta med en rekke spesialord som strengt tatt ikke hører hjemme i allmennordbøker, men som han antok at russiske brukere ville ha nytte av. For det andre ønsket han å hjelpe russiske lesere til å forstå norsk litteratur, derfor tok han med både dialektord, slang og utdaterte ord. For det tredje må vi huske på at ordbokens russiske målgruppe på den tiden hadde liten eller ingen kontakt med nordmenn. Daglig kommunikasjon med utlendinger var neppe aktuelt for sovjetborgere, og det var så godt som umulig for dem å krysse landegrensene. Arakins verk er altså en ordbok for folk som leser, ikke for folk som snakker sammen. Det viktigste er å forstå, ikke å uttrykke seg. Aleksej Perminow, tidligere førsteamanuensis ved Universitetet i Oslo, har omtalt ordboken som et produkt av den kalde krigen.<sup>4</sup>

SENJE

Med *Stor norsk-russisk ordbok* (2003) har vi fått et godt og pålitelig verktøy, slik at russere og nordmenn i dag kan studere hverandres språk uten å snuble i søplenisser og snilefisker. Likevel kan Arakins ordbok fortsatt glede oss, om den ikke kan tilføre oss kunnskap, og den kan faktisk utvide vårt norske ordforråd. Den kan gi oss det som i politiet kalles overskuddsinformasjon, altså opplysninger som man kommer over mens man egentlig leter etter noe annet, opplysninger som kan komme godt med i en annen sammenheng. For spesielt interesserte kan boken være et oppkvikkende supplement til andre ordbøker. Dette er en ordbok som har vært med på å krydre ordforrådet til generasjoner av norske slavister, så den er mer enn et tidsbilde og en kuriositet. Har man først blitt kjent med den, blir den lett en hjerneflue, og man kan ende som gaustadkandidat. Arakins ordbok er en skatt som ikke fortjener å bli glemt.

## Litteratur

### Ordbøker

- Arakin, V. D. (1963): *Norsk-russisk ordbok*. Moskva: Statsforlaget for ordbøker på fremmede språk og nasjonalitetsspråk.  
Arakin, V. D. (1998): *Stor norsk-russisk ordbok* (2. rev. utg.). Moskva: Levende språk.

---

4. Forsningsmagasinet *Apollon* 30.4.2008.

- Brynildsen, J. (1926): *Norsk-tysk ordbog*. Oslo: Aschehoug.  
Christiansen, H. (1955): *Norsk-dansk ordbog*. Oslo: Gyldendal.  
Danielsen, D. (1997): *Kryssordexperten*. Oslo: Orion.  
Gleditsch, U. (1952): *Det får 'n si. Norsk slangordbok*. Oslo: Nasjonalforlaget.  
Jakobsen, J. (1941): *Norsk-fransk ordbok*. Oslo: Gyldendal.  
Jenstad, T. E. & A. Dalen (red.) (1997): *Trønderordboka*. Trondheim: Tapir.  
*Norsk Riksmålsordbok* (1937–1957) (digital utg.) <[www.ordnett.no/butikk/norskeordboker.norskriksmålsordbok](http://www.ordnett.no/butikk/norskeordboker.norskriksmålsordbok)>  
*Nynorskordboka* (2006). Oslo: Det Norske Samlaget.  
Skougaard, J. & E. Løseth (1940): *Norsk-fransk ordbok*. Oslo: Aschehoug.

## Annen litteratur

- Ordbøker stemoderlig behandlet (30.4. 2008). I: *Apollon* (digital utg.) <<http://www.apollon.uio.no/artikler/2008/spraak-ordboker.html>>

Aud Anna Senje  
seniorrådgiver, cand.philol.  
Språkrådet  
Postboks 8107 Dep  
NO-0032 OSLO  
aas@sprakradet.no

## **Brugerne er allerede mobile!**

*Henrik Køhler Simonsen*

Smartphones have become an increasingly popular and indispensable personal device – also for medical doctors. Ten medical doctors were asked to look up five medical terms, and empirical data were collected during three tests, a think-aloud-session and interviews. The purpose of the empirical survey was to analyze the characteristics of the mobile user situation. On the basis of the data it is argued that modern dictionary app development should be based on a calibration between the nature of the mobile user situation, the mobile user, the mobile access methods, the information task, the nature and limitations of the mobile data and the information need of the user.

### **1. Indledning og problem**

Et stigende antal mennesker bruger avancerede smartphones, og tallet er støt stigende. Vi bruger smartphones til at navigere med, søge efter information med, se videoer med, bruge apps og tilgå sociale netværk med. Undersøgelser viser endda, at 83 % af os bruger vores smartphone, mens vi laver andre ting, jf. Google (2013:2).



Brugerne er allerede fuldt mobile og kan søge efter information uafhængig af deres fysiske placering, men spørgsmålet er, om leksikografien i tilstrækkelig grad er blevet mobil og har tilpasset sig det mobile medium, den mobile brugssituation og den mobile informationstilgang.

Formålet med denne artikel er på basis af en empirisk undersøgelse med otte læger at diskutere den mobile, punktuelle brugssituation, som her defineres som et «kort, usystematisk og tidsmæssigt afgrænset opslag ved hjælp af en mobiltelefon», jf. i øvrigt Tarp (2006:57-60), som diskuterer punktuelle og systematiske brugersituationer. Artiklen har til hensigt at diskutere hvorvidt og på hvilken måde der er forskel på en stationær, punktuelt informationstilgang (når brugeren er stillesiddende) og en mobil, punktuelt informationstilgang (når brugeren bevæger sig). I diskussionen vil fokus være på de kognitive leksikografiske situationer, jf. Tarp (2006:58-60).

## 2. Metode og empiri

I den empiriske undersøgelse, som blev gennemført i løbet af foråret 2013, blev otte læger bedt om at slå op i et lægevidenskabeligt opslagsværk via henholdsvis [www.medicin.dk](http://www.medicin.dk) og appen Medicin.dk. Forsøget omfattede i alt tre forsøgsopstillinger:

**Forsøgsopstilling 1 fokuserede på den stationære brugssituation med en PC**, hvor forsøgspersonen sad stille ved et bord og foretog søgeoperationer på en computer. Forsøget blev optaget «indefra» ved hjælp af screen recorder-software Debut,

med henblik på at kunne se hvordan forsøgspersonen rent faktisk søgte information på computeren ved hjælp af sitet [www.medicin.dk](http://www.medicin.dk), ligesom det hele blev optaget «udefra» ved hjælp af et digitalkamera.

**Forsøgsopstilling 2 fokuserede på den bevægelige brugssituation med en mobiltelefon**, hvor forsøgspersonen gik i en halvcirkel omkring et sygeleje med en patient, mens forsøgspersonen foretog søgeoperationer via appen Medicin.dk på en iPhone 4S. Testpersonernes søgninger på smartphonen blev optaget «indefra» ved hjælp af værktøjet Reflector, jf. <http://www.airsquirrels.com/reflector/>, som trådløst kan overføre en film til en PC i nærheden. På den måde blev der optaget en film af testpersonernes helt konkrete søgninger og opslag i appen. Formålet med denne forsøgsopstilling var at se, hvordan en hospitalslæge rent faktisk interagerer med en mobiltelefon i en situation, hvor lægen bevæger sig rundt omkring et sygeleje. Samtidig blev det hele optaget «udefra» ved hjælp af et digitalkamera med henblik på at indhente data om, hvordan lægen orienterer sig i det fysiske rum samtidig med at han foretager søgeoperationer.

**Forsøgsopstilling 3 fokuserede på den stationære brugssituation med en mobiltelefon**, hvor forsøgspersonen ligesom under forsøgsopstilling 1 sad stille ved et bord mens forsøgspersonen foretog søgeoperationer på en iPhone 4S. Igen her blev det hele optaget «indefra» ved hjælp af Reflector og formålet var også her at optage det hele indefra for at kunne se hvordan forsøgspersonen rent faktisk søger i en situation, hvor lægen sidder stille – f.eks. på sit kontor eller ved en patientkonference. Det hele blev igen optaget «udefra» med et digitalkamera.

