

Nordiska studier i lexikografi
15

Nordiska studier i lexikografi 15

Rapport från 15 konferensen
om lexikografi i Norden

Helsingfors 4–7 juni 2019

Redigerad av

Caroline Sandström

Ulla-Maija Forsberg

Charlotta af Hällström-Reijonen

Maria Lehtonen

Klaas Ruppel

Skrifter utgivna av Nordiska föreningen för lexikografi

Skrift nr 16

I samarbete med Institutet för de inhemska språken

Helsingfors 2020

Nordiska studier i lexikografi 15

© Nordiska föreningen för lexikografi och författarna, 2020

Redaktion:

Caroline Sandström

Ulla-Maija Forsberg

Charlotta af Hällström-Reijonen

Maria Lehtonen

Klaas Ruppel

Redaktionssekreterare: Caroline Elgert

Ombrytning: Caroline Elgert

Tryck: Edita Prima Oy, Helsingfors 2020

ISBN: 978-952-7359-03-7 (Nordiska studier i lexikografi 15)

ISSN: 1894-4663 (Nordiske studier i leksikografi, tryck)

ISSN: 2246-7823 (Nordiske studier i leksikografi, online)

ISSN: 0803-9313 (Skrifter udgivet af Nordisk Forening for Leksikografi)

Utgiven med stöd från:



Innehåll

Förord	1
Plenarföreläsningar	
<i>Anna Helga Hannesdóttir</i>	
Den tvåspråkiga lexikografins utmaningar förr och nu – eller: från lertavlor till Google Translate	5
<i>Nina Martola</i>	
Den tvåspråkiga ordboken – ett absurt åtagande	25
<i>Iztok Kosem</i>	
Abstrakt: Responsive dictionaries: bringing dictionaries (and lexicographers) and users closer together.	43
Temasektion	
<i>Santeri Junttila</i>	
Temasektion om etymologi	47
Sektioner	
<i>Øivin Andersen</i>	
Enspråklig ordbok for fremmedspråklige	51
<i>Sturla Berg-Olsen & Daniel Gusfre Ims</i>	
Eit dilemma om lemma.	61
<i>Kristín Bjarnadóttir & Kristín Ingibjörg Hlynisdóttir</i>	
Online Data on Icelandic Inflection. Descriptive or Prescriptive: “Why, for whom, by whom” and how?	71
<i>Kristian Blensenius & Lena Rogström</i>	
Att hantera grammatisk förändring i en deskriptiv ordbok.	81

<i>Erik Bäckerud, Pär Nilsson & Emma Sköldberg</i>	
Så används Svenska Akademiens ordböcker på nätet. Implicit och explicit feedback från användarna	91
<i>Karin Cavallin</i>	
Ngram i SAOB	103
<i>Hinrik Hafsteinsson & Einar Freyr Sigurðsson</i>	
A word frequency dictionary of Icelandic child-directed speech	115
<i>Poul Hansen</i>	
Hvordan har brugen af mobiltelefoner indvirket på efterspørgslen efter netordlister?	125
<i>Inger Schoonderbeek Hansen</i>	
<i>Jysk Ordbog</i> – af hvem, til hvem og hvorfor?	135
<i>Henrik Hovmark</i>	
<i>Ømålsordbogens</i> brugere – nu og i fremtiden	145
<i>Jonas Jensen, Henrik Lorentzen & Mette-Marie Møller Svendsen</i>	
Lad brugeren komme til orde! Hvordan får ordbogen mest muligt ud af brugerhenvendelser?	155
<i>Ellert Þór Jóhannsson & Simonetta Battista</i>	
Brugere og forskellig anvendelse af ONP	165
<i>Halldóra Jónsdóttir & Þórdís Úlfarsdóttir</i>	
Omdannelsen af en flersproget til en monolingval ordbog.	175
<i>Lennart Larsson</i>	
Total bruklighetsinskränkning i SAOB – nu och i framtiden	187
<i>Hanne Lauvstad</i>	
NAOBs etymologier	195

<i>Patrick Leroyer & Henrik Køhler Simonsen</i>	
Multimediale datatyper som leksikografisk merværdi	205
<i>Gunn Inger Lyse</i>	
Ut med <i>adamsslekt</i> og inn med <i>arveprinsesse</i> ? Leksikografiske metodar i revisjonen av <i>Bokmålsordboka</i> og <i>Nynorskordboka</i>	215
<i>Monica Martens</i>	
Redigeringsmiljøen för <i>Stora finsk-svenska ordboken</i> – återbruk, samarbete och vidareutveckling	225
<i>Johan Myking, Oddrun Grønvik & Stig Jarle Helset</i>	
Nettbrukaren i fokus: digitalisering og revisjon av <i>Norsk Ordbok a–h</i>	235
<i>Carina Nilstun</i>	
Den digitale ordbokens innvirkning på den praktiske leksikografien	245
<i>Christian-Emil Smith Ore</i>	
Å ta Hans Ross på ordet: Ross' ordbok i relasjon til Aasens med <i>Metaordboka</i> som verktøy	253
<i>Elina Palola & Petri Lauerma</i>	
The roles of Rasmus Rask and Gustaf Renvall in the publishing and editing of Christfrid Ganander's <i>Nytt Finskt Lexicon</i>	265
<i>Caroline Sandström & Harriet Eriksson</i>	
Suffixavledda verb av finskt ursprung i de svenska dialekterna i Finland	275
<i>Johan Schalin</i>	
Finnic <i>*litna</i> 'town, castle', a possible alternative origin.	285
<i>Bente Selback</i>	
«Å nei, det ordet er ikkje lov på nynorsk!» Eller ...? Om parallell redigering av to norske ordbøker.	297

<i>Henrik Køhler Simonsen</i>	
Augmented Writing: nye muligheder og nye teorier	307
<i>Ásta Svavarsdóttir</i>	
Islandske ordbøger i 1800-tallet og deres rolle i sprogstandardiseringen	317
<i>Mette-Marie Møller Svendsen, Nicolai Hartvig Sørensen & Thomas Troelsgård</i>	
Superordbog og salmesang: ordbogslinkning i praksis	327
<i>Ágústa Þorbergsdóttir</i>	
Tilblivelsen af fagordbøger og termsamlinger i Island.	337
<i>Lars Trap-Jensen</i>	
Normering af ord uden normering – i teori og praksis	347
Konferensprogram	357

Förord

Den 15 konferensen om lexikografi i Norden hölls 4–7 juni 2019 vid Helsingfors universitet. Temat för konferensen var *Ordböcker av vem, för vem och varför?* Under konferensen kunde aktörer inom nordisk lexikografi presentera pågående projekt och nya forskningsrön. Det fördes diskussioner om aktuella utmaningar och möjligheter som förändringarna i det digitala samhället har medfört. Tryggheten av resurser för lexikografiskt expertarbete med redigering, uppdatering och publicering av kvalitativa ordböcker var ett viktigt diskussionsämne. För såväl lexikografer som språkteknologer gav utbytet av erfarenheter och gemensamma diskussioner under konferensen goda förutsättningar att utveckla lexikografiskt arbete i Norden.

Konferensen arrangerades av Institutet för de inhemska språken i samarbete med Nordiska föreningen för lexikografi (NFL) och Helsingfors universitet. Huvudansvarig arrangör var Caroline Sandström från Institutet för de inhemska språken. I arrangörgruppen ingick Ulla-Maija Forsberg, Charlotta af Hällström-Reijonen, Lotta Jalava, Maria Lehtonen och Klaas Ruppel från Institutet för de inhemska språken samt Santeri Junttila från Helsingfors universitet. Konferensen hade sammanlagt 110 deltagare från Finland, Sverige, Norge, Danmark, Island, Färöarna, Tyskland, Estland och Slovenien.

Tre plenarföreläsningar som anslöt sig till temat hölls på konferensen: Anna Helga Hannesdóttir (Göteborg) *Den tvåspråkiga lexikografins utmaningar förr och nu*, Iztok Kosem (Ljubljana) *Responsive dictionaries: bringing dictionaries (and lexicographers) and users closer together* samt Nina Martola (Helsingfors) *Den tvåspråkiga ordboken – ett absurt åtagande*. Två av dessa, Hannesdóttirs och Martolas föreläsningar, är publicerade i denna volym. Av Iztok Kosems föreläsning finns abstraktet publicerat i volymen. Nytt för denna konferens var att plenarföreläsningarna på initiativ från NFL:s styrelse strömmades på nätet, vilket var möjligt tack vare ett stipendium från Svenska litteratursällskapet i Finland. Föreläsningarna finns tillgängliga på nätet (Hannesdóttir 5.6.2019 youtu.be/knY5bxp1r40, (Kosem 6.6.2019) youtu.be/vJGXwvK682E och (Martola 7.6.2019) youtu.be/9n0R43sxnY. Länkar till plenarföreläsningarna finns även på NFL:s webbplats nordisk-leksikografi.com.

På konferensen hölls, förutom de tre plenarföreläsningarna, 41 sektionsföredrag, sex posterpresentationer och en temasektion om etymologi med åtta presentationer. Ungefär en fjärdedel av bidragen var inriktade på konferensens tema, vilket innebär att ordböckernas situation i ett flertal lexikografiska miljöer i Norden belystes i föredrag och presentationer på konferensen.

I tillägg till det fackliga programmet arrangerades även socialt program under konferensen, vilket erbjöd möjligheter till kollegiala samtal och fördjupade kontakter mellan kolleger som träffats förut men också tillfällen att stifta nya bekant-

skaper. Konferensen inleddes med en mottagning på Institutet för de inhemska språken på Hagnäskajen 6 i Helsingfors. Efter den första konferensdagen hölls en mottagning på Helsingfors universitet. Följande eftermiddag gjordes en utflykt till Villa Hvitträsk, ett arkitektmuseum i Kyrkslätt. Utflykten avslutades med en middag på restaurang Haltia i Esbo. Lunch och kaffe avnjöts i Forsthuset på Helsingfors universitet i anslutning till föreläsningssalarna vilket ökade chanserna att upprätthålla kollegiala samtal under pauser. Konferensen gynnades också av varmt och vackert sommarväder.

I konferensrapporten ingår 30 artiklar som bygger på sektionföredrag eller posterpresentationer, två av de tre plenarföreläsarnas bidrag samt en presentation av temasektionen om etymologi. Merparten av artiklarna är skrivna på danska, norskt bokmål, nynorska eller svenska, men det finns också en handfull artiklar på engelska. Redaktionssekreteraren Caroline Elgert har redigerat artiklarna och gjort ombrytningen för rapporten. Bidragen har alla genomgått kollegial fackgranskning.

I redaktionen för konferensrapporten har Caroline Sandström, Ulla-Maija Forsberg, Charlotta af Hällström-Reijonen, Maria Lehtonen och Klaas Ruppel medverkat. När det gäller den tekniska utformningen av rapporten har redaktionen fått stöd av fil.mag. Jamima Löfström på Institutet för de inhemska språken.

Ekonomiskt erhöles bidrag för konferensen och arbetet med konferensrapporten från Nordplus (nordiska språk), Svenska litteratursällskapet i Finland, Vetenskapliga samfundens delegation, Helsingfors universitet, Institutet för de inhemska språken och Konestiftelsen. Vi riktar ett varmt tack till alla som genom ekonomiskt bidrag möjliggjorde konferensen och konferensrapporten. Vid planering och arrangemang har det kollegiala stödet och erfarenheterna från Nordiska föreningen för lexikografi varit ovärderliga.

Helsingfors juni 2020

Caroline Sandström
Ulla-Maija Forsberg
Charlotta af Hällström-Reijonen
Maria Lehtonen
Klaas Ruppel

Plenarföreläsningar

Den tvåspråkiga lexikografins utmaningar förr och nu – eller: från lertavlor till Google Translate

Anna Helga Hannesdóttir

The history of bilingual lexicography starts in the third millennium B.C. From the age of glosses carved in clay tablets through the media revolution, when printed books replaced manuscripts, little happened until the end of the 20th century. Then the concept of “translation equivalent”, hitherto solid as a rock, was questioned by the translation science. With the e-dictionaries bilingual lexicography is now at a crossing. The users take multiple access to all information coded in the dictionary for granted. The established theoretical dichotomy between source language vs. target language and between L1 and L2 perspective is no longer valid. And then there is Google Translate, which now – as shown below – provides adequate translations to an increasing extent. The future challenge for bilingual lexicography lies in adjusting the theory of lexicography and the corresponding methods to the possibilities offered by technology and the demands of the users.

1. Inledning

Den tvåspråkiga lexikografen är en utmaning i sig. Den går ut på att med de enskilda orden som utgångspunkt spegla den kultur, den livssyn och världsuppfattning som kodats i ett språk och förmedla allt detta på ett annat språk. Just denna verksamhet är den äldsta språkvetenskapliga genren, och i lexikografihistoriska sammanhang sträcker sig ”förr” inte sekler utan millennier tillbaka i tiden. Hur den nutida lexikografen klarar de utmaningar den står inför får framtiden utvisa. Här nedan följer först ett svep över lexikografins långa historia och den tidiga lexikografins olika drivkrafter – där vi finner såväl ideologiska som praktiska sådana – liksom de svårigheter som själva beskrivningsobjektet har erbjudit. Därefter diskuteras den moderna tvåspråkiga lexikografen, dess utgångspunkter och syften samt de utmaningar och möjligheter som den nu står inför.

När det gäller de lexikografiska genrernas status idag är den tvåspråkiga lexikografen sekundär till den enspråkiga lexikografen. Relationen mellan de båda genrerna brukar bland annat avspeglas i programmet för de stora lexikografiska konferenserna, så också i den konferens som dokumenteras i denna rapport. Lexikografins teoretiska påbyggnad är mycket sen, och den tekniska utveckling som vi nu är mitt uppe i innebär fantastiska möjligheter och – för den tvåspråkiga lexikografen – också en del utmaningar. Här nedan handlar det, om inget annat sägs, om två- eller flerspråkig lexikografi.

2. De första tre tusen åren – lexikografi och glossografi

De äldsta spåren av lexikografisk verksamhet dateras till 2400-talet före vår tideräkning. Vid arkeologiska utgrävningar i Mellanöstern har flera fynd gjorts av lertavlor med lexikografiska uppställningar. De hittills äldsta fynden påträffades på 1970-talet vid en utgrävning i trakten av Aleppo i norra Syrien. Bland hundratalens lertavlor med texter huvudsakligen av administrativ art fanns också tavlor där parallella spalter med eblaitiska ord och deras sumeriska motsvarigheter stämplats in: "världens äldsta lexika!" (Linnér 2002:45). Lertavlornas inskrifter är svårtolkade, de språk som där är dokumenterade är döda sedan länge. Lika som andra texter från den tiden och den delen av världen är de skrivna med kilskrift. Något yngre är de ordlistor där sumeriska ord förses med uttalsuppgifter och akkadiska motsvarigheter och, i vissa fall, med motsvarigheter även från andra språk (Snell-Hornby 1986). Trots att de texter som påträffats huvudsakligen innehåller information av administrativ natur förekommer även texter som anses ha använts i skolsammanhang.

2.1. Greker och romare träder in

Från det att de eblaitiska och sumeriska orden stämplades in i lertavlorna dröjer det närmare 2000 år tills den gamla grekiskan gör entré i lexikografihistorien. Främst rör det sig då om att för senare generationers greker förklara dialektala, ålderdomliga och på andra sätt svåra ord i de stora grekiska författarnas redan sekelgamla texter. Här är det snarare fråga om *glossografi*, dvs. uppteckningar av svåra ord med förklaringar på samma språk, än om tvåspråkig lexikografi (*Store norske leksikon*, s.v. *glosse*). Från 500-talet f.Kr. finns exempelvis glossarier över dunkla ord i Homeros då drygt 200-åriga texter (Collison 1982:26).

Den till sin natur enspråkiga glossografin med fokus på föråldrade, litterära ord spelade fortfarande en viktig roll under den första tiden av vår tideräkning:

[...] the pattern in the first-century glossaries was very similar to that of previous centuries: the preoccupation with Homer's language, with the difficult words used by poets and dramatists, and – above all – with literary usage, was only broken by Valerius Harpocration's quotations from the spoken word. (Collison 1982:33)

Det är alltså först på 100-talet e.Kr. som det samtida språket införlivas i glossografin genom greken Harpocration. Glossarier var dock inte bara en grekisk genre. Att också romarna sammanställde glossarier över gamla eller av andra orsaker svåra ord i latinet är välkänt. De upprättade också förteckningar över svåra enheter i främmande källspråk och försåg dem med förklaringar på latin.

Dessa ordlistor var inte avsedda att främja kommunikation med talare av främmande språk utan att underlätta läsningen av litterära texter, främst de grekiska. Samtidigt, dvs. under de första seklen av vår tideräkning, utvecklas enspråkiga latinska ordböcker.

2.2. Medeltidens lexikografi

Under de därpå följande seklen, fram till den medierevolution som boktryckarkonstens uppfinnande innebar, fortsätter den lexikografiska verksamheten i Europa efter i stort sett samma linjer. Utgångspunkten var i första hand att i ett slags bildningssträvan tillgängliggöra antika texter avfattade på ett främmande språk snarare än att främja aktiv personlig kommunikation. I samband med kristendomens utbredning skapades dock ett behov av hjälpmedel för kyrkans kontakt med lokala folkgrupper. Under medeltiden dominerade latinet i stora delar av Europa som kyrkans, kulturens och bildningens språk. Manuskript skickades kors och tvärs över kontinenten mellan kloster och skriptorier för kopiering och översättning, en verksamhet som bland annat genererade ordlistor mellan latinet och diverse folkspråk.

Från det nordiska området finns en del sådana ordlistor bevarade i de medeltida handskrifterna: mellan latin och isländska från sekelskiftet 1200 samt något senare på 1200-talet och från 1300-talet också mellan latin och norska respektive danska. Även latinsk–svenska gloslistor finns bevarade i diverse handskrifter. Den äldsta av dessa, innehållande ”en række verber og en gruppe fiskenavne”, återfinns i en handskrift från 1280-talet, Cod. Holm. B 59 (Skautrup 1960, spalt 360). I de allra flesta fall rör det sig bokstavligen om enkla gloslistor, några få latinska ord försedda med en ekvivalent på respektive folkspråk (Raschellà 2001).

Därmed är vi framme vid de första spåren av lexikografisk verksamhet i Norden: nästan 4000 år efter det att de eblaitiska–sumeriska ordlistorna stämplades in i lertavlorna. År 1286 färdigställde Johannes Balbus de Janua den latinska ordboken, encyklopedin och grammatiken *Catholicon*, ett arbete som under de följande århundradena flitigt kopierades och spreds. Tidigt under 1400-talet fanns en avskrift av *Catholicon* i Vadstena kloster; en avskrift som antas ha legat till grund för den första större ordlistan där svenska förtecknas (Neuman 1973). Denna latinsk–svenska ordlista finns utgiven i *Svenska fornscriftsällskapets samlingar* (SFSS 45)¹.

¹ Verk som endast nämns flyktigt i framställningen har inte tagits med i litteraturlistan.

3. De tryckta ordböckernas tid

Tryckerikonsten revolutionerade produktion och spridning av skriven text och naturligtvis också av lexikografiska uppställningar. Nu massproducerades och spreds ordlistor och ordböcker av olika slag istället för att, som tidigare, framställas genom den tidsödande manuella kopieringen av texter, som i sin tur ofta var avskrifter av ytterligare andra avskrifter. Av de nordiska språken är det danskan som först återges med hjälp av den nya tekniken. Den första latinsk–danska ordlistan, tillskriven bibelöversättaren Christiern Pedersen, trycktes i Paris 1510. Syftet med den var att ”åbne adgangen til antikkens litt[eratur]” (Karker 1967, spalt 638–640). Några decennier senare trycktes i Stockholm den första latinsk–svenska ordlistan, *Variarum rerum vocabula* (1538), en ordlista med vilken ”Nysvenskans språkvetenskapliga litteratur tager, om ock blott i ganska oegentlig mening, sin svaga början” (Noreen 1903:183).

Båda dessa arbeten har kontinentala föregångare; de innehåller latinska uppslagsord och danska respektive svenska motsvarigheter. Syftet med dem var didaktiskt, de var avsedda för undervisningen i latin och vände sig således huvudsakligen till den studerande ungdomen. I Sverige trycktes flera nya upplagor av *Variarum rerum*, och den översattes också till danska.

När den lexikografiska verksamheten väl kommit igång, ökade aktiviteten successivt. Under 1500-talet trycktes flera ordböcker mellan latin och danska, och i Sverige publicerades en svensk–grekisk–latinsk ordbok 1587. Där fungerade för första gången svenskan som källspråk.

3.1. Inte bara latin

Det fanns naturligtvis ett behov av hjälpmedel för kommunikation på andra språk än latin. I Sverige trycktes i slutet av 1620-talet en liten, praktiskt inriktad tysk–svensk parlör och ordlista. I inledningen marknadsför utgivaren sin produkt och vänder sig direkt till läsaren med en välformulerad varudeklaration och argument för nyttan av språkkunskaper och därmed också för behovet av denna bok:

Til läsaren: Käre läsare, thenna Book är så beqvem och nyttigh, och hennes bruuk så höghnödigt, at hon icke noghsamt kan beprijnsas. Ty thet är ingen, hwarken i Swerige eller i Tyskland, som här drifwer någhon Köp-handel, som icke behöfwer thesse tu Språken, som vthi thenna Book äre beskreffne. [---] Hwem hafwer någhon tidh bekommit fremmande Folcks wänskap medh itt Språk allena? [---] Hwem kan wäl regera Städer och Land, om han intet annat Språk eller Tungomål kan vthan allena sitt Modersmål? (*Samtaal på svenska och tyska*, s. [1–3])

Syftet med den lilla publikationen är uttryckligen att underlätta kommunikationen mellan tyskar och svenskar, bland annat i handelssituationer. Lexikografiska hjälpmedel behövdes också för att möjliggöra kommunikation vid mindre etablerad kontakt mellan talare av olika språk. Fragment av gloslistor som tillkommit i samband med språkkontakt mellan basker och islänningar har bevarats från 1600-talet: "A Basque-Icelandic glossary is undoubtedly an exotic item, it is perhaps surprising to find therefore, not one, but four of them [...]" (Etxepare & Miglio 2015). Dessa uppteckningar av glosor var avsedda att underlätta den direkta kommunikationen mellan lokalbefolkningen på Islands nordvästligaste hörn och de baskiska sjömän och valfångare som periodvis under 1600-talet bedrev fiske utanför Islands nord- och västkust. Ordförrådet betecknar huvudsakligen konkreta ord och vardagliga fraser.

På 1630-talet trycktes i Köpenhamn en liten ordlista där norska termer främst inom juridiken men även inom andra områden förklaras på danska (Bjelke 1634), och några år senare inlemmas finskan i en ordlista med latin, svenska och tyska (Schroderus 1637). Även om latinet i samband med reformationen förlorede mycket av den religiösa domänen till folkspråken behöll det länge sin roll som lexikografins metaspråk i ordböcker mellan två eller flera folkspråk.

3.2. Ideologiska utmaningar

Under 1600-talet utgjorde den världsbild som Bibeln förmedlade fortfarande utgångspunkten för kunskap och vetande. Synen på språket byggde på den bibliska skapelseberättelsen: språket var människan givet av Gud själv. Alla språk antogs på något sätt härstamma från ett gemensamt urspråk; det språk som Noa och hans söner spred över jorden efter att ha överlevt syndafloden. Mycken möda gick åt till att argumentera för det egna språkets speciellt nära släktskap med detta urspråk, och lexikografiska projekt startades med främsta syfte att bevisa denna relation.

Ett viktigt framsteg i danskars och svenskar positionering i den frågan gjordes när de isländska medeltidstexterna blev kända i Skandinavien. Kring dessa texter utvecklades en egen variant av den kontinentala renässansen, ett slags isländsk renässans med starka inslag av göticism och patriotism (Ralph 2016). I de isländska sagorna och eddorna beskrivs hjältars och asars djärvhet och ärorika bragder på ett kärnfullt och mustigt nordiskt urspråk. Innehållet betraktades som ett nordiskt kulturarv som ansågs vara väl i paritet med det klassiska antika. Med upptäckten av de isländska texterna gör ideologiska aspekter på allvar sitt intåg i den nordiska lexikografen. Då ställdes också danskar och svenskar inför samma utmaning som romarna tidigare inför de grekiska texterna: varken

danskar eller svenskar förstod det språk som dessa berättelser var nedtecknade på, alltså isländska. Texterna måste översättas och för det behövde man bland annat ordböcker.

En som tog initiativet till en ordbok över språket i de isländska medeltidstexterna var Ole Worm, en dansk antikvarie och läkare. Worm var en mångsidig och framstående forskare som undersökte och kartlade Danmarks fornlämningar, däribland runinskrifterna. Han var övertygad om att också de äldsta isländska texterna ursprungligen måste ha skrivits med runor. Runorna såg han som en förbindelselänk mellan den nordiska kulturen, de heroiska goternas kultur och den klassiska grekiska kulturen. Och visst anade man likheter mellan runorna och tecknen i det gotiska alfabetet – och till och med i det hebreiska? Så det var inte bara en ordbok över språket i de isländska handskrifterna det gällde; en forntida isländsk litteratur skriven med runor skulle vara ett starkt argument för nordbornas (främst danskarnas) klassiska arv och rättmätiga plats i mänsklighetens tidiga historia. Som Considine har påpekat var fornforskning av olika slag ett viktigt inslag i de nationsbyggen som pågick under 1600-talet: "[L]exicography and the study of the material heritage of landscape and monuments went together, and were associated with national identity" (Considine 2008:243).

År 1650 publicerades till slut den första ordbok som innehåller isländska: Ole Worms *Specimen Lexici Runici*. Trots att de isländska medeltidstexterna aldrig skrevs med runor återges det isländska lemmat med runor och en transkription med latinska bokstäver, följt av en motsvarighet på latin. Därefter ges belägg som hämtats ur de isländska medeltidshandskrifterna. Dessa förses med latinsk översättning. *Specimen* manifesterar hur drivkraften bakom ett tvåspråkigt ordboksprojekt i äldre tid kunde vara något helt annat än det vi nu förknippar med sådana projekt, dvs. att underlätta studier, översättning och aktiv kommunikation. Till skillnad från såväl dagens ordböcker som de kommunikativt och praktiskt inriktade baskisk–isländska gloslistorna var skandinavernas lexikografiska behandling av isländskan inte avsedd att främja kontakten med de samtida islänningarna. Både Worms *Specimen* och den svenska motsvarigheten, Verelius ordbok över det nordiska medeltidsspråket (1691), var först och främst avsedda att tillgängliggöra den isländska medeltidslitteraturen, på liknande sätt som de latinska ordlistorna över den antika grekiskan. Worms och Verelius ordböcker utgjorde också ett instrument i en statsideologi; de bottnade i en patriotisk strävan och självhävdelse, samtidigt som de utgjorde ett bidrag till tidens vetenskapliga framsteg (Ralph 2016).

3.3. De språkliga utmaningarna

Under 1600-talet och in på 1700-talet fanns det också rent språkliga utmaningar som lexikograferna hade att förhålla sig till. Framför allt var det ont om ord! Folkspråkens ordförråd var i stort behov av modernisering och utökning för att täcka nya begrepp. Särskilt torftigt var ordförrådet inom de domäner som folkspråken då nyligen övertagit från latinet. I göticistisk anda förordades att gamla, bortglömda ord skulle återuppväckas och tas i bruk, såväl ord från medeltidstexterna som ord hämtade ur reformationsbibeln, liksom dialektala och folkliga ord. Dessutom kunde nya ord skapas av inhemska element eller lånas in från andra språk. Problemet var naturligtvis högst relevant för lexikograferna, och dessa var också tongivande i diskussionen (Hannesdóttir 2007). För de svenska lexikograferna var också, förutom ordförrådets behov av förnyelse, det dåligt standardiserade svenska språket ett problem i sig. Ortografin var ännu inte reglerad, inte ens uppsättningen bokstäver i det svenska alfabetet var fastställd och inte heller alla detaljer i deras inbördes ordning. Dessa omständigheter var onekligen en komplikation för såväl författare som användare av alfabetiskt ordnade ordböcker.

Det är i 1700-talets tvåspråkiga ordböcker som den vetenskapliga beskrivningen av svenskan tar fart (Hannesdóttir 2000). Den lexikografiska beskrivningen utvecklas generellt. Ordförrådet konsolideras och informationen om de svenska uppslagsorden utvecklas genom att deras grammatiska egenskaper beskrivs. Med tiden kompletteras de också med diasystematiska uppgifter. Även den semantiska analysen av den främmande källspråksenheten skärps, och betydelsebeskrivningen i form av en svensk målspråkmotsvarighet förfinas. Denna kan redovisas i form av en ekvivalent eller en räkka av flera ekvivalenter, i form av parafraaser eller regelrätta definitioner av aristoteliskt slag. Och inte minst viktigt är att den lexikografiska beskrivningen av svenskan tidigt återspeglar just svenskans egenskaper medan grammatikornas beskrivning betydligt längre höll fast vid latinets beskrivningsmallar även för svenskan. Den tvåspråkiga lexikografin spelade en mycket viktig, och länge underskattad, roll både i svenskans standardisering och i den grammatiska beskrivningen (Hannesdóttir 2000).

Under 1700-talet vidgades världen, och nya språkkonstellationer tillkom i ordböckerna. De politiska förhållandena i Norden avspeglas i ordboksbeståndet. Förutom tidiga ordlistor mellan norska och danska genererade den danska missionen på Grönland en grönländsk–dansk ordbok. Också grammatikor över samiska varieteter och ordböcker mellan danska och samiska publicerades, dels för att öka kunskapen om samernas språk, dels för att underlätta kommunikationen mellan dem och den danskspråkiga administrationen. Inom det svenska riket tillkom ordböcker mellan svenska och finska och även mellan svenska och samiska.

4. Den enspråkiga lexikografins premiär

Under 1800-talets förra hälft publicerades de första enspråkiga danska och svenska ordböckerna. I Sverige inleddes den enspråkiga eran med ett par överambitiösa projekt (Hannesdóttir 1998:365–461). Första försöket utgjordes av Kindblads ordbok (1840). Denna avstannade en bit in i bokstaven *A* (mitt i artikeln *Afhackning*). Därefter tog författaren och skolmannen Almqvist vid. Hans ordbok (1842) omfattar orden på *A* och en del av ordförrådet på *B* (t.o.m. *Brambär*). Första fullständiga enspråkiga svenska ordboken, *Ordbok över svenska språket*, var färdig 1855. Upphovsmannen, Anders Fredrik Dalin, hade ett förflutet inom den tvåspråkiga lexikografen. Bland annat gav han på 1840-talet ut en stor fransk–svensk ordbok, som i stor utsträckning byggde på Franska akademiens ordbok. Hans arbete med de svenska motsvarigheterna till de franska uppslagsorden har avsatt tydliga spår i lexemindelning och betydelsebeskrivning i hans svenska ordbok (Malmgren 1988; Hannesdóttir 1991).

5. Lexikografihistoriens relevans för dagens tvåspråkiga lexikografi

I sin plenarföreläsning vid den sjunde Euralexkonferensen (1996 i Göteborg) konstaterade Atkins att "Change is not something that people tend to associate with dictionaries. [...] The dictionary of the present is at heart little different from the dictionary of the past" (Atkins 1996:515).

Kan då tusenåriga lösningar förklara något över huvud taget? Möjligen, det kanske inte bara är den hävdvunna artikelstrukturen, som följt genren sedan urminnes tider, som består. Även arvet efter de tidiga ordböckernas funktioner lever kvar i den tryckta ordbokens organisation. De tidiga ordböckerna användes för envägskommunikation, för att avkoda texter från sedan länge svunna tider, texter som avspeglade passerade kulturstadier och var avfattade på döda språk eller språkformer. Risken för kommunikativa kollapsar var liten. Även arbetet med översättning från språk som inte hade infödda talare utfördes företrädesvis i miljöer där målspråkmotsvarigheterna tenderade att standardiseras i form av översättningsekvivalenter (Raschellà 2001). Kanske är det först under de senaste tre-, fyrahundra åren som den tvåspråkiga lexikografen på allvar har varit avsedd att överbrygga den klyfta som uppstår i mötet mellan två levande språk som representerar var sin livssyn och världsuppfattning, inbäddade i respektive språks lexikaliska och strukturella system.

5.1. Den tvåspråkiga lexikografins teoretisering

Sent under 1900-talet akademiserades lexikografen som disciplin, och en teoretisk överbyggnad utvecklades också för den tvåspråkiga lexikografen. Ordboksfunktioner definierades och beskrevs. Den tvåspråkiga ordboken skulle anpassas till en specifik verksamhet såsom exempelvis avkodning, fri textproduktion, översättning osv. Också relationen mellan de avsedda användarnas modersmål (L1) och dess funktion som källspråk respektive målspråk i ordboken fokuserades. Vilket av ordbokens språk, källspråket eller målspråket, som var användarnas modersmål var avgörande för vilken information som lämnades om respektive språk och för hur denna presenterades. En ordbok för källspråkstalande användare (L1>L2) är således avsedd för produktion på det främmande målspråket (L2), medan en ordbok för målspråkstalande användare är avsedd för avkodning av text på det främmande källspråket (L2>L1). Denna funktionsuppdelning kan även i praktiken vara relevant för ordböcker mellan språk som har många talare, men den är inte praktiskt genomförbar i samma utsträckning för ordböcker mellan språk med små marknader och få talare. Där har det snarare varit multifunktionalitet som varit modellen: så många funktioner som möjligt i en och samma ordbok. En sådan lösning är långt ifrån renlärig men ekonomiskt nödvändig.

5.2. Den tvåspråkiga lexikografen och översättningsvetenskapen

I samband med att teorin utvecklades, betonades också lexikografens nära koppling till översättningsvetenskapen eller translatologin. Kärnan i den tvåspråkiga lexikografen är ekvivalens och ekvivalenter. Det lexikografiska ekvivalensbegreppet är förvillande likt ekvivalensbegreppet inom translatologin men samtidigt väsensskilt. Medan översättaren arbetar med texter, fokuserar lexikografen på ord. Denna diskrepans har sedan länge uppmärksamats och problematiserats av translatologer som exempelvis Snell-Hornby på 1980-talet och Koller från 1990-talet och framåt. Snell-Hornby förankrar föreställningen om en ”translation equivalent” i det uråldriga studiet av (mer eller mindre) döda språk, där kommunikation med infödda talare av språket i fråga inte var aktuell (Snell-Hornby 1986:214; se också Zgusta 1984). Förutsättningarna är fundamentalt annorlunda beskaffade när det gäller avkodning av texter på döda språk och även för användning av språk utan modersmålstalare, alltså som lingua franca, än vid konfrontationen med en ”fuzzy, shifting, multi-dimensional world of a living foreign language” (Snell-Hornby 1986:214).

Den inom lexikografin etablerade termen ”translation equivalent” är olycklig på många sätt, menar Snell-Hornby (1986) vidare. Både *translation* och *equivalent* kännetecknas av vaghet: ”translation involves texts and not words, and equivalence is for the most part an illusion” (Snell-Hornby 1986:214). Den tråden tog Adamska-Sałaciak upp i sin plenarföreläsning vid den fjortonde Euralexkonferensen (2010 i Leeuwarden). Hon problematiserade lexikografens uppgift: går den ut på att hitta den rätta ekvivalenten, dvs. som Atkins och Rundell formulerar saken ”to find a TL-word” (Atkins & Rundell 2008:503), eller är lexikografens uppgift att etablera en ekvivalensrelation mellan källspråksenheten och enheter i målspråket? Frågan är om ord kan översättas med ord, dvs. om ordekvivalens är något som ”finns” och som det gäller att hitta eller om ordekvivalens är en relation mellan lexikaliska enheters semantiska och pragmatiska egenskaper som det kommer an på lexikografen att etablera (Adamska-Sałaciak 2010:388). Ekvivalensbegreppet har, som bekant, en given parallell i det vanskliga begreppet synonymi.

5.3. Parallellkorpora och tvåspråkig lexikografi

Tillbaka till Atkins (1996) och hennes plenarföreläsning på den sjunde Euralexkonferensen. Hon konstaterade att framtiden för den tvåspråkiga lexikografin var ljus. Datorerna hade gjort entré, så också de stora och stadigt växande digitala korpusarna. Den lösning hon såg framför sig bestod i två separata databaser, en för varje språk, som i den virtuella ordboken processas, länkas och presenteras för användaren på ett sätt anpassat till dennes behov. Bland annat skulle dikotomin källspråk/målspråk därmed upphävas.

På de 20 år som gått sedan Atkins höll sin föreläsning har mycket hänt. Ordböcker som redan fanns i tryckt form har digitaliserats och gjorts tillgängliga på nätet. Också nya ordböcker har tillkommit, designade och utarbetade för att publiceras enbart i digital form, som e-ordböcker. En stor satsning här i Norden var arbetet med *Islex*-ordböckerna. Enheter i en isländsk lexikalisk databas försågs med ekvivalenter och översättningar på de andra nordiska språken: danska, bokmål, nynorska, färöiska och svenska samt finska. Samma isländska databas ligger också till grund för en isländsk–fransk ordbok, *Lexia*, som nu är under arbete (Davíðsdóttir 2018), och nu även för en enspråkig isländsk ordbok (Jónsdóttir & Úlfarsdóttir 2020).

Islex är på många sätt fantastiskt. Användargränssnittet är tilltalande och de genomtänkta, avancerade sökfunktionerna ger tillgång inte bara till lemmat – också ekvivalenterna är sökbara liksom de källspråks- och målspråksenheter som förekommer i språkproven. Samtidigt illustrerar ordboken tydligt den klyfta som föreligger mellan de lexikografiska aspekterna och de möjligheter som den

avancerade språkteknologin erbjuder. Här nedan kommer konflikten mellan den lexikografiska traditionen och de språkteknologiska möjligheterna att diskuteras ur ett teoretiskt perspektiv men även praktiskt förankrat i *Islex*-ordböckerna. Främst är det två problem som jag vill lyfta fram. Det första gäller representationen av de båda språkens lexikaliska system, det andra språkproven. Utgångspunkten är i båda fallen den isländsk-svenska ordboken *Islex.se*.

5.4. Lemmalistan och målspråkets ordförråd

I ett digitalt användargränssnitt kan alla enheter i den tvåspråkiga ordbokens båda språk göras sökbara och tillgängliga. Lexikografen idag utgår ifrån den teori som utvecklades främst på 1980-talet; med strikt funktionsuppdelning mellan ordböcker avsedda för avkodning, $L2 > L1$, respektive produktion, $L1 > L2$. För e-ordbokens användare är det dock inget argument att ordboken *egentligen* inte är avsedd att kunna fungera lika bra för alla lexikografiska handlingar. Ur användarens perspektiv är lexikografens distinktion mellan källspråk och målspråk irrelevant. Det som användaren har är ett *sökspråk*, dvs. det språk som är utgångspunkt för den specifika sökningen, och ett *träffspråk*, dvs. det språk man i det enskilda fallet vill ha information om (Hannesdóttir 2013, 2015). Sökspråket kan vara vilket som helst av de i den aktuella ordboken behandlade språken: det lexikografiska källspråket eller målspråket. På samma sätt är det med användarens träffspråk. Den information som användaren söker beror på om träffspråket i varje enskilt fall är dennes modersmål eller ett främmande språk, oavsett om det utgör ordbokens lexikografiska källspråk eller målspråk. En och samma ordbok används av talare av **båda** språken såväl för avkodning som för produktion. Hur lexikografen ska hantera detta är en teoretiskt såväl som metodologiskt avancerad utmaning.

Den tvåspråkiga lexikografen är ännu inte framme vid de lösningar som Atkins lanserade 1996. Distinktionen källspråk/målspråk lever kvar. Källspråket är ordbokens lexikografiska primärspråk. Det som från en lexikografisk utgångspunkt utgör ordbokens målspråk är på alla sätt sekundärt till källspråket (Hannesdóttir 2013, 2015). Främst märks detta i ordförrådet. Lemmalistan är sammansatt för att spegla den kultur, det samhälle och de livsvillkor som är kodade i källspråket. Målspråkets lexikaliska representation genereras däremot helt av källspråkets lemmaurval. Resultatet blir slumpmässiga glimtar av den kultur och den verklighetssyn som inbäddats i målspråket. Lexikaliska enheter som representerar områden som är viktiga i målspråkskulturen men inte i källspråkskulturen aktualiseras inte som ekvivalenter.

Konflikten mellan den lexikografiska traditionen och de språkteknologiska möjligheterna kan illustreras med representationen av svenskan såsom den akt-

ualiserats i *Islex.se*. De svenska enheterna är inte, såsom Atkins förespår, hämtade ur en svensk databas länkad till den isländska. De isländska uppslagsorden har genom traditionell lexikografisk analys försetts med svenska motsvarigheter. Därför lyser exempelvis svenskans ordförråd knutet till gruvsdrift och skogsbruk, luciafirande och skärgårdsliv samt delmängder av svenskans internationella ordförråd med sin frånvaro (Hannesdóttir 2013, 2015). Dessutom är relationerna mellan lemmat och ekvivalenten asymmetriska: ekvivalenten är en utsaga om lemmat men lemmat är inte en utsaga om ekvivalenten. Den denotativa, lexikografiska ekvivalensrelationen är i själva verket sällan ömsesidig.

5.5. Språkproven

I den elektroniska ordboken är det inte bara lemmat som är sökbart. Det ordförråd som förekommer i språkproven och deras översättningar kompletterar lemmalistan och ekvivalentuppsättningen genom att enheterna är sökbara. Språkproven får på så sätt en annan funktion i e-ordboken än i den tryckta. Det leder oss till det andra problemet jag vill framhålla: språkproven.

De språkprov som är valda eller konstruerade för att illustrera lemmats egenskaper och användning har en viktig funktion för den som använder ordboken för **avkodning** av det lexikografiska källspråket, L2>L1. För dessa användare aktualiserar översättningen den lexikaliska meningspotentialen och den mångfald alternativa formuleringar och uttryckssätt som modersmålskompetensen i det lexikografiska målspråket innebär. Förutom de kombinatoriska och kontextuella egenskaperna reflekterar språkproven också den semantiska variation som är knuten till olika kontexter, variation som uttrycks lexikaliskt i målspråket. Språkproven utgör i själva verket små texter, minitexter. Texter kan inte översättas ord för ord. En idiomatisk översättning av språkprovets semantiska innehåll kan med hänsyn till vilka lexem som aktualiseras ligga långt ifrån källspråksutformningen. Hur effektiva språkproven är vid produktion på målspråket som L2 är ännu dåligt utforskat.

Den tvåspråkiga lexikografins syfte är ”to help in translating from one language to another, or in producing texts in a language other than the user’s native one, or both” (Zgusta 1971:213). Frågan är om den tvåspråkiga lexikografen nu, 50 år senare, kan fylla båda dessa funktioner. Tvåspråkiga ordböcker kan fungera för avkodning: från L2 till L1. När det gäller produktion på det främmande språket är deras användbarhet begränsad.

6. Google Translate

Den lösning som Atkins 1996 såg ligga i framtiden, med enspråkiga databaser som länkas mellan språken, är nu redan verklighet.