Forsøgsopstilling 2 og 3 var udformet således, at de lignede typiske arbejdssituationer for hospitalslæger. Den stationære brugssituation kunne være de situationer, hvor hospitalslæger

forbereder konference eller under selve konferencen, hvor læger tilgår kompleks information med henblik på f.eks. kontrol af ordination etc. Den mobile brugssituation kunne være de situationer, hvor hospitalslæger er på vej til patienten eller står ved patienten og ønsker at tjekke data. Flere af testpersonerne bekræftede i øvrigt at forsøgsopstillingerne var særdeles autentiske og lignede situationer fra deres daglige arbejde.

Der blev således gennemført i alt tre tests og testpersonerne blev i alle tests bedt om at slå de samme fem medicinske termer op ved hjælp af henholdsvis websitet [www.medicin.dk](http://www.medicin.dk) og appen Medicin.dk. Under alle tre tests blev de otte testpersoner bedt om at «tænke højt», og mens de slog de fem medicinske termer op, gjorde de således verbalt rede for, hvad de gjorde, så og oplevede. Endelig blev lægerne interviewet før og efter forsøgene.

Den empiriske basis for denne diskussion omfatter således en masse både kvantitative og kvalitative data. Der blev optaget ca. 2½ times optagelser af forsøgene «udefra» ved hjælp af digi-talkamera og disse optagelser indeholder desuden tænke-højt-data samt interview-data. Der blev desuden også optaget ca. en times optagelse «indefra» med Screen Recorder, som viser hvordan forsøgspersonerne søger på websitet [www.medicin.dk](http://www.medicin.dk). Endelig blev der optaget ca. to timers optagelser «indefra» med Reflector, som viser hvordan forsøgspersonerne søger på og bruger appen Medicin.dk.

### 3. Teori

Tidligere og relaterede undersøgelser af direkte relevans for denne undersøgelse omfatter især Pedersen & Engrob (2008), Church & Smyth (2009), Kassab & Yuan (2013) og Ehrler et al. (2013).

Det første bidrag er Pedersen & Engrob (2008), som allerede i 2008 gennemførte en række forsøg for at besvare, hvilken interaktionsteknik der var mest anvendelig i forbindelse med søgning efter komplekse data under fysisk bevægelse. Pedersen & Engrob interesserede sig også for den mobile brugssituation og bad i deres forsøg otte studerende om at gå på et løbebånd mens de tilgik forskellige typer af data ved hjælp af en PDA. Pedersen & Engrob peger på, at brugerne anvender forskellige interaktionsteknikker i forskellige situationer.

Det andet særdeles relevante bidrag er Church & Smyth (2009). Church & Smyth har gennemført en række undersøgelser af brugernes mobile informationsbehov, og i denne artikel bad de 20 testpersoner om at deltage i en såkaldt dagbogsundersøgelse, jf. Church & Smyth (2009:249-251). De 20 testpersoner førte dagbog over deres mobile informationsbehov og -situationer. Church & Smyth fandt, at brugersituationer kan inddeles i fem overordnede kategorier og foreslog de fem kategorier: **Navigational, Informational, Transactional, Geographical** og **Personal Information Management**. Ifølge Church & Smyth (2009:251) er informationsbehovet ikke overraskende det mest fremherskende behov.

Den tredje relevante undersøgelse for denne artikel er Kassab & Yuan (2013), som undersøger hvorfor og hvordan 12 mobilbrugere anvendte deres smartphones. Kassab & Yuan

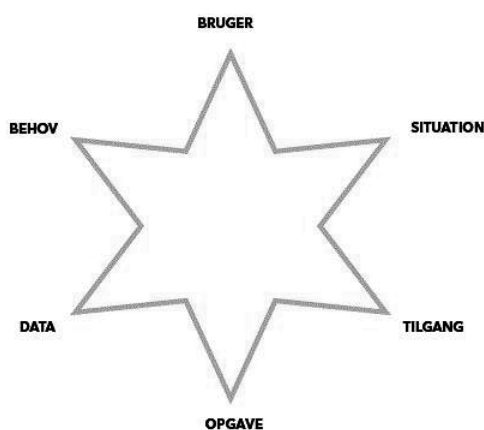
(2013) gennemførte en række interviews med de 12 testpersoner og fandt, at størstedelen af brugerne i stadigt stigende grad bruger deres smartphones til at tilgå sociale netværk samt til at foretage småindkøb. Dette resultat underbygges af Google (2013), som netop peger på, at den moderne mobilbruger i stor stil anvender mobiltelefon til at søge efter information samt især at tilgå sociale netværk med. Brugernes synes således i højere grad at bruge deres smartphones til kognitivt set «lette opgaver».

Endelig bidrager Ehrler et al. (2013) med en omfattende evidensbaseret undersøgelse af user interface-designet i håndholdte enheder i sundhedsvæsenet. Ehler et al. bad hele 93 sygeplejersker om at bruge forskellige data input-interfaces. Formålet var at måle kvaliteten af de data, der blev indtastet via forskellige data input interfaces. Ehrler et al. fandt, at langt størstedelen af de adspurgte testpersoner foretrak de helt simple data input interfaces, som i højere grad er tilpasset den komplekse brugersituation på det enkelte hospital. Denne konklusion er helt på linje med nærværende undersøgelse, som netop også peger på, at selve brugssituationen også spiller en afgørende rolle for hvor stor succes brugeren har med sin søgning, hvilket i øvrigt også påpeges af først Verlinde et al. (2010) og senere Simonsen (2011).

Netop Verlinde et al. (2010) og Simonsen (2011) spiller en central rolle for forståelsen af de faktorer, som både skal være til stede og skal tilpasses hinanden for at hjælpe brugeren i en given brugersituation, med et givent behov etc. Simonsens model kan med fordel anvendes både til analyse samt udvikling af informationsvidenskabelige produkter som f.eks. en app og data vil i analysen blive sat ind i en teoretisk ramme ved hjælp af netop stjernemodellen, der vises i figur 1 herunder.

Simonsen (2011) nævner især seks faktorer, som indbyrdes

skal tilpasses hinanden for at et værktøj er optimalt. Denne tilpasning omtaler Verlinde et al. (2010) som kalibrering, hvilket for så vidt også er meget beskrivende for, hvad der skal ske. De seks faktorer er henholdsvis **bruger, situation, tilgang, opgave, data** og **behov**, jf. figur 1 herunder. Figuren forsøger at illustrere, hvorledes disse seks faktorer indbyrdes skal tilpasses hinanden samt hænger sammen. Dette er desværre ikke altid tilfældet i leksikografiske produkter og sker, når der enten er for stor fokus på f.eks. data (termerne) eller på tilgangen (den teknologiske søgemetode) eller på brugeren. Dette kan i særlig grad iagttages i visse apps, som mere ligner teknologiske prestigeprodukter end effektive leksikografiske produkter. Visse apps synes at have en svær teknologisk slagside (alt for stor fokus på tilgang og søgefaciliteter) eller en datatung slagside (alt for stor fokus på encyklopædiske og/eller terminologiske data) og disse apps synes på ingen måde at have taget højde for den mobile brugsituations kendetegn.



Som det vil fremgå af diskussionen herunder, er det netop i den indbyrdes tilpasning mellem de seks faktorer, at der empirisk kan observeres uhenigtsmæssigheder.

Figur 1. Informationsvidenskabelig stjerne-model.

## 4. Resultater og diskussion

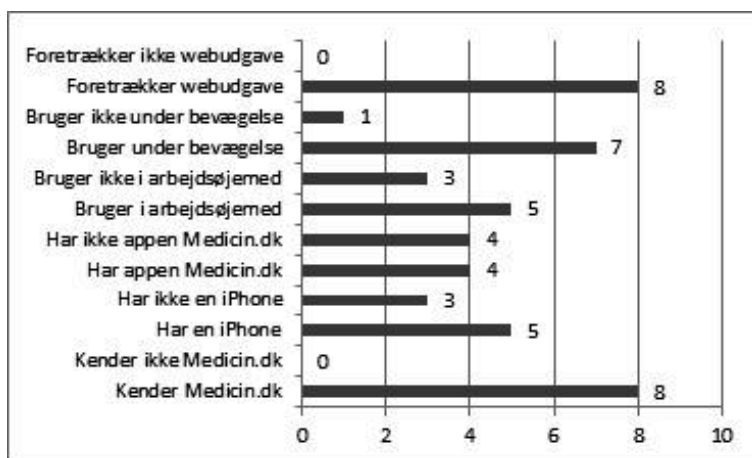
Kønsfordelingen blandt testpersonerne (TP) var fem kvinder og tre mænd og aldersspredningen var mellem 37 år og 61 år. Som figur 2 herunder viser, foretrak alle otte adspurgte læger web-sitet [www.medicin.dk](http://www.medicin.dk), da de blev stillet spørgsmålet «*På hvilken platform og hvilken brugersituation foretrækker du ved informationstilgang?*» Det er for så vidt både naturligt og på samme tid også tankevækkende. De otte testpersoner foretrak en brugersituation, hvor de kan sidde ved en skærm og tilgå medicinsk information. TP2 udtaler:

Jeg vil gerne have computeren hvis det skal være lidt mere dybdegående. Der har jeg nemmere adgang til andre sider også. Appen og iPhonen er god, hvis jeg står med et problem, som jeg ved jeg sagtens kan finde på [medicin.dk](http://www.medicin.dk). Men hvis jeg har brug for noget mere, så vil jeg hellere sidde ved computeren.

Dette nuanceres yderligere af TP5, som udtaler:

Det kommer meget an på, hvad man skal. Det er da rart at have adgang til appen, så man kan slå noget op i elevatoren på vej op til en patient. Men som arbejdsredskab kan jeg bedre lide webudgaven. Jeg arbejder meget administrativt, så der bruger jeg webudgaven. Men hvis jeg er på lægevagt, så bruger jeg min telefon og appen. Det kan også sagtens ske inde ved patienten. Hvis jeg er inde ved patienten og der ligger syv lægemidler, så slår jeg op på telefonen.

De to udsagn understreger i særlig grad behovet for en løbende kalibrering og stringent fokus på og forståelse af samspillet mellem brugerens behov, brugerens situation samt den måde, hvorpå brugeren kan søge data på, jf. Simonsen (2011). Det argumenteres således, at det er naturligt for læger at foretrække webudgaven, idet både deres behov og deres arbejdsituation tilsiger dette. Når behovet og situationen ændrer sig, så ændrer behovet sig og dermed også for type af værktøj sig tilsvarende, jf. f.eks. udsagnet fra testperson 5.



Figur 2. Oversigt over interviewdata – før og efter forsøg.

Denne observation synes desuden også at svare til observationer nævnt hos Kassab & Yuan (2013), som anfører, at brugerne i deres undersøgelse i større grad bruger deres smartphones til f.eks. at tilgå sociale medier etc. – dvs. kognitivt set relativt «lette opgaver». Som det yderligere fremgår af figur 2 ovenfor, så kender otte ud af otte læger Medicin.dk, syv ud af otte læger



anvender deres mobiltelefon under bevægelse og fem ud af otte læger bruger deres mobiltelefon i arbejdsøjemed. Det synes derfor at være typisk for populationen at anvende mobiltelefoner under bevægelse, hvilket da også i særlig grad understreges af TP3, da han blev spurgt «*Hvordan oplevede du informationstilgang under bevægelse?*». TP3 udtaler:

Bare jeg kan stå stille, så synes jeg faktisk, at det fungerer meget godt. Og jeg synes, at det er rart, at jeg kan tage den med ind til patienten, at jeg kan tilgå information alle mulige steder og at jeg kan bevæge mig - også i en patientsituation. Og i den moderne verden, så synes jeg, at det virker tillidsfuldt, at en læge slår ting op. Og patienter har stor forståelse og respekt for det. Det giver indtryk af en læge, som er velorienteret og som konsulterer information ved deres sengekant i stedet for blot at bilde dem ind, at de kan alt. Så det kunne jeg sagtens finde på.

Det er således højst relevant at se på den mobile brugssituation i forhold til den stationære brugssituation. Figur 3 nedenfor viser en opgørelse over, hvordan de otte testpersoners søgeperformans og -succes er blevet vurderet af forsøgslederen på en skala fra 1 til 10.

	Terbasmin		Tamoxifen		Antepsin		Tredaptive		Fludara	
<b>Opgave</b>	Finde information		Finde og uddrage information om bivirkninger mhp. at kunne informere patient om bivirkninger		Finde og uddrage information om dosering mhp. at kunne kontrollere ordinationsmængde		Finde og uddrage information om dosering mhp. at kunne informere patient om indtagelsesmåde		Finde og tjekke stavemåde mhp. at kunne skrive tekst	
<b>TP/Brug</b>	<b>Bevæge</b>	<b>Sidde</b>	<b>Bevæge</b>	<b>Sidde</b>	<b>Bevæge</b>	<b>Sidde</b>	<b>Bevæge</b>	<b>Sidde</b>	<b>Bevæge</b>	<b>Sidde</b>
T1	7	8	4	7	4	7	4	7	4	7
T2	3	4	2	5	2	5	2	5	2	5
T3	6	7	3	6	3	6	3	6	3	6
T4	7	8	4	7	4	7	4	7	4	7
T5	6	7	3	6	3	6	3	6	3	6
T6	2	3	2	3	2	3	2	3	2	3
T7	7	8	4	7	4	7	4	7	4	7
T8	7	8	4	7	4	7	4	7	4	7
<b>Total</b>	45	53	26	48	26	48	26	48	26	48

Figur 3. Oversigt over testdata – mobil og stationær.

Figur 3 viser fem kolonner med fem medicinske termer samt formålet med søgeoperationen og stammer fra forsøgsopstilling 2 og 3. De otte rækker viser for hver testperson et tal for henholdsvis den mobile brugssituation og den stationære brugssituation. De kvantitative vurderinger understøttes af TP6 som på spørgsmålet «Bruger du din mobiltelefon under bevægelse?» svarer:

Ja. Det er måske ikke så typisk, at jeg går, men hvis jeg nu står på en stue med en patient, der har en forgiftning f.eks. så kunne jeg finde på at bruge den.

En anden måde at beskrive den mobile, punktuelle brugssitua-

tion er ved at beskrive den som flygtig og noget, som foregår samtidig med noget andet. Brugeren skal desuden i en mobil brugssituation orientere sig i det fysiske rum samtidig med, at han skal interagere med en mobil enhed via oftest en semasiologisk tilgang (som kan være vanskelig at anvende på en lille skærm).

En mobil brugssituation kendetegnes desuden ved at være velegnet i flygtige og afsøgende brugssituationer, hvor brugeren vil tjekke eksisterende viden. Den synes ikke at være særlig velegnet til at understøtte komplekse kognitive og kommunikative leksikografiske funktioner.

## 5. Konklusion

Formålet med denne artikel var at diskutere den mobile, punktuelle brugssituation og diskutere forskellene mellem den stationære, punktuelle informationstilgang (når brugeren er stillesiddende) og den mobile, punktuelle informationstilgang (når brugeren bevæger sig).

Undersøgelsen pegede på flere markante forskelle mellem de to brugssituationer på basis af optagelser både udefra og indefra samt på basis af konkrete testdata samt interviews.

Undersøgelsen viste, at der er forskel mellem de to brugssituationer. Den stationære brugssituation er ikke overraskende hurtigere og resulterer i større og bedre informationstilgang end den mobile brugssituation, ligesom undersøgelsen viste, at ved kognitivt tunge opgaver (hvor brugeren skal tilegne sig kom-

pleks viden), foretrak testpersonerne PC'en, den stationære, mobile brugssituation og til sidst den mobile brugssituation.

Apps skal ligesom andre leksikografiske produkter udvikles på basis af en konstant kalibrering af følgende seks faktorer: **bruger, situation, tilgang, opgave, data** og **behov**. Kun på den måde undgår vi, at apps udelukkende bliver teknologiske prestigeprojekter.

Brugerne er i den grad allerede mobile. Men brugerne kan ikke løse alle opgaver ved hjælp af en mobiltelefon. Det er data ofte for komplekse til, ligesom telefonens ret begrænsende interface er for lille. En vej frem er at udvikle apps på basis af leksikografiske dyder og i højere grad udnytte moderne talegenkendelsesteknologi for at forbedre brugerens interaktionsevne.

## Litteratur

- Church, Karen & Barry Smyth (2009): Understanding the intent behind mobile information needs. I: Cristina Conati, Mathias Bauer, Nuria Oliver & Daniel S. Weld (red.): *UI 2009 International Conference on Intelligent User Interfaces*. Sanibel Island, Florida, USA, 247–256.
- Ehrler, Frederick, Magali Walesa, Evelyne Sarrey & Christian Lovis (2013): Evidence-based User-Interface Design. I: *Studies in health technology and informatics*, 57–61.
- Google (2013): *Our Mobile Planet*.  
<<http://services.google.com/fh/files/misc/omp-2013-dk-en.pdf>> (1 juli 2013).