Google Translate lanserades 2006. I media har verktyget främst uppmärksammats på ett negativt sätt: ”Tokiga översättningar när kommunerna använder Google Translate” (SR 1 augusti 2013) och liknande kommentarer är vanliga. Men verktyget har utvecklats sedan dess och kan nu fungera bra som utgångspunkt vid översättning av texter mellan språk som har många talare och till ett språk man behärskar tillräckligt väl för att genomskåda mindre lyckade översättningsförslag. Sådana förslag kan då även användas som utgångspunkt för egna lösningar.

Men hur är det med översättning mellan språk med färre talare, som de nordiska? För att ta reda på detta har jag jämfört Googles översättningar med en del av materialet i *Islex*, både enskilda ord och olika typer av språkprov, från isländska till svenska och omvänt. Språkproven i *Islex* är av olika slag; kollokationer som illustrerar lemmats selektionsrestriktioner, fraser avsedda att illustrera lemmats syntagmatiska egenskaper, fasta fraser och idiom. Studien är mycket liten och här redovisas den dessutom bara selektivt. Resultatet är ändå förvånande. Översättningarna håller ganska hög kvalitet. Alla är inte perfekta – men vilken översättning är det?

6.1. Google Translate gentemot *Islex.se*

Nedan ser vi sex isländska språkprov med svenska översättningar, först den översättning Google Translate (GT) presenterar, sedan de översättningar som finns i *Islex.se* (härefter även kallad ”ordboken”). Språkprovet i exempel (1) är avsett att illustrera användningen av verbet *dirfast*.

(1)	Islex	hvernig dirfistu að koma hingað inn?
	GT	hur vågar du komma in hit?
	Islex (sv)	hur vågar du dig att komma hit in?

Skillnaden mellan de båda översättningarna är hårfin. Ordbokens lösning fångar en emfatisk nyans i den isländska källtexten som går förlorad i GT. I andra avseenden är båda översättningarna acceptabla, alternativet *komma in hit* respektive *att komma hit in* är likvärdiga.

I exempel (2) återges översättningarna av ett exempel på hur lemmat *óathugaður* används. Detta adjektiv förekommer i princip bara i den konstruktion som här illustreras, <gera eitthvað> *að óathuguðu máli* <göra något> ’utan att ha undersökt saken’ eller ’oöverlagt’.

- | | | |
|-----|------------|--|
| (2) | Islex | ég dró ekki þessa ályktun að óathuguðu máli |
| | GT | jag drar inte denna slutsats oavsiktligt |
| | Islex (sv) | jag drog inte den här slutsatsen utan att ha undersökt saken |

GT:s översättning av *að óathuguðu máli* med *oavsiktligt* kan möjligen tangera betydelsen hos ’oöverlagt’, men den är inte ekvivalent med källtextens formulering. Verbets tempusbyte, presens *drar* istället för preteritum *drog*, är inte motiverat.

Exempel (3) illustrerar möjlig kontext för verbet *hella*: *ég hellti vatni á gólfíð*. Ordboken presenterar två översättningar.

- | | | |
|-----|------------|--|
| (3) | Islex | ég hellti vatni á gólfíð |
| | GT | jag hällde vatten på golvet |
| | Islex (sv) | jag hällde ut vatten på golvet
jag spillde vatten på golvet |

GT:s översättning återger språkprovets generiska betydelse. Istället för en generisk tolkning redovisas i de alternativa översättningarna i *Islex.se* den distinktion beträffande intentionalitet som svenskan erbjuder. Åsytar språkprovet ett avsiktligt hållande är det idiomatiska svenska uttrycket ”hälla ut vatten på golvet”, medan ”spilla vatten på golvet” avser den oavsiktliga handlingen.

Språkprovet i exempel (4) avser lemmat *landsmaður*. Enda skillnaden mellan GT och *Islex.se* är ekvivalenten till *landsmenn*.

- | | | |
|-----|------------|------------------------------------|
| (4) | Islex | landsmenn gengu til kosninga í vor |
| | GT | islänningar gick till val i våras |
| | Islex (sv) | invånarna gick till val i våras |

Medan ordbokens översättning tar fasta på den denotativa betydelsen, återger GT en pragmatiskt grundad ekvivalent: landsmännen förutsätts vara islänningar. Den kollektiva betydelse som uttrycks i *landsmenn* kräver dock bestämd form av *islänningar*. Av den svenska översättningen framgår inte att *landsmenn*, till skillnad från *invånarna*, avser ett lands alla invånare. Denna precisering impliceras dock av parentesen vid ekvivalenten till lemmat *landsmaður*: ”invånare (i ett land)”.

GT:s översättning i exempel (5) illustrerar svårigheten att med automatiska verktyg hitta rätt betydelse vid polysema ord, liksom att GT inte är finjusterat när det gäller valens och selektionsrestriktioner. Verbet *tjalda* är polysemt, vilket redovisas i *Islex*. Den svenska ekvivalenten till lexem 1 är *tälta* medan *täcka* är ekvivalent med lexem 2. De isländska lexemens konstruktionsmönster överensstämmer med ekvivalenternas.

(5)	Islex	við tjölduðum við fallegan læk
	GT	vi täckte en vacker ström
	Islex (sv)	vi slog upp tältet vid en vacker bäck

GT:s förslag är i och för sig en korrekt svensk utsaga. Som motsvarighet till källtexten är den dock inte helt lyckad. Den utgår inte ifrån den avsedda betydelsen av det polysema verbet *tjalda*, och ekvivalensrelationen mellan *ström* och substantivet *lækur* är på sin höjd partiell. I den generiska definitionen av *ström*, 'snabbt rinnande vattendrag' (SO, s.v. *ström*), ingår en intensitet som är fakultativ i det isländska *lækur*.

GT:s översättning av idiomat *láta álit sitt í ljós* (ex. 6) är helt korrekt och identisk med den ena av de två översättningar som ges i *Islex.se*.

(6)	Islex	láta álit sitt í ljós
	GT	uttrycka sin åsikt
	Islex (sv)	uttrycka sin åsikt säga sin mening

De översättningar som här har diskuterats utfaller ofta till ordbokens fördel. Det är dock lätt att konstatera att Googles översättningsverktyg i de flesta fall åstadkommer acceptabla och ibland till och med idiomatiska översättningar från isländska till svenska.

6.2. Google Translate översätter till isländska

I exempel (7)–(11) visas GT:s översättningar från svenska till isländska. Den ackumulerade isländska textmängd som Google Translate har tillgänglig kan antas vara avsevärt mindre än mängden tillgänglig svensk text. Översättningar av svensk text till isländska genom automatisk matchning av lexikon och strukturer torde därför bli mindre precisa än i det omvända fallet.

I exempel (7) visas hur GT översätter den svenska översättningen av ett språkprov i artikeln *þvo* i *Islex.se* tillbaka till den isländska källversionen i *Islex*.

(7)	Islex (sv)	hon tvättade handduken i varmt vatten
	GT	hún þvoði handklæð í heitu vatni
	Islex	hún þvoði handklæðið í heitu vatni

GT:s översättning är idiomatisk och korrekt i alla avseenden utom i fråga om formen på substantivet *handklæði*.

I exempel (8) är utgångspunkten den svenska översättningen av ett språkprov i artikeln *hlaupa*: ”hon sprang det fortaste hon kunde”. GT förmedlar det denotativa innehållet i översättningen på korrekt isländska: ”hún hljóp eins hratt og hún gat”. Den isländska text som ligger till grund för den svenska översättningen innehåller idiomat *eins og fætur toguðu*.

(8)	Islex (sv)	hon sprang det fortaste hon kunde
	GT	hún hljóp eins hratt og hún gat
	Islex	hún hljóp eins og fætur toguðu

Den svenska versionen, där det isländska idiomat har översatts med frasen *det fortaste (hon) kunde*, kan ses som ett försök att fånga konnotationer hos det isländska idiomat.

Det svenska uttrycket för metaforen *något far i någon* vid en oväntad humörväxling får i GT en misslyckad isländsk översättning (ex. 9). I stället för *hlaupa* (springa) föreslår GT det generiska *gerast* (som betyder ’hända’) följt av en felaktig preposition, *við*. För övrigt är översättningen korrekt.

(9)	Islex (sv)	han pratar inte med mig, vad har farit i honom?
	GT	hann talar ekki við mig, hvað hefur gerst við hann?
	Islex	hann talar ekki við mig, hvað hefur hlaupið í hann?

I exempel (10) kapitulerar GT inför den talspråkliga svenska frasen *skutta omkring* men lämnar ett förslag på en för övrigt korrekt isländska:

(10)	Islex (sv)	hon skuttade omkring av glädje
	GT	hún skuttade um af gleði
	Islex	hún hoppaði af kæti

GT presenterar ofta alternativa översättningar. I exempel (11) är den isländska översättning som först föreslås grammatiskt bristfällig på flera sätt. Ytterligare ett alternativ erbjuds dock, ett alternativ som är identiskt med den isländska källtext som lexikograferna översatt till svenska.

(11)	Islex (sv)	skådespelaren syns ofta på vita duken
	GT	leikari er oft séð á hvítum dúkur leikarinn sést oft á hvíta tjaldinu
	Islex	leikarinn sést oft á hvíta tjaldinu

De översättningar som här har redovisats utfördes av Google Translate i maj 2019. Vid en kontroll av översättningarna i december 2019 har de flesta av dem förändrats. Många har blivit bättre, andra inte. Som helhet visar dock denna begränsade undersökning av de korta, lösryckta textsträngar som språkproven utgör att Googles automatiska översättningsverktyg nästan är i paritet med lexikografernas översättning. Men medan ordboken erbjuder ett statistiskt antal färdigformulerade uttryck, är det bara användarens formuleringsförmåga som sätter gränsen för de förslag till översättningar som Google Translate tillhandahåller.

7. De framtida utmaningarna för tvåspråkig lexikografi

Den tvåspråkiga lexikografen utgår ifrån ordet, inte längre bara svåra ord utan också vardagliga och frekventa. Denna förändring beträffande fokus för den lexikografiska beskrivningen i ordböckerna är dock högst marginell i jämförelse med den senaste tidens utveckling. Språkteknologin och de digitala publikationsformerna har radikalt ändrat alla förutsättningar för den tvåspråkiga lexikografen. Den hävdvunna lexikografiska typologi som etablerades under det sena 1900-talet, med ett strikt användarperspektiv och konsistenta användningsområden för varje ordbok, är, ur användarens perspektiv, inte längre relevant.

Den tvåspråkiga lexikografins största utmaning består i att utveckla teorier och metoder för att gå användarna till mötes i genuint bidirektionella, multifunktionella verktyg där det enskilda ordets semantiska, grammatiska, syntagmatiska och pragmatiska egenskaper redovisas. En lexikografisk beskrivning byggd på en noggrann kontrastiv lexikalisk analys kan utgöra underlag för studier av olika slag, inte bara i kontrastiv lexikalisk semantik utan också kontrastiva pragmatiska, konstruktionsgrammatiska, syntaktiska och informationsstrukturella studier. Däremot har lexikografen inte tillgång till de verktyg som behövs för att i varje enskilt sammanhang kunna redovisa vilken enhet i målspråkets lexikon som aktiveras i varje unik kontext och inte heller den relevanta betydelsen i den aktiverade enhetens meningspotential. I fråga om att producera eller översätta **text** till ett främmande språk är Google Translate nog redan ett ordböckerna överlägset hjälpmedel.

Inom ramen för digitala humaniora inriktas nu språkteknologin i ökande grad på automatisk betydelseanalys och identifiering av betydelseförändringar. Effektiva metoder för att hitta sådana förändringar utvecklas i snabb takt. Redskapen för att automatiskt analysera och beskriva semantiska nyanser och semantiska processer brister fortfarande. När det gäller den kontrastiva analysen och beskrivningen av *ordet* och alla dess komplexa egenskaper saknar den tvåspråkiga lexikografin fortfarande sin like.

Litteratur

Ordböcker

Islex. Stofnun Árna Magnússonar í íslenskum fræðum. <www.islex.se>
Samtaal på tyska och svenska, som almennelighen brukas både i huushåld och annorstädes [u.å.]. Stockholm: Ignatium Meurer.
 SO = *Svensk ordbok utgiven av Svenska Akademien* (2009). Stockholm: Norstedts i distribution.

Annan litteratur

Adamska-Sałaciak, Arleta (2010): Examining Equivalence. I: *International Journal of Lexicography*, Volym 23:4, 387–409.
 Atkins, B. T. S. (1996): *Bilingual Dictionaries. Past, Present and Future*. I: Martin Gellerstam, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström & Catarina Röjder Pappmehl (red.): *Euralex '96. Proceedings I–II*. Göteborg: Göteborgs universitet, 515–546.
 Atkins, B. T. Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
 Collison, Robert L. (1982): *A History of Foreign-Language Dictionaries*. London: André Deutsch Limited.
 Considine, John (2008): *Dictionaries in Early Modern Europe. Lexicography and the Making of Heritage*. Cambridge: Cambridge University Press.
 Davíðsdóttir, Rósa Elín (2018): LEXIA – en islandsk-fransk online ordbog: Udfordringer og løsninger. I: *Nordiske studier i leksikografi* 14, 63–70.
 Etxepare, Ricardo & Viola G. Miglio (2015): The Newly-Discovered Fourth Basque-Icelandic Glossary. I: Irujo Xabier & Viola G. Miglio (red.): *Basque Whaling in Iceland in the XVII century. Legal Organization, Cultural Exchange and Conflicts of the Basque Fisheries in the North Atlantic*. Santa Barbara: Barandiaran Chair of Basque Studies, University of California, 345–393.
Google Translate. <translate.google.se>.

- Hannesdóttir, Anna Helga (1991): Relationer mellan Dalins olika ordböcker. I: Sven-Göran Malmgren & Bo Ralph (red.): *Studier i svensk språkhistoria 2*. Göteborg: Göteborgs universitet, 78–89.
- Hannesdóttir, Anna Helga (1998): *Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga*. Meijerbergs arkiv för svensk språkforskning 23. Göteborg: Göteborgs universitet.
- Hannesdóttir, Anna Helga (2000): Tvåspråkig lexikografi och språkplanering i 1700-talets Sverige. I: Lars-Erik Edlund (red.): *Studier i svensk språkhistoria 5*, 224–235.
- Hannesdóttir, Anna Helga (2007): Först och sist. Haqvin Spegel och ordförrådets förnyelse. I: *LexicoNordica 14*, 125–139.
- Hannesdóttir, Anna Helga (2013): Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi. I: *Nordiske studier i leksikografi 12*, 193–211.
- Hannesdóttir, Anna Helga (2015): What is a target language in an electronic dictionary? I: I. Kosem et al. (red.): *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Proceedings of the eLex 2015 conference. Ljubljana/Brighton: Trojiana, Inst. For Applied Slovene Studies/Lexical Computing Ltd, 236–249.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2020): Omdanneslen af en flersproget til en monolingval ordbog. I: *Nordiska studier i lexicografi 15*, 175–184.
- Karker, Allan (1967): Ordbøger. I: *Kulturhistorisk leksikon for nordisk middelalder*. Bd. XII. København: Rosenkilde og Bagger, spalt 638–639.
- Linnér, Sture (2002): *Europas ungtid. Nedslag i Europas kulturhistoria fram till cirka 500 f.Kr.* Stockholm: Wahlström & Widstrand.
- Malmgren, Sven-Göran (1988): Almqvist, Dalin och den svenska definitionsordbokens födelse. I: Gertrud Pettersson (red.): *Studier i svensk språkhistoria*. Lund, 154–164.
- Neuman, Erik (1973): *Inledning till utgåvan av latinskt-svenskt glossarium efter Cod. Ups. C 20*. Utg. Av Börje Tjäder. Samlingar utgivna av Svenska Fornskriftssällskapet 45:6. Uppsala: Almqvist & Wiksell.
- Noreen, Adolf (1903): *Vårt språk*. Bd 1. Lund: C. W. K. Gleerups Förlag.
- Ralph, Bo (2016): Patriotic lexicography – an obsolete genre? I: Victoria Domínguez-Ridrígues et al. (red.): *Words across History. Advances in Historical Lexicography and Lexicology*. Las Palmas: Universidad de Las Palmas de Gran Canaria, 331–340.
- Raschellà, Fabrizio D. (2001): Vernacular Gloss Writing in Medieval Scandinavia. I: Rolf Bergmann, Elvira Glaser & Claudine Moulin-Fankhänel (red.): *Mittelalterliche volkssprachige Glossen. Germanistische bibliothek 13*. Heidelberg: Universitätsverlag S. Winter, 587–599.
- Skautrup, Peter (1960): Glosser. I: *Kulturhistorisk leksikon for nordisk middelalder*. Bd. V. København: Rosenkilde og Bagger, spalt 359–360.

Snell-Hornby, Mary (1986): The bilingual dictionary – victim of its own tradition? I: R. R. K. Hartmann (red.): *The history of lexicography. Papers from the Dictionary research center seminar*. Amsterdam/Philadelphia: John Benjamins Publishing company, 207–218.

SR = Sveriges Radio: *Tokiga översättningar - när kommunerna använder Google Translate*. 1 augusti 2013. <www.sverigesradio.se/sida/artikel.aspx?programid=86&artikel=5605782>.

Store norske leksikon: *Glossarier, glosse*. <www.sn�.no/glosse>. (november 2019).

Zgusta, Ladislav (1971): *Manual of Lexicography*. Janua Linguarum. Series maior. 39. Prague: Academia, The Hague & Paris: Mouton.

Zgusta, Ladislav (1984): Translational equivalence in the bilingual dictionary. I: R. R. K. Hartmann (red.): *Lexeter '83 Proceedings*. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983, 147–154.

Anna Helga Hannesdóttir
professor
Göteborgs universitet
Box 200
SE-405 30 Göteborg
anna.hannesdottir@svenska.gu.se

Den tvåspråkiga ordboken – ett absurt åtagande

Nina Martola

Compiling a bilingual dictionary between structurally very different languages is a challenge. In the 1990's a comprehensive new Finnish-Swedish dictionary was compiled, and it is now under revision. In this article I discuss some issues that the lexicographers were confronted with. Using a monolingual Finnish dictionary as a basis is a good but not unproblematic method, since the users of the bilingual dictionary has other needs as to lemma selection and microstructure than those of the monolingual dictionary. Structural differences between Finnish and Swedish lead to lacking equivalency on a word basis more frequently than in dictionaries between related languages. The last aspect I discuss is the problem of normativity: should a bilingual dictionary be normative and if so – how?

1. Bakgrund

I mitten av 1980-talet togs initiativet till en ny omfattande finsk-svensk ordbok på dåvarande Forskningscentralen för de inhemska språken (sedan 2012 Institutet för de inhemska språken). Den storordbok som varit i användning ända sedan slutet av 1800-talet (Cannelin m.fl. 1976) var i mångt och mycket föråldrad. Vid den tiden hade arbetet med en ny, monolingval finsk ordbok inletts på Forskningscentralen, och att skapa en helt ny tvåspråkig ordbok utifrån en modern enspråkig finsk ordbok konstaterades vara ett bättre alternativ än att revidera den gamla tvåspråkiga, som det dessutom var rättighetsproblem kring.

Redigeringsarbetet kom i gång hösten 1986 och redaktionen bestod till att börja med av en ensam redaktör stöttad av ett redaktionsråd. Att redigera en storordbok med ett tilltänkt omfång på ca 90 000 ordartiklar, varav en del tämligen omfattande, visade sig föga överraskande vara en övermäktig uppgift för en enda redaktör. Vid årsskiftet 1989/1990 gick förlaget WSOY efter förhandlingar in i projektet, och redaktionen kunde utökas till fyra personer, två anställda av förlaget och två av Forskningscentralen. Så småningom tillkom två halvtidsanställda assistenter. Våren 1997 publicerades den första upplagan av *Stora finsk-svenska ordboken* (i fortsättningen SFSO).

Vägen fram till den färdiga ordboken var långt ifrån lätt. En stor del av problemen och svårigheterna berodde på att lexikografiskt arbete helt enkelt är svårt, och det är kanske extra svårt när det handlar om bilingval lexikografi mellan två

språk som är så strukturellt olika som finska och svenska. Men en del problem hade kunnat undvikas om erfarenheten av bilingval lexikografi hade varit större när projektet startade. Redaktionen var grön och ingen i redaktionsrådet hade erfarenhet av arbete med tvåspråkiga ordböcker, även om det i övrigt var en mycket kompetent grupp. I själva verket var den monolingvala kompetensen bättre representerad i gruppen.

Det är också värt att notera att det här var i den teoretiska lexikografins barn-dom. Den europeiska föreningen för lexikografi Euralex hade instiftats 1983 (Euralex) och Nordiska föreningen för lexikografi kom till 1991.

SFSO sammanställdes med den finska ordboken *Suomen kielen perussana-kirja* ('Finsk basordbok', i fortsättningen PS) som bas. Senare har PS bytt namn till *Kielitoimiston sanakirja* ('Språkbyråns ordbok', KS). I det nya namnet utnyttjas varumärket för finsk språkvård, *Kielitoimisto*. År 2004 kom en lätt reviderad upplaga av SFSO ut och efter det gjordes ingenting åt innehållet på åtskilliga år.¹

Vid årsskiftet 2016/2017 förhandlade sig Institutet för de inhemska språken till rätten att ge ut utvidgade utgåvor i egen regi och i april 2017 kunde den första versionen publiceras på webben. Efter det har det lagts ut nytt material med jämna mellanrum. Den version som är aktuell i skrivande stund är från augusti 2019². För den nya versionen används KS som förlaga.

2. Målgruppsanpassning

När ett ordboksprojekt initieras är det viktigt att definiera vilken eller vilka målgrupper ordboken ska vända sig till. Ett problem under arbetet med SFSO var strävan att beakta behoven hos en alltför brokig användarskara.

Tvåspråkiga ordböcker borde i princip finnas i två uppsättningar för varje språkpar, en uppsättning för användare med S1 som modersmål och en uppsättning för personer med S2 som modersmål. För stora språkpar som engelska-spanska, engelska-franska, engelska-tyska osv. är det också fallet. Ordböcker för språk med färre talare görs däremot ofta bara i en uppsättning och utformas i huvudsak för den ena språkgruppen. Ordböcker mellan svenska och språk som engelska, franska, spanska och tyska har vanligen i första hand gjorts för svenska användare, eftersom det är mycket vanligare att svenskar använder engelska än att engelsk-språkiga lär sig svenska (Svensén 2004:17ff.).

¹ Se även Cantell 2015.

² I april 2020 kom följande version och hösten 2020 kommer ordboken att publiceras med ny layout hos Språkinstitutet.

Samma typ av asymmetri har funnits och finns mellan svenska och finska. Ordböckerna har främst gjorts med tanke på finska användare medan förlagen i sin marknadsföring rätt ofta riktat sig också till andra användargrupper än den som ordboken egentligen är gjord för.

Helt symmetriska ordböcker är sällsynta men finns. Ett exempel är VanDale/Norstedts holländsk-svensk-holländska ordbokspär (Van Dale 1996, se även Laureys 1997).

I våra dagar är ordboksmarknaden ingen guldgruva och många förlag har helt och hållet slutat producera ordböcker. Målgruppsfrågan har också blivit mycket mer komplicerad i och med att ordböckerna är elektroniska och sökbara också på målspråket (Hannesdóttir 2016).

För SFSO:s del skulle huvudmålgrupperna vara två, dels svenska användare som behövde ett verktyg för översättning av olika slag (översättare, journalister, informatörer, lärare osv.), dels avancerade finska användare som behövde producera text på svenska av olika skäl. SFSO är inte så bidirektionell som VanDale-ordböckerna och har heller inte någon svensk-finsk pendang.³

Trots att målgruppen hade definierats som avancerade finska användare fördes det i praktiken många diskussioner om hur personer med bara elementära svenskkunskaper kunde tänkas tolka angivelserna och hur artiklarna borde utformas för att den användargruppen skulle få hjälp. Redaktionen (och redaktionsrådet) svävade alltså på målet. Ett exempel på detta är böjningsangivelserna för ettordsekvivalenter jämfört med tvåordsekvivalenter. I de senare används förkortningen *taiv* (av *taivutus* 'böjning') för att påminna användaren om att adjektivet ska kongruensböjas:

föremål -et =	sid a -an -or
gjutet föremål (<i>taiv -et =</i>)	stark sida (<i>taiv sid-an -or</i>)

Man kan fråga sig om användarna överlag uppfattar skillnaden i angivelsesätt. En person med rudimentära kunskaper i svenska får ju dessutom ingen hjälp med att välja rätt form på adjektivet.

Ett ännu mer påfallande fall är en förhållandevis stor grupp sammansatta adjektiviska uppslagsord som saknar svenska ekvivalenter och som därför åskådliggörs med nakna exempel. I den första upplagan av SFSO gavs genomgående

³ En svensk-finsk storordbok utarbetades senare vid förlaget WSOY. Den var tänkt som en pendang men kom till som en fristående ordbok. Under flera år fanns de två ordböckerna dock tillgängliga som ett par på webben mot licens. Språkinstitutet har fått rättigheterna bara till SFSO. Företaget Kielikone har köpt rättigheterna av förlaget Talentum och tillhandahåller en avgiftsbelagd utgåva av ordboksparet, veterligen oreviderade versioner men med maskinell översättning av ord som inte finns i lemmalistan.

exempelöversättningar med både *med*-konstruktion och *har*-konstruktion i artiklarna (exemplet *alkoholiongelmainen*). Det hade knappast gjorts om redaktionen mer entydigt haft avancerade finska användare i åtanke.

alkoholiongelmainen

► *alkoholiongelmainen henkilö*

en person med alkoholproblem

hän on alkoholiongelmainen

han har alkoholproblem

han har problem med alkohol

Det måste i varje fall sägas att en del av redaktionens vacklan kring målgruppen härrör från det yttre samhället. Det brister i den språkliga medvetenheten på sina håll och svenska texter produceras ibland av personer som inte borde åläggas den uppgiften.

3. Från enspråkig till tvåspråkig ordbok

Som det redan har framkommit utarbetades SFSO med den enspråkiga finska ordboken PS som bas. Grunden till beslutet var att redaktionen på så vis skulle ha tillgång till ett färdigt sammanställt, aktuellt ordförråd och till adekvata betydelsebeskrivningar som gäller för modernt språkbruk. Halva jobbet var så att säga gjort. Den tvåspråkiga ordboken skulle nå hög kvalitet med betydligt mindre tidsåtgång, eftersom det ”bara var att översätta PS-artiklarna”. I tillägg fanns det inga rättighetsproblem, eftersom PS redigerades vid Forskningscentralen för de inhemska språken.

Argumenten är utan vidare adekvata, men initiativtagarna hade kanske ändå hyst en viss övertro på fördelarna. Man gör inte en bilingval ordbok genom att översätta en monolingval ordbok.

3.1. Lemmaselektion

Varje ordbok utgör ju ett tämligen begränsat urval av alla de ord som förekommer i det språk den beskriver, och även om en ordbok strävar efter att beskriva det centrala ordförrådet kan urvalet i viss mån te sig slumpmässigt så fort man går ut över det centralaste kärnoförrådet.

Lemmaselektionen ska förstås göras med beaktande av den tilltänkta målgruppen (Bergenholtz 1991:49ff.). Målgruppen för en enspråkig, finsk ordbok för modersmålstalare är inte densamma som målgruppen/målgrupperna för en

bilingval ordbok. I en ordbok för modersmålstalare är sannolikheten större att användarna slår upp ”svåra” ord, t.ex. lågfrekventa ord och ord av främmande ursprung. För den som behöver kontrollera vad ett finskt ord eller uttryck motsvaras av på svenska är det knappast de främmande orden som vållar mest huvudbry. Gissningsvis är det inte heller i första hand lågfrekventa ord användarna stöter på och behöver få översätta, utan sannolikheten är högre att det är mer högfrekventa ord som vållar problem.

Det underlättar det redaktionella arbetet oerhört att ha en färdig ordbas att utgå ifrån, men för att resultatet ska bli optimalt borde en kompletterande lemmasektion göras med tanke på den nya målgruppen. För ett sådant arbete fanns det dock inga resurser inom ramen för SFSO-projektet. Det är tyvärr ofta de yttre ramarna som dikterar villkoren för lexikografiskt arbete.

Vi får också komma ihåg att redaktionen på 1990-talet hade mycket begränsade möjligheter att kontrollera frekvenser. Under arbetets gång fick redaktionen tillgång till Språkbankens dåvarande korpusar, men de hade inte tillnärmelsevis det omfång och de sökmöjligheter de har i dag, och användningen var inte alls så enkel som den nu är.

En annan sida av lemmasektionen är att en tvåspråkig ordbok borde få med ord som kan te sig triviala ur enspråkig synvinkel men som kan vara nog så besvärliga att översätta. Ett litet antal sådana ord kunde redaktionen komplettera med, via egna iakttagelser och tack vare kontakter med språkvårdare och översättare, men någon systematisk komplettering genomfördes inte.

Ett exempel på ett uppslagsord som lagts till i SFSO av kontrastiva skäl under uppdateringen 2017–2020 är *julkisuudenkipeä* till *julkisuus* ’offentlighet; publicitet’ och *kipeä* ’sjuk’:

julkisuudenkipeä

► *olla julkisuudenkipeä*

vara ute efter (sträva efter) publicitet

tycka om (ha behov av) publicitet

vilja vara känd

vilja synas och höras

vara publicitetshungrig

syssla med effektsökeri

Att *publicitetshungrig* (i den näst sista översättningen) inte angetts som ekvivalent trots att ordet bildningsmässigt rätt väl motsvarar det finska uppslagsordet har att göra med förekomst. Det finska ordet uppvisar vida högre frekvens än det svenska.

3.2. Betydelsebeskrivning

Ett annat viktigt argument för att använda den enspråkiga ordboken PS som bas för SFSO var de färdiga betydelsebeskrivningarna. Det underlättade arbetet oerhört, men att det skulle vara ”bara att översätta” eller att SFSO-redaktionen inte skulle få ”gå och peta i betydelsebeskrivningarna” var naturligtvis att dra argumentet för långt.

I en ordbok där två språk kontrasteras blir det ibland tydligare att strukturera om artiklarna enligt målspråkets förutsättningar. Sådana fall är t.ex. ord som på finska kan användas både som adjektiv och substantiv. Ett exempel på det är artikeln *rämäpää*. I KS är betydelseangivelsen ’våghalsig person’ med substantivet *ihminen* ’människa, person’ inom parentes. I SFSO har artikeln indelats i ett substantiviskt och ett adjektiviskt moment. Under uppslagsordet *rämäpäinen*, som moment två hänvisar till, finns ekvivalenterna *våghalsig*, *oförvägen*, *vådlig*, *vild*, *dumdristig*:

KS **rämäpää** rämäpäinen (ihminen).
Mitään pelkäämätön rämäpää.

SFSO **rämäpää**
 1 (s) våghals, vildhjärna, galenpanna
 2 (a) ks. *rämäpäinen*

På samma sätt hanteras exempelvis många nomen på *lainen* som avser grupptillhörighet av olika slag (*bulgarisk*; *bulgariska – norrländsk*; *norrlänning*; *norrländska – främling*; *främmande*).

3.3. Exempelurval

I en mer omfattande ordbok spelar exemplen eller språkproven en viktig roll. Deras uppgift är att åskådliggöra hur ett uppslagsord används. Språkproven har naturligtvis olika funktion i en monolingval och en bilingval ordbok (jfr Svensén 2004:341ff.).

I PS/KS kan man i huvudsak särskilja tre typer av exempel. I vissa artiklar förekommer morfologiska exempel som illustrerar sammansättningsbildning (däremot finns det inte exempel på avledningar). Sammansättningsexempel togs med bara i undantagsfall i SFSO.

För det andra finns det exempel som skulle bli triviala i en tvåspråkig ordbok och som därför uteslöts, t.ex. *hämäläiset ja karjalaiset* ’tavastlänningar och karrelare’ under *hämäläinen* ’tavastlänning’ och *punainen muste* ’rött bläck’ under *muste* ’bläck’.

Den överlägset största gruppen är exempel på mer eller mindre fasta fraser och de är förstås viktiga också i en tvåspråkig ordbok. Eftersom finska och svenska existerat så länge jämsides och representerar mycket likartade kulturer finns det många fraser som är exakta motsvarigheter till varandra på de två språken, ibland skiljer sig fraserna till bildningssättet trots att tankeinhållet är det samma och ibland uttrycks samma tankeinhåll med helt olika fraser. Ett antal fraser med *käsi* 'hand' får illustrera detta:

- | | | |
|----|--|---|
| a) | <i>kulkea käsi kädessä</i>
<i>elää kädestä suuhun</i>
<i>olla jkn oikea käsi</i> | gå hand i hand
leva ur hand i mun
vara ngns högra hand |
| b) | <i>pestä kätensä</i>
<i>levitellä käsiään</i>
<i>paiskata kättä jkn kanssa</i> | två sina händer
slå ut med armarna
skaka hand med ngn |
| c) | <i>olla käsi maassa</i>
"vara med handen i marken"
<i>hänellä on hätä kädessä</i>
"hen har nöden i handen"
<i>ottaa silmä käteen</i>
"ta ögat i handen" | vara stupfull (packad)
han sitter illa till
titta noga efter; se sig noga för |

Grupp a) illustrerar alltså fraser som motsvarar varandra till bildningssättet. En användare kan komma fram till rätt svensk fras genom att översätta den finska frasen ord för ord. I grupp b) motsvarar fraserna också varandra rätt väl, men en ord för ord-översättning skulle leda till ett oidiomatiskt resultat på svenska. I bildlig användning *tvår* man sina händer på svenska, inte *tvättar*, som *pestä* i första hand översätts med. Man *slår ut med* armarna (eller händerna) när man är uppgiven i stället för att *breda ut* dem (*levitellä* 'sprida ut; breda ut') och man *skakar* hand med någon på svenska, medan det finska verbet är *paiskata* 'slänga, kasta'. I grupp c) har vi exempel där *käsi* bara ingår i den finska frasen och där en ord för ord-översättning leder till ett obegripligt resultat på svenska.

För de finska användarna är alla de tre grupperna av frasexempel viktiga och ska därför vara med i den tvåspråkiga ordboken. En icke-modersmålstalare kan ju inte veta vilken kategori ett exempel hör till. För en svensk användare är grupp c) viktigast. Som svenskspråkig kanske man inte ens alltid förstår fraserna i den gruppen, och även om man gör det så kan det ofta vara svårt att komma på goda motsvarigheter på svenska.

4. Lexikografisk hantering av språkstrukturella skillnader

Varje tvåspråkslexikograf tvingas lösa en mängd problem som bottnar i olikheter mellan de två språken. För en lexikograf som arbetar med en ordbok mellan svenska och finska blir dock en viss typ av problem extra påtagliga, nämligen problem som härrör från de strukturella skillnaderna mellan källspråk och målspråk. Jag ska ta upp några fall.

4.1. Implicit och explicit upplevare

På finska realiseras den semantiska rollen⁴ UPPELVARE ofta med hjälp av genitiv eller partitiv. I motsvarande svenska utsagor utgör upplevarrollen syntaktiskt subjekt. I finskt talspråk är det vanligt att upplevaren inte uttrycks explicit om den som faller ett yttrande syftar på sig själv, och det här ger upphov till en (skenbar) asymmetri som diskuterades mycket inom SFSO-redaktionen. Borde vi lägga in en explicit upplevare också i det finska exemplet för att få full motsvarighet till den svenska översättningen, där upplevarrollen är obligatorisk? Redaktionen stannade för att ange upplevaren inom hakparenteser, men tillämpade det inte helt konsekvent och i de senare redigeringsomgångarna har principen frångåtts:

<i>juostessa [minun] tuli kuuma</i>	jag blev trött av språngmarschen
<i>[minun] täytyy tästä jo lähteä</i>	jag måste ge mig i väg
<i>[minua] väsyttää niin!</i>	så trött jag är!

I vissa fall kan det finnas en motsvarande ”upplevarlös” svensk motsvarighet:

<i>nyt tuli ajatuskatko</i>	nu står det still i huvudet
-----------------------------	-----------------------------

4.2. Översättningsmodeller

En nydanande tanke, som redaktionen hade och också genomförde, var att förse användarna med översättningsmodeller för vissa finska avledningstyper som kan vara svårhanterade vid översättning, bl.a. egenskapssubstantiv och verbalsubstantiv. Modellerna är placerade i omtexten och består av korta beskrivningar av strukturerna tillsammans med exempel på olika möjligheter att översätta dem. I artiklarna för uppslagsord av de här typerna hänvisas det till modellerna.

⁴ Semantiska roller förekommer i olika uppsättningar inom olika grammatiska inriktningar och rollerna definieras på varierande sätt, men rollen upplevare kan i korthet definieras som ”den som upplever något”, t.ex. *han* i *han avskyr skarvar*. För en översikt hänvisas till Martola (2007:27) och SAG (1:219f.).

I modellen för egenskapssubstantiv ges det t.ex. exempel på hur avledningen *pakollisuus* ”obligatoriskhet” ska översättas i strukturen *jäsenäänestysten pakollisuus* ”medlemsomröstningarnas obligatoriskhet”:

Man kunde överväga **obligatoriska** medlemsomröstningar.

Man kunde överväga att göra medlemsomröstningarna **obligatoriska**.

... att införa **obligatorisk** medlemsomröstning.

Han tyckte inte att medlemsomröstningarna skulle vara **obligatoriska**.

Ofta har de finska avledningarna ekvivalenter på svenska (*haittomuus* – ’oskadlighet’; *raitaisuus* – ’randighet’; *suodatus* – ’filtrering’; *testaus* – ’testning’) och dessa ges förstås i så fall i själva ordartikeln tillsammans med eventuella exempel. En modellhänvisning placeras sist i artikeln, som en påminnelse om att ekvivalenten/ekvivalenterna inte alltid kan ingå i samma typ av strukturer som det finska uppslagsordet. Även om t.ex. avledningen *oskadlighet* är ett gångbart svenskt ord låter ”materialens oskadlighet” inte idiomatiskt, medan *materiaalien haittomuus* är en fullt gångbar formulering på finska.

I många av artiklarna från 1990-talet står modellhänvisningen ensam. Ett sådant exempel är *sekalaisuus*, som är en avledning till *sekalainen* ’blandad’:

sekalaisuus *ks. käännösmalli 1*

Artiklarna med ensamma modellhänvisningar har kritiserats av användarna, eftersom de inte ger någon hjälp med just det svåröversatta ord som användaren slagit upp, och i den nu pågående revideringen är ett av målen att förse varje artikel med åtminstone ett naket exempel. När det gäller att hitta lämpliga exempel har man helt andra möjligheter som redaktör i våra dagar än på 1990-talet.

Jag ska inte fördjupa mig mer i översättningsmodellerna här – de finns närmare presenterade i Martola (1993, 2001) och Cantell (2015) – men kan avrunda med att konstatera att vi fått in en del kommentarer om att man som användare inte ”orkar kolla” och ”inte riktigt förstår sig på den konstiga hänvisningen” (som naturligtvis är en länk i den elektroniska versionen). Å andra sidan vet vi också att modellerna utnyttjas i översättningsundervisningen vid Helsingfors universitet (Jan Lindström, muntlig kommentar 20.4.2018), och vi hoppas förstås att det också finns användare som utnyttjar modellerna ibland.

4.3. Finska adverb på *sti* som uppslagsord

Adverb som anger sätt, mängd och intensitet bildas på finska bland annat med hjälp av ändelsen *sti* som fogas till adjektivstammen och till (lexikaliserade) particip (*haikea* 'vemodig' – *haikeasti* 'vemodigt'; *ikävyystynyt* utträkad – *ikävyestyneesti* 'utträkat'). Enligt VISK (§ 373) är det en produktiv ändelse.

Redaktionsrådet för SFSO hade i ett tidigt skede slagit fast en princip som löd ungefär så här: "Adverb på *sti* kan uteslutas som uppslagsord om de svenska ekvivalenterna skulle bestå av *t*-former av ekvivalenterna till motsvarande adjektiviska uppslagsord." Det innebar alltså att adverbet kunde utgå vid uppslagsordspar som *ammattilinen* – *ammattilisesti* när adjektivekvivalenten *professionell* motsvaras av "adverbet"⁵ *professionellt*.

Det här speglar en övertro på systematik i språk. Morfologiskt motsvarar *sti*- och *t*-formerna visserligen varandra, men funktionellt gör de det bara delvis. De finska *sti*-formerna fungerar nästan genomgående som bestämningar till verb och particip. Svenska *t*-former bestämmer däremot i hög grad adjektiv eller adverb – *professionellt trovärdig, otroligt mycket* – och bara en del av dem förekommer som bestämning till verb (däremot nog till particip som ju är adjektiviska). Om användaren tror att de adjektiviska ekvivalenterna kan förses med ändelsen *-t* och läggas in samma typ av struktur som de finska adverbena leder det lätt till övergenerering. I artikeln *luottavainen* t.ex. finns ekvivalenter som *förtroende-full, tillitsfull, förtröstansfull*, men *t*-formerna passar inte som motsvarigheter till adverbet *luottavaisesti*: en sats som *fatta beslut förtroendefullt* osv. låter inte riktigt idiomatisk på svenska.

En rätt omfattande uppgift när uppdateringsarbetet inleddes 2017 var att föra in de uteslutna adverbblemmarna i ordboken. Så här ser t.ex. artikeln *luottavaisesti* ut i den nya webbversionen:

luottavaisesti

► ***suhtautua varovaisen luottavaisesti jhk***

ha (känna) en viss tillförsikt inför ngt

ha en viss tilltro till ngt

hän suhtautuu aikuisiin luottavaisesti

hon litar på de vuxna

hon känner tilltro (tillit) till de vuxna

⁵ Redaktionsrådet använde termen *adverb* för de svenska *t*-formerna, vilket förekommit i skolgrammatik. SAG (2:681) definierar *t*-formerna som adjektiv som kan ha adverbial funktion. Den finska deskriptiva grammatiken VISK (§ 373) betraktar orden på *sti* som adverb men nämner att de adjektiv- och participavledda formerna på *sti* ligger på gränsen till att vara böjningsformer av adjektiv.

suunnata katseensa luottavaisesti tulevaisuuten

se framtiden an med tillförsikt

he katselevat luottavaisesti kohti kolmivuotiskautta

de ser fram emot treårsperioden med tillförsikt

näiden tietojen avulla voimme tehdä päätöksiä luottavaisesti

de här uppgifterna ger en tillförlitlig grund för våra beslut

I artikeln återfinns inga ekvivalenter alls, och så är fallet med rätt många artiklar av den här typen. Det visade sig också att ekvivalenterna till adverbena, där sådana finns, ofta är färre än för motsvarande adjektiv. Det är uppenbart att det bara är vissa betydelser av adjektiv med vitt betydelseomfång som tenderar att adverbialiseras. Men också när ekvivalenter ges i artiklarna behövs det ofta exempel för att visa att man på svenska många gånger använder sig av andra konstruktioner.

I själva verket kan en viss systematik iakttagas när man går igenom ett stort material. Den finska konstruktionen [V] [Adj-sti] motsvaras av åtminstone fem olika svenska mönster:

Exempel*laulaa kauniisti**vaikuttaa fataalisti jhk**suhtautua kielteisesti**kertoa ylpeästi**suhtautua kielteisesti***Svensk konstruktion*****exempel***[V] [Adj-t]
sjunga vackert[V] [på ett [Adj] sätt]
påverka ngt på ett fatalt sätt[V-sig] [Predikativ]
förhålla sig negativ
ställa sig negativ[V] [med] [S-egenskap]
berätta med stolthet[V] [Adj] [S]
ha en negativ inställning

Den här typen av motsvarigheter är något som eventuellt kunde visas i en översättningsmodell av den typ som diskuterades ovan, för i de enskilda ordboksartiklarna framträder det här mönstret inte särskilt tydligt. Alla *sti*-artiklar har inte försetts med exempel på samtliga typer av konstruktion.