- Kassab, Dima & Xiaojun Yuan (2012): Understanding the information needs and search behaviour of mobile users. I: *Information Research*, 17(4), paper 551.
- Pedersen, Anders Fritz & Jan Engrob Hyldgaard (2008): *Interaktion under bevægelse: Et komparativt studie af interaktionsteknikker til arbejde med komplekse data på håndholdte enheder*. Aalborg: Institut for Datalogi, Aalborg Universitet.
- Reflectorapp.com. <<http://www.airsquirrels.com/reflector/>> (2 januar 2012).
- Simonsen, Henrik Køhler (2011): Et informationsvidenskabeligt serviceeftersyn af Medicin.dk. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.) (2012): *Nordiska studier i lexicografi 11*. Lund: Svenska Akademiens ordboksredaktion, 563–574.
- Tarp, Sven (2006): *Leksikografien i grænselandet mellem viden og ikke-viden: Generel leksikografisk teori med særlig henblik på lærerleksikografi*. Doktorafhandling. Århus: Center for Leksikografi, Handelshøjskolen i Århus.
- Verlinde, Serge, Patrick Leroyer & Jean Binon (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, 1–17.

Henrik Køhler Simonsen  
ekstern lektor, Ph.d.  
Copenhagen Business School  
Dalgas Have 15  
2000 Frederiksberg  
[hks.abc@cbs.dk](mailto:hks.abc@cbs.dk)

# Tvåspråkiga morfemordböcker för inlärare

*Jonas Winnerlöv*

This article argues for a new type of bilingual learners' dictionary which would group the words along their root morphemes and thereby permit systematic vocabulary learning according to an etymological version of the keyword mnemonics method. The concept is illustrated by Swedish examples of these extended word root families.

## 1. Inledning

Många av dagens inlärorordböcker är tematiskt organiserade och syftar till att förmedla ett centralt ordförråd som inlärarna behöver för att behärska kommunikativa situationer. Den alternativa typ av inlärorordböcker som föreslås här bygger istället på morfologiska principer och syftar till att hjälpa inlärarna att åtminstone receptivt lära sig ett större antal ord. För svenska skulle Lexin-seriens cirka 20 000 ord kunna utgöra urvalet för en klassificering i ordsläkten utifrån det kärnord som tydligast illustrerar rotmorfemets prototypiska betydelse och som mnemotekniskt skulle utnyttjas som nyckelord för de övriga orden. Att på detta sätt göra den etymologiska dimensionen explicit i ordinläring är främst tänkt för en smal målgrupp av

akademiskt avancerade inlärare, men eftersom ordboksmaterialet inte behöver vara heltäckande och i många fall skulle kunna framställas genom att kombinera existerande källor, skulle produktionen av en serie tvåspråkiga ordböcker vara fullt möjlig även för relativt små källspråk som de skandinaviska språken.

## 2. Existerande inlärorordböcker

### 2.1. Allmänna inlärorordböcker med uppslagsfunktion

Allmänna inlärorordböcker är en stor produkt. Enspråkiga *Oxford Advanced Learner's Dictionary* (2008) påstår sig ha sålt 1 miljon exemplar av sin tryckta version. Detta pionjärverk har ett antal enspråkiga uppföljare på engelska och på andra språk, t.ex. *Langenscheidts Grosswörterbuch Deutsch als Fremdsprache*. En vidareutveckling av dessa ordböcker är de så kallade «semi-bilingual» där i princip samma enspråkiga struktur kompletteras med en översättning till målspråket i fråga. På svenska finns Lexin-serien, som utgår från en svensk ordbas på cirka 20 000 lemma med ett tjugotal invandarspråk som målspråk. Den finns nu gratis på Internet genom (det svenska) Språkrådets försorg. Dessa ordböcker är alfabetiska och har en ren uppslagsfunktion.

## 2.2. Inläraordböcker med basordförråd

Inläraordböcker kan också grunda sig på ett mindre urval ord som det verkligen är tänkt att man ska lära sig. Bara för tyska ord på den franskspråkiga marknaden finns det ett tjugotal sådana verk till salu, ännu fler om man också räknar särskilda verbböcker. Om en alfabetisk ordbok kan kallas semasiologisk (utgår från tecknen) är dessa ordböcker vanligtvis onomasiologiska (utgår från begreppen). De allra flesta är också tematiskt sorterade och tar avstamp i kommunikativa situationer, enligt nuvarande paradigm i andraspråksdidaktiken, bland annat formulerat av Europarådet (Council of Europe 2001).

Inläraordböcker kan vara frekvensbaserade, t.ex. Routledges aktuella tvåspråkiga serie med bland annat tyska (Jones & Tschirner 2006) och ett tiotal andra stora språk, alla med engelska som målspråk. Att dessa innehåller 4000–5000 ord är ingen slump. I litteraturen förekommer ofta uppgifter om att denna mängd motsvarar 95 procent av de löpande orden i en text, och att just 95 procents texttäckning skulle räcka för att förstå en text (Tschirner 2008). Även om dessa uppgifter grundar sig på ytterst få och tunna undersökningar och nästan liknar en akademisk vandringsägen är grundkonceptet ändå lovande, nämligen att man genom att behärska ett mer eller mindre begränsat antal ord kan börja läsa en text.

## 2.3. Morfemordböcker

En helt annan och betydligt smalare produkt är morfemord-



böcker. August (2009) hävdar att hans verk är det första i sitt slag på tyska, och att det ännu inte finns någon morfemordbok på engelska, men hänvisar samtidigt till den rikare traditionen i rysk och fransk lexikografi. För ryska finns till exempel Kuznetsova & Efremova (1986), som radar upp i princip hela det ryska ordförrådet efter rotmorfem men utan ytterligare förklaringar. För franska finns Picoche (1992), en traditionell etymologisk ordbok men där orden grupperas efter sina djupaste indoeuropeiska rötter. En intressant svensk produkt är Ernby (2008), med lättillgängliga beskrivningar som inte tyngs av processerna bakåt i tiden men som anger hur dagens ord hänger ihop, till exempel korshänvisningen mellan *gärna*, *begära* och *girig*. Och för engelska har faktiskt Bird (1990), slavist och engelsklärare i Hongkong, fört tillbaka det engelska ordförrådet till 2000 indoeuropeiska rötter.

#### 2.4. Kombination av inlärorordbok och morfemordbok

Som examensarbete för mina egna ryskstudier samkörde jag den frekvensbaserade inlärorordboken Brown (1996) med den ovannämnda morfemordboken Kuznetsova & Efremova (1986). Det visade sig till exempel att de 100 mest produktiva slaviska rötterna täcker in 3000 av 10 000 ord. Jag argumenterade också för att dessa «ordsläkten» (i kontrast till begreppet «ordfamiljer» som reserveras för de transparent härledningsbara ordbildningsmönstren enligt Bauer & Nation (1993)) skulle kunna utnyttjas för systematisk inläring.

### 3. Ordinlärning

#### 3.1. Etymonik

Generellt sett ställer sig forskare inom andraspråksinläringen (Second Language Acquisition/SLA) skeptiska till att föra in en etymologisk dimension i ordinlärning (Nation 2001), men det finns dokumenterade exempel på att etymologisk analys kan vara till hjälp vid inläring av idiom (Boers et al 2004) och etymologi som inlärningsredskap har vissa förespråkare (Ilson 1983, Pierson 1989, Brodsky 2008).

En bevisat effektiv inlärningsmetod är «key word mnemonics» (nyckelordsmetoden), först introducerad av Atkinson (1975), där man på grundval av en oftast helt slumpmässig formlighet mellan det främmande ordet och ett redan bekant ord mentalt skapar en färgstark bild som sammanför de två ordens betydelser, t.ex. för det franska ordet *avide* (glupsk, girig) kan man föreställa sig någon med *vidöppen* mun. Med den föreslagna metoden, kallad *etymonics* eller *etymonik* (teleskopord på *etymology/etymologi* och *mnemonics/mnemonik*) skulle man istället koppla orden till dess etymon eller till ett enklare ord på samma rot. Till det ryska grundordet för *hetta* (*zjar*) kan man koppla till exempel de ryska orden för *feber*, *iver*, *het*, *steka*, *grilla*, *köttstek*, *eldsvåda*, *brandman* m.m., som alla har samma rot och en uppenbar betydelsekoppling till grundordet. August (2009), som innehåller utförliga enspråkiga förklaringar och har ett klart pedagogiskt syfte, menar att orden i ett sådant ordsläkte är ömsesidigt upplysande («wechselseitig erhellend»).

### 3.2. Transparens- och kognateffekt

Den hjälp man i ordinlärningen får från att analysera ordbildningen kan man kalla transparens-effekt. Mest transparenta är de traditionella ordfamiljerna, med ett svenskt exempel *intressant/intressera/intresse*, där man av det ena ordet kan härleda det andra. De många sammansättningarna i germanska språk kan placeras på en skala från tämligen genomskinliga såsom *silversmycke* och *intensivkurs* till lexikaliserade ordbildningar som *grönsak* och *hemsöka*. Men även för de ord som inte kan begripas spontant finns det en ordbildningslogik som mnemotekniskt kan stödja ordinlärningen när man väl har fått betydelsen.

Den intralinguala transparens-effekten samspelar med den interlinguala kognateffekten, d.v.s. att man kan förstå ett ord i ett främmande språk genom likheten med ett besläktat ord (kognaten) i det egna språket (eller ett annat språk som man behärskar). I en färsk studie (Winnerlov 2013) har jag konstaterat att kognateffekten för svenska inlärare av tyska spontant är 70 procent och att det finns en potentiell kognateffekt för ytterligare 20 procent av kärnorden. För svenska inlärare av ryska uppskattade jag kognateffekten till 10 procent (Winnerlov 2000), och för svenska inlärare av franska, med engelskans hjälp, till 50 procent (Winnerlov 1999).

Ju mindre kognateffekten för en inlärare av ett visst språk är, desto större blir potentialen för att istället försöka utnyttja transparens-effekten. Om man inte kan hänga upp orden på liknande ord i sitt eget språk finns det desto större anledning att istället koppla dem till varandra. Målspråk för morfemordböckerna med svenska som källspråk vore alltså inte andra germanska språk utan snarare slaviska och romanska språk, samt

andra språkområden där svenska är ett universitetsspråk eller från vilka många högutbildade kommer till Sverige, t.ex. finska och språken i Baltikum, och kanske också stora världsspråk som kinesiska och arabiska.

I och med att svenska för dessa avancerade inlärare i regel vore tredjespråk (L3) efter engelska som andraspråk (L2), och i vissa fall även tyska, finns det anledning att utveckla ett pedagogiskt material som konsekvent utreder kognatpotentialen mellan dessa germanska systerspråk.

### 3.3. Systematisk ordinlärning

I den idag förhärskande språkundervisningsmodellen är det tänkt att orden ska läras integrerat i de kommunikativa övningarna. Mängden inlärorordböcker på marknaden är dock ett tecken på att många inlärare känner ett behov att gå systematiskt tillväga för att tillägna sig ett stort och främmande ordförråd, t.ex. franska inlärare av tyska. Routledges tvåspråkiga frekvensordböcker, t.ex. Jones & Tschirner (2006), har ett klart pedagogiskt syfte. Men att plugga in 5000 ord är en diger uppgift, i synnerhet i frekvensordning, där funktionsorden blandas med innehållsorden, vilka kan vara kognater eller helt transparenta och därför inte behöver läras in, i alla fall inte i ett receptivt perspektiv.

I en alternativ modell som ser till ordens strukturella egenskaper skulle inlärarna få 1) bekanta sig med ett dussin fasta kommunikativa formler, t.ex. hälsningsfraser, 2) kontrastiva uttalsövningar med höggradiga kognater, d.v.s. nästan identiska ord, 3) de grammatiska orden undervisade som system, 4) öva

upp förmågan att känna igen kognater av lägre rang 5) systematiskt lära in (med lämplig metod) ett antal kärnord som är konkreta och morfologiskt enkla och som ofta utgör naturliga serier, och slutligen 6) ge sig i kast med ordsläktena för dessa kärnord. Först när ett betydande receptivt ordförråd byggts upp skulle de kommunikativa övningarna sättas in.

### 3.4. Översättningsekvivalenter

Ordinlärningsexperiment i SLA-forskningen, men även i psykologiska studier och i den rena hjärnforskningen, visar att människor har förmåga att lära sig stora mängder ordpar. Även om det finns många uppfattningar om vilket som är det bästa sättet att förstå och lära sig ett främmande ord visar rönen på att en översättningsekvivalent är det effektivaste sättet att få inlärare att gripa betydelsen av ett ord (Laufer & Hadar 1997). Starka argument för översättningar i undervisningen framför av Augustyn (2013). I ett sammanhang där det finns en etymologisk dimension och många snarlika ord och fina betydelsenyanser finns det ännu större anledning att bygga systemet på översättningsekvivalenter till inlärares förstaspråk.

## 4. Ordsläkten

### 4.1. Metodkommentar

Poängen med ordsläkten (se 2.4) är att hjälpa inlärare från andra språkfamiljer att få struktur på ett främmande ordförråd. Som modersmålstalare ser man i regel orden som lexikaliserade enheter utan att vara medveten om deras etymologi och möjliga kopplingar mellan dem inom ordsläktena. Det är utifrånperspektivet som är den springande punkten. För en fackpublik är det ändå naturligt att presentera de ”egna” ordsläktena. Nedan kommer modellen med dess olika frågeställningar och gränsdragningsproblem att illustreras med svenska exempel, vilka dock i hög grad lär överensstämma med de övriga skandinaviska språken, tyska och nederländska. Modellen är dock giltig för fler språkfamiljer än den germanska och slaviska. Oavsett normer för grafiska ord är det typiskt att uttrycken för komplexa begrepp bygger på enklare, mer grundläggande ord.

Eftersom det ännu inte finns någon rotmorfemklassning tillgänglig för svenska har jag för detta konferensbidrag ad hoc tagit fram ordsläktena för några hundra svenska kärnord ur en svensk orddatabas med cirka 100 000 ord där de cirka 20 000 Lexin-orden (i en version från 1990-talet) är särskilt markerade. De första 100 kärnorden täcker in ungefär 3500 av Lexin-orden, och de 300 mest produktiva av dem jag tagit fram cirka 5400 ord.

## 4.2. Exempel på svenska ordsläkten

Kroppsdelarna har varit synnerligen produktiva kärnord. Ett urval av ordsläktet kring *hand* är: *handled, handtag, handduk, handstil, handboll, handske, handskas, händig, egenhändig, knapphändig, behändig, avhända sig, hantera, lätthanterlig, handla, handlingskraft, ansökningshandling, världshandel, kohandel, handelsrätt, behandla, behandlingsmetod, särbehandling, misshandla, förhandla, underhandla, avhandling, handledning, handläggning, handfull, handgriplig, handlag, handfallen, handplocka, omhändertagande, tillhandahålla, tumanhand, förhand, handvändning, järnhand, mellanhand, hantlangare, hantverkare, konsthantverk, enahanda.*

För *huvud* är några av de många lexikaliserade sammansättningarna: *huvudsaklig, överhuvudtaget, bakhuvud, ljushuvud, huvudlös, huvudvärk, huvudbonad, huvudkudde, huvudräkning, huvudbry, huvudstupa, huvudstad, huvudroll, huvudkontor, huvudrätt, huvudsats, huvudman, huvudförhandling, statsöverhuvud.*

När det gäller *mun* med ord som *mungipa, munkavle, livmodermun, munstycke, harmynt, munfull, munhuggas, inmundiga, muntlig, mynna, utmynna i, flodmynning* kan man ställa sig frågan hur trogen man i framställningen måste vara den faktiska etymologin. En inlärare skulle ju kunna se en koppling till *myndig* (förmögen att tala för sig?) med ord som *straffmyndig, omyndig, myndighetsdag, tillsynsmyndighet, myndighetsutövning, bemyndigande, förmyndare* fast detta ord egentligen har ett annat ursprung. Samma sak gäller *buk* (*bukhåla, bukspottkörtel, buklandning, buktalare, girigbuk*) som inte är besläktat med *bukta* (*bukta sig, utbuktning, bukt*) men faktiskt skulle kunna tolkas så på grund av en typisk buks buktning. Den enkla

lösningen på detta är att hålla framställningen strikt till den belagda etymologin men samtidigt påpeka att orden inte är besläktade för att ändå möjliggöra för inlärarna att göra sina egna mentala kopplingar.