5. Normering

Det sista ämnet jag ska ta upp till diskussion är normering. I och med att SFSO:s redaktion var förlagd till Forskningscentralen för de inhemska språken kom normfrågor att aktualiseras i betydligt högre grad än vad som hade varit fallet om ordboken hade getts ut på gängse vis av ett förlag. Inom Forskningscentralen verkade den officiella språkvården både för finska och för svenskan i Finland. Manuskriptet granskades därtill av Svenska språknämnden (numera Språkrådet inom Institutet för språk och folkminnen) i Stockholm.

Med tanke på hur mycket normfrågor diskuterades är det anmärkningsvärt att inte ett ord sägs i förord eller användaranvisningar om hur SFSO förhåller sig till normen, eller rättare sagt normerna, eftersom ordboken ju har både en finsk och en svensk norm att förhålla sig till.

Det fanns också en viss olikhet i synsätt mellan redaktionen, som i högre grad hade en deskriptiv ordbok i åtanke, och en del av redaktionsrådet och språkgranskarna, som starkare förespråkade ett normativt förhållningssätt. Resultatet blev en kompromiss, och SFSO fick alltså vissa normerande funktioner. Vi har alltså att göra med förtäckt normering (jfr Svensén 2004:29f.; Nikula 1992).

5.1. Källspråksnormering

De finska kollegor som ingick i redaktionsrådet och de som granskade manuskriptet tänkte på ordboken ur finsk synvinkel. Det var naturligtvis deras uppgift, men de hade ibland en tendens att förbise att SFSO hade en annan funktion än PS. De kom då och då, särskilt i början av arbetet, med kommentarer om att icke-rekommenderade ord eller uttryck borde strykas ur manuskriptet. De tänkte då inte på att SFSO-användarna behövde få hjälp med översättningen av finska ord och uttryck till svenska oavsett om dessa var normenliga eller inte, eller att det knappast är för att kontrollera gångbarheten på finska ord som man i första hand slår upp i en finsk-svensk ordbok.

PS och i dess efterföljd KS är deskriptiva ordböcker men med en del normativa funktioner. I artiklarna finns det i vissa fall angivet att ett uppslagsord eller ett uttryck inte följer språkvårdsnomen. Ett exempel är KS-artikeln *hälyyttää* 'larma':

hälyyttää pitää olla: hälyttää.
Sanaa ei suositeta käytettäväksi.

'ska vara: hälyttää'
'Ordet rekommenderas inte'

I SFSO ingår inga explicita normangivelser av det slaget, men den finska språkvårdens önskemål tillmötesgicks såtillvida, att det i de här fallen hänvisas till korrekt form:

hälyyttää ks. hälyttää

Hänvisningarna representerar en form av förtäckt normering, och frågan är i vilken mån de fungerar som norm. Användaren har ingen möjlighet att skilja dem från hänvisningar som görs mellan uppslagsord av andra skäl.

När det gäller icke-rekommenderade exempelfraser ges svenska översättningar/frasmotsvarigheter, eftersom det före de elektroniska ordböckernas tid inte fanns möjlighet att adressera en hänvisning till ett exakt ställe i en annan artikel. Här bör det understrykas att normdiskussionerna fördes med en tryckt ordbok för ögonen, en elektronisk version började planeras på förlaget på allvar först när den tryckta boken var klar.

Redaktionen var i själva verket skeptisk till behovet av normering av källspråket: SFSO var ju inte avsedd att användas för att kontrollera den finska normen. Numera, när ordboken kan användas också i riktningen svenska-finska börjar det bli aktuellt med en ny diskussion kring detta inom redaktionen (jfr Hannesdóttir 2016).

5.2. Målpråksnormering

För en bilingval ordbok som SFSO var det dock framför allt kring målpråket normeringsdiskussionerna fördes. Det rådde i alla fall allmän enighet om att målpråket skulle vara *normenligt* såtillvida att stavning och böjningsangivelser skulle följa SAOL. Användarna ska inte behöva gå till andra källor för att kontrollera att ett ord faktiskt skrivs på det sätt SFSO anger. Etiketteringen av de svenska orden och uttrycken med avseende på stil och bruklighet skulle också följa de svenska ordböckerna.

En finsk-svensk ordbok som i hög grad är tänkt att användas i finländska förhållanden måste beakta den finlandssvenska språkvarieteteten. Det var också den, dvs. de s.k. *finlandismerna*, som normeringsdiskussionerna i hög grad kom att handla om. En finlandism är ett ord eller uttryck som förekommer enbart eller huvudsakligen i finlandssvenska eller ett ord eller uttryck som förekommer i en annan betydelse i finlandssvenska än i sverigesvenska. Det är en definition i grova drag, för en utförlig diskussion hänvisas till af Hällström-Reijonen (2012:78–84).

Den finlandssvenska språkvården arbetar för att svenskan i Finland, åtminstone som den talas och skrivs i mer officiella sammanhang, inte ska fjärma sig för mycket från sverigesvenskt standardspråk, och i det arbetet har finlandiserna spelat en framträdande roll. Men även om synen på finlandismer ännu på 1990-talet var något striktare än vad den är i dag, var det i alla fall ingen kring SFSO som förespråkade att finlandiserna skulle uteslutas helt och hållet.

Redaktionen och redaktionsrådet försökte enas kring vissa allmänna principer för hur finlandismer skulle hanteras. Principerna finns beskrivna i Martola (1992) och i Cantell (2015) så jag redogör inte för dem här men konstaterar att det alltså var förtäckt normering som förespråkades: ”icke-önskvärda, icke-rekommenderade” finlandismer skulle inte tas med i ordboken med motiveringen att en alltför frikostig redovisning skulle bidra till ökad användning. De finländska användarna antogs tolka anförda finlandismer som ett slags påbud, att det är de ord som ”måste” användas i Finland i stället för de allmänsvenska ekvivalenterna.

Det övergripande argumentet var att det i praktiken inte finns några deskriptiva ordböcker utan att användarna tolkar också förment deskriptiva ordboksartiklar normativt. Nikula (1992) problematiserar begreppet ”deskriptiv”, men håller i princip med om påståendet. Ordböcker som används för produktion får en normerande effekt även om de är avsedda att vara deskriptiva, eller *icke-normerande*, som Nikula anser vore en bättre term. Hon gör emellertid en annan slutledning: eftersom det inte är så lätt att ändra på ordboksanvändarna vore det bättre med en ”förändring i riktning mot öppet i stället för förtäckt normerande ordböcker” så att man lämnar det avgörande beslutet om normvalet till användaren (Nikula 1992:49).

6. Avslutning

Det har gått trettio år sedan principerna för SFSO slogs fast. Mycket har hänt inom lexikografin under de åren, dels inom teoretisk lexikografi, dels vad gäller redigeringsmetoder och de verktyg som står till buds. När man sitter vid sin dator i dag och redigerar förundrar man sig över att det ändå blev så pass bra ordböcker som det blev på den tiden. Möjligheterna att kontrollera och dra slutsatser om språkliga fenomen är helt andra när man har tillgång till webben och omfattande korpusar. En stor mängd viktiga och adekvata källor av olika slag finns ett par klick bort och är elektroniskt sökbara. Men det betyder inte nödvändigtvis att ordboksarbete enbart blivit lättare. Dels drunknar man lätt i allt material, dels konfronteras man med en hel del problem som man som redaktör var lyckligt omedveten om på 1990-talet. Exempelvis blir problemet med motstridiga källor betydligt större när källorna är fler.

Också synen på språk har förändrats. Under redigeringen av första upplagan av SFSO förekom det en viss överbetoning av språkens systematik på bekostnad av funktion och användning. Den ovan diskuterade motsvarigheten mellan finska adverb på *-sti* och svenska neutrumformer (*iloisesti* ~ *glatt*) är ett sådant exempel. Likaså antogs det att sammansättningar som bildades lika, t.ex. *myyntiaika* ~ *försäljningstid* (till *myynti* 'försäljning' och *aika* 'tid') kunde anses självklara. Redaktionen skulle inte använda onödig tid på dem. Visserligen är en förhållandevis stor del av sammansättningarna rätt "självklara", men alla är det inte och det vet man ju inte förrän man granskat dem i kontext. *Myyntiaika* är exempel på en sammansättning som ofta får andra motsvarigheter än ekvivalenten *försäljningstid* vid översättning.

Det fanns också, åtminstone i viss utsträckning, ett systembetingat krav på att exemplen inte fick "halta". Om det behövdes en explicit UPPLEVARE i en svensk översättning skulle upplevaren läggas in också på finska – också när det hade varit mer idiomatiskt utan. Man befarade att "haltande" exempel skulle förvirra användarna.

På 1990-talet var osäkerheten kring användarna stor och redaktionen kan sägas ha gapat efter för mycket genom att försöka beakta alla möjliga olika användares behov. En ordboksredaktion borde definiera de tilltänkta användarnas behov bättre och slå fast principerna efter det. Å andra sidan måste dagens redaktion förhålla sig till problemet att ordboken faktiskt kan användas av en ännu brokigare skara och också i svensk-finsk riktning i och med att den finns gratis tillgänglig på internet. Det är ett stort problem att hantera framöver.

Tack vare att en hel del användarundersökningar genomförts de senaste trettio åren vet vi lite mer om ordboksanvändarna nu. Men för att en användarundersökning verkligen ska vara till konkret nytta vid redigeringen av en enskild ordbok borde en undersökning genomföras på just den specifika ordbokens målgrupp(er). Jag hoppas att någon intresserad lexikograf åtar sig att göra en användarundersökning på SFSO:s användare inför nästa större revidering.

En ordbok kan aldrig göra rättvisa åt det rika, dynamiska och mångfasetterade språket som ständigt är i rörelse och där betydelser ständigt modifieras och förhandlas. En ordbok är ju ett statiskt och fryst utsnitt av en bit av språket, hur omfattande den än är. I en tvåspråkig ordbok tenderar målspråket dessutom att förlora ännu mer av sin glans än källspråket, eftersom bara en del av dess resurser aktualiseras av källspråket. Målspråket ter sig otympligt när det ska silas genom källspråkets tvingande filter. På det viset är en tvåspråkig ordbok strängt taget en omöjlighet – ett absurt åtagande. Men trots allt kan ordboksanvändaren få mycket hjälp av det frysta utsnittet, och den tvåspråkiga ordboken behövs som ett redskap för språkanvändarna. Lexikografen få nöja sig med "gott nog" och i själva verket är "gott nog" inte illa alls.

Litteratur

Ordböcker

Cannelin, Aulis, Knut Cannelin, Lauri Hirvensalo & Nils Hedlund (1976): *Suomi-ruotsi suursanakirja. Finsk-svensk storordbok*. Helsingfors: WSOY.

KS = *Kielitoimiston sanakirja* (2018). Kotimaisten kielten tutkimuskeskus. <www.kielitoimistonsanakirja.fi>

PS = *Suomen kielen perussanakirja* (1990–94). Kotimaisten kielten tutkimuskeskuksen julkaisuja 55. Helsinki.

SAOL = *Svenska Akademiens ordlista*. 2015. <www.svenska.se>

SFSO 1997 = *Suuri suomi-ruotsi-sanakirja/Stora finsk-svenska ordboken*. (1997). Redaktion: Birgitta Romppanen (huv.red.), Ilse Cantell, Nina Martola & Mats-Peter Sundström. Forskningscentralen för de inhemska språken & WSOY. Helsingfors.

SFSO 2004 = *Suomi-ruotsi-suursanakirja/Stora finsk-svenska ordboken*. (2004). Forskningscentralen för de inhemska språken & WSOY: Helsingfors.

SFSO 2019 = *Suomi-ruotsi-suursanakirja/Stora finsk-svenska ordboken*. Institutet för de inhemska språken (innehåll) & Institutionen för svenska språket vid Göteborgs universitet (teknikplattform). <finsk-svenska.svenska.gu.se> (maj–november 2019).

Van Dale = Laureys, Godelieve, Hans de Groot m.fl. (1996): *Van Dale handwoordenboek Nederlands-Zweeds = Norstedts Nederländsk-Svenska ordbok & Van Dale handwoordenboek Zweeds-Nederlands = Norstedts Svensk-Nederländska ordbok*. Utrecht: Van Dale Lexicografie – Stockholm: Norstedts.

Annan litteratur

Bergenholtz, Henning (1991): Lemmaselektion in zweisprachigen Wörterbüchern. I: Gregor Meder och Andreas Dörner (red.): *Worte, Wörter, Wörterbücher. Lexikographische Beiträge im Essener linguistischen Kolloquium*. Tübingen: Niemeyer, 49–65.

Cantell, Ilse (2015): *Stora finsk-svenska ordboken. Ett lyft för tvåspråkig lexikografi i Finland*. I: Caroline Sandström, Ilse Cantell, Eija-Riitta Grönros, Pirkko Nuolijärvi och Eivor Sommardahl (red.): *Perspektiv på lexikografi, grammatik och språkpolitik*. Institutet för de inhemska språkens webbpublikationer 39. <scripta.kotus.fi/www/verkkojulkaisut/julk39/antologi2015.pdf>

Euralex – European Association for Lexicography <www.euralex.org>

- Hannesdóttir, Anna Helga (2016): Tvåspråkig lexikografi – ny teknik, nya funktioner. Den digitala ordboken som parallellkorpus. I: *Framtidens lexikografi. Meijerbergs arkiv för svensk ordforskning* 42, 29–58. <hdl.handle.net/2077/49808>
- af Hällström-Reijonen, Charlotta (2012): *Finlandismer och språkvård från 1800-talet till i dag*. Nordica Helsingiensia 28. Helsingfors: Finska, finskugriska och nordiska institutionen. <www.core.ac.uk/download/pdf/14922565.pdf>
- ISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen, Irja Alho (2004): *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seuran toimituksia 950. Helsinki: SKS.
- Laureys, Godelieve (1997): Den svensk-nederländska/nederländsk-svenska ordboken. En bidirektionell ordbok i ett dubbelt binationellt perspektiv. I: *Nordiska Studier i Lexikografi* 4. <www.docplayer.se/18268132-Nordiske-studier-i-leksikografi.html>
- Martola, Nina (1992): Arbetet med Finsk-svensk ordbok: Normeringsproblem visavi finlandssvenskan. I: *Nordiske Studier i Leksikografi* 1. <www.tidsskrift.dk/nsil/article/view/19766/17385>
- Martola, Nina (1993): Översättarhjälp i en ordbok. I: *Nordiske Studier i Leksikografi* 2. <www.tidsskrift.dk/nsil/article/view/19693/17311>
- Martola, Nina (2001): Översättningsmodeller i en ordbok? En liten användarundersökning. I: *Nordiska studier i lexikografi* 5 <www.tidsskrift.dk/nsil/article/view/19464/17090>
- Martola, Nina (2007): *Konstruktioner och valens. Verbfraser med åt i ett jämförande perspektiv*. Nordica Helsingiensia 9. Helsingfors universitet. <www.core.ac.uk/download/pdf/14915194.pdf>
- Martola, Nina (2011): Den finlandssvenska språkbanken och Korp. I: *Språkbruk* 4/2011. <www.sprakbruk.fi/-/den-finlandssvenska-sprakbanken-och-korp>
- Nikula, Kristina (1992): Deskriptiva ordböcker – finns dom? I: *Nordiske Studier i Leksikografi* 1. <www.tidsskrift.dk/nsil/article/view/19694/17312>
- SAG = Teleman, Ulf, Staffan Hellberg, & Erik Andersson (1999): *Svenska Akademiens grammatik*. Stockholm: Svenska Akademien & Norstedts ordbok.
- Svensén, Bo (2004): *Handbok i lexikografi: Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts akademiska förlag.
- VISK = Nätversionen av ISK. <scripta.kotus.fi/visk/etusivu.php>

Nina Martola
huvudredaktör, fil.dr
Institutet för de inhemska språken
Hagnäskajen 6
FI-00530 Helsingfors

Abstrakt:

Responsive dictionaries: bringing dictionaries (and lexicographers) and users closer together

Iztok Kosem

A recent study on the use of (monolingual) dictionaries across Europe (Kosem et al. 2019) has shown that up-to-date content is among the top three dictionary features valued by users. In fact, in countries where existing dictionaries are rather outdated, the users rank up-to-date content as high as reliability, sometimes even higher. The message is clear – users want dictionaries that reflect contemporary language and that react to the changes in language as quickly as possible. The advances in corpus tools and related lexicographic tools (e.g. tools for neologism detection) have enabled lexicographers to address these user demands rather successfully; however, this is mainly the case for languages with good existing dictionary infrastructures, such as English, German etc.

However, in countries such as Slovenia, one of the main problems is the outdatedness of dictionaries, and even the non-existence of, or poor accessibility to certain, more specialised dictionaries. This means that these dictionaries need to be made from scratch, which would traditionally also mean years (even decades) of waiting for the users. In order to address this issue, the team at the Centre for Language Resources and Technologies at the University of Ljubljana has tested and implemented several state-of-the-art approaches into the lexicographic process, such as automatic extraction of lexical data from the corpus (Gantar et al. 2016; Kosem et al. 2013), and crowdsourcing (Čibej et al. 2015). Furthermore, we have used semi-automatic methods for the creation of language resources such as Thesaurus of Modern Slovene (Krek et al. 2017a).

The implementation of the methods mentioned above had one important consequence for the dictionary-making process; namely, the data obtained with (semi-)automatic approaches was already of such quality that we decided to present it to the users. Two dictionaries, *Thesaurus of Modern Slovene* (Krek et al. 2017b) and *Collocations Dictionary of Modern Slovene* (Kosem et al. 2018), were published in this manner. We named them responsive dictionaries as they are able to quickly “respond” to changes, either because of completed lexicographic work and/or availability of new data (e.g. corpora). Another form of responsiveness is related to the users, as we offered them the opportunity to be involved in the dictionary-making process via voting on particular types of information, suggesting new content, or via specific crowdsourcing tasks.

In my talk, I will present both dictionaries, paying particular attention to the lexicographic process behind both dictionaries, and the contributions and feedback received from the users. In addition, as the dictionaries share certain types of data (collocations, examples), I will show how they were integrated into a large database where all the information from the two dictionaries, as well as from

other resources (including crowdsourcing tasks), is combined. I will also discuss some of the solutions that had to be implemented in the dictionary interface to provide the users with the necessary information about the reliability of dictionary data, and to enable their participation in a straight-forward manner. I will argue that the adoption of such a model of dictionary publication brings several benefits to both users and lexicographers. In conclusion, I will present future plans, including the implementation of various methods that are being developed within the ELEXIS (European Lexicographic Infrastructure) project.

References

- Čibej, Jaka, Darja Fišer & Iztok Kosem (2015): The role of crowdsourcing in lexicography. In: Kosem, Iztok et al. (eds.): *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Ljubljana: Trojina, Institute for Applied Slovene Studies; Brighton: Lexical Computing, pp. 70-83. <www.elex.link/elex2015/proceedings/eLex_2015_05_Cibej+Fiser+Kosem.pdf>
- Gantar, Polona, Iztok Kosem & Simon Krek (2016): Discovering automated lexicography = the case of Slovene lexical database. In: *International journal of lexicography* 29/2, pp. 200-225. <www.academic.oup.com/ijl/article/29/2/200/2413284/Discovering-Automated-Lexicography-The-Case-of-the?guestAccessKey=95f18766-f10f-4994-a6fa-448cf75ac55e>
- Kosem et al. (eds.) (2017): *Kolokacije 1.0: Collocations Dictionary of Modern Slovene*. <viri.cjvt.si/kolokacije>
- Kosem et al. (2019): The image of the monolingual dictionary across Europe. Results of the European survey of dictionary use and culture. In: *International Journal of Lexicography* 32/1, pp. 92-114. <academic.oup.com/ijl/article-abstract/32/1/92/5237031?redirectedFrom=fulltext>
- Kosem, Iztok, Polona Gantar & Simon Krek (2013): Avtomatizacija leksikografskih postopkov. In: Erjavec, Tomaž & Jerneja Žganec Gros (eds.): *Jezikovne tehnologije, Slovenščina 2.0, 1/2*. Ljubljana: Trojina, zavod za uporabno slovenistiko, pp. 139-164. <www.trojina.org/slovenscina2.0/arhiv/2013/2/Slo2.0_2013_2_07.pdf>
- Krek et al. (eds.) (2017b): *Sopomenke 1.0: Thesaurus of Modern Slovene*. <viri-cjvt.si/sopomenke>
- Krek, Simon, Cyprian Laskowski & Marko Robnik-Šikonja (2017a): From translation equivalents to synonyms: creation of a Slovene thesaurus using word co-occurrence network analysis. In: Kosem, Iztok et al. (eds.): *Proceedings of eLex 2017: Lexicography from Scratch, 19-21 September 2017, Leiden, Netherlands*. <www.elex.link/elex2017/wp-content/uploads/2017/09/paper05.pdf>

Iztok Kosem
 Jožef Stefan Institute (ELEXIS project)
 Centre for Language Resources and Technologies,
 University of Ljubljana
 iztok.kosem@ijs.si

Temasektion

Temasektion om etymologi

Santeri Junttila

Som en del av den 15 konferensen om lexikografi i Norden ordnades en etymologisk temasektion med åtta bidrag. Ansvarig för denna temasektion var arbetsgruppen för den etymologiska nätdordboken över det äldsta finska ordförrådet (*Etymologinen verkkosanakirja* eller EVE) under Santeri Junttilas ledning. Sektionen hölls den 5 juni med sju föredrag och en poster. I fokus för tre av föredragen stod etymologisk e-lexikografi. Två av föredragen inom sektionen har resulterat i artiklar som är publicerade i denna volym: Hanne Lauvstad (NAOB:s etymologier) s. 195–203 och Johan Schalin (Urfenniska **litna* (finska *linna*, estniska *linn*) ’borg, köping’) s. 285–295.

Mikko Bentlin från Greifswalds universitet (Tyskland) presenterade sitt projekt om Osmo Nikkiläs (1933–2002) manuskript över svenska lånord i finskan. Bentlin arbetar sedan år 2015 med att publicera Nikkiläs etymologier i Språkbankens wikimiljö sanat.csc.fi (sanat.csc.fi/wiki/Nikkilän_lainasanat). Han presenterade databasens struktur med hjälp av exempel från de redan färdigställda artiklarna.

I samma wikimiljö verkar också den etymologiska nätdordboken över det äldsta finska ordförrådet (EVE), som Santeri Junttila presenterade i sitt föredrag inom temasektionen. I Junttilas arbetsgrupp vid Helsingfors universitet forskar Petri Kallio, Sampsa Holopainen, Juho Pystynen och Juha Kuokkala. Projektet finansieras sedan 2018 av Konestiftelsen. Syftet med projektet är att ge etymologier för alla de ungefär 2 400 ord som finskan har ärvt från urfenniskan. Här ingår de vanligaste orden som *minä* ’jag’, *metsä* ’skog’ och *mennä* ’gå’, men också många sällsynta dialektala ord som *kuu* ’talg’ och *orasa* ’kastrerad galt’. Den viktigaste metodologiska innovationen är att branschens alla akademiska forskare öppet får delta i redaktionsarbetet samt publicera och diskutera etymologier på ordbokens wikisidor.

Vesa Jarva från Jyväskylä universitet talade om ”gradvis etymologi” hos finska dialektord (fi. *Suomen murteiden sanaston asteittaista etymologiaa*). Det finns en stor mängd ord i de finska dialekterna – i synnerhet i de östliga dialekterna – som till sitt yttersta ursprung är lånade från ryska. De har ändå utvecklats till ordfamiljer som i fråga om såväl fonetik som semantik uppvisar en onomatopoetisk karaktär. ”Gradvis” innebär i fråga om dessa ord att en exakt etymologi inte kan fastslås. Det kan finnas ett samband med ett visst lånord men orden beter sig som varierande, expressiva ord. Till exempel ordklungan *jaassikka* ~ *jaassiekka* ~ *jassakka* ~ *jassikka* ~ *jessikka* ~ *jässikkä* ~ *jässäkkä* härstammar

från det ryska *яуук (jasjtjik)* 'låda, ask', men den ursprungliga betydelsen har i de finska dialekterna påverkats av den deskriptiva betydelsen 'en undersäsig el. kraftig människa, klumpedump', jfr *jässi, jäsy* '(muskel)kraft'.

Miro Metsämuuronen från Helsingfors universitet talade om färöisk lånordsforskning i ett historiskt perspektiv. I sitt vetenskapshistoriska föredrag redogjorde Metsämuuronen för den färöiska lånordsforskningens inriktning och särintressen under olika perioder. Forskningsområdet uppstod genom Jakob Jakobsens studier i keltiska lånord i färöiskan, t.ex. fär. *Drunnur* 'Ende, Halestykke af et Kreatur' ← gäl. *Dronn* 'den bageste af Ryggen', och på samma spår fortsatte också Chr. Matras, t.ex. *blak* 'kärnmjök' < **blaðak* ← gäl. *blàthach* id. På senare tider har T. Jóansson forskat på de engelska inflytandena, t.ex. etymologin *dreggj* (ett fiskeredskap) ← eng. *dredge* id. De äldre danska lånorden är fortfarande i stort sett outforskade. Enligt Metsämuuronen kan lånetymologierna bidra till en ökad förståelse av färöisk ljudhistoria, men först efter en kritisk genomgång av materialet då det består av förslag av högst varierande kvalitet.

Johan Schalin och Juha Kuokkala från Helsingfors universitet höll en posterpresentation som behandlade en ny germansk lånetymologi, den urfenniska **šike-/šiko-* (finska *hiki* 'svett', *hiota* 'svettas'). Den sannolikaste källformen har varit urgermansk **sigan* ~ *sigōn-*, som har utvecklat sig till svenska *sege*, dalmålets *siga* 'vatten som tränger upp genom marken, källåder; vätska som rinner ur sår eller böld'.

Minerva Pihas tema var mytologiska lånord från de skandinaviska språken i sydsamiskan. De skandinaviska grannspråken har gett sydsamiskan ett stort antal lånord, men Pihas ordmaterial visar att de oftast har utvecklat mytologiska betydelser först i sydsamiskan. Eftersom den samiska mytologin inte i någon större utsträckning bär spår av skandinaviska element, verkar orden primärt ha lånats in via andra kontexter än de religiösa. Till exempel i *Hovrengaellies* 'åskguden' har förleden *Hovren-* lånats från fornnordiska *Þórr* (sv. Tor), medan grundordet *gaellies* är lånat från urnordiskan i betydelsen 'make'.

Följande dag, den 6 juni, organiserade EVE-arbetsgruppen en inskolnings-eftermiddag för alla konferensdeltagare som var intresserade av att ta del i diskussioner om etymologier och det redaktionella arbetet med ordbokens wikisidor. Inskolningen fick ett drygt dussin deltagare. Både temasektionen om etymologi och inskolningen var till stor nytta och glädje för arbetsgruppen.

Santeri Junttila
forskare, projektledare
Greifswalds universitet
Helsingfors universitet
santeri.junttila@helsinki.fi

Sektioner

Enspråklig ordbok for fremmedspråklige

Øivin Andersen

In this paper I describe some basic principles of editing the monolingual dictionary for refugees and immigrants in Norway (*Norsk ordbok for fremmedspråklige*). The basic structure of a dictionary entry is illustrated, both as a working document and a finished product. The application of prototype theory and lexical field theory are discussed. Basic principles of pronunciation coding and coding of grammatical information are also described.

1. Bakgrunn

Utgangspunktet for denne enspråklige ordboken, *Norsk ordbok for fremmedspråklige*¹, er den norsk-vietnamesiske ordboken (heretter NVO) som kom ut på Universitetsforlaget i 1983 (Khánh & Tràn 1983). Arbeidet ble påbegynt i 1979. Det mest spesielle med denne ordboken er dens skreddersydde tilpasning til de vietnamesiske flyktningenes behov i Norge. Vietnameserne selv tok initiativet til ordboken, men innså raskt at de trengte hjelp. Universitetet i Bergen bevilget driftsmidler til prosjektet og det ble etablert en arbeidsgruppe bestående av norske og vietnamesiske fagfolk.

Det første problemet i NVO-prosjektet var å etablere et ordtilfang tilpasset vietnamesernes behov. Et frekvensbasert ordutvalg ble presentert for lærere i norsk som andrespråk og fremmedspråk hvor de ble bedt om å krysse av for viktige ord og supplere i ordlisten. Ord fra ulike innvandrerbrosjyrer og lærebøker i norsk for innvandrere ble ekserpert og innlemmet i listen. Det ble også arbeidet terminologisk med fagtermer av særlig betydning for flyktninger og innvandrere. Det ble særlig fokusert på områdene medisin og jus. På disse to områdene ble det nedsatt oversettelsesgrupper bestående av sakkyndige personer. Norske medisinstudenter arbeidet sammen med vietnamesiske leger som i hovedsak var utdannet i Frankrike. Tilsvarende gruppe ble også etablert innen fagområdet jus.

For hele det norske ordtilfanget ble det laget synonymer og definisjoner som i etterkant ble oversatt til vietnamesisk. Oversettelsesprosessen ble utført ved etablering av norske forklaringer og definisjoner av oppslagsord og underoppslag, inkludert idiomatiske konstruksjoner. Disse ble diskutert med våre vietnamesiske medarbeidere med henblikk på vietnamesiske nærekvivalenter. Det er først og fremst dette som dannet utgangspunktet for videreføring av den norske delen

¹ Ordboken er fremdeles under utarbeidelse i skivende stund.

av NVO til en enspråklig ordbok for fremmedspråklige som er bosatt eller skal bosettes i Norge generelt. Dette gjelder særlig flyktninger som har fått permanent opphold i Norge.

Disse personene får et obligatorisk norskkurs og det blir stilt tøffe krav til norskkunnskaper når de skal integreres i det norske arbeidslivet. Derfor har de behov for en tilpasset ordbok både i kursopplæringen og senere ved den sosiolekttilpasningen som skjer når de blir plassert i forskjellige deler av Norge. Dette innebærer også en integrering i norsk kultur- og arbeidsliv.

Dette arbeidet har foregått siden siste halvdel av 80-tallet med diverse avbrudd, men fra 2011 ble revisjonsarbeidet og oppdateringsarbeidet med ordboken tatt opp og har vært drevet kontinuerlig frem til i dag. I det følgende skal jeg redegjøre for dette redigeringsarbeidet.

2. Prinsipper for redigering

Først vil jeg illustrere ordboksartikkelens oppbygging som arbeidsdokument. Som illustrasjon brukes substantivet **hagl**.

```

S HAGL
100 hagl
120 h_agel
140 s.n.
150 (hagl^et, -, -a/-ene)
312 nedbør som består av kompakte, runde iskorn
000 312 skal koordineres for alle ordene REGN, YR, SNØ,
HAGL, SLUDD. Dette er gjort per 6.7.15.
314 Haglene slo blomstene ned.
314 I Syden kan hagl være store som fugleegg.
316 hagl+byge s.fm.
322 liten blykule som brukes i ladning for haglgevær
324 Man skyter fugl og hare med hagl.
327 hagle s.fm.
327:haglgevær
327 hagl+gevær s.n.
327:våpen som lades med hagl

```

Figur 1: Ordboksartikkel som arbeidsdokument.

I arbeidsdokumentet i figur 1 har hver informasjonsdel sin egen separate tallkode: 100 er oppslagsordet, 120 er uttaleangivelse, 140 er ordklasseinformasjon, 150 er bøyningsparadigmet, 312 er oppslaget leksikalske definisjon for betydning 1. Koden 000 er reservert for interne kommentarer for ordboksmedarbeiderne. 314 er reservert for eksempelsetninger og fraser som ikke trenger videre forklaring. I

figur 1 er det tatt med to 314 eksempelsetninger. 315 brukes til ordtak (ordtak er ikke representert i oppslaget **hagl**) og idiomatiske uttrykk som er ikke-komposisjonelle, og som derfor forklares. Eksempel er oppslag som **snø**: 315: *snøen som falt i fjor*: 'noe som tilhører fortiden; noe som er avsluttet/uaktuelt/glemt'. 316 brukes for komposisjonelle sammensetninger som angis uten forklaring, mens 317 brukes for ikke-komposisjonelle sammensetninger med forklaring. 317 brukes også for å angi underoppslag som er avledninger dannet på hovedoppslaget, med forklaring. 322 angir oppslagets betydning 2, og 324, 325, 326, 327 har samme funksjon for betydning 2 som tilsvarende felt for betydning 1.

Hvis dette gjøres på en entydig og konsistent måte skal det ferdige produktet kunne genereres maskinelt som vist i figur 2.

<p>hagl [hagel] s.n.(hagl et, -, -a/-ene)</p> <ol style="list-style-type: none"> nedbør som består av kompakte, runde iskorn <i>Haglene slo blomstene ned. I Syden kan hagl være store som fugleegg.</i> hagl+bygge s.fm. liten blykule som brukes i ladning for haglgevær <i>Man skyter fugl og hare med hagl.</i> hagle s.f.m. haglgevær hagl+gevær s.n. våpen som lades med hagl hagle s.fm. haglgevær hag l+gevær s.n. våpen som lades med hagl
--

Figur 2: Ordboksartikkel som ferdig dokument.

2.1. Bruk av prototypeteori

I den internasjonale litteraturen har man diskutert bruken av prototypeteori i pedagogiske ordbøker (for en oversikt se Jiang & Chen 2015). Potensielt kan man tenke seg flere områder i ordartikkelens oppbygging hvor dette kan anvendes. Her skal jeg drøfte et sentralt anvendelsesområde: betydningsrekkefølge ved polysemi. Som illustrasjonseksempel vises substantivet *land*. Oppslaget har fire betydninger:

- del av jorden som ikke er hav/vann; fastland; jord
Den som har reist lenge til sjøs, synes det er godt å kjenne land under føttene igjen.
- område som utgjør en politisk og oftest geografisk enhet
Norden består av fem land: Danmark, Finland, Island, Norge og Sverige.
- stykke jord, særlig område som skal brukes til dyrking eller bygging
Staten kjøpte opp land til å bygge ny flyplass.
- (til forskjell fra 'by') område med spredt bebyggelse
Mange nordmenn har leilighet i byen og hytte på landet.

Intuitivt vil betydning 1 være den mest generelle og grunnleggende, og derfor klar kandidat for å komme først. Betydning 2 og 3 har et element av institusjonalitet i tillegg til det rent geografiske og betydning 4 er mer spesialisert enn 1. En slik rekkefølge gjør ordartikkelen mer motivert for brukerne. Men hvis betydning 1 er mest prototypisk vil man forvente at den også er den mest frekvente. For å sjekke dette brukte jeg det norske aviskorpuset. Dette korpuset er et monitorkorpus (McEnery & Wilson 2001:30f). Det går tilbake til 1998 og drives kontinuerlig av språk og språkteknologi-gruppen ved Uni Research i Bergen. Prosjektet heter Corpuscle. De tekniske redskapene er utviklet ved Clarinosenteret i Bergen (clarino.uib.no/korpuskel/page).

I desember 2019 hadde dette korpuset over 1,88 milliarder løpende ord-eksemplarer. Ordet *land* forekommer 546 203 ganger i korpuset. Jeg analyserte de 180 første forekomstene i kwik-konkordanser med hensyn til betydningsfordeling. Tidligere undersøkelser har vist at en manuell analyse av de første 180–200 forekomstene gir en tilstrekkelig pekepinn for de øvrige forekomstene, særlig ved høyfrekvente fenomener (cf. Andersen 2007). Betydning 1 forekom bare 13 ganger, betydning 2 forekom hele 161 ganger. Betydning 3 hadde 6 forekomster mens betydning 4 hadde 0 forekomster. En ren frekvensanalyse skulle derfor tilsi at betydning 2 skulle ha kommet først.

Frekvens og intuitiv prototypanalyse motsier hverandre, og det fins ingen enkle regler for hvordan man skal løse problemet.

Jiang & Chen (2015) undersøkte fem store enspråklige pedagogiske ordbøker for engelsk, blant annet med henblikk på struktureringen av prototypiske og ikke-prototypiske betydninger av oppslag. De kom frem til at prototyppløsninger kun ble sporadisk og intuitivt brukt. Forfatterne påpeker at årsaken til dette er at prototypeteorien mangler klare retningslinjer for leksikografisk anvendelse. Enda mer problematisk blir det å velge et prototypisk eksempel på bruk.

For fremmedspråklige innlærere vil sentral encyklopedisk informasjon kanskje være mer nyttig, som i betydning 2 under oppslaget **land** (se ovenfor). Selv når det er klart hva som er grunnbetydningen kan det være problematisk å rangere de andre variantene fordi de går i forskjellige retninger ut fra det sentrale. Under oppslaget **land** er det vanskelig å si om 2 eller 3 bør komme nærmest etter 1. De er omtrent like nær, men går i ulik retning. 3 er mer konkret enn 2, og knytter kanskje litt nærmere til betydningen av «jord, ikke sjø», men er ikke så klart avgrenset i et område som 2 er, og 1 inneholder noe om «del», altså avgrensning.

2.2. Arbeid med leksikalske felt

Et leksikon kan deles inn i en rekke leksikalske felt, også kalt semantiske felt (Bergenholtz et al. 1997:231). Bergenholtz et al. definerer et semantisk felt som en «samling av leksikalske enheter som er forent gjennom en viss betydnings-

messig samhörighet» (for oversikt over utvikling av teorien om leksikalske felt, se Öhman (1953)). Et eksempel på dette i vår ordbok er feltet «nedbør». Oppslaget **hagl** i figur 1 (ovenfor) inneholder en opplysningslinje som kun er til internt bruk for leksikografene. Feltet er merket med 000 og påpeker at definisjonsfeltet 312 skal koordineres for alle oppslagene *regn*, *yr*, *snø*, *hagl* og *sludd*. Dette feltet kan også mer presist defineres som et hierarkisk begrepssystem hvor begrepsrelasjonene utgjøres av overordning (genus proximum) og underordning (differentia specifica) (cf. Bergeholtz et al. 1997).

Genus proximum i vårt tilfelle er oppslaget **nedbør**. Definisjonsfeltet inneholder først en intensjonal definisjon: «vann (i en av sine mulige former) som faller fra skyer ned på jorden». Deretter følger en ekstensjonal definisjon som angir de relevante differentia specifica: *regn*, *sludd*, *snø* og *hagl*. De underordnede begrepene kan da defineres etter et bestemt mønster hvor genus proximum innleder definisjonen og differentia specifica utformes som en relativsetning som modifierer genus proximum. *sludd* defineres slik: «nedbør som består av både regn og snø». *snø* defineres som «nedbør som består av lette, hvite krystaller av is». *regn* defineres som «nedbør som består av vandrdåper» og *hagl* defineres som «nedbør som består av kompakte, runde iskorn».

Slike leksikalsk/semantiske felt bidrar til å gi ordboken mer systematikk og vil være motiverende og klargjørende for brukergruppen. Men i praksis er det ikke alle områder av ordboken som ligger til rette for en slik feltanalyse. Derfor har vi ikke satset på å gjøre ordboken til en begrepsordbok, som beskrevet i Bergeholtz et al. (1997:75).

2.3. Prinsipper for uttalemarkering

I Norge har det vært utgitt noen uttaleordbøker for norsk. Ifølge Vanvik (1985:9) har disse brukt «en norsk lydskrift som en ikke kan regne med blir forstått utenfor landets grenser». I stedet brukte han det internasjonale uttalesystemet IPA. Det er flere grunner til at denne løsningen ikke treffer den intenderte brukergruppen for vår ordbok. For det første baserte Vanvik seg på det han kalte standard østnorsk. Det har skjedd mye siden 1980-tallet. I stedet for å satse på en uttalenormering basert på et begrenset sosialt og geografisk område har tendensen i massemedia vært å stimulere våre ulike dialekter. Dermed blir en uttaleangivelse bortimot umulig å forsvare. For fremmedspråklige blir derfor våre betydelige dialekt- og sosiolektforskjeller en stor utfordring.

I fravær av aksepterte uttalenormer har vi utarbeidet et sett av prinsipper som vil være nyttige for vår brukergruppe. Hovedmålet vårt er å få frem én uttale som har stor utbredelse og som trygt kan brukes på tvers av dialektgrenser. For det første, aksentuering (ordtrykk) skal angis, også ved enstavelsesoppslag hvis oppslaget potensielt sett kan aksentueres, for eksempel [huus]. Her ser vi også

at lang vokal angis med dobbelskriving. For det andre, det skal i hovedsak være én uttale for hvert oppslag. I visse tilfeller kan man tillate to alternativer, som i oppslaget **annen** som dels uttales [an^hen] og dels [aa^hen]. For det tredje skal uttalen være så bokstavnær som mulig, og ingen fonetiske tegn skal brukes. Eksempelvis vil oppslag som **kirke** og **kino** transkriberes [kjirke] og [kjiino], siden bokstavkombinasjonen <kj> typisk uttales [ç], en palatovelar spirant. Da får man noen unntak hvor kombinasjonen <kj> skilles ved morfemgrense, som i *brekkjern* [br^hek-jern]. Dette markeres ved å innføre en bindestrek i uttaleangivelsen. Den velare nasalen [ŋ] angis som <ng> som i [sang]. I en del norske dialekter uttales [g] i denne posisjonen, dvs. i stavelseskoda. Et tostavellesord som *kongress* angis som [kång-gress].

Prinsippet om bokstavnær uttale tilsier også at vi unngår markering av de østnorske retro-fleksene, som i *ert* [æf]. Generelt bygger vi på de vestnorske konsonantene og de østnorske vokalene. Verbet *erte* uttalemarkeres derfor med [æ^herte]. For nærmere opplysninger om dette henvises til innledningen i Khánh & Tràn (1983:10f).

2.4. Grammatiske opplysninger

Når det gjelder ordklassebetegnelser følger vi i hovedtrekk Bokmålsordboken (ordbok.uib.no). Men i noen tilfeller avviker vi fra det. I vår ordbok har bøyning alltid forrang foran syntaktisk funksjon. For eksempel oppslaget **delvis**. Dette er registrert som adverb i Bokmålsordboken mens vi har registrert det som adjektiv fordi oppslaget har adjektivmorfologi, som i *delvise tilståelser*. Hvis oppslaget ikke har bøyning anvendes syntaktiske kriterier, som i oppslaget **dagevis**, for eksempel i frasen *i dagevis* hvor oppslaget opptrer som styring for en preposisjon, en typisk substantivisk funksjon.