En annan fråga är hur långt tillbaka i etymologin man ska gå när man sammanför orden: *gärna* är besläktat med och har en semantisk koppling till *begär* (*begära, begäran, begärelse, begärlig, maktbegär*) och *girig* (*girigbuk, hämndgirig, vetgirig, äregirig*) men eftersom roten uppträder i så olika form kan det vara tveksamt att sammanföra dem. För *hem* (*daghem, hemvist, hemlös, hemlängtan, hembränt, hembud, hemma, hemmablind, hemmamarknad*) är däremot roten oförändrad i besläktade *inhemsk, hemsk* och *hemlig* medan den semantiska kopplingen är mer långsökt, men just det är hela poängen med nyckelordsmetoden, och det finns alltså anledning att sammanföra dem.

Homografer måste naturligtvis skiljas åt, till exempel *nära* som i *närhet, närmande, förnärma, närvaro, närbild, när- gången, närpå, jordnära, närsynt och närstående* och *nära* som i *livnära sig, undernärd, näringsämne, näringskedja, näringsliv och restaurangnäring*.

Ibland behöver kärnordet i sig inte vara så centralt, i varje fall inte så frekvent, men ändå ligga till grund för en rad användbara ord: *skåda* (*genomskåda, skärskåda, överskådlig, svåröverskådlig, beskåda, åskådare, åskådliggöra, livs- åskådning, navelskådning, fågelskådare, skådeplats, skådespe- lare*) och *skrida* (*skrida fram, framskriden, framåtskridande, fortskridande, underskrida, överskrida, gränsöverskridande, skridsko*). Orden på *barm* har viktiga överförda betydelser (*barmhärtig, obarmhärtig, barmhärtighetsmord, erbarmlig, för- barma sig*) medan orden på betydligt frekventare *bröst* är mer konkreta (*bröstkorg, bröstvärta, barbröstad, bröstmjölk, bröst- cancer, kycklingbröst, bröstficka, skjortbröst, bröstsim*). Mycket



är vunnet för inläringen om kärnordet i sig är konkret och färgstarkt, t.ex. *troll* (*trolla, trolleri, trollkonstnär, trollslag, trolla fram, trollbinda, förtrolla, trollslända, charmtroll*).

För att framställningen ska bli användbar måste man se till att kärnorden inte kombineras med alltför många okända ord. För till exempel *makt* (*mäktig, maktlös, fullmakt*) finns till bland annat *maktkamp* och *maktskifte* där *kamp* (*kämpa, förkämpe, bekämpa*) och *skifte* (*skiftarbete, skiftnyckel, årsskifte*) själva är kärnord som bör ingå i ordboken och till vilka det räcker att göra en korshänvisning, liksom *maktpolitik* och *centralmakt* som är kombinerade med internationella kognater som förhoppningsvis är begripliga tack vare engelskan, men också ord som *maktgalen* där *galen* kan behöva en särskild referens.

En avgörande svårighet för åskådligheten är att många ordsläkten helt enkelt blir för stora, t.ex. *stå* (87), *ställa* (78), och *hålla* (73) bara ur Lexin-materialet. Man kan förstås tänka sig att begränsa urvalet till t.ex. de 5000 ord som maximerar texttäckningen men samtidigt skulle man då gå miste om ”stor-driftsfördelarna” i att orden fungerar ”ömsesidigt upplysande”. I många fall vill man snarare använda ett större urval som SAOL och för till exempel *skatt* (*överskatta, underskattning, uppskattningsvis, oskattbar*) komplettera Lexin-orden *källskatt, mervärdesskatt* och *kvarskatt* med andra ord såsom *skattetryck, skatteflykt* och *skatteväxling*, som dels är relevanta för någon som vill förstå det svenska samhället, dels innebär ytterligare en möjlighet att koda in orden *tryck, fly* och *växla*.

## 4.3. Exempel på ordboksartikel

Nedan följer ett exempel på hur ett ordsläkte rent konkret skulle kunna se ut i ordboksform, här för en fransktalande målgrupp.

<b>mat</b>	nourriture; aliment; repas
<b>maträtt</b>	plats, mets
<b>matlagning</b>	cuisine
<b>matsäck</b>	provisions, panier-repas
<b>vardagsmat</b>	cuisine de tous les jours; chose ordinaire
<b>matsal</b>	salle à manger; cantine
<b>matbröd</b>	pain
<b>matberedare</b>	robot de cuisine
<b>matsedel</b>	menu, carte
<b>matsilver</b>	argenterie
<b>matsked</b>	cuillère (à soupe), cuillerée
<b>matsmältning</b>	digestion
<b>matvrak</b>	goinfre, glouton
<b>matnyttig</b>	bon à manger ; utile
<b>mata</b>	nourrir, faire manger, alimenter
<b>matarkabel</b>	câble d'alimentation
<b>matarbuss</b>	autobus de correspondance
<b>mätta</b>	rassasier
<b>mätt</b>	rassasié, satisfait, en avoir assez mangé
<b>mättad</b>	saturé
<b>omättlig</b>	insatiable

I denna artikel finns t.ex. de tämligen marginella orden *matarbuss* och *matarkabel*. För en fransk inlärare är dessa dock «gratisord» (särskilt som buss och kabel har kognater i *bus* och *câble*) som åskådliggör de överförda betydelseerna av verbet *mata*.

## 5. Framställningssätt

Att skapa en fullständig morfemordbok med systematisk och vetenskaplig metod, såsom de ovannämnda böckerna för tyska och ryska, är naturligtvis ett mycket omfattande arbete. En tydlig pedagogisk tillämpning av ett sådant arbete kan visserligen bli ett argument för att framställa sådana för fler språk. Om syftet med morfemklassningen endast är att skapa en ordbas för inläroordsböcker kan man dock gå mer intuitivt till väga och «plocka russin ur kakan». En sådan inläroordsbok behöver ju endast omfatta de kärnord som har de viktigaste och de mest åskådliga ordsläkterna.

Dessutom skulle det enligt argumentationen ovan bara handla om de igenkännbara rotmorfemen med dess förutsägbara växlingar och inte om alla ord som etymologiskt hör samman. För att få fram sådana ordsläkten kan man helt enkelt söka på en bokstavssekvens i en given ordlista som finns i databasform. För *\*klar\** får man förutom de givna träffarna (*klarhet, klarna, oklar, förklara, klarlägga, självklar, klartecken*) t.ex. helt orelaterade *jäklar* och *mäklare* som dock lätt rensas bort manuellt, och också mer avlägset besläktade *deklaration* och *klarinet*.

Att gå från en enspråkig ordbas till en tvåspråkig ordboksprodukt vore tämligen enkelt. Översättningarna behöver inte vara uttömmande utan bara ge ordens mest prototypiska betydelser för modersmålstalare i enkel ordlisteform, och man skulle besparas det stora merarbete det är att göra en bidirektional ordbok.

Eftersom det främsta syftet med denna ordboksprodukt är ordinläring kan man förvänta sig att användarna arbetar aktivt med orden genom att söka ytterligare källor. Ett sätt att nå fram

till dessa användare skulle kunna vara särskilda ordkurser inom ramen för universitetens språkutbildning. En början vore ett dokumenterat språkinlärningsexperiment på grundval av den metod som föreslås i avsnitt 3.4 för att utröna hur detta faktiskt slår an hos målgruppen.

Ordboksmaterialet skulle vara medieoberoende. För att få full funktionalitet med sökbarhet och kors hänvisningar är databasformen idealisk, men delar av materialet skulle även kunna ges mer styrda elektroniska former som passar för mobiltelefonernas appar. Som läromedel kan man också tänka sig tryckta produkter, kanske ett välkommet tillskott i ordboks förlagens sortiment i takt med att övriga produkter blir helt elektroniska.

## 6. Slutord

Digitaliseringen har inte bara i grunden förändrat i vilken form ordboksprodukter används och det sätt på vilket det ursprungliga materialet kan framställas. Den har också gjort det möjligt att med rimliga insatser framställa nya typer av produkter utifrån ett givet material i elektronisk form. Genom att samköra en rotmorfemklassning av orden med ett urval inlärrrelevanta ord och förse dem med enkla översättningar till olika målspråk skulle man kunna skapa en serie inlärrordböcker för en smal målgrupp akademiskt avancerade inlärrare som är beredda att komplettera den kommunikativa undervisningsformen med systematisk ordinlärrning. Detta vore just den typ av inlärrare som inte är främmande för argumentet att latin är ett medel för

att behärska romanska språk, avancerad engelska och det internationella vetenskapsspråket. Min bedömning är dock att en rotmorfemklassning av slaviska och germanska språk blir betydligt tydligare och mer pedagogiskt användbar än ordsläkterna för latin (och romanska språk), enligt t.ex. Wahlgren (1976). Såsom pedagogiskt material behöver en sådan ordboksprodukt inte heller vara heltäckande. Redan ett fåtal väl valda ordsläkten kan få inlärarna att börja se ordbildningsmönster och tänka enligt denna etymologiska logik som kan vara till hjälp för att greppa en stor och främmande ordsfatt.

## Litteratur

### Ordböcker

- August, Gerhard (2009): *Wortfamilienwörterbuch der deutschen Gegenwartssprache*. Studienausgabe. Tübingen: Niemeyer.
- Brown, Nicholas J. (1996): *Russian Learners' Dictionary. 10,000 words in frequency order*. New York: Routledge.
- Ernby, Birgitta (2008): *Norstedts etymologiska ordbok*. Stockholm: Norstedts.
- Hornby, S. (2010): *Oxford Advanced Learner's Dictionary of current English* (8. ed.). Oxford: Oxford University Press.
- Jones, Randall & Erwin Tschirner (2006): *Frequency dictionary of German: Core vocabulary for learners*. London: Routledge.
- Kuznetsova, A.I. & T.F. Efemova (1986): *Slovar' morfem russ-*

- kogo jazyka*. Moskva: Russky Jazyk.
- Langenscheidts Grosswörterbuch, *Deutsch als Fremdsprache*. (1993): Berlin und München: Langenscheidt.
- Picoche, Jacqueline (1992): *Dictionnaire étymologique du français*. Paris: Dictionnaires Robert.
- Wahlgren, Bengt (1976): *Latinskt morfemlexikon*. Stockholm: Almqvist & Wiksell.

### Annan litteratur

- Atkinson, R. C. (1975): Mnemotechnics in second-language learning. I: *American Psychologist* 30, 821-828.
- Augustyn, Prisca (2013): No dictionaries in the classroom: translation equivalents and vocabulary acquisition. I: *International Journal of Lexicography* 26/3, 362-385.
- Bauer, L. & I. S. P. Nation (1993): Word families. I: *International Journal of Lexicography* 6, 253-279.
- Bird, Norman (1990): *A first handbook of the roots of English*. Jersey: Lapine Education and Language Services.
- Boers, F., M. Demecheleer & J. Eyckmans (2004): Etymological elaboration as a strategy for learning idioms. I: Paul Bogaards & Batia Laufer: *Vocabulary in a second language: selection, acquisition, and testing*. Amsterdam, Philadelphia: John Benjamins, 53-78.
- Council of Europe (2001): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Iison, Robert (1983): Etymological information: can it help our students?. I: *ELT Journal* 37/1, 76-82.
- Laufer, B. & L. Hadar (1997): Assessing the effectiveness of monolingual, bilingual and 'bilingualised' dictionaries. I:

- The Modern Language Journal* 81/2, 189-196.
- Nation, P. (2001): *Learning vocabulary in another language*.  
Cambridge: Cambridge University Press.
- Pierson, Herbert, D. (1989): Using etymology in the classroom.  
I: *ELT Journal* 43/1, 57-63.
- Tschirner, Erwin (2008): Das professionelle Wortschatzminimum im Deutschen als Fremdsprache. I: *Deutsch als Fremdsprache* 45: 195-208.
- Winnerlöv, Jonas (1999): "Les vrais amis. Le « tatut congénérique » des 5000 mots lexicaux les plus importants du français pour un apprenant avancé suédois." Mémoire de 60p. Département de français et d'italien. Université de Stockholm.
- Winnerlöv, Jonas (2000): *Att lära sig 10 000 ryska ord*. Specialarbete fördjupningskurs i ryska. Stockholm: Stockholms universitet, slaviska institutionen.
- Winnerlöv, Jonas (2013): *Wie viele Wörter bekommt man beim Deutschlernen „geschenkt“? Der Kognatenstatus der häufigsten deutschen Wörter aus Sicht schwedischer LernerInnen auf Universitätsebene*. Examensarbete för magisterexamen. Stockholm: Stockholms universitet.

Jonas Winnerlöv  
Svenska översättningsenheten  
Europaparlamentet  
L-2929 Luxemburg  
[jwinnerlov@europarl.europa.eu](mailto:jwinnerlov@europarl.europa.eu)

## Tolvte konferanse om leksikografi i Norden

Oslo, 13.–16. august 2013

### PROGRAM

#### **Tirsdag 13.8.2013 Språkrådet/Nasjonalbiblioteket**

**Inngang: Observatoriegata 1 B, 0254 Oslo (nær Solli plass)**

18.30–21.00	Registrering og uformell velkomstmottakelse med lett servering Elisabeth Eide, Nasjonalbiblioteket: <i>Foredrag om Slottsbiblioteket</i> , kl. 19.45–20.05
-------------	---

#### **Onsdag 14.8.2013 Forskningsparken ved Universitetet i Oslo**

**Adresse: Gaustadalleen 21, 0349 Oslo**

8.30	Registrering
9.00	Sal: Forum <b>Velkommen:</b> Jan Hoel, leder i arrangementskomiteen for NFL 2013 <b>Åpning:</b> Arnfinn Muruvik Vonen, direktør i Språkrådet <b>Kunstnerisk innslag:</b> Liv Ulvik
9.30	<b>Fellesforelesning.</b> Ordstyrer: Jan Hoel <i>Sameness of meaning in a bilingual dictionary</i> Arleta Adamska-Sałaciak, Faculty of English, Adam Mickiewicz University, Poznań



KONFERANSEPROGRAM

10.30–11.00 **KAFFE**

	Sal: Forum <b>Semantikk og leksikologi</b> Ordstyrer: Ruth Vatvedt Fjeld	Sal: Ciens Via <b>Digitalisering av ord-boksverk</b> Ordstyrer: Anki Mattisson
11.00	<i>Automatisk identifisering av semantisk förändring med hjälp av lexikala distributionella mönster</i> Karin Cavallin	<i>Fra bog til base: om implementering af nye edb-programmer ved Ømålsordbogen</i> Henrik Hovmark
11.30	<i>Semantisk nærhet og synonymi i parataktisk og paradigmatiske perspektiv</i> Jón Hilmar Jónsson	<i>Digitalisering og oppkoding av Norges største rettskrivningsordbok</i> Boye Wangensteen og Øystein Eek
12.00	<i>Att läsa ordboken baklänges. Om betydelseutveckling hos fast sammansatta partikelverb och presentationen i SAOB</i> Pär Nilsson	<i>Nya utmaningar för Ordbok över äldre finskt skriftspråk</i> Tiina Arpiainen, Taru Laanti, Maria Lehtonen og Sanna Nissinen

12.30–13.45 **LUNSJ**

KONFERANSEPROGRAM

	Sal: Forum <b>Informasjonstyper i ordbøker</b> Ordstyrer: Jón Hilmar Jónsson	Sal: Ciens Via <b>Digitalisering av ordboksverk</b> Ordstyrer: Lena Rogström
13.45	<i>Bruklighetsoppgifter i svenska ordböcker: framväxt av en lexikografisk tradition</i> Lars-Erik Johansson	<i>Brugerne er allerede mobile – er leksikografien? En undersøgelse af den mobile brugssituation</i> Henrik Køhler Simonsen
14.15	<i>Bruklighets- og stilmarkører i en historisk ordbok</i> Anki Mattisson	<i>Lemmatisering och länkning i Ordbok över Finlands svenska folkmål på nätet</i> Caroline Sandström, Outi Lehtinen og Risto Widenius
Saler: Vestibulen (postere) og Ciens Via (demoer) Ordstyrer: Emma Sköldberg		
14.45–15.30	<b>DEMO</b> <i>Ordbok över Finlands svenska folkmål på nätet</i> Susanne Bergström	<b>POSTER og DEMO</b> <i>Nydigitalisering av SAOB</i> Erik Bäckerud
	<b>POSTER</b> <i>Norsk Tegnordbok</i> Irene Greftegreff	<b>POSTER</b> <i>Det Norske Akademis Store Ordbok</i> Carina Nilstun