Et annet problem er representasjonen av oppslag som refererer til ulike språk. For eksempel er oppslaget adjektivet **tysk** definert som «som har sammenheng med, kommer fra eller tilhører Tyskland». *Han studerer tysk språk og litteratur. Det fins mange kjente tyske komponister*. Denne typen oppslag blir typisk registrert som adjektiver. Men disse kan også opptre som substantiver: *Han snakker et utmerket tysk*. Den mest nærliggende forklaringen er å analysere dette som en elliptisk konstruksjon hvor substantivet *språk* er utelatt. Men hva da med konstruksjoner som *Tysken hans er dårlig*? Her kan ellipseløsningen ikke anvendes. Jeg har ikke funnet noen norske enspråklige ordbøker hvor dette er representert. Vår nødløsning er å registrere disse hovedopplagene som adjektiver, og to substantiver som underoppslag, som for eksempel under *tysk* adjektiv: *tysk* s.m. tysk språk, *tysk* s.n. (bare i ubestemt form) tysk språk. Dette representasjonsproblemet gjelder alle oppslag som refererer til ulike lands kulturelle og politiske forhold. Som adjektiver oppfører de seg regelmessig som andre adjektiver, men når

de refererer til ulike lands språk oppstår representasjonsproblemen. Det gjelder adjektiver som **polsk**, **engelsk**, **fransk**, etc. Andre refererer kun til språk, som **latin** og **svahili**. For eksempel *Han snakker et godt svahili/polsk/fransk, men engelsken/tysken/russisken hans er dårlig*. Dette er altså et representasjonsproblem som omfatter en hel gruppe av ord i ordboken.

Når man skal gi bøyningsparadigmer i en ordbok brukes lineær segmentering så langt det er mulig. Dette anvender vi også i slike tilfeller hvor det bryter med morfologisk analyse, som i oppslaget **møbel**: møb|e|et, -|er, -|a/-|ene. Hvis lineær segmentering ikke er mulig angis formene fullt ut, som i synge, sang, sunget. De norske tonelagene angis ikke. Mange fremmedspråklige snakker fullt funksjonelt norsk uten tonelag. I flere norske dialekter mangler tonelagsmotsetning, og fenomenet er forholdsvis begrenset i norsk til forskjell fra tonespråk som vietnamesisk (hvor tonemarkeringen uttrykkes ortografisk).

3. Konklusjon

I denne artikkelen har jeg vist noen grunnleggende aspekter ved ordboksarbeid som har som mål å tilpasse ordtilfang, strukturering av ordboksartikler, bruk av prototypeteori og leksikalsk feltteori, prinsipper for uttalemarkering, og grammatiske opplysninger. Alle disse aspektene viser at leksikografer alltid må ha en analyse av intendert brukergruppe i tankene ved arbeidet. Dette gjelder særlig det skreddersydde ordutvalget, som beskrevet i seksjon 1 om bakgrunnen for dette arbeidet. NVO-prosjektet var tilpasset vietnamesernes behov i Norge på 1980-tallet. Siden den gang har norsk språk endret seg, særlig når det gjelder oppkomsten av nye ord. Dette har innvirket på ordtilfanget i den enspråklige ordboken for fremmedspråklige.

Det har vært viktig å unngå vanskelige ord og vendinger i synonymangivelse og i definisjonsutforminger. Standard uttalemarkering er viktig, og det er lagt stor vekt på å finne frem til uttalemarkeringer som er gangbare over hele Norge, og som unngår uttalevarianter som er vanskelig å kode på en enkel måte.

Litteratur

Ordbøker

- Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén (red.) (1997): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Khánh, Nguyễn Quốc & M.M. Trần Ly San (1983): *Norsk-Vietnamesisk Ordbok*. Elektronisk utgave 2008: <www.uib.no/lle/20667/norsk-vietnamesisk-ordbok-p%C3%A5-nett>
- Kjelsvik, Bjørghild, Terje Svardal & Christian-Emil Ore: *Bokmålsordboka*. Nettutgave 2016. <ordbok.uib.no>
- Norwegian Newspaper Corpus Bokmål*. Created by Norsk aviskorpus. Distributed by the CLARINO UiB Portal: hdl:11495/D9B5-0349-4330-0 <clarino.uib.no/korpuskel/clarino-metadata?identifier=avis-plain>
- Vanvik, Arne (1985): *Norsk uttaleordbok*. Oslo: Fonetisk institutt, Universitetet i Oslo.

Annen litteratur

- Andersen, Øivin (2007): Deverbal Nouns, Lexicalization and syntactic Change. I: *Nordic Journal of Linguistics*. 30.1, s. 55–86.
- Jiang, Guiying & Qiaoyon Chen (2015): A Micro Exploration into Learners's Dictionaries: A Prototype Theoretical Perspective. I: *International Journal of Lexicography*. Vol. 30. No. 1. s. 108–139. Oxford: Oxford University Press.
- McEnery, Tony & Andrew Wilson (2001): *Corpus Linguistics*. 2.utgave. Edinburgh University Press.
- Öhman, Suzanne (1953): Theories of the "linguistic field". I: *Word* 9.2, s. 123–134.

Øivin Andersen
 professor, allmenn lingvistikk
 Institutt for lingvistiske, litterære og estetiske studier
 Universitetet i Bergen
 oivin.andersen@lle.uib.no

Eit dilemma om lemma

Sturla Berg-Olsen & Daniel Gusfre Ims

This article discusses different criteria for classifying related word forms either as separate lemmas or as variant forms with equal status in one dictionary entry. The question of which of these two solutions should be chosen is especially relevant for Norwegian, as the official orthography of the two Norwegian written standards in a large number of cases defines two or more spellings as equally correct forms of the same word. We examine three principles put forward in 2011 by the committee responsible for a revised orthography of Nynorsk, showing that these principles, if followed by the letter, could lead to the establishment of a large number of separate, but otherwise parallel, entries. To address this issue a revised principle was passed in 2019 by the advisory board for orthography under the auspices of the Language Council of Norway.

1. Lemma og variantformer

Nordisk leksikografisk ordbok definerer *lemma* som «uttrykkssiden hos det språklige tegn som en ordboksartikkel innledes med» (Bergenholtz et al. 1997:172). Ein ser altså for seg at kvar artikkel omhandlar eit (abstrakt) språkleg teikn, og lemmaet er ei form som har uttrykksside og som blir brukt som oppslagsform i artikkelen. Noko denne definisjonen ikkje nemner, men som dei fleste vil vere kjende med, er at ein artikkel stundom kan innleiast med meir enn éi form. Døme på dette vil bli gjevne nedanfor.

I alle ordbokprosjekt må ein gjere ein del prinsipielle val når det gjeld lemma-status. Gjeve at ein har eit sett ord eller ordformer som deler visse trekk, må ein ta stilling til om dei skal reknast som ovringar av eitt og same språklege teikn eller som ovringar av separate språklege teikn. Det er mogleg å setje opp ei rekke ulike kriterium for lemmastatus; Svensén (2009:96–101) nemner fleire slike kriterium, mellom anna desse tre (her attgjevne med vår eigen ordlyd):

- **Semantikk:** Ulike former som har same tyding(ar) eller klart nærskylde tydingar, blir rekna som variantar av same lemma.
- **Etymologi:** Ulike former som har sams historisk opphav, blir rekna som variantar av same lemma.
- **Grammatikk:** Ulike former som har same tyding(ar) og sams historisk opphav, blir rekna som variantar av same lemma om dei òg har same grammatiske eigenskapar.

I dei fleste tilfelle fell desse kriteria saman, og det byd på få problem å identifisere kva som skal reknast for variantar av eit lemma og dermed behandlast i same ordbokartikkel. Eit eksempel er formene *kviss* og *quiz* i norsk (nynorsk og bokmål): Dette er to ulike skrivemåtar av det som er eitt og same ordet etymologisk sett. Det knyter seg ikkje ulike tydingar til dei to formene, og dei har eksakt same grammatiske eigenskapar. I norske rettskrivingsordbøker finn ein konsekvent *kviss* og *quiz* i same artikkel, som valfrie og jamstilte variantformer. Eit lemma av denne typen, med to variantformer, kan òg kallast eit *di-lemma*. Slike di-lemma finn ein i ordbøker for dei aller fleste språk, men dei er nok spesielt hyppige i ordbøker over bokmål og nynorsk, der ein òg kan finne artiklar med mange fleire enn to jamstilte variantformer.

Eit anna og mindre klart eksempel er formene *heim* og *hjem* i bokmål. Vi kan vurdere dette paret ut frå dei tre nemnde kriteria for lemmastatus henta frå Svensén:

Semantikken til dei to formene er i hovudsak overlappande; *heim* tyder ikkje noko anna enn *hjem* om ein analyserer korleis formene blir brukte i bokmål. Ein del språkbrukarar vil nok knyte litt ulike stilvalørar til dei to formene, og det er truleg at somme vil kunne bruke begge orda etter visse kriterium som dels knyter seg til tyding. Om vi tillèt oss å sjå litt stort på det, kan vi likevel rekne *heim* og *hjem* som synonyme.

Etymologisk sett kan ein diskutere i kva grad dei to formene har same opphav. Nøkkelspørsmålet er korleis ein bør definere «same opphav» i saker som dette. *Heim* er nedervt frå norrønt *heimr*, medan *hjem* har opphav i dansk, som jo er del av grunnlaget for dagens bokmål. Slik sett kan ein seie at formene har opphav i ulike språk. Går ein lenger tilbake i tid, er det klart at norrønt/gammalvestnordisk og dansk har eit sams opphavsspråk. Ein kan likevel med rette spørje seg om det er rimeleg å rekne eit såpass fjernt fellesstadium som «sams opphav».

Svenséns *grammatiske* kriterium talar klart imot å rekne *heim* og *hjem* som ovringar av eitt og same språklege teikn – dei har ulike genus og ulik bøyning, og såleis ulike grammatiske eigenskapar.

Kvar ordbok og ordliste som tar mål av seg til å spegle den offisielle rettskrivingsnorma for bokmål, vil trass i det som er nemnt ovanfor, føre opp *heim* og *hjem* som variantformer i same artikkel. Den direkte årsaka til dette er å finne i det vi vil leggje til som eit sjølvstendig kriterium:

- **Normstatus:** Ulike former som rettskrivingsnorma definerer som valfrie, blir rekna som variantar av same lemma.

Det er lang tradisjon for å føre slike valfrie former som variantformer, og i praksis ser det ut til at dette fjerde kriteriet veg tyngre enn dei tre kriteria til Svensén – eller i alle fall at det kan vege opp for at eit eller to av dei første tre kriteria ikkje er oppfylte. Hovudmotivasjonen for at ein fører *heim* og *hjem* (og *melk* og *mjølk*, *hval* og *kval*, *selv* og *sjøl* osv.) som variantar, er eit ønske om å vise fram

den valfridommen som finst i norma. Det gjeld både bokmål og nynorsk, sjølv om eksempla her alle er frå bokmål. Det er elles verdt å merke seg at former som i norma er definerte som valfrie, per definisjon må vere med i offisielt godkjende ordbøker. Ei bokmålsordbok som har artikkel på *hjem*, men ikkje nemner at *heim* er eit fullverdig alternativ til *hjem*, vil ikkje kunne få godkjenning til bruk i skolen. Om dei to formene hadde vore definerte som separate lemma, hadde det vore ei anna sak, for utvalet av lemma i ei ordbok er ikkje regulert av normeringsmyndigheitene. Lemmalista er det leksikografane sjølve som set opp etter dei kriteria dei finn føremålstenlege.

2. Prinsipp om to oppslagsord i arbeidet med 2012-rettskrivinga for nynorsk

Den siste store rettskrivingsreforma for nynorsk blei vedteken i 2012 (den siste store for bokmål var i 2005). Innstillinga frå rettskrivingsnemnda blei lagd fram i 2011 i dokumentet *Ny rettskriving for 2000-talet. Innstilling til ny rettskriving for det nynorske skriftmålet* (Riise et al. 2011). Her tok nemnda til orde for å dele fleire oppslagsord med to eller fleire variantformer i ulike artiklar. Nemnda la visse prinsipp til grunn for denne oppdelinga:

Orda bør oppfattast som separate ord og ikkje som formvariantar når:

- dei har (til dels) ulikt semantisk innhald (døme: lut: pengeinnskot; aksje; del / lott: del av fortene)
- dei har opphav i ulike språk (døme: bok f. (om tre, frå norrønt) / bok m. (frå dansk))
- dei har ulik bøyning (døme: drynja dryn, drunde [drunte], drunt [drund] / drønna drønner, drønte, drønt)

[...] Når nemnda har vurdert om ord skal først opp som ordvariantar eller eigne oppslagsord, har ho teke omsyn til skriftbilete, tyding og bøyingsmønster i vår tid. Nemnda har ikkje primært sett på etymologi.¹ (Riise et al. 2011:42)

Framlegga frå rettskrivingsnemnda blei vedtekne i 2012, og med det blei det òg vedteke for 58 ord at dei skulle først som to oppslagsord. Sjølv om kriteria nemnda la til grunn i grunngjevinga si, ikkje høyrer med til sjølve rettskrivingsvedtaka, vil vi i det følgjande drøfte kva følgjer oppdelingane fekk for *Nynorskordboka*, og kva følgjer det vil kunne få om kriteria blir følgde i arbeidet med både *Nynorskordboka* og *Bokmålsordboka* framover.

¹ I eit tidlegare utkast (februar 2011) var òg dette med: «Her skal helst fleire av kriteria verka samstundes, men eit minimumskrav er at to av dei er dekte».

2.1. Semantikkriteriet (ulik tyding)

Det å føre ord med ulike tydingar som separate oppslagsord er langt på veg sjølv-sagt. Det har vore lagt til grunn som eit ibuande prinsipp i rettskrivingsordbøkene tidlegare òg, slik at ord av same opphav skil lag som ulike oppslagsord når det lèt seg gjere å identifisere éi eller fleire bitydingar der berre éi av formene er, eller kan tenkjast å vere, i bruk.

I utgangspunktet må det vere ei sak for leksikografien og ordbokredaksjonen å avgjere når det er grunnlag for ulike oppslagsord. Men det finst gråsoner, og særleg for bokmål ville det kunne få tydelege normeringskonsekvensar om ein legg stor vekt på semantikkriteriet i føringa av det som nå er jamstilte former (jamfør drøftinga av *heim/hjem* over).

2.2. Etymologikriteriet (opphav i ulike språk)

Sjølv om etymologikriteriet er eit av hovudkriteria hos Svensén (2009), har dei norske rettskrivingsordbøkene før 2012 ikkje lagt dette til grunn for å skilje mellom former av norsk og dansk opphav. I rettskrivingsinnstillinga er det i mange tilfelle ikkje gjort greie for kva kriterium som er lagde til grunn for kvart lemma som skal splittast i to oppslagsord, og det er heller ikkje heilt klart om høvesvis dansk og norrønt opphav er rekna som ulikt opphav. For *bok/bók* står norrønt og dansk som ulikt opphav, mens det om *drynje/drønne* står følgjande: «Dei har same opphavlege tyding og *same etymologiske opphav*, ved at det eine kjem frå norrønt og det andre truleg frå dansk.» (Riise et al. 2011:42, vår utheving.)

Vi har to hovudinnvendingar mot å rekne dansk eller norrønt opphav som eit tilstrekkeleg kriterium for å splitte lemma med variantformer. For det første vil ein i mange tilfelle få artiklar som er heilt identiske i ordforklaringane og bruksdøma. Slik ser artikkelen for *bok* (treet) ut i *Nynorskordboka* i dag:

II **bok** f. (norrønt *bók*; same opphav som *I bok*)

1 (tre av) slekt i bøkefamilien; *bøk*; *Fagus*

2 ved, vyrke av *bøk* (1)

Bøk har separat artikkel, men han er heilt parallell med artikkelen for *bok*:

bøk m1 (norrønt *bók*; same opphav som *I bok*)

1 (tre av) slekt i bøkefamilien; *bok* (II,1); *Fagus*

2 ved, vyrke av *bøk* (1)

Problemet med dette i eit rettskrivings- og brukarperspektiv er at det ikkje kjem tydeleg fram at det er heilt valfritt i rettskrivinga å bruke den eine eller den andre forma (jamfør *kviss/quiz*, som er ført som eitt oppslagsord).

For det andre vil det i særskilt mange tilfelle ikkje vere lett å slå fast om ei form i dagens norsk har opphav i norrønt eller dansk. For det nemnde verbparet *drynje/drønne* er det enkelt å slå fast at *drynje* har opphav i norrønt, og at *drønne* har eit dansk motsvar *drøne*. Samtidig er *drønne* ei heilt endefram lydutvikling i mange norske dialektar. Dette ser vi tydelegare om vi jamfører med andre liknande verbpar som *grysje/grøsse* og *gryppe/grøppe*. Til det første ordparet kan vi finne ein dansk parallell i det danske *grysse*, som er rekna som dialektalt i *Ordbog over det danske Sprog*. Til det andre finn vi eigentleg ikkje nokon dansk parallell i det heile. Det er følgjeleg rimeleg å meine at fleire av desse parallellformene i nynorsk helst må reknast å ha opphav i ulike variantar av norsk utvikla frå norrønt.

2.3. Grammatikkriteriet (ulik bøyning)

Ulik bøyning ser ut til å ha vore det kriteriet som oftast låg til grunn i dei tilfella der rettskrivingsnemnda i 2011 fremja å splitte oppslagsord. Det gjeld til dømes fleire verbpar som *drynje/drønne*, der *drynje* har kortform i presens og vokalveksling frå presens til preteritum og perfektum (*drynja/drynje – dryn – drunde – har drunt*), mens *drønne* er eit regelrett svakt verb som kan bøyast både som a-verb (*drønna/drønne – drønna – har drønna*) og e-verb (*drønna/drønne – drønner – drønte – har drønt*).

I eit rettskrivings- og brukarperspektiv meiner vi det òg er problematisk å splitte mange variantformer i ulike oppslagsord etter grammatikkriteriet aleine. Igjen ser vi at vi vil måtte få mange identiske artiklar. Slik ser dei to artiklane for *grysje* og *grøsse* ut i *Nynorskordboka* i dag:

grysje verb (samanheng med *II grue* og *II grus*)

- brått skjelve eller kulse i kroppen
grysje av uhyggje, skrekk / grysje ved tanken
- upersonleg:
det grys i meg

grøsse verb, v1, v3 (samanheng med *II grue* og *II grus*)

- brått skjelve eller kulse i kroppen
grøsse av uhyggje, skrekk / grøsse ved tanken
- upersonleg:
det grøsser i meg

I tillegg til at det ikkje kjem fram for ordbokbrukarane at det er valfritt å skrive *grysje* eller *grøsse* (med tilhøyrande bøyingsmønster), vil det å rekne slike verbpar som ulike lemma òg gjere det til eit reint redaksjonelt val kva ord som blir med i ei ordbok eller i ei ordliste. Ordlistar og ordbøker som viser den offisielle rettskrivinga, skal ha med alle valfrie rettskrivingsvariantar av eit oppslagsord.

Men det er opp til den enkelte redaksjonen kor mange og kva ord som skal vere med. Med andre ord vil ein utstrekt bruk av grammatikkriteriet potensielt kunne føre til at mange former som er vedtekne som gjeldande rettskriving, ikkje blir viste fram.

3. 2012-kriteria praktiserte for bokmål

Sjølv om 2012-reforma berre galdt nynorsk, meiner vi dei same kriteria for kva ein reknar som variantformer, bør liggje til grunn for bokmål og nynorsk. Ikkje minst gjeld dette for *Bokmålsordboka* og *Nynorskordboka*, som vender seg til den same norske allmenta, og som av brukarane blir oppfatta som tvillingordbøker eller som éin ressurs med to delar (Rauset 2019). Om vi hadde lagt etymologikriteriet, slik det var formulert i innstillinga frå 2011, til grunn for oppføring av oppslagsord i *Bokmålsordboka*, ville ei rekkje former som i dag er variantformer, blitt omgjorde til ulike oppslagsord. Særs mange ord i bokmål ovrar seg i variantpar, der den eine forma har opphav i dansk og den andre i norsk folkemål (frå norrønt). Nokre døme er *brott/brudd*, *bu/bod*, *fjøl/fjel*, *fram/frem* og *grein/gren*. Desse formene har eit sams etymologisk opphav (lenger tilbake), men likevel opphav i det vi reknar som ulike språk.

Grammatikkriteriet, ulik bøyning, ville òg vere relevant for særs mange ord i bokmål, dersom ein praktiserte det i vid tyding. I ytste konsekvens kunne ein til og med argumentere for å splitte opp par på bakgrunn av formvariasjon som *hoppa/hoppet*, *bygde/bygget*, *boka/boken* og *dyra/dyrene* (men i praksis er det vel få eller ingen som vil gå inn for dette).

Følgjeleg vil anten etymologikriteriet eller grammatikkriteriet kunne gjelde dei aller fleste oppslagsorda som har variantformer i *Bokmålsordboka*. I ein stor del av tilfella er det tale om både ulikt opphav og ulik bøyning.

Innvendingane våre i 2.2 og 2.3 mot ein praksis som fører til ei lang rekkje heilt parallelle artiklar, usynleggjering av valfridomen i rettskrivinga og i verste fall utelating av variantformer (som blir rekna som lemma), blir etter vårt syn vel så relevante for bokmål og bokmålsordbøker som for nynorsk og nynorskordbøker.

4. Reviderte prinsipp

Som vist over er det fleire uheldige sider ved dei prinsippa som rettskrivingsnemnda i 2011 sette opp og brukte som rettesnor for splitting av ei rekkje ordartiklar i nynorske ordbøker og ordlister. Av spørsmål som har kome inn til Språkrådets svarteneste, ser vi òg at denne føringsmåten i nokre tilfelle bryt med folks oppfatning av at det er snakk om to variantar av same ord. Ein spørjar skriv til dømes: «Kvifor står ikkje ”å venne” og ”å venje” saman som variantar av same ordet i

ordboka? Det verkar lite logisk på meg.» Eit anna aspekt ved prinsippa i Riise et al. (2011) er at dei stundom blir oppfatta som vedtekne reglar. Formelt sett er det ikkje rett, for da styret i Språkrådet og seinare Kulturdepartementet gjorde vedtak om den nye rettskrivinga for nynorsk i 2012, var det berre endringsframlegga i innstillinga dei vedtok, ikkje grunngevingane for desse framlegga. I praksis blir dei likevel gjerne oppfatta som bindande reglar.

På denne bakgrunnen og fordi spørsmålet om val av føringsmåte var høgaktuelt i samband med eit nyleg påbyrja prosjekt for å revidere *Bokmålsordboka* og *Nynorskordboka*, kom det i mars 2019 opp ei sak i Språkrådets fagråd for normering og språkobservasjon om reviderte prinsipp for skiljet mellom lemma og variantformer. Fagrådet vurderte det slik at rettskrivingsnemnda i 2011 hadde gode grunnar for å føreslå splitting av ein del ordbokartiklar, men at nemnda gjekk noko langt i formuleringa av etymologikriteriet og grammatikkriteriet (attgjevne i del 2 over). Om ein følgjer prinsippa i Riise et al. (2011) konsekvent, kan ein argumentere for å splitte svært mange artiklar. Det gjer at ein i mindre grad får vist fram valmoglegheitene som rettskrivinga gjev. Splitting fører òg til at utvalet av former i ordbøker og ordlister blir overlate til leksikografane, og konsekvensen kan bli at former som er meinte å vere rettskrivingsvariantar, ryk ut av ordbøkene utan at det er gjort noko vedtak om det av normeringsmyndigheitene. Som nemnt ovanfor vil det òg ha store potensielle konsekvensar om prinsippa frå 2011 skulle leggjast til grunn i bokmål (sjølv om rettskrivingsnemnda rimeleg nok ikkje hadde dette i tankane da ho formulerte prinsippa). Fagrådet gjorde dette vedtaket i saka:

Jamstilte former med hovudsakleg same tyding og same etymologiske opphav (eller nærskyldt opphav) bør i Bokmålsordboka og Nynorskordboka reknast som variantformer av same oppslagsord, i samsvar med rettskrivingstradisjonen for bokmål og nynorsk. (Språkrådet 2019:6)

Vedtaket innebar ei delvis reversering av dei prinsippa som blei formulerte i 2011, men det blei presisert at reversering av sjølve vedtaka om splitting av visse artiklar må vurderast frå tilfelle til tilfelle, og mange av dei splitta artiklane vil etter alt å døme bli ståande. Mest aktuelt vil det vere å gå tilbake til føring i éin artikkel for verbpar som *venne/venje*, *drynje/drønne* og *gryse/grøsse*.

Rettskrivingsnemnda av 2011 hadde sjølv sagt rett i observasjonen om at «[f]ramtidige ordlister og ordbøker toler mykje større ordtilfang og fleire ordartiklar enn dagens ordlister, fordi dei i stadig større grad blir digitale» (Riise et al. 2011:42). Å splitte tidlegare di-lemma er i mange tilfelle på sin plass, særleg der det faktisk er skilnader i tyding eller stilistikk som ein best kan yte rettferd ved å behandle variantformene kvar for seg. Men å føre rettskrivingsvariantar separat i elles heilt parallelle artiklar har lite for seg sjølv i ei elektronisk ordbok der plassomsyn ikkje spelar inn.

Litteratur

Ordbøker

Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén: *Nordisk leksikografisk ordbok* (1997). Oslo: Universitetsforlaget.

Bokmålsordboka. Språkrådet og Universitetet i Bergen. <ordbok.uib.no> (november 2019).

Nynorskordboka. Språkrådet og Universitetet i Bergen. <ordbok.uib.no> (november 2019).

Ordbog over det danske Sprog. Det Danske Sprog- og Litteraturselskab. <www.ordnet.dk/ods/> (november 2019).

Annan litteratur

Rauset, Margunn (2019): *Bokmålsordboka og Nynorskordboka – einægga, toegga eller siamesiske tvillingar? I: LexicoNordica* 26, 155–175.

Riise, Grete et al. (2011): *Ny rettskriving for 2000-talet. Innstilling til ny rettskriving for det nynorske skriftmålet*. Oslo: Rettskrivingsnemnda for nynorsk og Språkrådet. <www.sprakradet.no/upload/Rettskrivingsnemnda/Innstilling%20nynorsk.pdf> (november 2019).

Språkrådet (2019): *Referat frå møte i fagrådet for normering og språkoobservasjon 6. mars 2019*. Upublisert.

Svensén, Bo (2009): *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

Sturla Berg-Olsen
seniorrådgjevar, dr.art.
Språkrådet
Postboks 1573 Vika
NO-0118 Oslo
sturla.berg.olsen@sprakradet.no

Daniel Gusfre Ims
seksjonssjef, cand.philol.
Språkrådet
Postboks 1573 Vika
NO-0118 Oslo
daniel.ims@sprakradet.no

Online Data on Icelandic Inflection. Descriptive or Prescriptive: “Why, for whom, by whom” and how?

Kristín Bjarnadóttir & Kristín Ingibjörg Hlynsdóttir

The topic of this paper is the production of a prescriptive subset of the Database of Modern Icelandic Inflection, which was originally designed for use in language technology and is descriptive in nature. The data has been available to the general public online from 2004, and users have expressed the need for prescriptive data. Such data is also needed in some language technology projects, such as spell checking and language production of any kind. The method of classification used to produce the prescriptive subset, the DMII Core, is described.

1. Introduction

The Database of Modern Icelandic Inflection (DMII) contains paradigms of Icelandic words. The DMII was originally produced for use in language technology (LT) and lexicography, but the data was also to be made available to the general public online. Work on the DMII started in 2002 and the first version of the website was opened in 2004. The original DMII was descriptive in nature (cf. Bjarnadóttir 2012, Bjarnadóttir et al. 2019).¹

In 2017, a grant from the Icelandic Language Technology Fund (Máltæknisjóður) made restructuring of the original DMII database possible. A part of the restructuring is classifying the data according to acceptability, with reference to Icelandic language standards. The result of this work is an enhancement of the downloadable LT data, and the DMII Core, which is a prescriptive subset of the DMII intended for learners of Icelandic. The topic of this paper is the creation of the DMII Core, specifically the classification system used to mark the prescriptive data.

The paper is structured as follows: Section 2 contains a description of the DMII and the need for prescriptive data. Section 3 contains a description of variants in Icelandic morphology. Section 4 briefly describes Icelandic standards of language use. Section 5 contains a description of the DMII Core. Section 6 describes the classification system. The conclusion is in Section 7.

¹ The work described here was done at The Árni Magnússon Institute of Icelandic Studies by the authors of this paper.

2. DMII and the Need for Prescriptive Data

The DMII data is accessible online at the website of The Árni Magnússon Institute for Icelandic studies², and it is also downloadable for use in LT. At the time of writing, DMII contains over 292,000 headwords and over 6.2 million inflectional forms. The online DMII is extensively used by the Icelandic public as a reference on inflection, and it has been popular from the start in 2004. In the year starting on June 1 2018, there were 201,207 users and 1,714,726 pageviews, and 9% growth from the previous year.

From the start, the DMII was to be a description of Icelandic inflection “as is”, i.e., without regard to correct spelling, grammar, vocabulary, etc. For LT analysis, the key factor is inclusive data, showing all varieties, ranging from context-bound obsolete word forms to non-standard inflectional variants and non-standard compound formations and words. As clearly stated on the website, the DMII was from the beginning meant to be descriptive, and it was neither to be an exhaustive source of Icelandic vocabulary nor an authority on the acceptability of vocabulary, inflectional variants or spelling. In spite of rather forcefully phrased instructions to the contrary, online users still expect online paradigms published by an accredited authority on the Icelandic language, such as the Árni Magnússon Institute for Icelandic Studies, to be “correct”, i.e., prescriptive. The expectation is that all word forms appearing on the website are “best usage”, all spelling variants are “good”, and all words are acceptable. Some users even expect the DMII to be exhaustive, assuming that the absence of a word indicates that it is not an acceptable Icelandic word.

To assist the users of the original website, the paradigms include usage notes. These explain the choice between inflectional variants, such as constrictions of the use of individual inflectional forms. The notes are in Icelandic only, and they make the website unsuited for any but native or near-native speakers of Icelandic, as the users have to make the final choice between variants themselves. Fig. 1 shows a paradigm with notes.

The problems learners of the language face when using the online DMII are not trivial. The data is both too complex, as when addressing instances of obscure usage, and not instructive enough, in not presenting simple information on acceptability. The paradigm for *hönd* is a case in point, as the inflectional variants are acceptable to different degrees.

² www.arnastofnun.is/en

hönd kvenkynsnafnorð

Athugið: Upprunalega beygingin er **hönd** í nefnifalli og þolfalli eintölu og **hendi** í þágufalli. Dæmi: *hafa e-ð við höndina, hafa e-ð í hendi.*
 Í merkingunni 'ríthönd' er þágufallsmyndin **hönd** oft notuð: *Bréfið var skrifað með hönd Jóns Jónssonar.* Í sömu merkingu bregður fleirtölunni **handir** fyrir: *Þá tekur við önnur hönd eða aðrar handir.*
 Orðmyndin **hendi** í nefnifalli og þolfalli eintölu er algeng í talmáli, sérstaklega í íþróttamáli í merkingunni 'snerting við bolta', en er ekki talin æskileg.
 Fleirtölumyndin **höndur** er gömul í málinu en er nú sjaldséð.

Eintala		Fleirtala	
	án greinis	með greini	
Nf.	hönd / hendi	höndin / hendin	Nf. hendur / höndur / handir
Pf.	hönd / hendi	höndina / hendina	hendurnar / höndurnar / handirnar
Þgf.	hendi / hönd	hendinni / höndinni	Þgf. hendur / höndur / handir
Ef.	handar	handarinnar	hendurnar / höndurnar / handirnar
			Þgf. höndum
			Ef. handanna

Figure 1: The paradigm for *hönd* ‘hand’ on the DMII Website.³

3. The DMII and Icelandic Morphology

The morphology of Icelandic is rich, both in inflection and word formation, and variant inflectional forms are possible for very many inflectional categories. The number of inflectional variants in each pattern varies, from one set as in the noun *bátur* ‘boat’ where the dative singular is *báti/bát*, to almost a full set of duplicate inflectional forms in the adjective *þögull* ‘silent’, e.g., *þögulan/þöglan* (masc.acc.sg.indef.), *þögulum/þöglum* (dat.sg.indef.), etc. There are 56,143 sets of inflectional variants in the database, most of which contain two variants, as in *báti/bát*, with three variants found in only 129 sets, as is the case for *hönd*,

³ Translation of the note (‘Athugið’): The original inflection is nom. & acc.sg. **hönd**, dat. sg. **hendi**, e.g. *hafa eitthvað við höndina* ‘have sth. at hand’, *hafa e-ð í hendi* ‘have sth. in hand’ (i.e., ‘be sure of sth.’). In the the sense ‘handwriting’ the dat.sg. **hönd** is common: *Bréfið var skrifað með hönd Jóns Jónssonar* ‘The letter was written in JJ’s handwriting. In the same sense, the plural **handir** is also found, e.g. *Þá tekur við önnur hönd eða aðrar handir* ‘Then we have another scribe or scribes.’ The variant **hendi** in nom. & acc. sg. is common in colloquial language, especially in sports in the sense ‘a touch of a hand with a ball’. This is not considered good usage. The plural **höndur** is old-fashioned and now very rare.

e.g., *hendur/höndur/handir*. Two variants of the same inflectional category can be equally valid in (almost) any context, as in *báti/bát*, but the choice between variants can also be context-sensitive or semantically restricted, as in the dative singular *hönd* in the meaning ‘handwriting’, where the unmarked dative is *hendi*, as shown in Fig. 1.

4. Icelandic Standards for Language

Icelandic language standards appear in the “Rules of spelling and punctuation” (*Ritreglur* (2016) & *Reglur um greinamerkjasetningu* (2018)), and in “The Dictionary of Spelling” (*Stafsetningarorðabókin*) (Sigtryggsson 2016). The work on the prescriptive version of the DMII starts with these standard references.

There is no comprehensive standard for Icelandic inflection. Icelandic dictionaries show only a small subset of inflectional forms, i.e., nom.sg. & pl., and gen.sg. (3 forms out of 16), and for verbs the principal parts are shown (4-5 out of over a hundred). The remainder of the paradigms are not always deductible from the available forms. Grammar books also contain fragmentary descriptions, as the inflectional system is described “top-down”, with few examples. The result is that producing the paradigms for the DMII involved considerable research where the standards were of no help.

5. The DMII Core

The DMII Core is a subset from the DMII, designed to meet users’ demand for data from the DMII in a prescriptive context, especially for use in schools and for learners of the language. It is created for third party publication, and it is accessible through a RESTful API,⁴ open for everyone. The API allows users to send simple queries and receive full paradigms in JSON-format as a response. In order to make the transition from descriptive to prescriptive, the core vocabulary of the DMII has been checked against the standards of good usage mentioned above.

The DMII Core contains a subset of the vocabulary of current Icelandic, i.e., common non-domain specific words, and a selection of named Icelandic entities, i.e., personal names, common place names, etc. The spelling is standardized. The sources of the vocabulary are *Íslensk nútímamálsorðabók* [*The Modern Icelandic Dictionary*] (Jónsdóttir & Úlfarsdóttir 2018), containing approximately 50,000 headwords, with additions from the 50,000 most frequent lemmas in The Icelandic

⁴ API: application program interface. REST: Representational State Transfer, an architectural pattern for creating web services. A RESTful service implements that pattern.

Gigaword Corpus (Steingrímsson et al., 2018). The overlap between these two sources is great, and the total number of paradigms in the DMII Core now stands at 58,086, out of 293,783 in the DMII.

The paradigms in the DMII Core have been simplified as much as possible, without omitting equally valid variants, showing only the best forms, or the variants not limited by specific usage restrictions. The classification system created for this use is described in Section 6.

6. The Classification System

The classification system used to create the DMII Core extends to both the vocabulary and individual inflectional forms. It was created by analysing all the handcrafted notes on individual paradigms in the DMII as they stood in March 2018 when work on a new version of the DMII was started. The results were then formalized, both with regard to the Icelandic standard (as for spelling, word formation variants, and inflectional variants), usage considerations, the grammatical features involved, and considerations of style and/or register.

The classification system is in three parts, two of which apply to both words and variants, i.e., genre/register (Section 6.1) and correctness grade (Section 6.3). Variants are also graded according to their relative value (Section 6.2). The parts of the classification system are interdependent, especially the correctness grade which is largely estimated on the basis of the other three.

6.1. Genre or Register

Genre or register is used to sort words and inflectional forms according to style, age and other comparable features. The classification functions as a tool for finding the relevant vocabulary for inclusion in the Core, omitting marginal data. The division in genres/register is as follows:

- Included in the DMII Core: derogatory, obscene, informal (and unmarked words and inflectional variants).
- Not included in the DMII Core: rare, old-fashioned, obsolete; poetic language, regional; unadapted loan words.

6.2. Value of Inflectional Variants

As stated above, there are 56,143 sets of inflectional variants in the DMII. Most of the sets contain two variants each, but 129 sets contain three variants. The variants have been evaluated and graded, to indicate their status with respect to

the other variant in each set. The classification also includes the marking of inflectional forms only found in fixed expressions or idioms. The value determines inclusion or exclusion in the Core, and they are as follows:

- Equal: There are no restrictions on the use of either variant in a set. Example: The two genitive singular forms of the masculine noun *bekkur* ‘bench’, *bekklar/bekks*, are equally valid, according to *The Dictionary of Spelling*. Both are included in the DMII Core.
- Dominant/yielding: One variant in a set is acceptable without restriction, the other one is not. Example: The masculine noun *refur* ‘fox’ has sets of two variants in the nominative and accusative plural: *refir/refar* (nom. indef.), *refirnir/refarnir* (nom. def.), and *refi/refa* (acc. indef.), *refina/refana* (acc. def.). The first variant in each set is dominant and thus included in The DMII Core. The other variants are yielding, as they are obsolete in Modern Icelandic and only used in idioms such as *til þess eru refarnir skornir* ‘that’s why the fox are cut’ (i.e., ‘that is why a certain action was taken’). These are omitted from the DMII Core.
- Obsolete inflectional forms, mostly found in idioms or fixed expressions. Example: *skjöldu* (acc. pl.) of the masculine noun *skjöldur* ‘shield’, used in a number of idioms. The corresponding modern inflectional variant is *sildi*. These are omitted from the DMII Core.

Frequency is used to determine the value of variants, when the standards are of no help and the distribution seems to be independent of context.

6.3. Correctness Grading of Words and Inflectional Variants

The Correctness grade is used in the work on the Core to mark a word’s or variant’s correctness according to prescriptive grammar rules and standardized spelling. Grades range from 0 to 4. The only grade included in the DMII Core is “1”. It should be noted that the equal value of variants is quite common and in that case both forms are assigned the value “1” and included in the DMII Core. The grading is as follows:

- 1) Not applicable, dependent on genre or register, cf. Section 5.1.
- 2) Correct, default value, most words and inflection forms.
- 3) Used, not as good. Mostly used of common words and inflectional forms not approved of in the standards.
- 4) Not good. Used of words and inflectional form directly in contradiction to the recommendations of the standards.
- 5) Very bad. Used of words and inflectional forms deemed errors in the standards.

In the production of the DMII Core, the Icelandic standards are used to determine the grade of a word or inflectional variant according to correctness, as far as possible. The goal is to find the items to mark with “1”. One problem is that the data in the standards is simply too scarce to answer all questions when it comes to individual words or inflectional forms. The *Rules of spelling and punctuation* are generalizations, with relatively few examples, and they are not always easy to interpret or extrapolate. The problem of scarcity also applies to *The Dictionary of Spelling*, as the vocabulary is small.

6.4. Use of the Classification System

The aim in the creation of the DMII Core is to produce a prescriptive, simplified version of the inflectional paradigms, containing only the most common (and correct) variants, as in the 2nd column in Table 1.

	Core	A	DMII	A	B	C
Nom.sg.	<i>hönd</i>	1	<i>hendi</i>	3	Informal	
Nom.sg.def.	<i>höndin</i>	1	<i>hendin</i>	3	Informal	
Acc.sg.	<i>hönd</i>	1	<i>hendi</i>	3	Informal	
Acc.sg.def.	<i>höndina</i>	1	<i>hendina</i>	3	Informal	
Dat.sg.	<i>hendi</i>	1	<i>hönd</i>	0		Meaning
Dat.sg.def.	<i>hendinni</i>	1	<i>höndinni</i>	0		Meaning
Gen.sg.	<i>handar</i>	1				
Gen.sg.def.	<i>handarinnar</i>	1				
Nom.pl.	<i>hendur</i>	1	<i>höndur</i>	0	Old-fash.	Yielding
			<i>handir</i>		Old-fash.	Yielding
...						

Table 1: Grading: A: Correctness grade, B: Genre/register, C: Value.

The difference between the DMII and the Core (cf. the paradigm for *hönd* in Section 2) is demonstrated in Table 1. The column headed ‘Core’ only contains acceptable variants, i.e., Correctness grade 1 in the 3rd column (A), unmarked for genre, register or value. The DMII contains the complete set of data, including Correctness grade 0 or 3 (A), and variants marked for genre, register or value (B & C). The omission of a variant does not necessarily imply that it is wrong, and for a full exposition of variants the users are referred to the DMII Website.

7. Conclusion

The classification system described above has been used to mark all sets of inflectional variants currently in the DMII database, and all headwords in the DMII Core. The simplified DMII Core is made for language learners and for use in teaching material. This does not in any way detract from the importance of access to the complete set of data on the DMII Website. In turn, work on the DMII Core has led to better usage notes on the website, with more information on prescriptive data found in the classification system, with extensive cross-referencing of less than optimal instances of usage to standardized forms.

The DMII Core API is open for everyone to use. It is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License (CC BY-ND 4.0).

Acknowledgements

The authors wish to thank the Icelandic Language Technology Fund for the grant which made the restructuring of the DMII possible. Thanks are also due to our DMII colleagues Samúel Þórisson, Steinþór Steingrímsson and Trausti Dagsson, and to the editors of the *Dictionary of Spelling*, Jóhannes B. Sigtryggsson, and the *Dictionary of Modern Icelandic*, Halldóra Jónsdóttir and Þórdís Úlfarsdóttir, for being very generous with their time.

References

- Bjarnadóttir, Kristín (2012): The Database of Modern Icelandic Inflection. In: *LREC 2012 Proceedings: Proceedings of "Language Technology for Normalization of Less-Resourced Languages"*, *SaLTMiL 8 -- AfLaT 2012*, pp. 13-18.
- Bjarnadóttir, Kristín, Kristín Ingibjörg Hlynsdóttir & Steinþór Steingrímsson (2019): DIM: The Database of Icelandic Morphology. In: *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pp. 146-154.
- DMII = Kristín Bjarnadóttir (ed.): *Database of Modern Icelandic Inflection [Beygingarlýsing íslensks nútímamáls]*. Reykjavík: The Árni Magnússon Institute for Icelandic Studies. <www.bin.arnastofnun.is>
- Jónsdóttir, Halldóra, & Þórdís Úlfarsdóttir (eds.) (2018): *Íslensk nútímamálsorðabók* Reykjavík: The Árni Magnússon Institute for Icelandic Studies. <islenskordabok.arnastofnun.is>
- Ritreglur* [Rules on Spelling] (2016). Published by The Ministry for Culture and Education, Reykjavík.

Reglur um greinarmerkjasetningu [Rules on Punctuation] (2018). Published by The Ministry for Culture and Education, Reykjavík.

Sigtryggsson, Jóhannes Bjarni (ed.) (2016): *Stafsetningarorðabókin* [The Dictionary of Spelling], 2nd ed. Reykjavík: The Árni Magnússon Institute for Icelandic Studies.

Steingrímsson, Steinþór, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson & Jón Guðnason (2018): Risamálheild: A Very Large Icelandic Text Corpus. In: *Proceedings of LREC 2018*, pp. 4361-4366.