15.30 **KAFFE og noe å bite i**

17.00–18.00	<p><b>Omvisning på Akershus slott – veterantrikk dit fra John Colletts plass kl. 16.10</b></p> <p>Middag etter omvisningen for egen regning. Bord er bestilt på en restaurant nær Akershus festning. Meld deg på liste ved registreringsranken!</p>
-------------	---

**Torsdag 15.8.2013 Forskningsparken ved Universitetet i Oslo**

**Adresse: Gaustadalleen 21, 0349 Oslo**

9.00	<p>Sal: Forum</p> <p><b>Fellesforelesning.</b> Ordstyrer: Caroline Sandström</p> <p><i>Leksikografisk tradition og fornyelse: digital revolution eller organisk reformisme?</i></p> <p>Lars Trap-Jensen, Det danske Sprog- og Litteraturselskab, København</p>	
	<p>Sal: Forum</p> <p><b>Lemmaseleksjon</b></p> <p>Ordstyrer: Øystein Baardsgaard</p>	<p>Sal: Ciens Via</p> <p><b>Rapport fra ordboksprosjekter</b></p> <p>Ordstyrer: Dagfinn Worren</p>
10.00	<p><i>Fællesdanske ord i en dialektordbog</i></p> <p>Asgerd Gudiksen</p>	<p><i>Planlegging og gjennomføring av eit stort vitenskapleg ordboksprosjekt</i></p> <p>Åse Wetås</p>
10.30	<p><i>Nyord i norsk</i></p> <p>Boye Wangenstein</p>	<p><i>Fra begreb til bog</i></p> <p>Sanni Nimb, Thomas Troelsgård, Lars Trap-Jensen, Henrik Lorentzen, Nicolai Sørensen, Liisa Theilgaard</p>

KONFERANSEPROGRAM

11.00 **KAFFE**

11.15	<i>Korpuslingvistikk, neologismar og ei tradisjonsrik ordliste</i> Margunn Rauset	<i>Kvensk ordbok på nett – hvem har nytte av den?</i> Mervi Haavisto, Kaisa Maliniemi, Leena Niiranen, Pirjo Paavaliemi og Trond Trosterud
11.45	<i>Et talesprogsbaseret lemmabegreb</i> Carsten Hansen	<i>Ordrelasjoner fra kollokasjoner</i> Lars Johnsen og Arne Martinus Lindstad

12.15–13.15 **LUNSJ**

KONFERANSEPROGRAM

	Sal: Forum Forts. <b>Lemmaseleksjon</b> Ordstyrer: Halldóra Jónsdóttir	Sal: Ciens Via Forts. <b>Rapport fra ord-boksprosjekter</b> Ordstyrer: Henrik Hovmark
13.15	<i>Leksikaliseringsprosesser i laust samansette verb: partisipp eller adjektiv?</i> Bjørghild Kjelsvik	<i>Bokmålsordboka og Nynorskordboka – tvillingar på 27</i> Marit Hovdenak
13.45	<i>Fypike gjorde selbureis i røytemåned. Om Norsk-russisk ordbok av Vladimir Arakin: utvalg, kilder, formål og målgruppe</i> Anna Senje	<i>Från aspekt till övergripande – en nätbaserad lista över svenskt akademiskt ordförråd</i> Håkan Jansson, Judy Ribbeck og Emma Sköldbberg
14.15	<i>Svenska fasta liknelser i en studieordbok</i> Aleksej Aljosjin	<i>LBK2013: Korpusressurs med mange anvendelsesmuligheter</i> Rune Lain Knudsen og Ruth Vatvedt Fjeld

14.45–15.00      **KAFFE**

KONFERANSEPROGRAM

<p>Saler: Vestibylen (postere) og Ciens Via (demoer)                  Ordstyrer: Arnfinn Muruvik Vonen</p>	
<p>15.00–15.45</p>	<p><b>POSTER</b>  <i>En korpusbaseret undersøgelse af interjektioner i talesprog</i>                  Jessie Leigh Nielsen</p>
	<p><b>POSTER</b>  <i>Ein studie av fleirordsuttrykk i norsk</i>                  Gyri Smørdal Losnegaard</p>
	<p><b>POSTER</b>  <i>Automatisk kollokasjonsanalyse</i>                  Ruth Vatvedt Fjeld, Rune Lain Knudsen, Eszter Horvati og Annalisa Miccoli</p>
	<p><b>DEMO</b>  <i>Brugerinterface for den flersproglige ordbog ISLEX</i>                  Halldóra Jónsdóttir og Þórdís Úlfarsdóttir</p>
<p>15.45–16.15</p>	<p>Sal: Forum                  Ordstyrer: Ruth Vatvedt Fjeld  <i>OENOLEX Bourgogne. Lær at sette ord på vin. Nyt vinordbogs-koncept, nye veje for brancheordbøger</i>                  Patrick Leroyer og Asta Høy</p>
<p>19.00</p>	<p>Konferansemiddag på Ekebergrestauranten</p>

**Fredag 16.8.2013 Forskningsparken ved Universitetet i Oslo**

**Adresse: Gaustadalleen 21, 0349 Oslo**

KONFERANSEPROGRAM

9.00	Sal: Forum <b>Fellesforelesning.</b> Ordstyrer: Jón Hilmar Jónsson <i>Samspeilet mellom leksikon og syntaks</i> Victoria Rosén, Universitetet i Bergen	
	Sal: <i>Forum</i> <b>Leksikografisk historie</b> Ordstyrer: Lars Trap-Jensen	Sal: Ciens Via <b>Leksikografisk terminologi</b> Ordstyrer: Åse Wetås
10.00	<i>Marcus Pauli Sadeleri – den finske medarbeitaren i Schroderus Lexicon Latino-Scondicum</i> Petri Lauerma	<i>Samvirket mellom ordbank og ordbok, og spørsmålet om namn på delane</i> Oddrun Grønvik og Christian-Emil Ore
10.30	<i>Ivar Aasen og Universitas Regia Fredericiana</i> Kjetil Gundersen og Dagfinn Worren	<i>Leksikografiens grunnenhet – hva heter det, dette vi lager ordbokartikler om?</i> Petter Henriksen

11.00 **KAFFE**

KONFERANSEPROGRAM

	<b>Forts. Leksikografisk historie</b> Ordstyrer: Lars Trap-Jensen	<b>Fagleksikografi</b> Ordstyrer: Henrik Lorentzen
11.15	<i>Petter Dass' ordförråd – hans kilder og hans forfatterskap som kilde for ordböcker</i> Hanne Lauvstad	<i>Från dilation till uppskov. Vilken roll spelade ordböcker i etableringen av ett juridiskt ordförråd?</i> Lena Rogström
	<b>Terminologi og leksikografi</b> Ordstyrer: Victoria Rosén	<b>To- og flerspråksleksikografi</b> Ordstyrer: Henrik Lorentzen
11.45	<i>Terminologiprojektene CLARINO og Maritim ordbok og deres relevans for ordboksarbeid</i> Marita Kristiansen, Gisle Andersen og Vemund Olstad	<i>Tvåspråkiga morfemordböcker för inlärare</i> Jonas Winnerlöv

12.15–13.15      **LUNSJ**



KONFERANSEPROGRAM

	Sal: <i>Forum</i> <b>Flerordsenheter og leksikografi</b> Ordstyrer: Victoria Rosén	Sal: Ciens Via Forts. <b>To- og flerspråksleksikografi</b> Ordstyrer: Marit Hovdenak
13.15	<i>Skjematiske fleirordskonstruksjonar i Norsk Ordbok</i> Sturla Berg-Olsen	<i>Språkproven i ISLEX – problem och potential</i> Anna Helga Hannesdóttir
13.45		<i>Nordisk miniordbok – et kritisk blick</i> Ruth Vatvedt Fjeld og Janne Bondi Johannessen

14.15 **KAFFE**

14.30	Sal: <i>Forum</i> <b>Generalforsamling</b>
15.30	<b>Avslutning</b>