Kristín Bjarnadóttir
associate Research Professor
The Árni Magnússon Institute
for Icelandic Studies
Reykjavík
kristinb@hi.is

Kristín Ingibjörg Hlynsdóttir
researcher
The Árni Magnússon Institute
for Icelandic Studies
Reykjavík
kih4@hi.is

Att hantera grammatisk förändring i en deskriptiv ordbok

Kristian Blensenius & Lena Rogström

This paper studies the representation of more or less novel constructional types in five Swedish dictionaries of different types and, to a certain extent, different ages. Based on a number of constructional types, the aim is to find out which of them have been entered in the dictionaries and how the types are treated from a normative and a descriptive point of view. The study also discusses the normative role of the contemporary dictionary in relation to the constructional types.

1. Normativa anspråk i ordböcker. En problematisering

Kategorierna *brukslexikografi* och *dokumentationslexikografi* diskuteras bl.a. av Bergenholtz et al. (1997:169), Svensén (2004:1f.) och Josephson (2018:249). Brukslexikografin ska tillvarata språkanvändarens behov av ordböcker i den dagliga kommunikationen, och dokumentationslexikografin tillgodoser det kollektiva behovet av att kunna dokumentera ett språk. Inom lexikografin kan dessa funktioner närmast liknas vid vad man benämner *normativ* respektive *deskriptiv* lexikografi, i betydelsen 'som gör resp. inte gör normativa anspråk' (se Nikula 1992). Åtskillnaden mellan normativ och deskriptiv lexikografi är dock inte knivskarp, genom att ordböcker förväntas erbjuda användaren både en rådgivande och en vägledande funktion samtidigt som de också ska vara beskrivande och dokumenterande.

Den ordboksanvändare som söker vägledning om stavning och böjning av ord har sedan lång tid tillbaka kunnat vända sig till den inofficiella normen på området, *Svenska Akademiens ordlista* (SAOL). Här anges också bruklighetsinskränkningar för såväl uppslagsord (ett ord som *mellis* är t.ex. "vardagligt" enligt SAOL 14) som böjningsformer (t.ex. den vardagliga presensformen *brer* till **breda**¹). Ordböcker går annars ofta en balansgång mellan mer eller mindre uttalad deskriptivitet och "förtäckt" normativitet (jfr Nikula 1992:49 och Josephson 2018:249) i fråga om bl.a. stavning och böjning. Man kan då också fråga sig hur de hanterar andra informationskategorier där det saknas en specifik hållning i moderna ordböcker, såsom för flerordsuttryck som *mena* (*på*). Användaren kan ju tänkas vilja ha vägledning i "det goda språkbruket" (Larsson 2004:211) på konstruktionsområdet likaväl som på andra områden. Till saken hör också att

¹ Uppslagsord markeras hädanefter med fetstil.

företrädare för den officiella språkvården rådfogar ordböcker vid besvarandet av språkfrågor, vilket kan tänkas tillskriva ordböcker en normativ funktion även om intentionen har varit mer deskriptiv.

Vårt syfte med denna undersökning är att diskutera fem exempel på flerordsuttryck som genomgår förändring i svenskan och undersöka hur dessa hanteras i några moderna svenska ordböcker. Ämnet har aktualiserats med anledning av att *Svensk ordbok utgiven av Svenska Akademien* (SO, 2009) står inför en ny revidering och olika innehållskategorier ses över. Vi undersöker flerordsuttryckens utbredning i korpusar (se avsnitt 2.1 nedan) och studerar dels vilka av dem som tas upp och hur de behandlas i ordböckerna, dels hur ordböckerna verkar förhålla sig till flerordsuttryck på ett mer generellt plan. Slutligen försöker vi ge förslag till åtgärder i revideringen av SO.

Vi kan förvänta oss att finna flerordsuttryck bland uppslagsorden (som flerordslemman), språkproven och bland flerordsuttryck som utgörs av formaliserade mönster i form av så kallade *konstruktionsuppgifter*. Ett exempel på en konstruktionsuppgift är *belöna ngn (med ngt) (för ngt/SATS)*, som ges i ett särskilt avsnitt i artikeln till uppslagsordet **belöna** i SO. Jämför liknande termer som *skjematiske fleirordskonstruksjonar* (Berg-Olsen 2013) och *fraseoskabeloner* (Farø & Lorentzen 2009). Kategorin är lexikografiskt intressant såtillvida att den berör både syntax och semantik och därmed hör hemma i såväl ordböcker som grammatikor. Konstruktionsuppgifter är inte särskilt ingående diskuterade i lexikografisk litteratur ur ett normeringsperspektiv, men se bl.a. Nikula (1992) och Trap-Jensen (2002a) om normativitet/deskriptivitet generellt, och Trap-Jensen (2002b) om normativitet/deskriptivitet i *Den Danske Ordbog* (DDO). Se även Vannebo (1998:365) om syntaktiska uppgifter i ordböcker.

Dispositionen av artikeln är denna: I avsnitt två beskrivs material och metod för undersökningen. I avsnitt tre redovisas undersökningen, och i fjärde avsnittet presenteras några slutsatser.

2. Material och metod

Undersökningen gäller som sagt fem flerordsuttryck som samtliga är stadda i någon form av förändring. De har alla det gemensamt att de består av verb som växlar i fråga om huruvida de står med eller utan formord. Tablå 1 nedan anger i den första spalten det ”vedertagna” flerordsuttrycket, dvs. det som brukar anses som den allmänt gängse och korrekta, och i andra spalten återfinns de nyare motsvarigheterna tillsammans med ungefärlig ålder. För uppgifter om ålder hos de nyare uttrycken har *Svenska Akademiens ordbok* (SAOB) och Kungliga bibliotekets dagstidningsarkiv² nyttjats.

² <www.tidningar.kb.se/> (läst 2019-10-20).

Vedertaget uttryck	Nyare uttryck	
<i>kommer att</i> + verb	<i>kommer</i> + verb	(åtm. sedan 1589)
<i>eftersom</i>	<i>eftersom att</i>	(åtm. sedan 1850)
<i>mena att</i>	<i>mena på att</i>	(åtm. sedan 1840)
<i>visa att</i>	<i>visa på att</i>	(åtm. sedan 1865)
<i>indikera/indicera att</i>	<i>indikera/indicera på att</i>	(åtm. sedan 1956)

Tablå 1: Undersökta flerordsuttryck.

Några av varianterna på flerordsuttryck är gamla och ser i vissa stilarter ut att hålla på att ersätta de äldre varianterna, medan andra är uppstickare i språket. Det äldsta ”nyare” flerordsuttrycket, med verbet *komma* i presens + infinitivverb utan *att* (*Barn kommer gå i skolan i minst nio år*), är gammalt men fortfarande inte helt vedertaget i ordböcker. Uttrycket *eftersom att* innehåller i stället ett tillagt *att*, här dock som subjunktion (*Hen åkte buss eftersom att tåget var för dyrt*). På används också allt flitigare och har länge funnits som tillägg till verben *mena* och *visa*. Ett senare uttryck, troligen bildat i analogi med de två äldre, är *indikera på att* (vi illustrerar fortsättningsvis med denna stavning) som blir frekventare i vissa typer av texter.

Vi undersöker dels hur flerordsuttrycken används i ett korpusunderlag (se 2.1 nedan), dels hur uttrycken behandlas i fem svenska ordböcker med något olika ålder och inriktning, utgående från SO och dess föregångare NEO. Samtliga har det gemensamt att de ger uppgifter om flerordsuttryck, även i form av konstruktionsuppgifter.

- *Nationalencyklopedins ordbok* (NEO, 1995–1996)
- *Svensk ordbok utgiven av Svenska Akademien* (SO, 2009)
- *Svensk handordbok. Konstruktioner och fraseologi* (SHO, 1966)
- *Svenskt språkbruk. Ordbok över konstruktioner och fraser* (SSB, 2003)
- *Lexins svenska lexikon* (Lexin, 2013)

NEO gavs i mitten av 1990-talet ut i anslutning till *Nationalencyklopedin* och är numera inkluderad i Nationalencyklopedins internetjänst NE.se. Ordboken inkluderas framför allt i egenskap av jämförelsematerial med SO, eftersom SO är en vidareutveckling ur samma lexikaliska databas som NEO bygger på (se Malmgren 2009 för en historik). SO och dess föregångare är primärt deskriptiva definitionsordböcker, även om SO innehåller normativa inslag främst i form av s.k. stilrutor (Sköldberg 2017:123).

Två konstruktionsordböcker är också inkluderade: en äldre, SHO, och en yngre, SSB. Dessa är intressanta såtillvida att de har som specifik uppgift att hantera just flerordsuttryck. Som en representant för inlärorordböcker, vilka inte sällan lägger vikt vid flerordsuttryck, fungerar Lexins digitala utgåva från 2013. SSB har liksom SO en deskriptiv intention (SSB s. 15), medan den äldre SHO har en klart normativ intention: man vill ”varna för felaktigt eller mindre lämpligt språkbruk” (SHO s. VII). Lexins inriktning på området är inte klargjord: i användaranvisningen på nätet³ anges bara att den är en inlärorordbok för den som lär sig svenska.

SO utgår alltså från den databas som i en tidigare version låg till grund för NEO. Detta verkar i viss mån även gälla SSB: i dess förord sägs t.ex. att ”Exemplen har hämtats från Internet och stora pressdatabaser som Presstext, men många har också sitt ursprung i konkordanserna hos Språkdata vid institutionen för svenska språket vid Göteborgs universitet” (SSB s. 24). Även Lexin har på olika sätt kopplingar till Göteborgs universitet sedan 1970-talet då ordboksunderlaget utarbetades vid dåvarande Språkdata (Malmgren 2012:455).

2.1. Korpusunderlaget i undersökningen

Vi har utnyttjat två olika korpusmaterial i korpusverktyget Korp (Borin et al. 2012) när vi excerperat våra belägg. Den första korpusen utgörs av konventionellt skriftspråkligt material som kan kallas traditionell sakprosa. Denna korpus består av akademiska texter (delkorpusen Akademiska texter: 25 milj. tokens) och tidningstexter (delkorpusarna GP, Press och Webbnyheter: ca 580 milj. tokens). Totalt utgör materialet alltså drygt 600 miljoner tokens. Den andra korpusen består av bloggar och olika diskussionsforum som Flashback, Familjeliv och Twitter (delkorpusen Sociala medier). Denna korpus är betydligt större: 10 miljarder tokens.

En typ av material som växt snabbt som underlag för lexikala korpusar under senare tid är bland annat bloggar och diskussionsforum av olika slag. Dessa har till skillnad från konventionella skrivna texter producerats direkt för webben och har ofta en mycket mer tillfällig, ibland dialogisk, karaktär och påminner mer om talat språk (jfr t.ex. Bockgård 2012:210). Vi kallar denna typ av skriftspråk för *oredigerad text* i undersökningen.

Oredigerad text håller på olika sätt på att inlemmas i arbetet med SO. I fråga om vilket textmaterial man bygger ordboken på, anges att ”etablerat, redigerat skriftspråk hämtat från tidningar och romaner utgör en betydande del” (Sköldbberg 2017:125). Det står dock klart att man gärna baserar kommande upplaga också på andra material: ”Det är svårt att hävda att inte texter publicerade på åt-

³ <lexin.nada.kth.se/lexin/doc/Lexin.pdf> (läst 2019-10-20).

minstone vissa sociala medier ska in i underlaget för den nya upplagan [av SO]” (Sköldberg & Hannesdóttir 2017:332). Eftersom det verkar som om det sker snabba förändringar av flerordsuttrycks grammatiska struktur (åtminstone strukturen i de uttryck vi valt att studera) i oredigerad text och därmed förknippade skrivna textgenrer, är dessa uttrycks behandling i bland annat SO en intressant fråga.

3. Undersökningen – flerordsuttryck i olika korpusar

I undersökningen av de utvalda flerordsuttrycken har vi först excerperat de båda delkorpusarna, och sedan har vi listat vilka av de två varianterna av respektive flerordsuttryck som är mest frekvent i respektive delkorpus. Resultaten redovisas i tabell 1. I redovisningen anges både absolut antal belägg och hur stor andel de ”nya” flerordsuttrycken utgör i relation till de vedertagna flerordsuttryckens antal i procent.

	Korpus 1 (Akad. + Tidn.)		Korpus 2 (Soc. medier)	
	Antal	Andel ”nya” uttryck	Antal	Andel ”nya” uttryck
kommer att + verb	391 945		2 459 293	
kommer + verb	35 574	9,01 %	4 183 298	170 %
eftersom	283 840		4 973 116	
eftersom att	558	0,19 %	203 408	4,0 %
mena.* att	78 475		551 985	
mena.* på att	132	0,16 %	32 970	5,9 %
visa.* att	119 074		536 850	
visa.* på att	1 260	1,05 %	33 943	6,3 %
indi.era.* att	2 523		13 553	
indi.era.* på att	16	0,6 %	2 238	16,5 %

Tabell 1: Förekomst av flerordsuttryck i två korpusar.

Som förväntat är de nyare flerordsuttrycken bättre företrädna i korpus 2, som innehåller sociala medier. Allra störst är skillnaden för *kommer att* + V, där användningen av *kommer* utan infinitivmärke med en andel av 170 % är mycket mer frekvent än den vedertagna varianten. *Eftersom att*, som är vanlig i talat språk och talspråkliga genrer, används nästan inte alls i korpus 1. På liknande sätt är det med *mena på att*, som har procentuellt sett fler belägg i korpus 2. *Visa*

på att är något vanligare än *mena på att* i korpus 1, och även denna konstruktion har fler belägg relativt sett i korpus 2. Den nyaste av konstruktionerna, *indikera på att*, har stor andel av beläggen i korpus 2, även om det absoluta antalet förekomster inte är lika stort som för de andra konstruktionerna.

Man kan konstatera att de båda varianterna av respektive flerordsuttryck förekommer i olika utsträckning i olika material och att de därmed är genreberoende. Får då detta genomslag i vårt ordboksmaterial? I tablå 2 nedan återges om, och i så fall hur, den nyare varianten av flerordsuttrycket är beskriven i ordböckerna. Varianterna *visa på* och *indikera på att* tas inte upp i någon av ordböckerna. *Kommer + V* tas upp i konstruktionsuppgiften i de tre nyare ordböckerna. SO väljer också att notera i en stilruta att uttrycket förekommer och att det numera måste anses acceptabelt. Stilrutan används igen för att påpeka att man inte ska skriva *eftersom att*, som SO berör som enda ordbok men avfärdar (därför anges ”Nej” i tablåen). Uttrycket *mena på* tas upp av tre av ordböckerna som partikelverb, i två av dem som ett s.k. sublemma (som ”nästan” har lemmastatus) och ingen reser något varnande finger för detta uttryck mer än att brukligheten inskränks till ”vardagligt”.

	kommer + V	eftersom att	mena på	visa på	indikera på
NEO	Nej	Nej	Nej	Nej	Nej
SO	Konstr.: komma (att +) V Stilruta: ”måste numera anses acceptabel, om än mer informell [...]”	Nej Stilruta: ”Eftersom bör inte följas av att. [...]”	Eget lemma, partikelverb: <vard.> Konstr.: mena på SATS	Nej	Nej
SHO	Nej	Nej	Sublemma, partikelverb (vard.)	Nej	Nej
SSB	Konstr.: ngn kommer att göra ngt el. ngn kommer göra ngt	Nej	Sublemma, partikelverb (vard.)	Nej	Nej
Lexin	Konstr.: ngn/ngt kommer (att +) VERB	Nej	Nej	Nej	Nej

Tablå 2: Förekomst av flerordsuttryck i ordböckerna.

4. Slutsatser och framåtblick

Ett första svar på frågan om hur flerordsuttryck behandlas i ordböcker är att de behandlas ganska olika beroende på typ, vilket framgår av tablå 2 ovan. Det är generellt de två äldsta varianterna som tas upp: *kommer* + verb, som inte berörs i NEO men väl i efterföljaren SO, samt i SSB och Lexin, och *mena på*, som tas upp i SO och SSB och faktiskt även i den äldre SHO. SO avråder i en stilruta från *eftersom att*, men för övriga flerordsuttryck ges ingen information. Verben *kommer* och *mena på* behandlas dock på olika sätt: *kommer* utan *att* ges i konstruktionsuppgift under **komma** i de ordböcker som tar upp uttrycket, medan **mena på** behandlas som en egen uppslagsform, och endast som partikelverb. Under det enkla **mena**, som tas upp i samtliga ordböcker, ges bara mönstret *mena att*.

I fråga om *kommer* + V gentemot *kommer att* + V ger ordböckerna inga tilläggsupplysningar. Båda varianterna betraktas visserligen av språkvården som acceptabla i dagens språkbruk (Språkriktighetsboken 2005:360), men eftersom många fortfarande verkar lägga märke till *att*-löst bruk⁴ vore någon kommentar om de båda varianternas likvärdighet kanske på sin plats. Stilskillnaden mellan *mena* och *mena på* kunde också uttryckas tydligare, förslagsvis även under det enkla **mena**.

De olika typerna av redovisning medför att flerordsuttrycket som kategori ofta blir svårt att tolka för användaren. Större konsekvens i angivandet av flerordsuttryck skulle kunna göra informationen lättare att tyda. Därtill tas som synes många flerordsuttryck inte upp alls i de undersökta ordböckerna, vilket kan tolkas som att ordböckernas förhållningssätt till flerordsuttryck är starkt avvaktande. I förhållande till hur frekventa de nyare flerordsuttrycken trots allt är i både redigerade och icke redigerade typer av texter kan man hävda att uttrycken bör uppmärksammas mer systematiskt. Exempelvis borde *kommer* + V ha inkluderats i samtliga ordböcker (möjligen undantaget den äldre SHO) i kraft av sin höga frekvens. Det är också något förvånande att de båda konstruktionsordböckerna, särskilt den nyare SSB, inte har med fler varianter på flerordsuttryck. En tolkning av ordböckernas behandling av variantuttrycken är att ordböckerna fungerar som grindvakter för talspråkliga varianter som används i skriftspråket, och så kanske det ska vara? Om man betraktar användarnas inställning till ordböckers normerande funktion är det kanske en fördel att de inte tar med varianter på flerordsuttryck, om man tänker sig att användarna vill ha klara besked utan alternativ. Om man i stället ser till ordböckers deskriptiva funktion som speglingar av språkutvecklingen, ska de naturligtvis återge fler varianter, gärna med upplysningar om i vilken genre och stilistisk funktion varianten kan brukas.

⁴ <www.gp.se/debatt/kommer-att-vara-kvar-1.4124737> (läst 2019-10-20).

I omarbetningen av SO, Sveriges idag största definitionsordbok som fortfarande görs på officiellt uppdrag av en språkvårdande instans i form av Svenska Akademien, finns det flera utmaningar, inte minst eftersom inriktningen har en deskriptiv intention och utarbetandet av ordboken i ökad utsträckning bygger på ett oredigerat textunderlag. Det kan vara rimligt att skilja mellan den officiella och redigerade prosan och den oredigerade i fråga om flerordsuttryck, och i ett val mellan två liknande konstruktioner som fortfarande betyder samma sak är det rimligast att välja den som är frekventast och mest accepterad (om en betydelse-skillnad uppstår kan saken hamna i ett annat läge). Mycket vanliga varianter, såsom *kommer + V* och *eftersom att*, bör tas med, men då gärna med bruklighets-uppgifter av typ ”formellt”, ”vardagligt” m.fl.

Klart är i alla händelser att förändrade konstruktioner kommer att ta större plats i språket i takt med att nya skriftliga genrer etableras. Språkvårdande instanser bör diskutera också dessa förändringar mer strategiskt och hitta ett förhållningssätt till förändringarna, gärna tillsammans med lexikograferna. När språkvården hänvisar till lexikografin lägger man en viss press på lexikograferna att ta ett språkvårdande ansvar som kanske inte är deras att ta?

Litteratur

Ordböcker

- DDO = *Den Danske Ordbog*. <www.ordnet.dk/ddo> (oktober 2019).
 Lexin = *Svenska ord*, 4 uppl. <lexin.nada.kth.se/lexin> (oktober 2019).
 NEO (1995–1996) = *Nationalencyklopedins ordbok*. Höganäs: Språkdata Göteborg & Bokförlaget Bra Böcker AB.
 SAOB = *Svenska Akademiens ordbok*. Gleerups: Lund. <www.saob.se> (oktober 2019).
 SAOL 14 (2015) = *Svenska Akademiens ordlista*. 14 uppl. <www.svenska.se/saol> (oktober 2019).
 SHO (1966) = Johannisson Ture & Karl Gustav Ljunggren: *Svensk handordbok. Konstruktioner och fraseologi*. Stockholm: Esselte Studium.
 SO (2009) = *Svensk ordbok utgiven av Svenska Akademien*. <www.svenska.se/so> (oktober 2019).
 SSB (2003) = *Svenskt språkbruk. Ordbok över konstruktioner och fraser*. Utarbetad av Svenska språknämnden. Stockholm: Norstedts ordbok.

Annan litteratur

- Berg-Olsen, Sturla (2013): Skjematiske fleirordskonstruksjonar i *Norsk Ordbok*. I: *Nordiske Studier i Leksikografi* 12, 80–94.
- Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén (red.) (1997): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Borin, Lars, Markus Forsberg, & Johan Roxendal (2012): Korp – the corpus infrastructure of Språkbanken. I: *Proceedings of LREC 2012*. Istanbul: ELRA, 474–478.
- Bockgård, Gustav (2012): ”Jovars, det duger tills vidare”. En interaktionell analys av konstruktionsfamiljen diskurspartikel + vars. I: Jan Lindström & Sofie Henricson (red.): *Språk och interaktion* 3. Helsingfors: Helsingfors universitet, 175–232.
- Farø, Ken & Henrik Lorentzen (2009): De oversete og mishandlede ordforbindelser – hvilke, hvor og hvorfor? I: *LexicoNordica* 16, 75–102.
- Josephson, Olle (2018): *Språkpolitik*. Språkrådets skrifter 25. Stockholm: Morfem.
- Larsson, Lennart (2004): Svenskt språkbruk. Ordbok över konstruktioner och fraser. Utarbetad av Svenska språknämnden. I: *Språk och stil* NF 14, 209–219.
- Malmgren, Sven-Göran (2009): Från Nationalencyklopedins ordbok (1995–96) till Svensk ordbok utgiven av Svenska Akademien (2009). Med tillbaka- och sidoblickar. I: *LEDA-Nyt* 47, 14–20.
- Malmgren, Sven-Göran (2012): Från Svenska ord (Lexin) 3 till Svenska ord 4. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i leksikografi 11. Rapport från Konferensen om lexicografi i Norden, Lund 24-27 maj 2011*. Lund: Nordiska föreningen för lexicografi, 454–465.
- Nikula, Kristina (1992): Deskriptiva ordböcker – finns dom? I: Ruth Vatvedt Fjeld (red.): *Nordiske Studier i Leksikografi. Rapport fra Konferanse om lexicografi i Norden 28.-31. mai 1991*. Oslo: Nordisk förening for lexicografi, 43–53.
- Sköldberg, Emma (2017): Innehållet i Svensk ordbok utgiven av Svenska Akademien – eller kampen mellan norm och bruk. I: Saga Bendegard, Ulla Melander Marttala & Maria Westman (red.): *Språk och norm. Rapport från ASLA:s symposium, Uppsala universitet 21–22 april 2016*. Uppsala: ALSA, 123–129.

- Sköldberg, Emma & Anna Helga Hannesdóttir (2017): Svenska ord – men vilka? Om uppslagsorden i *Svensk ordbok utgiven av Svenska Akademien*. I: Emma Sköldberg, Maia Andréasson, Henrietta Adamsson Eryd, Filippa Lindahl, Sven Lindström, Julia Prentice & Malin Sandberg (red.): *Svenskans beskrivning 35. Förhandlingar vid trettiofemte sammankomsten Göteborg 11–13 maj 2016*. Göteborg: Göteborgs universitet, 329–340.
- Språkriktighetsboken* (2005). Utarbetad av Svenska språknämnden. Stockholm: Norstedts.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Trap-Jensen, Lars (2002a): Descriptive and normative aspects of lexicographic decision-making: the borderline cases. I: Anna Braasch & Claus Povlsen (red.): *Proceedings of the 10th EURALEX International Congress*. København: Center for Sprogteknologi, 503–508.
- Trap-Jensen, Lars (2002b): Normering og deskription i Den Danske Ordbog – mere eller mindre? I: *LexicoNordica* 9, 63–79.
- Vannebo, Kjell Ivar (1998): Grammatikk i ordbøker og normeringsspørsmålet. I: Ruth Vatvedt Fjeld & Boye Wangensteen (red.): *Normer og regler: festskrift til Dag Gundersen 15. januar 1998*. Oslo: Nordisk forening for leksikografi, 360–367.

Kristian Blensenius
postdoktor, ph.d.
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
kristian.blensenius@svenska.gu.se

Lena Rogström
universitetslektor, ph.d.
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
lena.rogstroem@svenska.gu.se

Så används Svenska Akademiens ordböcker på nätet. Implicit och explicit feedback från användarna

Erik Bäckerud, Pär Nilsson & Emma Sköldberg

This study presents and analyses search strings and user data for different Swedish lexicographical websites. The underlying empirical material was sourced from two relatively new websites, www.saob.se and the joint dictionary portal www.svenska.se, which collects and provides a single point of entry to three Swedish monolingual dictionaries financed by the Swedish Academy. Statistics are presented on the most common search strings, when and where the searches take place, and what devices and digital platforms that are commonly used while visiting the sites. In addition, the study addresses a number of questions and other forms of feedback received from dictionary users. Furthermore, the study provides suggestions and examples of how the collected data can be utilised in upcoming updates and revisions of the dictionaries.

1. Inledning

I mars 2016 lanserades den nya webbplatsen för Svenska Akademiens ordbok www.saob.se och i september 2017 blev ordboksportalen www.svenska.se publik. I den senare går det att söka i tre ordböcker på en gång, SAOL14 (2015), SO (2009) och SAOB (1898–). Sedan lanseringarna av webbplatserna har antalet sökningar stadigt ökat, men hur utnyttjar användarna egentligen Svenska Akademiens verk? I den här artikeln granskas och jämförs statistik över hur sökningarna i Akademiens ordböcker utförs. Dessutom analyseras frågor och synpunkter som skickats in till ordboksredaktionerna från användarna. Syftet är därmed att kartlägga var, när och hur Akademiens ordböcker används på nätet. Resultaten av studien är relevanta inte bara för ordboksredaktörer och för dem som forskar om ordboksanvändning, utan även för lexikologer och språkvetare på ett allmännare plan. Resultaten är förstås även viktiga vid revideringar av de tre ordböckerna och vid utvecklingen av en ny version av ordboksportalen svenska.se.

1.1. Metod och material

Webbenkäter, intervjuer, loggfiler m.m. kan användas när man vill närma sig ordboksanvändare, och exempelvis Tarp (2008) konstaterar att en kombination av undersökningsmetoder är att föredra om man vill kunna dra så långtgående slutsatser som möjligt av sina resultat. I studien som ligger till grund för föreliggande artikel har därför olika metoder utnyttjats, och tre olika delstudier har genomförts. Först har allmän statistik över nätrafiken granskats, dvs. sådan som ringar in allmänna mönster om hur ordböckerna används, tid och plats för sökningarna och frekventa sökord m.m. (se avsnitt 2.1 nedan). Därefter har de mest frekventa söksträngarna som inte leder till träff i någon av ordböckerna, s.k. icke-träffar, analyserats (avsnitt 2.2). Slutligen har frågor och synpunkter som skickats in från användarna till redaktionerna via kontaktformulären på webbplatserna studerats (avsnitt 2.3). Föremålet för studierna representerar vad Abel & Meyer (2013) kallar *implicit feedback* (delstudie 1–2) och *explicit feedback* (delstudie 3).

Materialet i samtliga delstudier utgörs av feedback från år 2018. I några fall lyfts en särskild vecka från detta år fram (nämligen vecka 50, dvs. 10–16 december) för att illustrera ordboksanvändning under en kortare period (se vidare avsnitt 2.1).

1.2. Tidigare studier

Det finns endast ett fåtal tidigare användarstudier som behandlar Svenska Akademiens ordboksmaterial. Rogström (2001) undersöker loggfiler för SAOB:s äldre webbplats (den s.k. OSA-versionen), Törnqvist (2010) diskuterar kortfattat e-postbrev till SAOL-redaktionen och Holmer, Hult & Sköldberg (2015) undersöker, med hjälp av en webbenkät, hur SAOL-appen används. I övrigt är utbudet magert. Det finns således en tydlig lucka för vår studie att fylla. Den kan även sägas komplettera större användarstudier som har genomförts på senare tid både inom och utanför Norden (se t.ex. Müller-Spitzer (ed.) 2014, Trap-Jensen, Lorentzen & Sørensen 2014, Hult 2016, Wolfer et al. 2018 och Kallas et al. 2019, för att bara nämna några).

2. Resultat och analys

I det följande presenteras de tre delstudierna i tur och ordning.

2.1. Allmän nättrafik – när, var, hur och vilka ord?

Den statistik som ligger till grund för föreliggande avsnitt har tagits fram med verktyget Google analytics och avser s.k. *besök*.¹ I genomsnitt läser en användare på saob.se ca 2 sidor under ett besök och på svenska.se ca 4 sidor per besök.

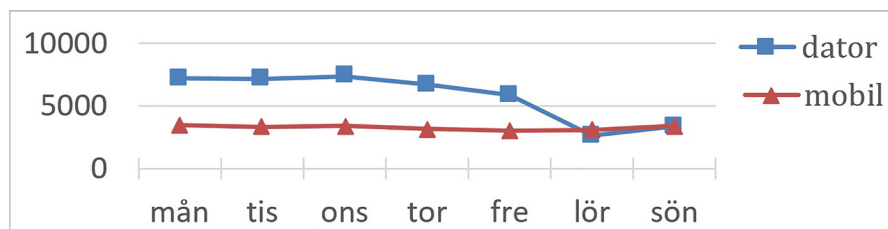
Totalt sett hade saob.se under hela 2018 ca 1,8 miljoner besök och svenska.se ca 2,5 miljoner besök. För vecka 50 uppgick antalet besök per dag till ca 6 800 på saob.se och 9 000 på svenska.se.² Under denna vecka noterades besök från 96 olika länder på svenska.se. De flesta av dessa kom, föga förvånande, från Sverige, men påfallande många härrörde från Finland (drygt 8 procent av alla besök). Användningen av båda webbplatserna har stadigt ökat sedan starten, och när 2019 års statistik ska summeras kommer antalet besök (med dagens takt) uppskattningsvis att ha ökat med 45 procent sedan 2018. För att få ett referensvärde och för att sätta ovanstående statistik i ett sammanhang, har en jämförelse med de danska ordböckerna på ordnet.dk genomförts. Under vecka 50, 2018 hade DDO ca 142 000 besök per dag, vilket är betydligt fler än de sammanlagda besöken hos saob.se och svenska.se. (Dock har användningen av mobilappar inkluderats i statistiken från ordnet.dk, vilken alltså exkluderats i föreliggande studie; jfr fotnot 2 nedan.) Intressant att notera är att SAOB:s systerordbok, ODS, liksom SAOB hade ca 6 800 besök per dag under samma period.³

Webbplatserna skiljer sig åt när det gäller läsarnas användning av dator (stor skärm) eller mobil enhet. Till saob.se kommer majoriteten av besöken från mobiler medan det omvända gäller för svenska.se. Båda webbplatserna är anpassade för mobila enheter. En undersökning av bruket under vecka 50 visar att användarna av svenska.se som sagt främst utnyttjar datorer (stor skärm), men detta gäller i huvudsak under vardagarna. Bruket av stor skärm minskar tydligt på helgen (se figur 1 nedan). Användningen av mobila enheter är däremot ungefär lika stor under veckans alla dagar.

¹ Ett besök påbörjas i och med att en sida på webbplatsen visas för en användare och pågår till dess att användaren varit inaktiv i 30 minuter. Skulle samma användare läsa en ny sida efter att 30 minuter passerat, räknas detta som ett nytt besök.

² SAOL14 (2015) och SO (2009) föreligger även som mobilappar. Användningen av dessa har inte räknats med i undersökningen. Vi saknar uppgifter om hur många artiklar som faktiskt läses i apparna. Däremot kan det noteras att SAOL-appen sammanlagt hade ca 66 000 besök per dag under vecka 50, 2018.

³ Ett stort tack riktas till Henrik Lorentzen på Det Danske Sprog- og Litteraturselskab för tillhandahållande av jämförelsematerial.



Figur 1: Antal besök med dator respektive mobil per dag i svenska.se under vecka 50, 2018.

De flesta besökare kommer till de båda webbplatserna genom någon sökmotor på internet, företrädesvis Google. Hela 86 procent av besöken på saob.se sker via en sökmotor. För svenska.se är motsvarande siffra drygt 50 procent. Detta är inte alls förvånande då saob.se är väl indexerad av Google – söker man efter ett ord som är lågfrekvent i allmänspråket kommer ofta SAOB:s artiklar upp bland de första svaren. Detsamma gäller inte för svenska.se i lika hög grad (i synnerhet eftersom SAOL och SO innehåller betydligt färre uppslagsord än SAOB).

Av tabell 1 nedan framgår för det första de vanligaste söksträngarna i saob.se och i svenska.se under hela 2018 (de två vänstra spalterna i tabellen), och för det andra de vanligaste söksträngarna som gett träff under vecka 50 samma år (de högra spalterna).

	saob.se hela 2018	svenska.se hela 2018	saob.se v. 50, 2018	svenska.se v. 50, 2018
1	antilopfamiljen (1702)	hen (3914)	slejsa (228)	fjät (190)
2	hej (1356)	få (2719)	romarespråket (212)	stuva (154)
3	grusöken (1169)	gå (2503)	snotterbär (168)	hen (100)
4	salpetersyresalt (1168)	z (2422)	barometerglas (147)	som (77)
5	sparkskida (1107)	dag (2395)	jordmätning (135)	gå (75)
6	pelarrum (1040)	som (2348)	kittelflickare (65)	se (73)
7	jordperiod (1011)	man (2255)	tenntråd (65)	lussekatt (70)
8	brax (996)	hej (2235)	kalvill (54)	träl (70)
9	kejsartitel (992)	se (2139)	kompromisspolitik (54)	var (69)
10	feminism (954)	vara (2120)	prokrustessäng (52)	komma (68)

Tabell 1: De vanligaste söksträngarna i saob.se och i svenska.se, dels under hela 2018, dels under vecka 50, 2018 (antal sökningar inom parentes).

Som synes skiljer sig sökorden tydligt åt med avseende på vilken webbplats man söker på. Topporden i saob.se – både under hela 2018 och under vecka 50 – utgörs i stor utsträckning av substantiviska, tämligen specifika (och i allmänspråket ovanliga) sammansättningar, som t.ex. *antilopfamiljen* och *snotterbär*.

De återkommande söksträngarna på svenska.se under de två undersökta perioderna består nästan uteslutande av enkla ord, och flera av söksträngarna (t.ex. *man* och *var*) är homografa. Det är förstås omöjligt att utifrån rena loggfiler veta vilka ord användarna i själva verket haft för avsikt att söka information om och om de blev tillfredsställda med de uppgifter de fick fram (jfr Tarp 2008:22). Båda listorna för svenska.se innehåller dock flera högfrekventa, oregelbundna och polysema verb (se t.ex. *få*, *gå*, *vara*), vilket möjligen indikerar att sökningarna sker i produktions syfte snarare än receptionssyfte. Av sökningarna på svenska.se framgår det också tydligt att vecka 50 är Lucia-veckan. De ålderdomliga orden *ffät* och *stuva* ingår i luciasången och många bakar dessutom lussekatter inför jul. Dessa 10-i-topp-listor stöder således Wolfer et al.:s (2014:286–289) slutsatser att antalet sökningar på enskilda ord varierar över tid, att sökfrekvenser kan spegla händelser i det omgivande samhället samt att användarna slår upp ord som är frekventa i allmänspråket (vilket motsäger resultaten från de Schryver et al. 2006). Precis som i saob.se-listan från hela 2018 finner man i motsvarande svenska.se-lista söksträngen *hej*, en interjektion som tycks vara lite av ett testord i olika svenska webbdataböcker (se t.ex. Hult 2016 om de vanligaste sökorden i *Lexins svenska lexikon*). Värt att notera är även att pronomenet *hen* hamnar i topp 10 på svenska.se. Detta ord gjorde på allvar sitt inträde i svenska ordböcker genom SAOL14 (2015) och det tycks fortfarande väcka stort intresse.

Det är ofta svårt att avgöra varför vissa söksträngar har varit populära och särskilt saob.se-sökningarna förvånar. Många av dem tycks emellertid vara kopplade till korsordsanvändning, och av de olika länkar som leder till saob.se att döma kommer många användare från olika korsordssajter. Ett ord som var mycket populärt under vecka 50 i saob.se är exempelvis *barometerglas*. Just det substantivet ingick i Mästarkrysset (vilket tillhandahålls av Aller media) under den aktuella perioden och det diskuterades sedan på korsordsforumet webbkryss.nu.

2.2. Återkommande icke-träffar

Ur loggfilerna kan man ta fram listor med söksträngar som inte har lett till träff i ordböckerna, s.k. icke-träffar. Sådana listor är mycket användbara vid revideringar av ordböckers lemmalistor (se Hult 2016:95, 155 med anförda referenser). För att avgränsa oss diskuterar vi i det följande de 500 vanligaste icke-träffarna under 2018 i SO, dvs. den av de tre ordböckerna som uppdateras just nu.

Bland söksträngarna finns det en stor mängd återkommande, till synes tillfälliga, (två)bokstavskombinationer som är obegripliga för oss, t.ex. *ox* (622 sökningar), *ga* (548) och *sä* (263).⁴ Vidare finns det en del böjningsformer (t.ex. *dåligare*, 88) och flerordsuttryck (t.ex. *på grund av*, 364) som är upptagna i SO, men som användarna ändå inte fått träff på. Även ett stort antal förkortningar påträffas här (t.ex. *ca*, 878). Dessa icke-träffar ger således nyttig information om hur sökfunktionerna på webbsidan kan förbättras. Det finns också flera, mer lexikologiskt intressanta icke-träffar. För det första utgörs en stor andel av dem av svenska ord, varav några är ganska nyetablerade, t.ex. *funktionsvariation* (136) och *filterbubbla* (93). Vissa av dessa ord har en slangartad prägel (t.ex. *jalla*, 304). För det andra finns det engelskklingande ord och förkortningar, och frågan är om de numera också har konventionaliserats i svenskan. Detta gäller i synnerhet ursprungligen engelska ord med försvenskad stavning (jfr t.ex. *najs*, 332, med *cringe*, 481). Det finns också, utifrån de stavningssätt som anges i SO och SAOL14, ett antal felstavade ord i materialet, t.ex. *priviligerad* (1778), *mail* (1242) och *orginal* (260). Dessa icke-träffar illustrerar den gränsdragningsproblematik som så ofta är närvarande vid ordboksarbete. Eventuellt hör dessa ord inte hemma i en normativ ordlista som SAOL, men kanske ska de behandlas i en deskriptiv ordbok av SO:s slag? (se Sköldberg 2017). Slutligen finns det också flera egennamn (t.ex. *Zlatan*, 109, och *Fortnite*, 105) i materialet. Dessa pekar på att den traditionella svenska arbetsfördelningen mellan allmänna ordböcker, namnlexikon och encyklopedier när det gäller beskrivningen av proprier inte alltid upplevs som helt självklar – eller adekvat – bland användarna (se vidare bl.a. Trap-Jensen & Lorentzen 2017 om sökningar på namn i DDO).

2.3. Explicit feedback från användarna

Som en sista delundersökning har frågor och synpunkter som kommit in från ordboksanvändarna analyserats. Samtliga meddelanden som via kontaktformulär på de olika webbsidorna har skickats till redaktionerna under hela 2018 har granskats och sorterats i olika kategorier.

Efter att spam, nonsens- och redaktionsinterna meddelanden räknats bort återstår sammanlagt 85 meddelanden skickade till saob.se och 311 till svenska.se. Dessa har vart och ett kunnat föras till någon av nedanstående fem kategorier (se tabell 2).

⁴ Möjligen avser sökningar på *ox* sammansättningsförleden av *oxe* (som i t.ex. *oxfilé* och *oxpytt*).

Kategori		<saob.se>		<svenska.se>	
		Antal	(%)	Antal	(%)
1	Jag hittar inte ordet/betydelsen	27	32	114	37
2	Jag behöver hjälp att använda ordboken/ordböckerna	12	14	70	22,5
3	Vad är rätt? (Normering)	-	-	40	13
4	Ändra det här! (Felaktigheter)	41	48	58	18,5
5	Övrigt	5	6	29	9
Totalt		85	100	311	100

Tabell 2: Meddelanden från användarna, sorterade per kategori.

För både svenska.se och saob.se gäller ofta (inte oväntat) att användare kommenterar eller frågar om ett specifikt ord (se kategori 1 i tabell 2). Cirka 1/3 av meddelandena handlar om detta, och det är den enskilt största kategorin av meddelanden för svenska.se. Främst undrar användarna varför ett ord inte finns med i ordboken (lemmalucka). Inte sällan handlar det om nya ord som ännu inte hunnit komma med i aktuella upplagor, och ofta är orden som efterfrågas också rimliga (utifrån gällande urvalsprinciper) att ta med i framtiden. Flera kommentarer handlar dock om ord, uttryck eller betydelser som i själva verket finns med i det aktuella verket. Användaren har då i regel sökt på böjda former av ord eller skrivit in flerordsförbindelser i sökrutan och därmed inte fått träff (jfr avsnitt 2.2 om icke-träffar ovan). Detta är, som redan påpekats, ett tydligt område att arbeta vidare med. Redan nu finns en "Menade du"-funktion, som vid en icke-träff ger förslag på stavningsmässigt närliggande ord, men den fungerar dåligt för flerordsförbindelser. I fråga om böjda ord har SAOL kommit längst, med samtliga böjningsformer lemmatiserade. Detsamma gäller inte för SO eller SAOB.

Till kategori 2 hör strukturella och tekniska frågor kring hur ordböckernas information är strukturerad och hur beskrivningspråket ska tolkas. Användarna vill ofta ha hjälp med att uttyda uttalssystemen eller SAOB:s många förkortningar (inte minst i etymologiparenteserna). Hit har också förts frågor och kommentarer om hur man söker fram en viss informationstyp, exempelvis: "Jag vill söka fram alla andra konjugationens verb", eller "Hur får jag tag i alla ord som

etiketterats som 'nedsättande'?. Flera frågor handlar även om att man vill kunna ladda ner hela ordböksinnehåll (vanligen SAOL:s) för att använda i egenkonstruerade appar eller liknande.

Till svenska.se vänder sig också användare för att få vägledning i normerande frågor (kategori 3 i tabell 2). Anledningen är rimligen att SAOL återfinns i den portalen och att ordlistan beskrivs som den inofficiella normen för svensk stavning och böjning. Ett exempel på en sådan fråga är: ”'rookies' anges i SAOL som pluralform till 'rookie', men är -s verkligen en svensk pluraländelse?”. Motsvarande meddelanden skickas inte till saob.se.

Ibland ger de olika ordböckerna på svenska.se olika bud gällande samma uppslagsord. Exempelvis anger SAOL *driftskostnad* (med fog-s) som uppslagsform, medan SO anger *driftkostnad* (utan fog-s). SAOB upptar båda formerna. Denna typ av skillnader mellan verken kan förvirra användarna, vilket en avsändare påpekar: ”Vad heter det egentligen? Bestäm er!”.

Den klart största andelen meddelanden som skickats till saob.se handlar om fel eller ändringsförslag i artiklarna (se kategori 4 i tabell 2). Det gäller då skriv- eller stavfel, rena sakfel eller andra fel beträffande innehållet snarare än formen. Väldigt många kommentarer till såväl saob.se som svenska.se har att göra med SAOB:s föråldrade metaspråk. Allra vanligast är kommentarer gällande SAOB:s stavning av kort ä-ljud (dvs. *ä* i ord som idag stavas med *e*, t.ex. *värksamhet*, *rägn* och *pängar*). Den typen av stavningar användes i ordbokens definitioner ända fram till 1960-talet. Som en direkt följd av denna feedback inleds numera (sedan januari 2019) kontaktformuläret på saob.se med en upplysning om ordbokens stavningsprinciper i detta avseende.

Många synpunkter och frågor av innehållsmässigt slag rör också kontroversiella ord. Vissa definitioner är inte politiskt korrekta idag och vissa språkexempel är olyckligt formulerade. Ett exempel som sticker ut i det här avseendet är språkprovet ”mesig tjejfotboll” som återfinns under uppslagsordet *mesig* i SO och som många användare har reagerat på. Kritiken plockades upp i riksmidia och även Svenska Fotbollförbundet har protesterat mot formuleringen i verket (se t.ex. Svenska Dagbladet 19/2 2018). Just detta språkexempel har ändrats inför kommande upplaga av SO. Ändringen är dock inte synlig för användarna ännu eftersom svenska.se inte uppdateras löpande.

Kategori 5, Övrigt, utgörs av sådana meddelanden som inte kunnat föras till någon av de fyra övriga grupperna. I flera fall handlar det om allmänna berömmande eller kritiserande kommentarer.

3. Sammanfattning och framåtblickar

Av den feedback som presenterats och diskuterats ovan kan vi dra några slutsatser och peka på ett par möjliga utvecklingslinjer. För det första har Svenska Akademiens ordböcker på nätet potential att nå ut till fler. Sedan lanseringen av webbplatserna har antalet besök stadigt ökat, och framför allt borde svenska.se i mobilen ha potential att användas mer (inte minst visar jämförelsen med den danska användarstatistiken att detta är möjligt). Vidare bör ordboksredaktörerna tillmötesgå användarna i större utsträckning – t.ex. kan böjda former, avledningar och flerordsförbindelser göras sökbara i ännu högre grad. Det går också att förtydliga vilka funktioner de tre ordböckerna på svenska.se är tänkta att fylla. När ordböckerna beskriver samma ord eller betydelse på olika sätt blir det i nuläget ibland otydligt för användarna. Det måste således t.ex. framgå att SAOL är mer normativ än de andra två verken.

Meddelanden till redaktionerna tydliggör också att mycket av beskrivningspråket i ordböckerna är svårt att förstå. Framför allt gäller detta i SAOB:s artiklar. Med hjälp av feedbacken blir det klarare i vilken ände man rimligen ska börja när tiden har kommit för revidering av den historiska ordboken. Även de mest akuta fallen, som ofta rör artiklar med kontroversiella ord, går att ringa in med hjälp av användarnas synpunkter. När nätordböckerna så småningom inte längre speglar innehållet i tryckta versioner av samma verk (något som i nuläget fortfarande gäller för samtliga ordböcker på svenska.se) finns möjligheten att revidera kontinuerligt och att uppdatera webbsidorna fortare.

Studiet av loggfilerna som rör svenska.se bekräftar Wolfer et al.:s (2014) iakttagelse att ordboksanvändare tenderar att slå upp frekventa och socialt relevanta ord. I dagsläget finns på saob.se både ett länkat Instagram-konto och funktionen ”Dagens ord”, vilken slumpvis väljer ut och lyfter fram ett ord ur ordboken. Det vore rimligt att införa motsvarande s.k. ”add-on-material” även på svenska.se, där man dessutom kunde tänka sig att regelbundet lyfta fram utvalda grupper av tillfälligtvis aktuella ord (se vidare Holmer, von Martens & Sköldberg 2015 om tankarna bakom Dagens ord i SO-appen).

Avslutningsvis kan det konstateras att de två webbsidor som diskuteras i föreliggande artikel båda är relativt unga, och att dra nytta av feedback från användarna blir därför också till stora delar en ny erfarenhet för de olika ordboksredaktionerna. Användarnas synpunkter kommer utan tvekan att bli värdefulla inför kommande revideringar.

Litteratur

Ordböcker

- DDO (2019) = *Den Danske Ordbog*. Tillgänglig via <www.ordnet.dk/ddo> (februari 2020).
- ODS (2019) = *Ordbog over det danske Sprog*. Tillgänglig via <www.ordnet.dk/ods> (februari 2020).
- SAOB (1898–) = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1–. Lund.
- SAOL 14 (2015) = *Svenska Akademiens ordlista*. 14 uppl. Stockholm: Norstedts.
- SO (2009) = *Svensk ordbok utgiven av Svenska Akademien*. Stockholm: Norstedts.

Annan litteratur

- Abel, Andrea & Christian M. Meyer (2013): The dynamics outside the paper: user contributions to online dictionaries. I: *Proceedings of eLex 2013 conference*, 17-19 October 2013, Tallinn, Estonia. Tallinn, 179–194.
- Holmer, Louise, Ann-Kristin Hult & Emma Sköldberg (2015): Spell-checking on the fly? On the use of a Swedish dictionary app. I: Iztok Kosem et al. (eds.): *Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton, 356–371.
- Holmer, Louise, Monica von Martens & Emma Sköldberg (2015): Making a dictionary app from a lexical database: the case of the Contemporary Dictionary of the Swedish Academy. I: Iztok Kosem et al. (eds.): *Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton, 31–50.
- Hult, Ann-Kristin (2016): *Ordboksanvändning på nätet En undersökning av användningen av Lexins svenska lexikon*. Göteborgsstudier i nordisk språkvetenskap 27. Göteborg.
- Kallas, Jelena, Svetla Koeva, Iztok Kosem, Margit Langemets & Carole Tiberius (2019): *A survey of user needs*. Rapport utg. av Elexis (European Lexicographic Infrastructure) <www.elexis.eu/wp-content/uploads/2019/02/ELEXIS_D1_1_Lexicographic_Practices_in_Europe_A_Survey_of_User_Needs.pdf> (februari 2020).
- Müller-Spitzer, Carolin (ed.) (2014): *Using online dictionaries*. Berlin/New York: de Gruyter.
- Rogström, Lena (2001): Användarstrategier i OSA-databasen. I: Gunilla Byrman et al. (red.): *Korpusar i forskning och undervisning. Rapport från ASLA:s höstsymposium, Växjö, 11–12 november 1999*. Uppsala, 255–265.

- de Schryver, Gilles-Maurice, David Joffe, Pitta Joffe & Sarah Hillewaert (2006): Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography. I: *Lexikos* 16, 67–83.
- Sköldberg, Emma (2017): Innehållet i Svensk ordbok utgiven av Svenska Akademien – eller kampen mellan norm och bruk. I: *Språk och norm. Rapport från ASLA:s symposium, Uppsala universitet 21–22 april 2016*. Uppsala, 123–129.
- Tarp, Sven (2008): Kan brugerundersøgelser overhovedet afdække brugernes leksikografiske behov? I: *LexicoNordica* 15, 5–32.
- Trap-Jensen, Lars & Henrik Lorentzen (2017): Navnestof i ordbøger – nye tider, nye måder. I: *LexicoNordica* 24, 109–129.
- Trap-Jensen, Lars, Henrik Lorentzen & Nicolai H. Sørensen (2014): An odd couple – corpus frequency and look up frequency: what relationship? I: *Slovenščina 2.0*, (2/2014), 94–113.
- Törnqvist, Lars (2010): Brukarmedverkan i webbordböcker. I: Kristinn Jóhannesson et al. (red.): *Bo65. Festskrift till Bo Ralph*. Meijerbergs arkiv för svensk ordforskning 39. Göteborg, 383–390.
- Wolfer, Sascha, Alexander Kopenig, Peter Meyer & Carolin Müller-Spitzer (2014): Dictionary users do look up frequent and socially relevant words. Two log file analyses. I: *Proceedings of the XVI EURALEX International Congress*. Bolzano/Bozen, 281–290.
- Wolfer, Sascha, Iztok Kosem, Robert Lew, Carolin Müller-Spitzer & Maria Ribeiro Silveira (2018): Web-based Exploration of Results From a Large European Survey on Dictionary Use and Culture: ESDexplorer. I: *Lexikos* 28, 440–447.

Erik Bäckerud
IT-ansvarig
Svenska Akademiens ordboksredaktion
Dalbyvägen 3
SE-224 60 Lund
erik.backerud@svenskaakademien.se

Pär Nilsson
ordboksredaktör, fil. lic.
Svenska Akademiens ordboksredaktion
Dalbyvägen 3
SE-224 60 Lund
par.nilsson@svenskaakademien.se

Emma Sköldberg
lektor, docent
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 Göteborg
emma.skoldberg@svenska.gu.se

Ngram i SAOB

Karin Cavallin

The purpose of this study is to identify and eventually analyze reoccurring strings in the definitions of SAOB. Until recently, the digital version of SAOB consisted more or less of plain text, with some typographical information. Structure and informational categories in SAOB had not been marked in the xml code, and hence been practically unreachable for digital management. In order to improve readability, usefulness and adaptability of SAOB we have digitally identified different components of the text in the dictionary (Bäckerud 2016:102). During 2019 we, among other things, identified our language samples and the definitions. The definitions in SAOB contain several reoccurring strings. Within lexicography we know them as *collocations*. Since these strings usually lack other collocational features, we prefer to call them *reoccurring ngrams*. These ngrams can be quite formulaic, very long and differ from ordinary written and spoken Swedish. Ngram tools like *Ngram Statistics Package* (Banerjee & Pedersen 2003) and *The Sketch Engine* (Kilgarriff et al. 2004) have been used for a long time in lexicography to e.g. identify collocations. Such tools are here proposed to be used for purely lexicological purposes, on the dictionary text per se.

1. Inledning

Svenska Akademiens ordbok, SAOB, har redigerats under mer än hundra år. Ambitionen är att beskriva hela det nysvenska ordförrådet, från 1521 och framåt. Beskrivningen utgörs bland annat av etymologi, formredovisning (genus, plural, böjningar och dylikt), autentiska citat och deras respektive källa, samt definitioner.

En redigering som pågår under mer än ett sekel skapar givetvis ett eget sätt att beskriva betydelser av ord, med mer eller mindre formelartade uttryckssätt. Dessa uttryckssätt förändras också över tid, och kan skilja sig mellan olika redaktörer. För våra läsare och för nya redaktörer kan detta formelartade språk te sig ogenomträngligt. Olikheter i uttryckssätt kan också ge ett spretigt intryck när man jämför artiklar som behandlar likartade fenomen.

2. Syfte och problemställning

Lexikografens uppgift är att beskriva språket. I denna studie undersöks lexikografens eget språk, dvs. definitionsspråket och dess formler. Det primära syftet med studien är att få fram data som underlättar både den fortsatta redigeringen och den nära förestående revideringen av SAOB. Därtill tillkommer ett rent lexikologiskt syfte, nämligen att möjliggöra närstudier av definitionsspråket i SAOB.

3. Forskningsöversikt

Vanligtvis använder man korpusar för att få underlag vid redigering av en ordboksartikel. Här kommer istället SAOB själv, eller snarare SAOB:s definitionstexter, användas som en korpus. Mig veterligen har inga kvantitativa, *datadrivna*¹ studier gjorts på själva definitionsspråket i ordböcker. Nilssons licentiatavhandling (Nilsson 2019) är ett exempel på en *hypotesdriven* studie över hur fenomenet *bildlighet* beskrivs i just SAOB. Det har ännu inte gjorts någon studie över hur SAOB:s definitionsspråk skiljer sig från språket i en välbalanserad korpus. Det finns indikationer på att SAOB:s definitionsspråk är ganska avvikande.

Vid redigering av en ordboksartikel är det vanligt att man undersöker hur ordets närkontext ser ut. Till exempel kan översikter i kollokationsverktyg, som *WordSketches* i *SketchEngine* (Kilgarriff et al. 2004) eller KORP:s *ordbild* (Borin et al. 2012), användas. Kollokationsverktyg är datorprogram som utgår från frekvensberäkningar av vilka strängar som samförekommer med varandra. Dessa strängar motsvaras oftast av det vi känner som *ord*, men verktygen bryr sig egentligen inte om ord. Det som räknas är samförekomst av strängar avskilda med mellanrum eller skiljetecken. Dessa strängar kommer härefter kallas *ngram*.

Dessa frekvensberäkningar är generellt uppbyggda av ngramberäkningar där resultaten är viktade, utslätade (*smoothed*²) med hjälp av associationsmått, *Measures of Association*. Samförekomst mellan strängar som förekommer ofta och tillsammans med det mesta presenteras som mindre relevant än strängar som förekommer mer sällan, men mer eller mindre uteslutande tillsammans (Banerjee & Pedersen 2003:376). Utslätningen görs för att alltför frekventa strängar (såsom konjunktioner och prepositioner) annars lätt tar över som samförekommande och för att de mest semantiskt eller syntaktiskt relevanta samförekomsterna ska komma högt upp som kollokationsförslag.

¹ *Bottom-up* är kanske den mer gängse termen för *datadriven*.

² Det saknas tyvärr bra svenska termer för många begrepp inom språkteknologi och datavetenskap.

Denna ngramstudie skiljer sig från andra tillämpningar av ngram³ på ett par punkter:

- En ordboks definitionstext används som korpus.
- Det är väldigt höga, återkommande ngram, som är av intresse. Vanligtvis består kollokationer av upp till 3–4 ord.
- Utslätning tillämpas inte. Vid högre ngram än 4gram minskar frekvensen i sig, och gör utslätning tämligen irrelevant.⁴

4. Metod och material

För att undersöka definitionsspråket i SAOB krävs att SAOB:s definitionstext har märkts upp. Detta har gjorts under 2019 i samband med det stora digitaliseringsprojektet vid SAOB (beskrivet i Bäckerud 2016). Hela SAOB består av ca 200 miljoner tokens, medan definitionstexten utgör ca 9 miljoner tokens. En definition i SAOB kan bestå av noll⁵ till flera definitionsled. Vid en utförligare definition separeras varje definitionsled med semikolon. I denna studie undersöks endast ngram som ingår i eller utgör ett definitionsled.

Eftersom SAOB har bibehållit den redigeringstradition som grundlades redan för hundra år sedan skiljer sig SAOB i vissa avseenden från ordböcker som är skrivna under ett senare och kortare tidsspänn. Exempelvis saknas en enhetlig struktur och olika informationskategorier har inte nödvändigtvis en bestämd plats. För definitionsleden innebär detta att exempelvis etymologiparenteser och brukligheter kan ligga insprängda mitt i ett definitionsled. Detta medför att definitionsleden behöver rensas innan det utförs beräkningar på dem. Etymologiparenteser har plockats bort.⁶ Bruklighetsmarkörer har däremot inte plockats bort. Dessa är svåra att utskilja maskinellt från övrig definitionstext. Definitionsled som egentligen utgör en hänvisning⁷, såsom sådana som inleds med 'jfr', samt innehåller strängar längre än ett tecken som innehåller versaler, har rensats bort. Parenteser har tagits bort. Se vidare i avsnitt 4.2.

³ Den mest kända tillämpningen av ngram torde vara Google Books NGram Viewer, www.books.google.com/ngrams.

⁴ Personlig mejlkontakt med Ted Pedersen, 2016-08-02.

⁵ Det förekommer definitioner som är helt tomma, samt definitionsled som enbart utgörs av en hänvisning.

⁶ Eftersom etymologiparenteserna är omgärdade av hakparenteser ([]) var detta enkelt. I skrivande stund har etymologiparenteserna hunnit märkas upp inom ramen för taggningprojektet.

⁷ SAOB kallar det för hänvisning. I andra ordböcker kallas det exv. för korsreferens.

4.1. Perioder

SAOB kan delas in i fyra huvudsakliga perioder, baserade på bland annat omfång och ambitionsnivå (se vidare Larsson 2014). Den första perioden sträcker sig från SAOB:s början 1883 till ca 1920, period 2 från ca 1920–1960, period 3 1960–2002, och period 4 från ca 2002 och framåt. För att kunna studera eventuella förändringar och olikheter över tid har i princip denna indelning använts även i denna studie. För enkelhets skull har initial bokstav inom tidsintervallet delats in som hörande till en viss period, och inte exakt efter redigeringstidpunkten. Tidsintervallen motsvaras ungefär av artiklar med begynnelsebokstav A–E, F–Q, R–S respektive T–V.⁸

Antal tokens i definitionstexten skiljer sig åt mellan perioderna. Period 1 har ca 1,2 miljoner tokens, period 2 ca 3,5 miljoner, period 3 ca 3,5 miljoner och period 4 (till och med band 37, UTSUG–VRETA) ca 700 000 tokens. Detta är helt i linje med respektive periods ambitionsnivå.

4.2. Det statistiska verktyget Natural Statistics Package, NSP

För att beräkna ngrammen har definitionstexten matats in i verktyget Natural Statistics Package, NSP (Banerjee & Pedersen 2003). I skrivande stund är hänvisningarna i SAOB inte taggade, utan ligger inom räckvidden för definitionstexten. Dessa har därför hjälpligt filtrerats bort med heuristiska metoder⁹ med de i NSP inbyggda funktionerna *stop*, *token* och *non-token* i NSP. Under hösten 2019 planeras hänvisningarna att taggas. Vissa specifika förkortningar som förekommer i SAOB, som ”o.” och ”l.”, har också specialbehandlats.

NSP är inte helt anpassat för syftet att söka efter så höga ngram som studeras i denna studie. Frågan är om något ngramverktyg är anpassat för det. Vid höga ngram växer komplexiteten exponentiellt, och inte ens en kraftfull dator med stort minne klarar av att genomföra hela körningen.¹⁰

5. Resultat

Resultaten från denna studie presenteras för respektive ngram i form av 10-i-topp-listor i tabellform. Jag kommer särskilt redovisa 1gram, 3gram, 5gram och 7gram. Urvalet beror dels på utrymmesskäl, dels på att variationen mellan 10-i-topp-listorna inte visade sig vara särskilt stor vid en ökning med bara ett token. Det krävdes en ökning av minst två token för att det skulle bli tydliga och intressanta skillnader mellan tabellerna.

⁸ Larsson (2014) tar i sin studie hänsyn till att vissa artiklar på exv. E och F inte är skrivna i alfabetisk-kronologisk ordning. Det görs inte i denna studie.

⁹ Definitionsled som inleds med ”jfr” har filtrerats bort.

¹⁰ NSP kördes på en virtuell dator med 16 kärnor och 112 GB primärminne (D14_v2). Ett ngramverktyg med mindre minnesåtgång hade, om det finns, varit lämpligare.

Tabellen över 1gram visar förstås inga formler eller relevanta samförekomster, men visar ändå lite intressanta saker. Det första värt att notera är att de mest frekventa 1grammen genom perioderna är ganska lika. Värt att notera är att dödtecknet (†) är utanför 10-i-topp-listan under period 3. Det är också tydligt att materialet är olemmatiserat. ”Af” och ”av” syftar givetvis på samma semantiska enhet, men denna har genomgått en stavningsreform under 1900-talet.

	Hela SAOB	Period 1	Period 2	Period 3	Period 4
1	i 457984	i 49664	i 190802	i 178831	i 38687
2	om 207138	af 19299	om 77336	om 89348	om 22485
3	av 168829	om 17969	av 72814	av 79340	av 14652
4	som 154297	till 17527	som 65032	som 60179	med 14073
5	med 136652	med 16974	† 55224	med 53395	som 12732
6	till 127713	† 16608	till 53416	till 45932	se 11641
7	† 112921	som 16354	med 52210	på 39260	till 10838
8	ngt 107444	på 13068	ngt 48636	ngt 38385	ngt 10387
9	på 106175	en 12482	på 45612	att 36530	på 8235
10	att 86256	ngt 10036	en 36910	se 35397	† 7895

Tabell 1: 1gram.

I tabellen för 3gram börjar uttryck som liknar den så kallade saobitiskan dyka upp. Mest frekvent är flerordsuttrycket ”i fråga om”, som torde vara någorlunda, men inte lika, frekvent även i en vanlig, balanserad korpus. På resterande platser dyker det mest upp uttryck som är typiska för SAOB:s sätt att beskriva ords betydelser på. Vid 3gram har den absoluta frekvensen sjunkit avsevärt, men på grund av platsbrist redovisas inte siffrorna här.

	Hela SAOB	Period 1	Period 2	Period 3	Period 4
1	i fråga om	i fråga om	i fråga om	i fråga om	i fråga om
2	i sht i	i sht i	i sht i	i sht i	i sht i
3	i vissa trakter	i sht om	i vitter stil	i vissa trakter	i sht förr
4	i sht förr	i högre stil	i vissa trakter	i sht förr	numera i sht
5	i vitter stil	i sht förr	i sht förr	i vitter stil	fråga om att
6	i skildring av	numera i sht	i sht om	i skildring av	i sht om
7	numera i sht	för att beteckna	numera i sht	numera i sht	i skildring av
8	i sht om	på grund af	sht i vitter	sht i vitter	senare led i
9	sht i vitter	i motsats till	i högre stil	fråga om att	om person som
10	med tanke på	sht i skriftspr	för att beteckna	i sht om	om ngt sakligt

Tabell 2: 3gram.

Vid 5gram dominerar det formelartade språket. De mest frekventa 5grammen motsvarar framförallt det som används för att beteckna bruklighet, dvs. om eller när, eller i vilket sammanhang, ett beskrivet ord används. I tabellen förekommer bl.a. uttryck som betecknar inom vilken social eller regional sfär den beskrivna aspekten av ordet används, men även mer metaspråkligt språk, som ”stundom svårt att skilja från”. Det är inte självklart att skilja ett ords användningsområde från dess betydelse, varför konceptet *bruklighet* är problematiskt i sig. Det är därför inte givet att bruklighet bör filtreras bort, om det hade gått att göra på ett enkelt sätt.

Hela SAOB	Period 1	Period 2	Period 3	Period 4
i sht i vitter stil	i sht i fråga om	i sht i vitter stil	i sht i vitter stil	i sht i fråga om
i sht i fråga om	i sht i högre stil	i sht i fråga om	i sht i vissa trakter	i sht i skildring av
stundom svårt att skilja från	stundom svårt att skilja från	i sht i högre stil	i sht i skildring av	stundom svårt att skilja från
i sht i vissa trakter	i sht i vitter stil	stundom svårt att skilja från	stundom svårt att skilja från	i fråga om att ngt
i sht i högre stil	som har till yrke att	ofta svårt att skilja från	ngn gg i skildring av	i fråga om att ngn
i sht i skildring av	ofta svårt att skilja från	för att beteckna att ngn	numera i sht i skildring	numera i sht i skildring
numera i sht i vitter	person som har till yrke	om var särskild av de	i sht i fråga om	språkprov svårt att skilja från
numera i sht i skildring	i sht i religiös stil	för att beteckna att ngt	numera i sht i vitter	vissa språkprov svårt att skilja
ngn gg i skildring av	med syftning på en efterföljande	sht i vitter stil i	numera i sht i vissa	i vissa språkprov svårt att
sht i vitter stil i	ngn gg i högre stil	numera i sht i vitter	sht i vitter stil om	ersatt av bestämning inledd av

Tabell 3: 5gram.

Då denna studie primärt är en förstudie i väntan på en än mer uppmärkt SAOB, går jag inte in närmare i detalj på varje ngrams specifika frekvenstabell. Vid högre ngram än 7gram och större mängd tokens i periodfilerna räcker inte RAM-minnet till att köra igenom befintliga data. Endast de perioder som har minst antal tokens i sig kan hanteras av NSP för 7gram och uppåt. I appendix återfinns de 10 mest frekventa ngrammen för 7gram och 10gram.

Bland 10grammen finns ett tydligt exempel där redaktörerna brustit i konsistens. Talen *ett* till *tolv* borde beskrivas på snarlika sätt i SAOB. Bland 10grammens 10-i-topp-lista ser vi att beskrivningssättet varierar på två olika sätt: *tal som i serien av hela tal har sin plats*, respektive *som i serien av hela tal har sin plats mellan*.

Detta föranledde mig att skärskåda samtliga i SAOB beskrivna tal¹¹ och jag kunde då finna ännu mer variation. *TVÅ*, *TRE*, *SJU* och *TIO* beskrivs med *tal som i serien av hela tal har sin plats mellan*, medan övriga beskrivs på ett antal olika sätt:

- EN/ETT ”det första grundtalet (i skrift ofta betecknat 1 l. I)”
- FYRA ”det tal som i talräckan har sin plats mellan tre o. fem”
- FEM ”det tal som i talräckan har sin plats mellan fyra o. sex”
- SEX ”tal som i talserien har sin plats mellan fem o. sju”
- NIO ”det tal som i talserien har sin plats mellan åtta o. tio”
- ELVA ”grundtal som följer närmast efter tio”
- TOLV ”grundtal angivande det antal som med en enhet överstiger elva; tal som i talserien har sin plats mellan elva o. tretton o. som utgör bas för duodecimalsystemet”

6. Diskussion och analys

Bruklighetsmarkörer är ofullständigt uppmärkta i dagsläget och skymmer sikten för den rena definitionstexten i ngrammen. I princip hade dessa kunnat manuellt sorteras bort från 10-i-topp-listorna, men då skulle resultaten bli svåra att upprepa. Dessutom skulle parallelliteten med 1gram och 2gram bli skev. Som nämndes i avsnitt 5 är det inte helt självklart var gränsen går mellan en regelrätt betydelsebeskrivning och bruklighetsmarkör. Denna studie tydliggjorde en oklarhet inom SAOB som var mer omfattande än man tidigare insett.

Det är en klar nytta med att beräkna ngram på SAOB:s definitionsspråk. Typiska formler tydliggörs och kan därmed underlätta vid upplärning av nya redaktörer. Framöver skulle man även, givet att en jämförande ngramanalys görs på en balanserad korpus, kunna jämföra språket i SAOB med vanlig skriven text.

NSP byggdes inte för syftet att beräkna långa ngram, utan för kortare ngram, 1–4gram. Det är sällan någon intresserat sig för längre ngram än så. I denna studie har jag tittat på de tio mest frekventa 1–10grammen. I många fall kan det troligtvis räcka med att titta på resultat från de lägre ngrammen i SAOB, och därigenom hitta längre återkommande ngram.

¹¹ Alla grundtal utom ÅTTA är till dags dato beskrivna i SAOB.

7. Avslutande kommentarer

Det som återstår att studera är relationen mellan definitionstext och annan text. Ju mer uppmärkt allt blir, desto mer korrekt kan man se fördelningen av olika informationskategorier. Man måste också fundera på hur relevant och användbart det är att veta att en viss formulering återkommer så få gånger som bara två. Är orden semantiskt nära besläktade är det förstås motiverat.

För att kunna ta ett samlat grepp vid betydelsebeskrivningarna vid revideringen kommer SAOB-redaktionen förmodligen framförallt att ta hjälp av interna hänvisningar för att hitta de artiklar som är relaterade till varandra. Jag tror dock att ngramanalysen kan vara ett komplement till detta, genom att man med den kan dels hitta fler relaterade artiklar, dels skapa en utgångspunkt för hur vår betydelsebeskrivning ser ut språkligt.

Det torde även vara möjligt att ta bort vissa innehållstunga ord ur ngramsökningarna, för att få fram ännu fler typiska formler.

7.1. Lemmatisering och normalisering

Definitionstexten i SAOB är tämligen enhetlig vad gäller former. Oftast skrivs definitionerna i presens, och det är sällan ord böjs i bestämd form eller plural. Dock förekommer det förstås en del böjningar. Det förekommer exempelvis plurala verbformer ända in på 1970-talet i definitionstexten i SAOB, trots att sådana mer eller mindre försvann ur svenska språket i övrigt under 1940-talet. Därtill tillkommer exempelvis ett antal i dag föråldrade stavningar som *hv* (*hvarpå*), *fv* (*öfverläggning*) och *f* (*af*) där vi i dag endast skriver *v*, och *dt* (*landtman*), där vi i dag bara skriver *t*. Det förekommer föråldrade former som *konung* (*kung*), *fader* (*far*), *moder* (*mor*). I onormaliserad text räknas *mor* och *moder* som två olika strängar. Strängarna *av* och *af* förekommer frekvent. En normalisering skulle gett en mer rättvisande bild av exempelvis frekvensen för alla de ngram som *af* eller *av* ingår i. För denna studies del hade det varit önskvärt med en normalisering av redaktionstexten, men det hade varit ett svårt och mycket omfattande arbete.

7.2. Användningsområden

Det finns tre huvudsakliga användningsområden för resultaten i denna studie. Dessa är att:

- strama upp definitionstexten vid revideringen,
- träna upp nya redaktörer som snabbare kan tillgodogöra sig viktiga formler,
- rent lexikologiskt kunna studera hur SAOB har sett ut genom åren.

Förfaringssättet är givetvis tillämpligt på andra ordböcker, under förutsättning att definitionsspråket i den aktuella ordboken är tillräckligt uppmärkt. Ett sidoreultat från studien har redan bidragit till att strama upp brukligheterna i SAOB inför revideringen, då denna studie synliggjort att SAOB-redaktionen haft oklara gränser mellan vad som ska läggas i den rena definitionstexten och vad som bör placeras som en uppgift av mer kronologisk, frekventiell, sociologisk eller regionalpråklig natur, alltså om *bruket*.

Framöver skulle man kunna jämföra med en balanserad korpus och därmed kunna se mer exakt hur svenskan i SAOB skiljer sig från svenskan i stort. Detta kan dels ge rent lexikografiska insikter, dels ge uppslag till vad man kanske bör förändra för att underlätta för läsaren vid revideringen. Det skulle kunna röra sig om en uttolkning av vissa uttrycksätt, eller en förändring eller förenkling av vissa uttrycksätt. Därtill tillkommer det rent pedagogiska syftet. Nya redaktörer får chans att få en överblick av definitionsspråket i SAOB, och kan därmed tillgodogöra sig uttrycksättet i SAOB snabbare.

Litteratur

Ordböcker

SAOB (1889–) = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1–37. Lund, Sverige.

Annan litteratur

Banerjee, Satanjeev & Ted Pedersen (2003): The Design, Implementation, and Use of the n-gram Statistic Package. I: *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, s. 370–381. Mexico City, Mexiko.

Borin, Lars, Markus Forsberg & Johan Roxendal (2012): *Korp – the corpus infrastructure of Språkbanken*. Proceedings of LREC 2012, s. 474–478. Istanbul, Turkiet.

Bäckerud, Erik (2016): Indexeringen av SAOB. I: *Nordiska Studier i Lexikografi* 13, s. 101–109. Köpenhamn, Danmark.

Kilgarrieff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell (2004): The Sketch Engine. I: *Proceeding of Euralex*, s. 105–116. Lorient, Frankrike.

- Larsson, Lennart (2014): En ”mer l. mindre stor” stor ordbok – om variationerna i SAOB:s omfång och ambitionsnivå. I: *LexicoNordica* 21. Tidskrift om leksikografi i Norden, utgitt av Nordisk forening for leksikografi. s. 61–80.
- Nilsson, Pär (2019): *Bildliga betydelser i SAOB: Om beskrivningen av betydelseutvecklingsmekanismer analyserad ur ett kognitivt semantiskt perspektiv*. Licentiatavhandling 2019, 150 s. Lund, Sverige.

Appendix

Hela SAOB

numera i sht i vitter stil i 61
 i vissa språkprov svårt att skilja från 50
 numera i sht i vitter stil om 50
 i vissa fall utan bestämd avgränsning från 47
 sht i vitter stil i fråga om 44
 i sht i vitter stil i fråga 44
 i vissa fall svårt att skilja från 41
 numera i sht i vitter stil till 38
 i sht i vitter stil i utvidgad 33
 i sht i vitter stil i mer 31

Tabell 4: 7gram, Hela SAOB.

Period 1	Period 2
med syftning på innehållet i en hel 8	i sht i vitter stil i fråga 18
med syftning på en efterföljande relativ bisats 8	sht i vitter stil i fråga om 18
syftning på innehållet i en hel sats 8	i sht i vitter stil i mer 17
numera i sht i högre stil i 6	i fråga om det ljud som uppkommer 16
den relativa bisatsen motsvarar en nutida konsekutiv 5	i sht i vitter stil i överförd 15
bisatsen motsvarar en nutida konsekutiv att sats 5	i sht i vitter stil som har 15
hvarvid den relativa bisatsen motsvarar en nutida 5	i sht i vitter stil i utvidgad 13
relativa bisatsen motsvarar en nutida konsekutiv att 5	numera i sht i vitter stil i 13
† i ställning efter det styrda ordet 5	i fråga om det förhållande att ngt 12
närmande sig karaktären af ett gm sin 4	numera i sht i fråga om utländska 12

Tabell 5: 7gram, Period 1 och Period 2.

Period 3	Period 4
numera i sht i vitter stil i 47	i vissa språkprov svårt att skilja från 45
i vissa fall utan bestämd avgränsning från 47	i vissa språkprov utan klar avgränsning från 11
numera i sht i vitter stil om 42	förled i ssgr betecknande ngt som har 8
i vissa fall svårt att skilja från 30	det som betecknas med det styrda ledet 6
numera i sht i vitter stil till 30	om ngt som vittnar om sådan egenskap 6
sht i vitter stil i fråga om 25	som i ngt avseende påminner om en 6
i sht i vitter stil i fråga 25	ssgr betecknande ngt som har avseende på 5
i sht i vitter stil i utvidgad 20	i ssgr betecknande ngt som har avseende 5
i sht i vitter stil om person 19	man får hand om ngt för att 4
med objektet ersatt av bestämning inledd av 18	det som anges med det styrda ledet 4

Tabell 6: 7gram, Period 3 och Period 4.

Period 1	Period 4
hvarvid den relativa bisatsen motsvarar en nutida konsekutiv att sats 5	göra så att man får hand om ngt för att 4
mindre närmande sig karaktären af ett gm sin ställning i 4	tal som i serien av hela tal har sin plats 3
närmande sig karaktären af ett gm sin ställning i satsen 4	om växt vars blad innehåller ett enzym som kan användas 3
sig karaktären af ett gm sin ställning i satsen betonadt 4	utgörande det led som i sitt sammanhang betecknar detta okända 3
så stor mängd af ngt som man kan omfatta med 3	som i serien av hela tal har sin plats mellan 3
af alla de enskilda om hvilka den tillagda bestämningen gäller 3	simmare som gjort de näst bästa resultaten vid proven till 2
bygdemålsfärgad framställning i fråga om vissa af seden föreskrifna föräringar 2	det ljud som uppkommer då luften sätts i rörelse av 2
jordbruk som icke är förenadt med boskaps-skötsel af ngn betydenhet 2	på att tidrymden omfattar den tidpunkt som anges med rektionen 2
i olika färger allt efter som ljuset kommer från ena 2	en plats som är belägen på nästan samma höjdnivå som 2
på det steg som utfyller det kolumnmellanrum på ett ark 2	om sladd som förbinder telefonlur med den övriga delen av 2

Tabell 7: 10gram.

Karin Cavallin
fil. mag., redaktör och språkteknolog
Svenska Akademiens ordbok
Dalbyvägen 3
SE-224 60 Lund
karin.cavallin@svenskaakademien.se

A word frequency dictionary of Icelandic child-directed speech

Hinrik Hafsteinsson & Einar Freyr Sigurðsson

In this paper we present the making of a word frequency dictionary of Icelandic child-directed speech. The material consists of transcribed video recordings of two boys during their language acquisition period and their fathers (and, to a lesser degree, their mothers). The older boy, born in 1982, is the father of the younger one, born in 2011, which means that through the video recordings we get access to the spoken language of three generations within the same family. Data containing child-directed speech is important for research in language acquisition. Both the dictionary and the tools made in the project are freely available online, which hopefully facilitates Icelandic language acquisition research.

1. Introduction¹

We present the making of a word frequency dictionary of Icelandic child-directed speech.² The product itself is twofold. On the one hand, it is a dictionary which contains concrete information on children's input in language acquisition. It is available at www.github.com/hinrikur/BKL and can be used in research on child-directed speech. On the other hand, to be able to make the dictionary, we have developed methods and tools (Python scripts) which are, just like the dictionary, freely available online.

When finished, the dictionary will be the product of two transcribed corpora—the Gunnar Corpus (S.J. Sigurðsson 2019) and the Kalli Corpus (Sigurðsson & Árnadóttir 2019)—of approximately 50 hours in total of video recordings of two children, born 30 years apart, during their language acquisition period. The videos, recorded in most part by the children's fathers, contain conversations between the children and their parents. By focusing on what the children hear (child-directed speech) we gain information on the input in language acquisition and are able to, e.g., assess changes in Icelandic that are currently underway.

¹ We thank two anonymous reviewers and the editors for their comments on the paper.

² The dictionary is a part of the project “Í beinan karllegg: Skráning talmáls þriggja ættliða” (‘Patrilineal Descent: Transcribing Spoken Language of Three Generations’) which was funded by The University of Iceland Research Fund in 2019.

2. Why child-directed speech?

The term *child-directed speech* (or motherese or caregiver speech) is often used when highlighting the fact that parents, for example, talk in a particular way to their young children (such as in infant-directed speech; see discussion in, e.g., de Boer 2012). We use the term to focus on the fact that we are investigating not what the children themselves say, but the input they receive from their parents and others, which may differ in various ways from other types of speech or texts.

Even though children's input in language acquisition has been studied in detail cross-linguistically, few studies exist on the input in the acquisition of Icelandic (see discussion in Nowenstein 2014). However, Icelandic language acquisition is well researched and, as a matter of fact, a word frequency dictionary of Icelandic child language does exist (Einarsdóttir et al. 2019). There are also various Icelandic child language corpora (which also contain child-directed speech), most notably the Ari/Kari Corpus (Strömqvist, Ragnarsdóttir et al. 1995), the Birna Corpus (Gíslason et al. 1983), the Dóra Corpus (Pálsdóttir 1979), the Einarsdóttir Corpus (Einarsdóttir 2018), the Eva Corpus and the Fía Corpus (Sigurjónsdóttir 2000, 2007). Nonetheless there is a clear lack of child-directed speech data.

A variety of changes are currently underway in Icelandic syntax. It is important to study language acquisition to better understand these changes, and child-directed speech, which reflects the input in acquisition, is well suited for this. Some changes in Icelandic are more lexically rooted than others, such as dative substitution, where accusative subjects of experiencer verbs are replaced by dative subjects (e.g., Svavarsdóttir 1982, Jónsson & Eypórsson 2003, Yang 2016 and Nowenstein 2017). Speakers have to memorize, word by word, which experiencer verbs take an accusative subject. Some such verbs, e.g., *dreyma* 'dream', *langa* 'want' and *vanta* 'need', are frequently used in child-directed speech. As other such verbs, e.g., *bresta* 'lack', *hrylla við* 'be horrified by' and *sundla* 'feel dizzy', may be used much less frequently, it is important to explore to what degree they are used. If it turns out that (i) the number of experiencer verbs that take an accusative subject in standard Icelandic is much lower in child-directed speech than the number of experiencer verbs that take a dative subject and (ii) the frequency of many or most of the accusative subject verbs that do occur in child-directed speech is low, then we can better understand the ongoing change of dative substitution.

It should be emphasized that even though a change like dative substitution has been investigated in great detail, linguists have had very limited access to the input of children acquiring Icelandic. We believe that the dictionary, as well as the corpora which it is based on, is a step in the right direction.

Word frequency can potentially also help us understand how dative subjects in passives are acquired. A recent study by Sigurðsson, Nowenstein & Sigurjónsdóttir (2018) suggests that children acquire dative subjects that originate as

indirect objects in passives earlier than dative subjects that originate as direct objects. That is surprising and may indicate that ditransitives, such as *gefa* ‘give’, *selja* ‘sell’ and *senda* ‘send’, are more frequent in the input than dative-taking monotransitives, such as *hjálpa* ‘help’, *hrinda* ‘push’ and *kasta* ‘throw’. This can be explored further, using our word frequency dictionary.

3. The videos

The video recordings feature two boys, Gunnar (b. 1982) and Kalli (b. 2011). The videos were recorded in most part by the children’s fathers and are from the periods 1987 to 1990 (when Gunnar was 5–8 years old) and 2013 to 2018 (when Kalli was 2–7 years old).

Through the transcriptions we have access to the language of three generations within one and the same family, as Gunnar is, in fact, Kalli’s father: On the one hand we have Gunnar (as a child) and his father and, on the other hand, Gunnar (as a grown-up) and his son, Kalli. This adds an interesting diachronic dimension to the project. Also, this opens a window into Kalli’s and Gunnar’s language acquisition. We observe Gunnar as a grown-up serving as the input of Kalli and Gunnar’s father as the input of young Gunnar during the boys’ language acquisition. It should be noted that the recordings also feature the boys’ mothers talking to them even though the fathers conducted the vast majority of the recordings. The dictionary therefore also includes the mothers’ speech.

The goal (pending on more funding) is to transcribe 50 hours—10 from Gunnar’s acquisition period and 40 from Kalli’s acquisition period. We have currently transcribed approximately 6 hours of video recordings.

4. The making of the dictionary

4.1. Building the corpora

The corpora consist of video recordings transcribed using the multimedia annotation tool ELAN (version 5.4). This tool enables detailed text annotation for video and audio files and is used extensively in the fields of conversation and discourse analysis (see, e.g., Brugman & Russel 2004). For the purposes of this project we focused on the basic text annotation functionality in ELAN, where each annotation contains roughly a single utterance by a speaker, and each speaker has her own specific tier in the annotated data. We also opted for a simple transcription scheme, where notation of detailed discourse features is omitted. A screenshot displaying this is shown in Figure 1.

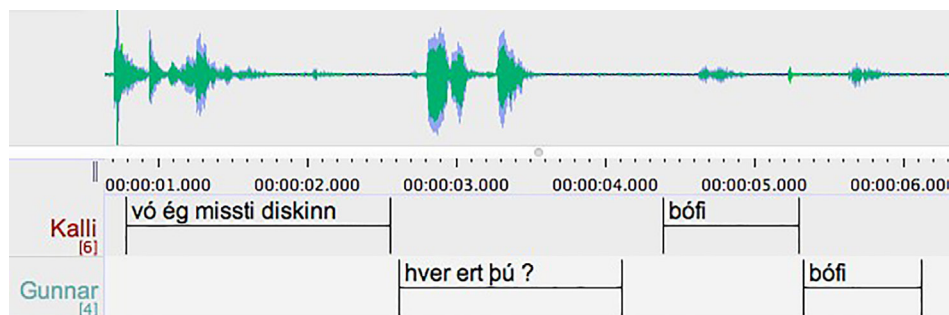


Figure 1: Screenshot from ELAN.

Even though Figure 1 shows only a part of the features ELAN offers, it gives an idea of how the program works. At the top we have a waveform window that displays the sound in the recording at a given time. At the bottom, the speaker-specific tiers are visible, one for each speaker in the recording. At the beginning of this recording (from 2016), Kalli drops a CD he is looking at on the floor and says: *Vó ég missti diskinn* ‘Whoa, I dropped the disc’. In the accompanying video recording—which can be played through ELAN in real time during the transcription—Kalli is dressed up with something covering his head and his father, Gunnar, asks him: *Hver ert þú?* ‘Who are you?’ which Kalli replies to: *Bófi* ‘A thug’. Gunnar then repeats Kalli’s reply. In this recording, Kalli’s and Gunnar’s speech does not overlap, i.e., they do not speak at the same time, which the tier-specific annotations clearly capture.

As the goal is to build a dictionary of child-directed speech, we need young Gunnar’s and Kalli’s input. The transcriptions in ELAN are saved as ELAN Annotation Format (EAF) files, a variant of the XML file schema. When working with these transcriptions we use the Pypmi package (Lubbers & Torreira 2018) for Python (and additional scripts written in the project) to extract all utterances made by the adult speakers. This is relatively straightforward, due to the speaker-specific tiers in the annotation data. By counting only the words uttered by young Gunnar’s and Kalli’s parents, we get a total number of 11,927 words of input in the current data, with 1,910 words uttered by young Gunnar’s parents and 10,017 by Kalli’s parents.³

Once we have extracted the texts from the boys’ parents from the ELAN-transcriptions, they need to be processed further before word frequency can be calculated. In our approach we need the lemmas of the words, i.e., the morphologically

³ This difference in size between the datasets is expected, as the length of transcribed Kalli recordings is currently greater than that of transcribed Gunnar recordings. As mentioned in Section 3, the total length of the recordings of Kalli is much greater than of Gunnar (40 vs. 10 hours), so this relative difference can be expected to remain throughout the project as the data is expanded upon with further transcriptions.

neutral form of each word. To achieve this, the text is first sent to an automatic Part-of-Speech (PoS) tagger. As Icelandic is a language with rich inflection, this greatly eases the lemmatization process, especially regarding morphological ambiguity between word forms. Several freely available PoS taggers exist that can tag Icelandic texts with good results, such as IceTagger (Loftsson 2008), IceStagger (Loftsson & Östling 2013) and, most recently, ABLtagger (Steingrímsson et al. 2019), a neural network-based tagger. We opted for using ABLtagger, as it has been reported to give the best accuracy in tagging Icelandic texts. For the lemmatization itself, we used Nefnir (Daðason 2018), which has been shown to be a very reliable lemmatizer, given that the source text is accurately PoS tagged (Ingólfssdóttir et al. 2019).

To illustrate the output of the lemmatization step, Table 1 shows the PoS tagged and lemmatized output for the sentence in (1):

- (1) langar þig að halda á þessu
 want you to hold on this
 ‘Do you want to carry this?’

langar	sfg3en	langa
þig	fp2eo	þú
að	cn	að
halda	sng	halda
á	aþ	á
þessu	faheþ	þessi

Table 1: PoS tagged and lemmatized text.

In Table 1 we see three columns. The first one shows the original text, the second column shows the PoS tags and the column on the right shows the lemmas. For example, *langar* ‘want’ is the first word in the sentence and it is tagged as “sfg3en” which stands for verb-indicative-active-third person-singular-present tense. Its lemma is *langa* (the infinitival form of the verb).

4.2. Extracting word frequency data

With the 11,927 lemmatized words at hand, we counted the occurrences of each lemma using a simple Python script,⁴ giving us detailed information on the frequency of each word in the dataset. There are various factors that need to be

⁴ As previously noted, all scripts written in this project in order to make the dictionary are available online: www.github.com/hinrikur/BKL.

taken into account in this step. There might, for example, be inaccuracies in the tagging and lemmatization step which have an effect on the end word frequency values. We may also assume that there will be errors in the original text itself, as the annotations have not been proofread in any fashion yet and will inevitably contain some typographical errors. The interactions of such errors in each step of the process may also produce wrong lemmas that skew the final word frequency values. These minor issues must eventually be manually checked and corrected, but leaving them aside for now, we can discuss the word frequency data itself.

The objective in creating the word frequency dictionary is to provide data on the frequency of different words in the input Kalli and Gunnar receive, and thus much-needed information on child-directed speech as a whole. Besides this main objective, the dictionary will enable comparison of the input of two Icelandic children roughly 30 years apart from each other, coupled with the children being father and son. As of now the dictionary achieves both of these objectives to an extent. Due to the relatively small size of the current dataset, especially for Gunnar's input, we will not make any decisive claims based on this data yet. We can, however, make broad observations regarding the content of the boys' input. For example, (2) and (3) below show the ten most frequent verbs in Gunnar's and Kalli's input, respectively—starting with the most frequent verb, which is *vera* in both cases.

(2) vera, segja, fara, koma, sjá, ætla, vilja, eiga, gera, taka
 be say go come see intend want own do take

(3) vera, fara, gera, koma, eiga, ætla, halda, geta, verða, segja
 be go do come own intend think/hold can become/must say

It is apparent from these excerpts from the dictionary that the most frequent verbs in the two inputs are relatively similar, despite the difference in size of the two datasets.

If these datasets indeed reflect the content of each boy's language input, we would expect them to be broadly compatible with more general data on Icelandic. This seems to be the case when we compare the excerpts in (2) and (3) to the one in (4), which shows the ten most frequent verbs in *Íslensk orðiðnibók* ('Icelandic Frequency Dictionary', IFD; Pind et al. 1991:612), which is based on a corpus of roughly 590,000 word tokens.

(4) vera, hafa, koma, verða, segja, fara, geta, taka, eiga, gera
 be have come become/must say go can take own do

From these examples we can infer that despite the relatively small size of our datasets, and even though the text type we are working with—child-directed speech—is different from the text types that the IFD consists of, our word frequency dictionary does describe its source data in an expected, systematic fashion, just like dictionaries based on larger corpora. Therefore, we are confident that our work could be expanded upon, given a larger dataset and furthermore that our production methods produce applicable results.

5. By linguists, for linguists

The dictionary is a part of the project “Í beinan karllegg: Skráning talmáls þriggja ættliða” (‘Patrilineal Descent: Transcribing Spoken Language of Three Generations’), in which the focus is language acquisition. Our hope is that the dictionary will be immediately useful in research on Icelandic language acquisition—that is, it is aimed at linguists.

The second author of this paper (alphabetically ordered), who was awarded the grant, is a linguist whose theoretical interests include language acquisition. Furthermore, the first author holds a B.A. degree in general linguistics and is currently an M.A. student in language technology. The dictionary is therefore not only made for linguists doing research, but also made by linguists.

In addition to this we hope that the dictionary, the scripts written in the project and the transcripts, which we aim at making freely available, will become useful in various language technology projects.

6. Current status and future directions

We now have a corpus of 22,698 words, of which almost 12,000 are child-directed speech. Much more transcribing is needed, however, as we have only transcribed approximately 6 out of 50 hours of video recordings. The word frequency dictionary (the most recent version at each time) and the scripts written specifically in order to make the word frequency dictionary are available online, at www.github.com/hinrikur/BKL. We also aim at making all the transcriptions available online, at www.github.com/einarfs.

The Gunnar Corpus and the Kalli Corpus provide new Icelandic acquisition data, including child-directed speech. In the near future we also wish to gather more recordings, old and new, from more speakers. Hopefully, this is just the start.

References

- de Boer, Bart (2012): Infant-directed speech and language evolution. In: Maggie Tallerman & Kathleen R. Gibson (eds.): *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press, 322–327.
- Brugman, Hennie & Albert Russel (2004): Annotating Multimedia / Multi-modal resources with ELAN. In: *Proceedings of LREC 2004*.
- Daðason, Jón Friðrik (2018): Nefnir. <www.github.com/jonfd/nefnir>.
- Einarsdóttir, Jóhanna Thelma (2018): *Gagnabanki Jóhönnu Einarsdóttur um málsýni* (GJEUM). [Einarsdóttir Corpus.]
- Einarsdóttir, Jóhanna Thelma, Anna Lísa Pétursdóttir & Íris Dögg Rúnarsdóttir (2019): *Tíðni orða í tali barna*. Reykjavík: Háskólaútgáfan.
- ELAN (Version 5.4) [Computer software] (2018): Nijmegen: Max Planck Institute for Psycholinguistics. <tla.mpi.nl/tools/tla-tools/elan>.
- Gíslason, Indriði, Randa Mulford & Ásgeir S. Björnsson (1983): Upp vek þú málið mitt. In: Sigurjón Björnsson (ed.): *Athöfn og orð. Afmælisrit helgað Matthíasi Jónassyni áttæðum*. Reykjavík: Mál og menning, 107–114.
- Ingólfssdóttir, Svanhvít Lilja, Hrafn Loftsson, Jón Friðrik Daðason & Kristín Bjarnadóttir (2019): Nefnir: A high accuracy lemmatizer for Icelandic. In: *Proceedings of the 22nd Nordic Conference on Computational Linguistics*. Linköping University Electronic Press, 310–315.
- Jónsson, Jóhannes Gísli & Þórhallur Eypórsson (2003): Breytingar á frumlagsfalli í íslensku. In: *Íslenskt mál 25*, 7–40.
- Loftsson, Hrafn (2008): Tagging Icelandic text: A linguistic rule-based approach. In: *Nordic Journal of Linguistics* 31, 47–72.
- Loftsson, Hrafn & Robert Östling (2013): Tagging a morphologically complex language using an averaged perceptron tagger: The case of Icelandic. In: *Proceedings of NODALIDA 2013*. Linköping University Electronic Press, 105–119.
- Lubbers, Mart & Francisco Torreira (2018): *pypmi-ling: a Python module for processing ELAN's EAF and Praat's TextGrid annotation files*. (Version 1.69.) Retrieved from: <pypi.python.org/pypi/pypmi-ling>.
- Nowenstein, Iris Edda (2014): *Tilbrigði í frumlagsfalli á máltökuskeiði. Þágufallshneið og innri breytileiki*. M.A. thesis, University of Iceland.
- Nowenstein, Iris Edda (2017): Determining the nature of intra-speaker subject case variation. In: Höskuldur Thráinsson, Caroline Heycock, Hjalmar P. Petersen & Zakaris Svabo Hansen (eds.): *Syntactic variation in Insular Scandinavian*. Amsterdam: John Benjamins, 91–112.
- Pálsdóttir, Margrét (1979): *Máltaka barns*. B.Ed. thesis, Kennaraháskóli Íslands.
- Pind, Jörgen (ed.), Friðrik Magnússon & Stefán Briem (1991): *Íslensk orðtíðnibók*. Reykjavík: Orðabók Háskólans.

- Sigurðsson, Einar Freyr & Hlíf Árnadóttir (2019): *The Kalli Corpus*.
- Sigurðsson, Einar Freyr, Iris Edda Nowenstein & Sigríður Sigurjónsdóttir (2018): *The acquisition of dative subjects in L1 Icelandic*. Talk given at GALANA 8, Bloomington, September 28th, 2018.
- Sigurðsson, Sigurður Júlíus (2019): *The Gunnar Corpus*.
- Sigurjónsdóttir, Sigríður (2000): *The Eva Corpus*.
- Sigurjónsdóttir, Sigríður (2007): *The Fía Corpus*.
- Steingrímsson, Steinþór, Örvar Kárason & Hrafn Loftsson (2019): Augmenting a BiLSTM tagger with a Morphological Lexicon and a Lexical Category Identification Step. In: *Proceedings of Recent Advances in Natural Language Processing*, 1161–1168.
- Strömqvist, Sven, Hrafnhildur Ragnarsdóttir et al. (1995): The Inter-Nordic Study of Language Acquisition. In: *Nordic Journal of Linguistics* 18, 3–29.
- Svavarsdóttir, Ásta. (1982): „Þágufallssýki.“ Breytingar á fallnotkun í frumlagssæti ópersónulegra setninga. In: *Íslenskt mál* 4, 19–62.
- Yang, Charles (2016): *The Price of Linguistic Productivity. How Children Learn to Break the Rules of Language*. Cambridge, MA: The MIT Press.

Hinrik Hafsteinsson
 M.A. student
 University of Iceland
 Sæmundargata 2
 IS-102 Reykjavík
 hih43@hi.is

Einar Freyr Sigurðsson
 research lecturer, Ph.D.
 Árni Magnússon Institute for Icelandic Studies
 Laugavegur 13
 IS-101 Reykjavík
 einar.freyr.sigurdsson@arnastofnun.is

Hvordan har brugen af mobiltelefoner indvirket på efterspørgslen efter netordlister?

Poul Hansen

Developers of online dictionaries need to understand how searches on mobile phones differ from searches on desktops. In this longitudinal study, the shift from desktop to mobile searches on three different dictionaries is presented. Results show that mobile applications may produce a new pattern of dictionary usage. Differences related to activity volumes and the speciality of the dictionaries were also found.

1. Indledning

Mobil internettrafik er i dag et område, hvor udviklingen sker meget hurtigt takket være store fremskridt inden for teknologi og softwareapplikationer. Statistik over brugen af henholdsvis mobilenheder og computere i perioden 2016-2019 giver os interessante data om den igangværende proces. På globalt plan overstiger antallet af søgninger på internettet fra mobilenheder nu trafikken fra computere (Enge 2019). I store dele af verden er brugstiden for mobilenheder nu ved at passere den tid, man bruger på at se TV (Rodriguez 2018).

Stillet over for disse fakta er der grund til at se nærmere på de eventuelle følgevirkninger for dem, som udbyder ordbogsressurser på nettet. Der er naturligvis en række spørgsmål, som en ordbogsudvikler gerne vil have svar på. Hvilken betydning har den kraftige vækst i søgninger via mobiltelefoner? Er det i det hele taget noget, der får konsekvenser for de ordbøger, vi har på nettet? I hvilken udstrækning påvirker det i så fald netordbøgerne? Kan vi regne med øget eller mindsket trafik? Er der noget, som vi bør lægge særlig vægt på for at give brugerne af mobiltelefoner en bedre oplevelse af ordbøgerne – såvel hvad angår struktur, hastighed som det visuelle indtryk? Det er vanskelige spørgsmål, som er svære at svare entydigt på. Det, vi umiddelbart kan gøre for at blive klogere og få større indsigt i, hvad der egentlig er ved at ske, er at dykke ned i de data, vi har om aktiviteten på nettet og analysere forløbet i de resulterende tidskurver med henblik på at identificere eventuelle afvigelser eller trender i udviklingen.

I denne artikel præsenteres fund af forandringstendenser i tre udvalgte ordliste-ressurser i en 13-års periode fra 2006 og frem til i dag (2019), hvor ordsøgningen i overvejende grad sker fra mobiltelefoner. Målet er at identificere aspekter, der er vigtige for udformning af strategier i arbejdet med at udvikle ordbøger på nettet.

Undersøgelsen forsøger for det første at besvare følgende spørgsmål:

- I hvilken grad påvirkes aktiviteten volumenmæssigt af, at flere brugere søger fra mobiltelefoner?
- Er der forskelle med hensyn til ordlisters fagområder?
- Er der tale om konsistente og karakteristiske forandringsmønstre?

For det andet vil artiklen sætte resultaterne i relation til den leksikografiske funktionsteori (præsenteret i bl.a. Fuertes-Olivera & Tarp 2014) og diskutere deduktiv og induktiv teoriudvikling.

2. Data

De tre analyserede ordlister indgår i et større materiale og er udviklet af artikelforfatteren bl.a. med støtte fra Nordplus. Ordlistematerialet er blevet beskrevet i et tidligere arbejde (Hansen 2016). Det foreligger i dansk-svensk version på en svensk hjemmeside (www.danska-svenska.se) og i en svensk-dansk version på en dansk hjemmeside (www.dansk-og-svensk.dk). Artiklen er baseret på, at man skal kunne finde ordlisterne, når man indtaster søgeord i browserens søgefelt. I ordlisterne er der medtaget opslagsord, som relativt sjældent er blevet oversat mellem dansk og svensk i andre kilder, og de findes somme tider kun i disse ordlisteresurser, hvilket giver en relativ høj placering i søgeresultaterne. Ordlisterne er desuden søgemaskineoptimerede hver for sig (nøgleord: svensk + dansk + søgeord).

Svensk-dansk bandeordliste omfatter ca. 200 lemmer, som er sorteret og arrangeret således, at de danner tematisk matchende grupper mellem dansk og svensk. Materialet er skabt med henblik på at være støttemateriale ved tolkning og oversættelse mellem dansk og svensk. Bandeordene er indsamlet ved ”wild-card”-søgninger i chatrooms og lignende fora på Google samt ved gennemgang af relevante artikler i nyhedsmedier og tidsskrifter. Ordlisten er placeret på den danske webside (dansk-og-svensk.dk). I ordlisten indgår også supplerende excerpter fra Ljung (1984), Kotsinas (1998), Gänger (2003) og Doggelito & Kotsinas (2004).

Svensk-dansk socialordliste omfatter ca. 750 lemmer og er en grundordliste, der hovedsageligt er udarbejdet på basis af myndighedstekster, sagsbehandlinger, spørgeskemaer og excerpteret materiale i forbindelse med et større oversættelsesprojekt inden for socialvidenskab. Ordlisten er placeret på den danske webside (dansk-og-svensk.dk). I ordlisten indgår også excerpter fra Kolding & Reintoft (1992).

Dansk-svensk hverdagsordliste omfatter ca. 7.000 lemmer og er en fortegnelse bestående af en grundordliste, som oprindeligt er konstrueret på basis af

bl.a. fortegnelsen i *Svensken i Danmark* (Malmström 1975) med indhentet tilladelse fra udgiveren, samt ordudvalget i ældre fortegnelser, fx over turistord. Materialet er herudover suppleret med nyere ord, bl.a. ord noteret i det daglige oversættelsesarbejde samt analyser af logfiler med hyppige søgeord og indkomne spørgsmål fra hjemmesidens brugere. Ordlisten er placeret på den svenske web-side (danska-svenska.se).

3. Måling af ordbogbrugernes aktivitet

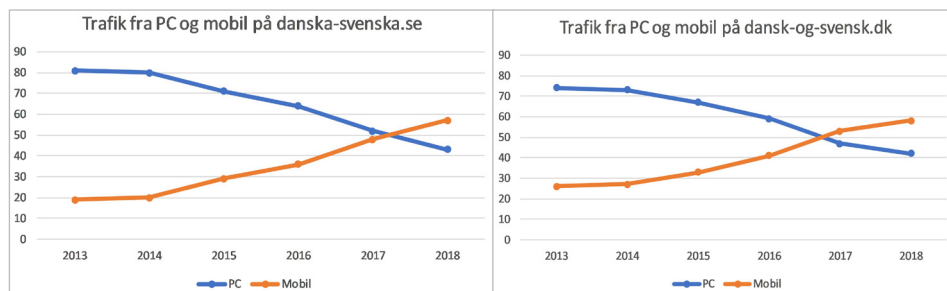
De analyserede data er hentet fra operatøren one.com, som står for webhosting af de to websider, hvor det analyserede ordbogsmateriale ligger. Data består af de registrerede månedlige aktivitetsdata for de tre ordlister i en 13-års periode fra og med januar 2006 til og med oktober 2018 (for visse dog kortere observationstid pga. senere startår). Til analysen valgtes parameteren antal eksponeringer (antal gange, der er blevet vist et link til hjemmesiden på skærmen i et søgeresultat) ud fra det synspunkt, at denne parameter numerisk ligger meget højere (op til omkring 6.000 om måneden) end antal klik. Tallene for antal klik er betydeligt lavere og giver ofte et så sparsomt materiale, at egentlige statistiske analyser og påvisning af signifikante sammenhænge bliver langt mere usikker. Trafikken til websiderne er hovedsageligt foregået ved at brugeren skriver ord i Googles søgefelt, og er derfor stærkt koblet til ordsøgningsaktiviteter.

Datamaterialet består af webhosterens egen bearbejdning af de såkaldte rå logfiler i form af de sammenstillinger af trafikdata, som webhosteren gør tilgængelig for alle kunder, der har hjemmesider hos dem. Hermed inkluderes også aktivitetsdata fra andre browsere og adgangsveje end fx Google.

4. Resultater

4.1. Udviklingen i trafik fra PC og mobiltelefoner

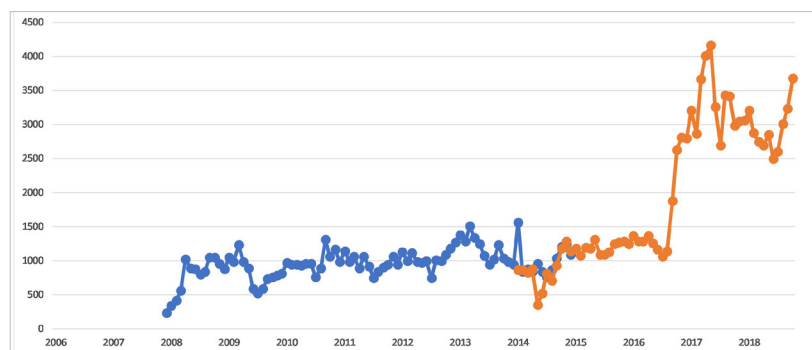
I perioden 2013-2018 er der sket et markant skift i aktiviteten fra computere og mobiltelefoner, både på den danske og den svenske webside (figur 1). Den overvejende del af den registrerede aktivitet stammer nu fra mobiltelefoner på begge websider. Det fremgår, at trenden er meget tydelig og konsekvent. Skiftet på den danske side skete i efteråret 2016 og nogle måneder senere, i foråret 2017, skete skiftet på den svenske side. Aktiviteten fra mobiltelefoner viser ringe tegn på stagnation i væksten eller tilbagegang til det tidligere forhold. Alt tyder på en klar irreversibel udviklingstendens.



Figur 1: Årligt antal visninger på mobiltelefoner og computere i procent.

4.2. Udviklingen i aktiviteten på de enkelte ordlister

Udviklingen for Svensk-dansk bandeordliste (figur 2) er frem til midten af 2016 karakteriseret af dels årstidsvariation, dels en lineær langtidstrend, som har været relativt konstant fra starten af 2008. Efter midten af 2016 sker der imidlertid en kraftig vækst i aktiviteten. Dette indtræffer samtidig med at trafikken på websiden fra mobiltelefoner overstiger trafikken fra computere (se figur 1). Bemærk, at antal visninger er gengivet med blå og orange linje i figur 2-4. Den blå linje repræsenterer websideoperatørens statistikrapportering frem til 2015, og den orange linje angiver data fra et nyt og mere avanceret beregnings- og rapporteringssystem, som blev taget i brug i løbet af 2014.

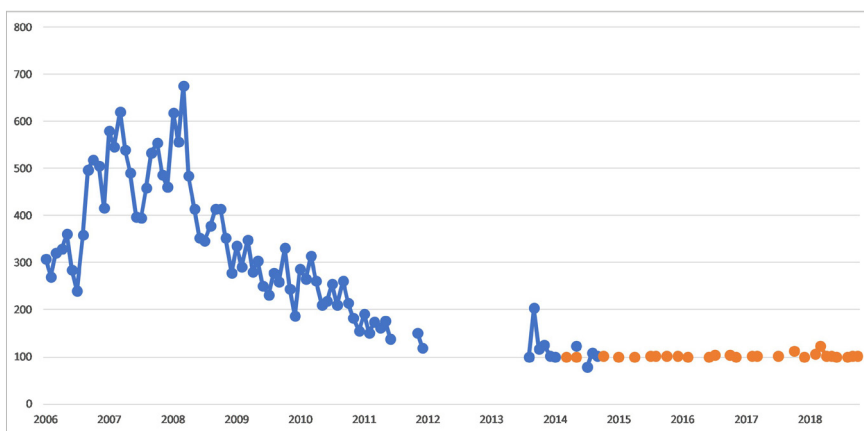


Figur 2: Antal visninger for Svensk-dansk bandeordliste på den danske webside.

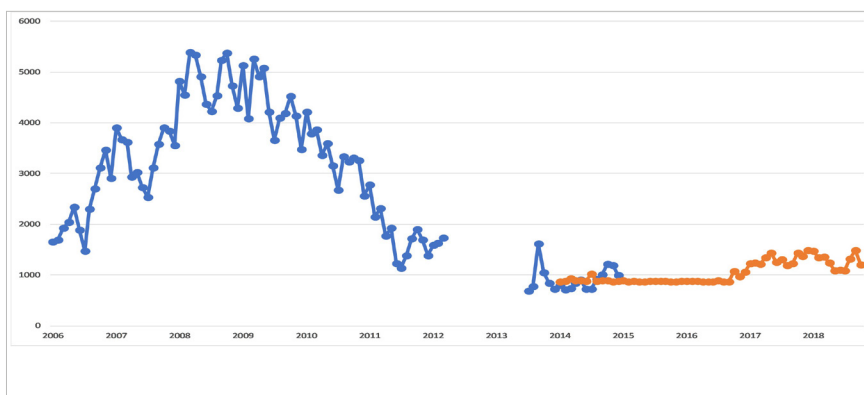
Svensk-dansk socialordbog 2006-2018 (figur 3) er karakteriseret af en relativt høj aktivitet mellem 2006 og 2008. Efter 2008 har der imidlertid været en langsom og konstant aftagende aktivitet. I den sidste del af registreringsperioden foreligger der sandsynligvis meget lav eller ingen trafik fra personer, eftersom der mangler tydelig sæsonvariation. Den aktivitet, som trods alt er registreret på

lavt niveau, skyldes formodentlig de regelmæssigt afsøgende (dvs. ikke-sæsonvarierende) søgeroboter. Trenden fortsætter helt frem til de sidste registreringer i 2018. Bemærk, at der både i figur 3 og figur 4 ses et 13 måneders afbrud i måleserierne i perioden 2012 til 2013. Dette skyldes et alvorligt havari hos operatøren, hvor data ikke kunne genskabes.

Dansk-svensk hverdagsordliste (figur 4) er en af de ordlister, der har været mest populær på den svenske webside med op til 6.000 visninger pr. måned. Aktiviteten kulminerede mellem 2008 og 2009 og er derefter aftaget kraftigt. Formodentligt er aktiviteten 2011-2016 hovedsageligt forårsaget af søgeroboter. I efteråret 2016 ses imidlertid en periode med højere aktivitet, som tydeligvis er sæsonbetonet og derfor indikerer bevidst aktivitet fra personer.



Figur 3: Antal visninger for Svensk-dansk socialordbog på den svenske webside.



Figur 4: Antal visninger for Dansk-svensk hverdagsordliste på den svenske webside.

5. Diskussion

Undersøgelsen har vist forandringer i aktiviteten i bandeordlisten (figur 2) og hverdagsordlisten (figur 4). Dette er øjensynligt sammenfaldende med det tidspunkt, hvor aktiviteten fra mobiltelefoner overstiger aktiviteten fra computere på de websider, hvor ordlisterne findes. Der kan også konstateres tegn på, at de observerede forandringer efter årsskiftet 2017 skaber helt nye mønstre i forhold til tidligere. Alt taget i betragtning tyder dette på, at den forholdsvis større anvendelse af mobiltelefoner har medført ny adfærd med hensyn til ordsøgning i netordlister.

5.1. Volumenmæssige aktivitetsforandringer

Med hensyn til undersøgelsens første spørgsmål om eventuelle volumenmæssige forandringer kan det konstateres, at der er tegn på forøget aktivitet. Især Svensk-dansk bandeordliste (figur 2) har fået en kraftigt forhøjet aktivitet, men samme vækstmønster kan i nogen grad også ses for Dansk-svensk hverdagsordliste (figur 4).

Det er bemærkelsesværdigt, at væksten i aktiviteten for fx Svensk-dansk bandeordliste har været så kraftig og slet ikke står i proportion til den gradvise forandring i det indbyrdes forhold mellem trafikken fra computere og mobiltelefoner. At bandeordlisten har fået en så kraftig vækst kan skyldes, at der er opstået en pludselig interesse, som har forårsaget en gennembrudseffekt på mobilinterfacet. En bidragende årsag til et større spredningspotentiale kan bl.a. være, at relevant indhold i ordlisten fremtræder tydeligt allerede ved visningen af linket til resursen, dvs. inden brugeren klikker aktivt, og at ordlisten med sin enkle struktur og overskuelighed har en god visning i et lille display og måske derfor tiltaler mobilbrugere mere end andre alternative resurser.

5.2. Forandringer med hensyn til fagområder

Undersøgelsen indikerer, at der er opstået forskelle efter årsskiftet 2017 med hensyn til, hvilket fagområde ordlisterne dækker. Der har været betydeligt kraftigere aktivitetsvækst for Svensk-dansk bandeordliste end for Dansk-svensk hverdagsordliste i forhold til tidsforløbet før 2017. Dette tyder på, at bandeord har fået en forholdsvis større popularitet i mobilinterfacet end hverdagsordene. En mulig forklaring på dette kan måske søges i, at bandeordlisten er relativt unik på nettet som bilingval ordliste mellem dansk og svensk. For hverdagsord er der langt flere (og bedre) alternativer og valgmuligheder.

Socialordlisten danner kontrast til de to andre ved, at den ikke havde nogen signifikant vækst efter årsskiftet 2017, hvilket taler imod forekomsten af en generel aktivitetsvækst på tværs af alle fagområder. Der er snarere tale om en specifik vækst for hvert fagområde.

5.3. Konsistente og karakteristiske forandringer over tid?

Flere faser kan identificeres i måleserierne, fx opstartsfasen, stagnationsfasen, nedgangsfasen og stilsandsfasen. På detaljeret niveau er der en del tilsyneladende vilkårlig variation (afvigelser fra normen) fra måned til måned. Denne vilkårlighed i forløbene virker fremtrædende, og det er nærliggende at tolke det som en usikkerhed med hensyn til de konklusioner, der kan drages af tidsforløbene. Men her er det meget vigtigt, at man ikke bedømmer graden af vilkårlighed med øjet alene. Samvirkende baggrundsfaktorer kan give mønstre, der opfattes som vilkårlige. I et studie, som netop er publiceret (Hansen 2019), er det blevet vist, at en kombineret model af årstidsvariation og langtidstrend er i stand til at forklare op til 90 % af den tidsvariation, der ses for socialordlisten i perioden 2006-2012 i denne undersøgelse (figur 3). Her er vilkårligheden (uforklarede afvigelser i forløbet) altså betydeligt lavere end det, vi umiddelbart ser med øjet.

5.4. Mulige fejlkilder og bias

Ved tolkning af resultatet er det vigtigt at overveje, i hvilken grad interne aktiviteter og ændringer på hjemmesiderne har påvirket aktivitetsmønstrene, herunder at det er artikelforfatteren selv, som har forfattet og udgivet ordlisterne. En egeninteresse kan ikke afvises, men projektet er drevet fra et neutralt standpunkt med et *objektivt sigte*. Materialet ligger åbent på nettet og er kontrollerbart, og resultaterne er mulige at verificere/falsificere hos andre aktører med lignende data. Med hensyn til aktiviteter har der siden 2011 været meget begrænsede ændringer og tillæg, både i materialet og på de to hjemmesider, som rummer materialet. Det skal også bemærkes, at ordlisterne blev udviklet ikke-kommercielt inden for rammen af et godkendt projekt med støtte fra Nordplus. I henhold til planen blev materialet indledningsvist tilgængeliggjort på artikelforfatterens hjemmeside. Indtil nu har det ikke været muligt at få det publiceret eller tilgængeliggjort på anden måde.

5.5. Mulighed for empirisk-induktiv teoriudvikling?

Det er åbenbart, at måleserierne i denne undersøgelse beskriver en dynamisk udvikling, tilsyneladende pga. bagvedliggende faktorer, som forandres succesivt. Det minder i høj grad om de successionsforløb, som er kendt fra økologiske systemer, hvor faktorer som klima, konkurrence og nishedannelse samvirker og ændres i et tidsperspektiv, dvs. en type af videnskabsområde, der præges af induktiv teoriudvikling snarere end deduktiv teoriudvikling. Inden for leksikografien, især funktionsteorien, har man et socialvidenskabeligt videnskabsteoretisk perspektiv og en slagside til deduktiv tænkning. Man forholder sig derfor gerne kritisk til naturvidenskabens induktive metode. De lange måleserier og relativt store datamængder i nærværende undersøgelse gør det imidlertid hensigtsmæssigt og rimeligt at arbejde med et naturvidenskabeligt induktivt ideal og tage udgangspunkt i deskriptive analyser af de observerede data. Desuden har den foreliggende type af data en kvalitet og kvantitet, der giver de grundlæggende forudsætninger for videre brug af statistiske værktøjer og modeller til at estimere, vurdere og forstå bagvedliggende faktorerens virkninger og interaktioner, fx sæsonvariationens og nye søgeplatformes effekt. Det er sandsynligt, at man via dette kan få ny indsigt i konsekvenser og mønstre.

Empirisk-induktive undersøgelser af tidsserier har i hvert fald et potentiale til med tiden at kunne give vigtige bidrag til leksikografisk forskning og stille krav om ændring og videreudvikling af den socialvidenskabeligt funderede leksikografiske funktionsteori.

6. Sammenfatning

For ordbogsudviklere er det vigtigt at forstå, hvordan ordbøgerne virker og fungerer på nettet set i forhold til brugernes behov for information. Denne undersøgelse viser, at udgivere af netordbøger nu for alvor bør tænke på at tilpasse indholdet og funktionen til brugere af mobiltelefoner. Der blev fundet tegn på en ny brugeradfærd i forhold til tidligere mønstre for specifikke fagområder, og det er en forandring, der bør udforskes nærmere. Udfordringen er først og fremmest at forbedre ordbogsressurserne for at imødekomme de tilsyneladende nye brugerbehov. Men der kan også forudses en større interesse for at vurdere og evt. prognosticere udviklingen på lang sigt via indsamlede data.

Litteratur

Ordbøger

- Doggelito, Dogge & Ulla-Britt Kotsinas (2004): *Förortsslang*. Stockholm: Norstedts ordbok.
- Gänger, Hasse (2003): *Hajaru klyket? Stockholmsk-svensk, svensk-stockholmsk*. Sportförlaget i Europa.
- Kolding, Hans-Egon & Hanne Reintoft (1992): *Social ABC*. Aktuelle Ordbøger.
- Kotsinas, Ulla-Britt (1998): *Norstedts svenska slangordbok*. Stockholm: Norstedts ordbok.
- Ljung, Magnus (1984): *Om svordomar i svenskan, engelskan och arton andra språk*. Akademilitteratur.
- Malmström, Sten (1975): *Svensken i Danmark*. Sten Malmström/Föreningarna Nordens Förbund.

Anden litteratur

- Enge, Eric (2019): *Where is the Mobile vs. Desktop Story Going?* <www.perficientdigital.com/insights/our-research/mobile-vs-desktop-usage-study> (juni 2019).
- Fuertes-Olivera, Pedro & Sven Tarp (2014): *Theory and Practice of Specialised Online Dictionaries*. Berlin: Walter de Gruyter.
- Hansen, Poul (2016): Hvordan har netbrugernes præference for forskellige emnespecifikke ordbøger ændret sig i perioden 2006-2014? I: Asgerd Gudiksen & Henrik Hovmark (red.): *Nordiske studier i leksikografi* 13. København: Nordisk forening for leksikografi, 131-141.
- Hansen, Poul (2019): Statistisk modellering og prognostisering af efterspørgslen efter netordlister. I: *LexicoNordica* 26.
- Rodriguez, Ashley (2018): *The internet is finally going to be bigger than TV worldwide* <www.qz.com/1303375/internet-usage-will-finally-surpass-tv-in-2019-zenith-predicts> (juni 2019).

Poul Hansen
 fil. dr., translør
 Box 4158
 SE-227 22 Lund
 poul@danska-svenska.se

Jysk Ordbog – af hvem, til hvem og hvorfor?

Inger Schoonderbeek Hansen

The focus of this paper is on *Jysk Ordbog*, founded 1932 by professor Peter Skautrup, covering the western Danish dialects (spoken on the peninsula of Jutland), published on the website www.jyskordbog.dk. Due to the strong standardization of Danish since the beginning of the 20th century, the dictionary presents historic versions of the dialects that are hard to recognise by users nowadays. Two main questions arise: 1) who are the supposed user groups, and 2) how do the editors work on bridging the gap between the dictionary's user groups and the historical topic?

1. Jysk Ordbog anno 2020

Det primære formål med denne artikel er at give en status over *Jysk Ordbog*. Ordbogen udarbejdes på Peter Skautrup Centret for Jysk Dialektforskning på Aarhus Universitet; det var professor Peter Skautrup der initierede dataindsamlingen i 1932 ved det nyoprettede universitet. *Jysk Ordbog* var tiltænkt at skulle erstatte H. F. Feilbergs imponerende men (som Feilberg medgiver i titlen) stadig mangelfulde *Bidrag til en ordbog over jyske almuesmål* (1896-1914). Publikationen af *Jysk Ordbog* startede med udgivelse af fire prøvehæfter, *a-ankel*, i perioden 1970-1979. Trykning af ordbogen blev imidlertid vurderet som økonomisk uoverkommelig, ikke mindst på grund af de påtænkte dialektgeografiske kort og illustrationer. I mellemtiden er redaktionsreglerne i flere omgange blevet forenklet, og siden 2000 bliver *Jysk Ordbog* publiceret online (www.jyskordbog.dk/jyskordbog/ordbog.html). Siden 2017 foregår redaktionen i værktøjet iLEX, hvorfra ordbogen opdateres med nye artikler, inkl. kortmateriale, hver uge. I dag består staben af tre fastansatte redaktører (hvoraf én bl.a. også underviser), én sekretær og et antal frivillige, der bl.a. hjælper med at digitalisere nogle af vores data.

I nedenstående redegørelse står to spørgsmål centralt: 1) hvem er ordbogens tilsigtede brugere, og 2) hvordan arbejder redaktørerne med at bygge bro mellem ordbogens brugergrupper og dens historiske emne? Artiklen kommer først og fremmest ind på hvem der redigerer *Jysk Ordbog*, og hvilke data der ligger til grund for den. Dernæst belyses hvem *Jysk Ordbog* gennem tiden har været tilsigtet, dvs. fra den første prøveredaktion til onlinepublikationen i dag. Endelig diskuteres ordbogens eksistensberettigelse og mulige input til dens fortsættelse, og overlevelse.

2. Jysk Ordbog – af hvem?

Indsamlingsarbejdet til Jysk Ordbog, der stod på i nogle årtier, havde to hovedformål. Dels havde det et konserveringshensyn: ”Hvis ikke vi tar dette arbejde op nu, vil det om en halv snes år være forgæves at forsøge, for så vil de værdifuldste meddelere være døde” (Skautrups note til medstifterne i 1932; Hansen 2011:300). Dels havde det et leksikografisk hensyn, nemlig en mere fuldstændig optegnelse af de jyske dialekter i forhold til Feilbergs ordbog (jf. Hansen 2011:306). Skautrup forudså imidlertid også at det daværende Institut for Jysk Sprog- og Kulturforskning kunne udføre andre opgaver, herunder udgivelse af bl.a. dialektmonografier og ældre jysk dialektlitteratur (Jysk Ordbog 1970, Bd. I, hæfte 1:V).

En stor del af de data der blev samlet ind, findes i dag i seddelsamlingen (14 skabe med kartoteksskuffer i træ) der indeholder ca. 3,1 millioner kopier af eller excerperter fra kilder der hovedsageligt dækker perioden 1700-1930. Den består af både skriftlige og ”mundtlige” kilder inden for bl.a. jysk skønlitteratur (fx St. St. Blicher, Jeppe Aakjær, Jens Thise), jyske folkeminder og folkeeventyr (fx indsamlet ved Evald Tang Kristensen) samt dialektbeskrivelser og dialektordbøger, herunder ”punktoptegnelser” (fx ved Casper Schade, Arne Espegaard, Ella Jensen, Peter Skautrup). Seddelsamlingen suppleres med spørgelistebesvarelser fra meddelere født hovedsageligt 1880-1920, indsamlet af centret i perioden 1949-2014, og med ca. 1.000 båndoptagelser af meddelere fra Jylland født i årtierne omkring 1900, indsamlet i perioden 1960-1990. Endelig bliver dataene fra seddelsamlingen suppleret med H. F. Feilbergs ordbog som konsulteres ved redaktion af ethvert lemma (de fleste af Feilbergs kilder er desuden excerperet til seddelsamlingen).

Jysk Ordbogs datagrundlag må siges at være overordentlig omfangsrigt, men imidlertid også heterogent (Bjerrum 1975:100) hvorfor der i redaktionen tages højde for princippet om dataenes repræsentativitet (jf. van Keymeulen 2017:52). For som Bjerrum med rette skriver i sin anmeldelse af Jysk Ordbogs prøvehæfter: ”[ordbogen] er [...] ikke en registrant over alle et ords eller en betydnings forekomststeder; at et sted ikke er nævnt, behøver ikke at betyde, at ordet ikke findes det pågældende sted [...], ordet er blot ikke optegnet i de til rådighed stående samlinger” (Bjerrum 1975:102). Desuden er perioden for indsamlingsarbejdet samt tidsangivelser som ”vor tid” (Jysk Ordbog 1965:5) og ”yngre”, ”ældre” i ordbogen i dag vigtige opmærksomhedspunkter for redaktionen af en ordbog med en så relativt lang historie som Jysk Ordbog¹.

¹ Opmærksomheden på ordbogens og dens datas tidsperspektiv understøttes af Henrik Hovmarks konklusioner når han reflekterer over begreberne diakroni og synkroni i forbindelse med udarbejdelsen af Ømålsordbogen (Hovmark 2006).

3. Jysk Ordbog – til hvem?

Det er først i den nyeste onlineversion (herefter Jysk Ordbog 2017) Jysk Ordbogs tilsigtede brugergrupper ekspliciteres:

Jysk Ordbog henvender sig til alle brugere med interesse for de jyske dialekter fra ca. 1700 til ca. 1930 og for landbefolkningens kultur- og samfundsforhold i samme periode (før landbrugets industrialisering). [...] Sprogforskere, historikere og museumsfolk ville kunne bruge ordbogen som indfaldsvej til relevante emner med henblik på efterfølgende benyttelse af ordbogens kilder (se www.jyskordbog.dk, under Velkommen til ordbogen).

Derimod bliver målgruppen overhovedet ikke nævnt i ordbogens prøvehæfter 1970-1979 (herefter Jysk Ordbog 1970). De daværende redaktionsregler favoriserede imidlertid i høj grad fagfolk med ordbogserfaring (filologer). Artiklerne har ofte omfattende redegørelser for ordenes etymologi og evt. forekomst i ældre dansk – som i *Ordbog over det Danske Sprog*, hvor Skautrup 1924-28 var redaktør (jf. ODS XXVIII:103). Betydningsangivelser og forklaringer var undertiden vidtløftigt formuleret, med omfattende anvendelse af forkortelser og faglige betegnelser. Den anvendte lydskrift var stort set kildernes egen, nu og da meget detaljeret og vanskeligt sammenlignelig med andre kilders. De valgte citater havde ikke kun til opgave at uddybe betydningsangivelsen eller forklare kontekst, men også at påvise ordets forekomst i alle dele af det angivne brugsområde.

Den første onlineversion ordbogen fra 2000 (herefter Jysk Ordbog 2000) blev publiceret i, var baseret på de nye redaktionsregler fra 1980, dvs. den tilstræbte større enkelhed og overskuelighed for bedre at nå ikke-fagfolk. I dag tjener ordbogen et polyfunktionelt sigte: den er primært målrettet den almindeligt interesserede bruger (reception og oversættelse af jyske tekster) og dernæst den interesserede detaljeorienterede kollega (sprogforskning). Den kan også anvendes til undervisningsbrug (NLO:201).

4. Sammenligning af Jysk Ordbog 1970, 2000 og 2017

I det følgende sammenlignes de tre forskellige udgaver af Jysk Ordbog med udgangspunkt i pronomenet *a* 'jeg' (se tabel 1).

Udgave	1970	2000	2017
Format	Trykte prøvehæfter	Onlineversion 1.0	Onlineversion 2.0
Omfang (antal linjer)	166	51	5-43 ²
Grafisk materiale	Signaturkort, illustrationer	Fladeudfyldte kort	Fladeudfyldte kort
Henvisninger	Tekstuelle referencer	Tekstuelle referencer	Link + tekstuelle referencer

Tabel 1: Pronomnet *a* 'jeg': overordnede forskelle mellem Jysk Ordbog 1970 samt 2000 og 2017.

Som tabel 1 viser, er der temmelig stor forskel på hvor omfangsrig de tre udgaver er, og ikke mindst i hvor høj grad artiklen *a* muliggør differentieret brug. Fx har dialektgeografiske kort fra start af været én af Jysk Ordbogs forcer. Marie Bjerrum skriver:

Jysk Ordbog er overordentlig righoldig på både ord og betydninger; kortene, som er af behageligt tilsnit med hensyn til størrelse og signaturtæthed og -tydelighed, er et fint illustrationsmateriale, som giver en god oversigt over de enkelte betegnelsers udbredelse, langt bedre end en opremsning af forekomststeder i de enkelte artikler ville have kunnet gøre det (Bjerrum 1975:101).

De kort Bjerrum omtaler, er ikke de få skraverede, men de dengang almindelige signaturkort (fx vedr. udtale af *age* eller synonymmer af *abraham*). Redaktørerne har siden skønnet at signaturkort er vanskelige at overskue og fortolke for den almindelige bruger hvorfor der i ordbogens onlineversioner, Jysk Ordbog 2000 og 2017, bliver bragt standardiserede, fladeudfyldte kort som indeholder oplysninger redaktøren har fortolket (én- eller tofarvede kort vedr. udbredelse, flerfarvede kort vedr. udtale). Det har indtil videre ikke været muligt at finde ressourcer til at bringe hverken illustrationer eller synonymkort i onlineordbogen (jf. Bjerrums rosende omtale af dem i 1975:101).

² Er artiklen udfoldet, fylder den 5 linjer; er den fuldt udfoldet, fylder den 43 linjer.

I tabel 2 vises den drastiske ændring af både ordbogens makro- og mikrostruktur der blev foretaget med overgangen til onlinepubliceringen så en bredere brugergruppe kunne nås (jf. citatet i afsnit 3). Det gælder en forenkling af homonymindekset: Pronomenet *a* er i onlineversionerne kun ét ud af fire homonymer på grund af forbedret henvisningsstruktur; fx er *III a* med henvisning til verbet *æ*, hvis almindelige udtale er [a·], flyttet til *I.æ* verb. Artiklernes hoved er blevet mere overskueligt; udtalen er visualiseret på udtalekort, der popper op i separate vinduer ved siden af artiklen, og eksempler, samt undtagelser og kommentarer til udtalekort bringes i et notefelt, som kan foldes ud eller ej. For pronomenet *a*'s vedkommende er særlig brug i oblik form placeret i såkaldte ”plankeafsnit”, dvs. dobbelte, lodrette streger nede i kroppen hvor der redegøres for specifik syntaktisk brug, betydningsnuancer mv. Artiklens hals er rensset for historik uden særjysk relevans, og i kroppen skelnes tydeligere mellem betydninger, alle med rigsdansk, og plankeafsnit. Desuden er eksempelmaterialet i form af citater kraftigt reduceret; de skal så vidt muligt bidrage til belysning af ældre landbokulturs livsvilkår og tænke måde. Og endelig er skriftbilledet forenklet væsentligt idet der anvendes typer fra samme skriftfamilie i tydeligt adskilte varianter, fx kursiv for lydskrift og antikva for oversættelser. Dette bevirker at helhedsbilledet er mere ensartet i onlineversionerne end i 1970-versionen.

Udgave	Omfang		Omfang	
	1970	(antal linjer)	2000 og 2017	(antal linjer)
Homonym	”a pron.” er 1 ud af 10 homonymer		”a pron.” er 1 ud af 4 homonymer	
Hoved	udtalegengivelse + 2 delkort; beskrivelse af oblik form; 9 særafsnit ³	87	udtalegengivelse + link til 1 kort; beskrivelse af oblik form, inkl. reference til noter	8-9
Noter			1 note	3
Hals	6 historiske former, reference til <i>jeg</i>	4	4 ældre jyske former	3
Krop	3 betydninger: alm., etisk dativ, oblik	74	1 betydning, inkl. 2 ”plankeafsnit”	35

Tabel 2: Pronomenet *a* ’jeg’: sammenligning af makro- og mikrostrukturen i Jysk Ordbog 1970 samt 2000 og 2017.

³ Ét særafsnit om enklisestød og otte særafsnit om påfaldende syntaktisk brug af pronomenets *a*'s bøjningsformer, alle med fyldige eksempler.

De reviderede redaktionsregler og -praksis (jf. Arboe & Sørensen 2001:11) lagde større vægt på ordbogens primære målgruppe: *alle* interesserede. Det har ud over de nævnte forenklinger medført reduktion af ordbogens fagterminologi, afkald på et særskilt forkortelsessystem og øget henvisning til relevant faglitteratur. I forbindelse med udviklingen af redaktionsredskabet iLEX havde redaktionen igen den almindelige bruger i fokus: Ordbogens onlineportal blev gjort mere informativ, og ordbogsfelter og link til det forskellige kort- og kildemateriale fremgår tydeligt og er i overensstemmelse med praksis på andre onlineportaler. Så tilgodeses den almindelige bruger ved at feltet BETYDNING altid er foldet ud; det tjener brugerens behov for at kende et ords betydning. Derimod kan den mere detaljeorienterede bruger folde feltet UDTALE og BØJNING ud hvorefter der er adgang til ordbogens kortmateriale, oversigt over områder, kilder, etymologi mv. Ydermere har brugerne adgang til den resterende anden halvdel af ordbogen da de endnu ikke færdigredigerede artikler er tilgængelige via såkaldte ”artikelstubbe”, dvs. forkortede opslagsord med indikation af ordklasse, ultrakorte angivelser af ordets betydning(er) og udbredelse (jf. Bøggild-Andersen et al. 1989-1990). Jysk Ordbog består af ca. 24.200 færdigredigerede ordbogsartikler (A-K), og der er 60.643 tilgængelige artikelstubbe (L-Å) (februar 2020). På grundlag af dem forventes ca. 20.000 artikler redigeret, resten vil enten blive medtaget i plankeafsnit under andre opslagsformer, eller helt droppet (fx ad hoc-materiale).

5. Jysk Ordbog – hvorfor?

Jysk Ordbogs konserverende formål, som det fremgår af de omfattende samlinger af ældre jysk dialekt (seddel-, bånd- og spørgelistesamlingerne, se afsnit 2), bidrager positivt til dens samfundsmæssige forpligtelse om at værne om en vigtig del af dansk kulturarv (jf. Grønvik 2016:51 f.). Den fortløbende redaktion stiller essensen af dette materiale til rådighed for offentligheden mens samlingerne er tilgængelige for alle der ønsker at undersøge enkelte emner nærmere, blot man møder op på Centret i Århus! Det omfattende datamateriale (herunder de 3,1 millioner sedler) er nemlig ikke blevet digitaliseret, og det egner sig formentlig heller ikke til uredigeret at blive lagt online; materialet fordrer leksikografisk tolkning (jf. Alfort 2016:13) og kulturhistorisk perspektivering.

Der er en række udfordringer forbundet med det store ordbogsprojekt. Dels er der ordbogens objekt, jysk dialekt, dels ordbogens organisatoriske virkelighed. I takt med de stærke standardiseringstendenser i Danmark (jf. Pedersen 2003) er det jysk ordbogen beskriver (perioden 1870-1920), ikke længere (i samme udprægede grad) genkendelig for ordbogens brugere i dag. Ressourcerne til de tre dialektordbøger Danmark er rig på (*Bornholmsk Ordbog*, *Ømålsordbogen* og

Jysk Ordbog), er endda geografisk og organisatorisk spredte: De udarbejdes ved tre forskellige centre der huses af to forskellige universiteter. Fælles for projekterne er imidlertid at de dokumenterer (traditionel) sub-standard, som, på nationalt plan, hverken har den anerkendelse eller prioritering standardsprog har, bl.a. til at modtage økonomisk støtte. Så har den danske stat ingen særlig pulje til understøttelse af danske dialekter (fx som regionale sprog). Derimod har sub-standardens sønderjysk i Nordtyskland international anerkendelse (Moseley 2010) og dermed også bedre mulighed for støtte.

Længerevarende projekter som Jysk Ordbog egner sig ikke til ekstern finansiering, for diverse fonde tildeler generelt midler til kortere projekter med en politisk eller dagsaktuel diskurs. Endelig er hele leksikografien under pres, for når en ordbog én gang er publiceret online, skal der ikke flere ressourcer til? Helt så enkelt er det ikke: ”Computere er notorisk uintelligente, så uden [leksikografisk] viden og indsigt bliver værktøjerne aldrig tilfredsstillende” hvorfor der er ”stor brug for folk med leksikografisk og anden sproglig viden og kunnen inden for den sprogteknologiske branche” (Alfort 2016:13).

6. Afslutning

Hvordan kan Jysk Ordbog komme i mål? Det er ressourcekrævende at lave ordbog, at rekruttere nye ordbogsredaktører, og karrierevejen på universiteterne er vanskelig, lang og uvis. De omfattende og unikke samlinger dialektordbøgerne i Danmark, herunder Jysk Ordbog, bygger på, bør udnyttes endnu bedre. Der bør være en større, og fornyet, offentlig bevågenhed til projektet og dets samlinger, *positivt* formidlet: mindre fokus på projektets lange tidsperspektiv, større fokus på projektets unikke indblik i jysk almuelliv. Således udnyttes Jysk Ordbogs og *Ømålsordbogens* båndsamling for tiden i to forskellige ph.d.-projekter: Yonatan Ungermann Goldshtein, Aarhus Universitet, og Rasmus Puggaard Hansen, Universitetet i Leiden, Nederlandene. Trods begejstringen for den viste interesse i og det intelligente arbejde med materialet lægger det uvilkårligt, og ved siden af redaktionsarbejdet, beslag på de faglige ressourcer, akkurat som rekruttering gør det.

Det nye redigeringsværktøj iLEX (jf. Sørensen 2019) har medført øget optimering af redaktionen og af brugerhensynet. Udbyttet for brugerne igennem en frugtbar ordbogsoplevelse kan utvivlsomt løbende forbedres ved mindre justeringer af webportalen og redaktionsreglerne, især hvis ændringerne kunne baseres på brugerundersøgelser.

Litteratur

Ordbøger

- Jysk Ordbog. Kilder og fortegnelser m.m.* (1965). Institut for Jysk Sprog- og Kulturforskning (udg.). Århus: Universitetsforlaget i Aarhus.
- Jysk Ordbog* (1970-1979). Bind I, hæfte 1-4. Prøvehæfte. Institut for Jysk Sprog- og Kulturforskning (udg.). Århus: Universitetsforlaget i Aarhus.
- Jysk Ordbog* (2000 ff.). Rasmussen, Ove, Viggo Sørensen, Torben Arboe, Inger Schoonderbeek Hansen & Nina Grøftehaug (red.). Grundlagt af Peter Skautrup 1932. Århus: Peter Skautrup Centret, Institut for Kommunikation og Kultur, Aarhus Universitet. <www.jyskordbog.dk/jyskordbog/ordbog.html> (februar 2020).
- NLO = *Nordisk Leksikografisk Ordbog* (1997). Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén (red.). Oslo: Universitetsforlaget AS.
- ODS = *Ordbog over det danske sprog* (1956). Grundlagt af Verner Dahlerup. Bd. XXVIII: Liste over forkortelser med lydskrifttavle og en efterskrift. København: Det Danske Sprog- og Litteraturselskab, Gyldendalske Boghandel.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omkringliggende øer* (1992 ff.). Bd. 1-12 samt Tillægsbind. (i de senere år:) Gudiksen, Asgerd, Henrik Hovmark & Karen Margrethe Pedersen (red.). Center for Dialektforskning (tidl. Institut for Dansk Dialektforskning), Institut for Nordiske Studier og Sprogvidenskab, Københavns Universitet. København: C.A. Reitzels Forlag (bd. 1-8 og Tillægsbind), Odense: Syddansk Universitets Forlag (bd. 9-12).

Anden litteratur

- Altoft, Esben (2016): Ord og IT. I: Asgerd Gudiksen & Henrik Hovmark (red.): *Nordiske Studier i Leksikografi* 13. København: (I samarbejde med) Nordisk Forskningsinstitut, Københavns Universitet, 13-25.
- Arboe, Torben & Viggo Sørensen (2001): Jysk Ordbog – præsentation af og kommentar til netudgaven. I: *LEDA-Nyt* nr. 32: 8-15.
- Bjerrum, Marie (1975): Anmeldelse af Jysk Ordbog I 1-2. I: *Danske Studier* 1975: 98-103.
- Bøggild-Andersen, Boel, Nina Grøftehaug, Anker Leo Jensen, Ester Andersen & Hanne Bønløkke Andersen (1989-1990): *Indeks til JYSK ORDBOGs seddelsamling*. Bd. I-IV. Århus: Institut for Jysk Sprog- og Kulturforskning, Aarhus Universitet.

- Grønvik, Oddrun (2016): Vitskappleghet og samfunnrelevans for store ordboksverk. I: Asgerd Gudiksen & Henrik Hovmark (red.): *Nordiske Studier i Leksikografi* 13. København: (I samarbejde med) Nordisk Forskningsinstitut, Københavns Universitet, 13: 27-61.
- Hansen, Inger Schoonderbeek (2011): Livslange indsatser – om arbejdet hen imod Jysk Ordbog. I: Torben Arboe & Inger Schoonderbeek Hansen (red.). *JYSK, ØMÅL, RIGSDANSK mv. Studier i dansk sprog med sideblik til nordisk og tysk*. Festskrift til Viggo Sørensen og Ove Rasmussen. Århus: Peter Skautrup Centret for Jysk Dialektforskning. <www.jysk.au.dk/publikationer/oversigt> (februar 2020).
- Hovmark, Henrik (2006): Ømålsordbogen mellem synkroni og diakroni. I: *LexicoNordica* 13: 129-145.
- Keymeulen, Jacques van (2017): The Dialect Dictionary. I: Charles Boberg, John Nerbonne & Dominic Watt (red.): *The Handbook of Dialectology*. Oxford: John Wiley & Sons, Inc.: 39-56.
- Pedersen, Inge Lise (2003): Traditional dialect of Danish and the de-dialectalization 1900-2000. I: *International Journal of the Sociology of Language* 159: 9-28.
- Sørensen, Viggo (2019): *Redaktionsregler for Jysk Ordbog i iLEX-regi* <www.jysk.au.dk/publikationer/retningslinjer-for-jysk-ordbogs-redaktion> (februar 2020).
- Moseley, Christopher (ed.) (2010): *Atlas of the World's Languages in Danger*. 3. ed. Paris: UNESCO Publishing <www.unesco.org/new/en/culture/themes/endangered-languages/atlas-of-languages-in-danger>.

Inger Schoonderbeek Hansen
 lektor, ph.d.
 Peter Skautrup Centret, Aarhus Universitet
 Trøjborgvej 88
 DK-8000 Aarhus C
 jysis@cc.au.dk

Ømålsordbogens brugere – nu og i fremtiden

Henrik Hovmark

This article discusses present and possible future users of ØMO (the Dictionary of Danish Insular Dialects), an academic dictionary describing the dialects and culture of peasants and fishermen on Zealand, Lolland-Falster, Funen and surrounding islands 1750-1950. Half of the dictionary has been published since 1992 ('a'-'march') in 12 printed volumes thus far and web publication is in the planning. Users have traditionally been defined as fellow researchers and non-experts with a special interest in dialects, local history etc. Web publication will obviously lead to a higher number as well as a broader range of users: the qualifications of many future users will be more diverse and less specialized. We discuss how to make the content more accessible to non-specialists without losing sight of specialized information and researchers' interests and needs.

1. De "almindelige brugere" banker på

Ømålsordbogen (ØMO) er en videnskabelig dokumentationsordbog der giver udførlige beskrivelser af de traditionelle danske dialekter på Øerne (Sjælland, Lolland-Falster, Fyn og omliggende øer) i perioden ca. 1750-1950. Ordbogen giver detaljerede sproglige beskrivelser, men rummer også omfattende etnologiske og kulturhistoriske oplysninger om den bonde- og fiskerkultur som dialekterne var indlejret i. Projektet blev grundlagt i 1909, og første bind udkom i 1992 efter en længere indsamlingsproces, klargøring af data og udarbejdelse af redaktionsprincipper. Herefter er der udkommet 12 bind (a-march), dvs. lidt over halvdelen af ordbogen.

ØMO er som nævnt en videnskabelig ordbog, af samme type som tilsvarende nationale ordbogsprojekter der blev grundlagt i samme periode (fx Ordbog over det danske Sprog), og ordbogens brugere har primært været tænkt som fagfolk. For ØMOs vedkommende har man dog også tidligt haft en gruppe af "ikke-professionelle", "særligt interesserede" tæt inde på livet. Tidsskriftet Danske Folkemaal (1927-1999) henvendte sig i de første par årtier således i høj grad til dialekttalende meddelere og alment dialektinteresserede og indeholdt fx spørgelister (jf. Pedersen 2000, Gudiksen & Hovmark 2009). Denne ikke-professionelle brugerreds bliver også nævnt som en mulig og ønskværdig brugergruppe i ordbogens præsentationshæfte fra slutningen af 1990'erne: "Ordbogen henvender sig til en kreds af fagfolk, først og fremmest sprogforskere, etnologer, histo-

rikere og museumsfolk, men den vil forhåbentlig også få en kreds af ikke-professionelle brugere”. Og i en præsentation af ordbogen på hjemmesiden (ca. 2010) konstateres det at ”[ØMO] har også fået en kreds af ikke-professionelle brugere med interesse for dialekter og landbokultur”.

I 2012-2018 gennemgik ØMO en større teknisk omstilling med nye programmer, digitalisering og omlægning til en databasebaseret redaktionsproces. Udgivelsesformatet er stadig en trykt bog, men data er allerede klargjort til webpublicering som forhåbentlig kan blive virkelighed i en ikke alt for fjern fremtid. Webpublicering vil uundgåeligt give flere brugere og ikke mindst flere ikke-professionelle brugere, vel at mærke også brugere med langt færre forkundskaber og med langt mere diffuse interesser end de særligt interesserede.

Jeg vil i det følgende se på hvordan ØMO evt. kan tilpasse sig en langt bredere brugerkreds. Der vil i fremtiden uundgåeligt skulle tages mere hensyn til brugere med få forkundskaber, men ordbogen – og den bagvedliggende base – må også stadig tænke på de mere eller mindre professionelle brugere som har stor interesse i ordbogens undertiden meget specialiserede oplysninger.

2. Hensynet til ”almindelige brugere” uden forkundskaber

2.1. Redskaber som kan gøre teksten mere overskuelig og læselig

Det er umiddelbart forholdsvis let at gøre en trykt ordbog mere brugervenlig når den publiceres på nettet. For det første kan man opløse de mange forkortelser o.l. der traditionelt og typisk har kendetegnet ældre, trykte ordbøger med en overflod af komprimerede oplysninger. De første skridt i denne retning er allerede taget i ØMOs bind 12, hvor tvetydige, sjældent brugte eller meget ukurante forkortelser er blevet opløst (fx ”udtr” = udtryk, ”konkr” = konkret).

Opløsning af alle typer forkortelser vil dog ikke nødvendigvis føre til større brugervenlighed i alle sammenhænge. Det typiske eksempel er litteraturforkortelser som helt indlysende vil give u hensigtsmæssigt meget støj i en artikel, også i en netudgave. Her er det mere oplagt at tænke i en mere dynamisk forkortelsesstrategi i form af fx mouseover eller links. Et interessant tilfælde i ØMO er de mange lokalisings- og udbredelsesangivelser – alle oplysninger i ordbogen lokaliseres, på detaljeret niveau (sogn, underlokalitet) eller på mere generaliseret niveau (egne, herreder, (dialekt)områder). En typisk og meget enkel angivelse kan fx være: ”nvS(spor opt)”. Denne angivelse vil i uforkortet, opløst version se således ud: ’dialektområdet Nordvestsjælland (sporadisk optegnet (og muligvis almindeligt))’. Hvis man medtænker at angivelser af denne type også kan omfatte længere rækker af sogne, egne, underområder mv. inde i den enkelte parentes, og at der kan anføres flere dialektområder fortløbende (fx nvS, øF = ’dialek-

tområderne Nordvestsjælland og Østfyn'), er det indlysende at en opløsning af disse forkortelser vil blive meget omfattende. På den anden side er oplysningstypen helt central i netop denne type ordbog, og informationen vil også være interessant for helt almindelige brugere uden forkundskaber – kan en betydning eller udtale fx lokaliseres til den egn som man selv eller ens familie kommer fra? Det er derfor umagen værd at gøre informationen mere tilgængelig, men en helt anden type formidling vil være mere oplagt og effektiv her, nemlig en kobling til kort e.l., sådan som det fx ses i Norsk Ordboks netudgave.

Umiddelbart er det også let og effektivt at gøre en længere og ofte fortløbende artikel mere overskuelig ved at bryde teksten op i flere nye afsnit og gøre (under)afsnit mere tydelige grafisk og/eller med ekspliciterende overskrifter og fikspunkter (fx UDTALE, BØJNING, GRAMMATIK, EKSEMPLER o.l.; jf. fx Ptasznik & Lew 2019; Hovmark 2018). Dette er også påbegyndt i bind 12, hvor etymologiske og saglige, dvs. kulturhistoriske afsnit, indledes fast med henholdsvis "Etym:" og "Ency:". Artiklerne i ØMO er imidlertid ofte så lange og har så mange betydninger og betydningsniveauer at det er hensigtsmæssigt at tænke i yderligere strategier.

En strategi kan være at brugeren som udgangspunkt ikke præsenteres for den fulde artikel, men at dele der kræver specialistviden er foldet sammen under en tydelig, klikbar overskrift. Udtale- og bøjningsoplysningerne i ØMO forudsætter fx et kendskab til lydskrift, oplysningerne er meget komprimerede og kan også være meget omfattende – og en korrekt afkodning af afsnittet forudsætter desuden et kendskab både til strukturalistisk teori og metode og til ØMOs egne normaliserings- og visningsregler som tilmed kun er sparsomt forklaret i ordbogens præsentationshæfte (jf. Lindegård Hjorth 1993:127-29). Man kan overveje at lade dette afsnit være skjult for en almindelig bruger når en artikel åbnes (jf. tilsvarende strategi i netudgaven af Norsk Ordbok). Man skal dog være opmærksom på at oplysningerne om udtale (og bøjning) også kan være interessante for almindelige brugere, det ville blot kræve en radikalt anderledes formidlingsform der fx inddrog lyd. På (meget) længere sigt er det spørgsmålet om netop dette afsnit vil have bedst af at blive opdelt i en alment tilgængelig version, og en version der henvender sig til specialister og fagfæller. En sådan løsning er ikke umulig at forestille sig i en netudgave.

Det behøver dog ikke nødvendigvis at være specialiserede afsnit der foldes sammen for at reducere et overvældende omfang af oplysninger – eller for at gøre alle artikler mere overskuelige. Man kan fx vælge kun at vise hovedafsnit og betydningsdefinitionerne. Det er således tilfældet i Jysk Ordbogs nye netudgave (jf. figur 1). En sådan strategi kræver til gengæld at den almindelige bruger tydeligt kan se at artiklen gemmer på flere oplysninger – i Jysk Ordbog markeret med et plus, og tilsvarende et minus ved afsnit som kan foldes sammen og føre brugeren tilbage til en mere overskuelig artikel.

have¹ subst.

UDTALE OG BØJNING +

NOTER +

ETYMOLOGI MV. +

BETYDNING –

1 = JORDVOLD, GÆRDE, DIGE +

2 = INDHEGNET (EL. INDDIGET, INDGRØFTET) JORDSTYKKE +

3 = (IND)MARK, DYRKET JORD OMKRING GÅRD(E) +

Figur 1: Artiklen *have*¹ i Jysk Ordbog som den vises umiddelbart efter søgning.

Det der møder brugeren i Jysk Ordbog, er i realiteten en oversigt – og dermed er vi fremme ved den anden strategi. Oversigter lægger sig som et mellemlid mellem brugerens søgning og det egentlige søgeresultat eller svar, og det er derfor en forudsætning at de er umiddelbart forståelige og tilstrækkeligt vejledende, ellers vil de stik imod hensigten blive en barriere. En væsentlig udfordring vil være at få betydningsbeskrivelserne til at fungere tilfredsstillende uden det eksempelmateriale som en almindelig bruger netop har så stor gavn af for at forstå definitionerne der pr. definition er metatekst og jævnligt kan have abstrakt karakter. En anden og særlig udfordring vil være de ”definitioner” som består af grammatiske oplysninger i stedet for betydningsbeskrivelser. Denne type betydningsinddelere på definitionsplads er ikke helt usædvanlige i den ældre, videnskabelige ordbogstradition. De findes adskillige steder i ØMO og på forskellige niveauer i strukturen, men de ses fx også i NAOB, den reviderede netudgave af Norsk Riksmålsordbok.

Både ØMO og NAOB benytter også egentlige oversigter ved lange artikler som en hjælp til læseren, dvs. at der er indsat en oversigt før selve den fulde artikel begynder – en slags variant af modellen hos Jysk Ordbog (jf. figur 2):

1 OM LEVENDE VESEN MED BEN

1.1 bevege seg raskt til fots uten at begge føtter berører bakken samtidig; springe

1.1.1 FORELDET bevege seg, fare, gli (raskt) (ved hjelp av fremkomstmiddel som føttene er festet til); gå

1.1.2 MED OBJEKT

1.1.3 OFTE MED OBJEKT ELLER ADVERBIAL SOM UTTRYKKER DISTANSE ELLER TYPE LØPING løpe for å drive idrett, sport; delta i løpekonkuranse (i en gitt distanse)

1.2 OFTEST MED RETNINGSBESTEMMELSE sette av gårde i sprang; springe av gårde; skynde seg av gårde (uten at

Figur 2: Udsnit af oversigten ved artiklen *løpe* i NAOB.

Indholdet i oversigten i NAOB består af indholdet i den fulde artikels definitioner, evt. i en reduceret udgave. Betydningsbeskrivelser er således almindelige i oversigten, men som nævnt kan indholdet også bestå af ren grammatisk information, fx ”med objekt”, dvs. en konstruktionsoplysning. Disse oplysninger kan være specificerede og tydeliggjorte, fx ”ofte med objekt eller adverbial som udtrykker afstand eller type løping”, men ekspliciteringen vil ikke nødvendigvis være til megen hjælp for en almindelig bruger, da der er tale om en abstrakt metabeskrivelse af et typisk sprogligt udtryk. Det konkrete sproglige udtryk som man kan formode at brugeren står med og gerne vil vide mere om (fx *løpe maraton*, *løpe 400-meter*, *løpe om kapp*), og som en almindelig bruger ville have langt større glæde af at møde som et fikspunkt i oversigten, gengives derimod ikke. Der mangler med andre ord nogle typeeksempler i oversigten. Oversigterne i ØMO er generelt født med detsamme problem hvis de skal fungere som en hjælp for en almindelig bruger. Det interessante er at de manglende typeeksempler i NAOB faktisk findes i den udfoldede betydningsbeskrivelse (*løpe maraton* osv.), ligesom man også har etableret en oversigt over de flerordsforbindelser som har egen, selvstændig betydningsbeskrivelse (fx *løpe løpsk*, *løpe sammen*). Eksemplerne synes at bekræfte undersøgelser der peger på at ”formidlende mellemlid” mellem søgning/opslag og (korrekt) søgeresultat er komplicerede at arbejde med og ikke nødvendigvis kommer til at fungere som den hjælp man havde forestillet sig (Ptasznik & Lew 2014).

2.2. Hvad stiller man op med grammatiske oplysninger?

De store dokumentationsordbøger rummer mange grammatiske oplysninger af høj grundvidenskabelig værdi. Men oplysningerne er ikke lette at formidle bredt på nettet, og man kunne fristes til at anlægge den samme strategi her som for udtale og bøjning, nemlig at de som udgangspunkt var foldet sammen. Det er dog ikke helt let i fx ØMO. De grammatiske oplysninger kan stå mange forskellige steder i en artikel. Desuden udgør de ikke sjældent det eneste indhold i en definition (fx ”som adv”) eller supplerer den ganske snævert med valens- eller konstruktionsoplysninger. Det er derfor ikke umiddelbart muligt af folde oplysningerne sammen. Formidlingsmæssigt udgør de derfor en udfordring, men det er samtidig værd at tænke over hvordan disse oplysninger kan komme til deres ret i en publikationsform der også vil have mange brugere uden specialiserede kundskaber. Mange af oplysningerne vil forblive tekniske og uforståelige for menigmand (fx ”i tryktabsforb med retnings-ang adv” eller ”med overgangens resultat angivet ved adj”; *løbe* vb. bet. 24), men andre kan godt tænkes aktiveret bredere, som vist ovenfor fx ved hjælp af mere konsekvent brug af (type)eksempler (jf. ”i medial ell passiv form (*løbs*, *løbes*)”; ØMO, *løbe* vb. bet. 7).

2.3. Formidling af det kulturhistoriske stof

Et særkende ved ØMO er de saglige oplysninger om bondekultur, dvs. etnologiske og kulturhistoriske beskrivelser af arbejdsgange, redskaber mv. både ude og inde (fx høst og bagning). Hertil kommer omfattende oplysninger om tro, skikke, traditioner mv. Det er rimeligt at antage at disse oplysninger kan appellere til en bred brugerkreds, og det er også en informationstype som kan formidles på mange forskellige måder. Det ses allerede på en temaside om netop bagning på Institut for Nordiske Studier og Sprogvidenskabets hjemmeside (dialekt.ku.dk/dialekter/broed-og-bagning), og det ses også på et nyt, svensk site om traditionel madkultur (matkult.se). Begge steder blandes oplysninger om ordforrådet med baggrundsforklaringer, folkløse, kort, billeder og lyd. En sådan multimedial formidling er også oplagt på en hjemmeside for ØMO, men det ligger længere ude i fremtiden og er ikke realistisk i første omgang. Her vil vi koncentrere os om mulighederne for at formidle det kulturhistoriske stof mere effektivt inden for den eksisterende ordbogsramme.

Krumtappen i organiseringen og formidlingen af det kulturhistoriske stof i ØMO er de såkaldte saglige afsnit, dvs. mere samlede fremstillinger af de enkelte (del)emner og det dialektale ordforråd. Ordbogens mangeårige leder Poul Andersen giver allerede i 1939 en beskrivelse af disse afsnit, deres rolle og værdi:

Naar det gælder Brugeren, har man et værdifuldt Hjælpemiddel i kortfattede saglige Fremstillinger af Fremgangsmaaden ved Bondelivets forskellige uden- og indendørs Arbejder, med Anførelse af hele det dertil knyttede Ordstof. [...] Disse saglige Beskrivelser skal orientere Brugeren, der søger Oplysning om det enkelte Ord paa alfabetisk Plads og som fra denne Plads henvises til den el. de saglige Beskrivelse(r), hvori Ordet forekommer. (Andersen 1939:26)

Inden man begynder at udgive ordbogen, forredigerede man en del saglige emner og oprettede også synonymkredse med hovedsynonymer. Informationerne er altså allerede til en vis grad organiserede og forarbejdet gjort, men mange ting er alligevel overladt til den enkelte redaktør under redaktionen, og nogle emner er kun nødtørftigt redigeret. I en webudgave er det derfor oplagt og sine steder også nødvendigt at gennemgå organiseringen af information inden for de enkelte emner med henblik på at gøre den mere klar og lettilgængelig for en bredere brugerkreds: Hvor mange saglige afsnit skal der være? Hvilket hierarki skal der være? Hvornår skal der være henvisning til et sagligt afsnit? I øjeblikket befinder det store saglige afsnit om bagning sig under opslagsordet *bage*, men det betyder ikke at alle oplysninger om bagning findes her. Tværtimod er de saglige afsnit karakteriseret ved en mængde henvisninger, til ord, men også til andre relevante saglige afsnit under andre opslagsord. Under *bage* henvises der fx til

bageovn hvor der findes yderligere saglig information, og under *bageovn* gives der en henvisning til *bageovnsfæl* som imidlertid blot nævnes under *bageovn*, men ikke forklares. Heldigvis giver den nye database allerede nu mulighed for at opmærke ikke blot de saglige afsnit, men også de ord og betydninger som hører til bestemte emner, dvs. at det bliver muligt at etablere en slags subkorpora for de enkelte emner – eller domæner. Denne opmærkning vil kunne bruges til formidlingsmæssige formål, fx til at styre det interne henvisningshierarki så en bruger ikke drukner i links. Derudover vil opmærkningen også give mulighed for helt nye typer søgninger og udtræk fra basen relateret til emner eller domæner, af specialiserede fagfæller eller måske med mere sprogteknologiske formål.

3. Hensynet til de (halv)professionelle brugere

Et karakteristisk træk ved de kulturhistoriske oplysninger er at alle opmærkningerne i databasen om emner e.l. ikke nødvendigvis vil blive vist, hverken i en trykt udgave eller på nettet. Databasen er ved at udvikle sig til et selvstændigt, grundlæggende informationsbibliotek, hvorfra der kan genereres forskellige standardiserede visninger (fx en bog eller en webpublikation) eller mere specialiserede søgeresultater.

I ØMOs tilfælde er det tydeligt at omlægningen til redigering i databasen giver øgede muligheder for at aktivere ordbogens og samlingernes oplysninger på nye måder. Det gælder almindelige brugere, men det gælder måske især specialister og måske også sprogteknologer, i form af søgninger, men også i form af berigelse af databasens – og den hidtidige ordbogs – oplysninger. Vi har set hvordan der kan lægges emneoplysninger ind på forskellige niveauer, og andre opmærkninger kan tænkes hvis de skønnes at være relevante for bestemte brugere. Et eksempel kan være tilføjelse af valensoplysninger ved flerordsforbindelser. ØMOs samlinger er karakteriseret ved at rumme meget fraseologisk stof som jævnlige dokumenterer systematiske, danske konstruktioner der ikke er beskrevet i andre sammenhænge. Dette materiale vil med andre ord være af interesse for en generel beskrivelse af dansk sprog og af (mulige) valensmønstre aktuelt og historisk. Visningsformatet for flerordsforbindelser i den trykte bog giver imidlertid ikke mulighed for at vise valensoplysninger i nævneværdig grad og slet ikke hvis oplysningerne skal kunne bruges på højt specialiseret niveau, evt. også i sprogteknologiske sammenhænge. Man kan tænke sig en skjult opmærkning her med specialiserede valensoplysninger.

Men hvis de meget specialiserede oplysninger helt eller delvis skal ligge som søgbare oplysninger i databasen og ikke nødvendigvis vil være tilgængelige i mere alment formidlet form på nettet, bliver det igen relevant at tænke i en opdeling mellem halv- og halvprofessionelle brugere. De halvprofessionelle eller

særligt interesserede har nok visse forkundskaber, men de vil ikke have særlig specialistviden. Alligevel vil de have behov for at kunne finde relevant information på nettet i en umiddelbart tilgængelig form. ØMOs udtaleoplysninger kan bruges som eksempel. Mange af oplysningerne her er meget vanskeligt tilgængelige for en helt almindelig bruger, men nogle af de mere specialiserede oplysninger om udtalen af bestemte ord i bestemte egne vil givetvis have interesse for halvprofessionelle brugere. Det gælder i øvrigt også bøjningsoplysninger, fx bøjningsmønstre ved både stærkt- og svagtbøjede verber.

Samlet set kan man sige at fremtiden rummer stærkt øgede muligheder for at nå ud til en langt bredere brugerkreds, ofte på mere målrettede måder. Samtidig er det værd at notere at udfordringen ikke blot består i at komme almindelige brugere med få eller slet ingen forkundskaber i møde – her er der umiddelbart mange velkendte tiltag man kan gøre, og meget af stoffet rummer også et stort formidlingspotentiale hvad angår almindelige brugere. Udfordringen består også i at gøre ordbogens mange specialiserede informationer yderligere tilgængelige for brugere med mere eller mindre specialiserede forkundskaber.

Litteratur

Ordbøger

Jysk Ordbog. <www.jyskordbog.dk> (juni 2019).

NAOB = *Det Norske Akademis ordbok*. <www.naob.no> (juni 2019).

Norsk Ordbok (i-åværig). <no2014.uib.no> (juni 2019).

ØMO. *En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer 1* - (1992-). København: Universitets-Jubilæets danske Samfund og Københavns Universitet.

Anden litteratur

Andersen, Poul (1939): Udvalg for Folkemaals Virksomhed 1933-38 (med Udkast til Plan for den kommende ”Ordbog over de danske Ømaal”). I: *Danske Folkemaal* 12, 17-64.

Gudiksen, Asgerd & Henrik Hovmark (2009): Måske husker De noget alle andre har glemt. I: Asgerd Gudiksen et al. (red.): *Dialektforskning i 100 år*. København: Afdeling for Dialektforskning, Nordisk Forskningsinstitut, Københavns Universitet, 13-64.

- Hovmark, Henrik (2018): Udflytning og nedarvning som formidlingsstrategi i ordbøger. I: Ásta Svavarsdóttir et al. (red.): *Nordiske Studier i Leksikografi* 14. Skrifter udgivet af Nordisk Forening for Leksikografi 15. Reykjavík, 129-140.
- Lindegård Hjorth, Poul (1993): Ømålsordbogen (anmeldelse). I: *Danske Studier* 1993, 125-132.
- Pedersen, Inge Lise (2000): Fra folkemål til multietnolekt: Kontinuitet eller brud. I: *Danske Talesprog* 1, 5-29.
- Ptasznik, Bartosz & Robert Lew (2019): New-line and run-on guiding devices in print monolingual dictionaries for learners of English. I: *Lexikos* 29, 1-19.
- Ptasznik, Bartosz & Robert Lew (2014): Do Menus Provide Added Value to Signposts in Print Monolingual Dictionary Entries? I: *International Journal of Lexicography* 27(3), 241-258. <www.doi.org/10.1093/ijl/ecu010> (juni 2019).

Henrik Hovmark
ordbogsredaktør, lektor, ph.d.
Ømålsordbogen
Institut for Nordiske Studier og Sprogvidenskab
Københavns Universitet
Emil Holms Kanal 2
DK-2300 København S
hovmark@hum.ku.dk

Lad brugeren komme til orde! Hvordan får ordbogen mest muligt ud af brugerhenvendelser?

Jonas Jensen, Henrik Lorentzen & Mette-Marie Møller Svendsen

Over the course of the last decade, The Danish Dictionary (*Den Danske Ordbog*, DDO) has received more than 12,000 emails from its end users. The amount of messages makes it impossible for the editors to thoroughly answer each individual enquiry. Hence, the editors of DDO have been requesting tools to help them process the emails more efficiently. This study aims to map the most common types of enquiries based on a representative sampling of emails from the month of March 2019. On that basis the study will outline a best practice for each of the categories to which the enquiries belong: When do we provide the user with an in-depth response? Under what circumstances can an enquiry simply be ignored? When should the user be redirected to The Danish Language Council, to our own IT department, or to a third party? And equally important: To what extent can the process be automated? In short: How can the day-to-day enquiries contribute to the lexicographers' work, and how can the workflow be optimised?

1. Indledning

Denne artikel handler om henvendelser fra brugere af onlineudgaven af *Den Danske Ordbog* og om hvordan redaktionen dels forholder sig til henvendelserne, dels udnytter dem bedst muligt til at forbedre ordbogen og til at sikre en god dialog med brugere.

I artiklens første afsnit ridser vi baggrunden op og gør rede for motivationen for den undersøgelse vi har foretaget af de indkomne mailhenvendelser. I afsnit 2 kommer vi ind på den herskende terminologi på området og på den metode vi har brugt, hvorefter vi i afsnit 3 fremlægger og diskuterer resultaterne af undersøgelsen. I afsnit 4 vurderer vi hvilke perspektiver undersøgelsens resultater har for redaktionens håndtering af brugerhenvendelser, og sluttelig præsenterer vi to løsningsforslag i afsnit 5.

Som nævnt drejer vores undersøgelse sig om brugerhenvendelser angående Den Danske Ordbog (DDO). DDO er en ordbog over moderne dansk sprog svarende til tiden efter 1950 og med hovedvægten på de seneste ca. 30 år. Den blev udarbejdet i 1990'erne og starten af 2000-tallet og udkom som trykt bog i 6 bind fra 2003 til 2005 (se bl.a. Lorentzen 2004 og Lorentzen & Trap-Jensen 2006).

Den første netversion kom i 2009, er siden blevet udvidet og suppleret og er også udkommet som applikation til smartphone og tablet. I øjeblikket rummer den ca. 100.000 opslagsord.

Den Danske Ordbog har været tilgængelig online siden november 2009, dvs. mere end 10 år. Når man tænker på at ordbogen har vældig mange brugere hver dag, kan det ikke overraske at der også kommer mange henvendelser per mail. I løbet af de 10 år har redaktionen modtaget ca. 12.000 mails. De udgør en fuldstændig uoverskuelig og uhåndterlig mængde, så redaktionen har et brændende ønske om at sortere og håndtere denne store bunke post. For at komme nærmere et svar på hvordan vi håndterer brugernes mails, har vi undersøgt alle mails fra en hel kalendermåned (i alt 125 stk.), og derefter har vi etableret en typologi over henvendelsernes karakter. Det vil blive uddybet i de følgende afsnit.

2. Baggrund, terminologi og metode

2.1. Baggrund

Der kan være mange grunde til at DDO's brugere tager kontakt til ordbogen. Det kan være for at lufte en holdning til de nye ord der bliver optaget i ordbogen ("der er ingen grund til at accelerere vores sprogs forfald!"), et spørgsmål til grammatik ("skriver man stresssymptom med eller uden bindestreg?") eller for at spørge til rigtigheden af sjove, sproglige finurligheder ("kan det virkelig passe at Ørsted indførte ordet sommerfugl på dansk?"). Derfor er det mange forskellige emner der bliver berørt i redaktørernes interaktion med brugerne.

Når det gælder digitale ordbøger, er brugerne ikke længere passive forbrugere, men aktive bidragsydere i form af korrekturrettelser, forslag til indhold med videre (Meyer & Abel 2017; Lew 2014). Brugere kan komme med input til ordbogens indhold, fx til revisioner, og det kan være givende for redaktionen. På den måde kan den holde sig bedre opdateret og lægge nyt digitalt indhold op hurtigere end hvis det kun var leksikograferne der egenhændigt skulle forsøge at holde trit med samtlige sproglige tendenser i samtiden.

2.2. Terminologi

For at kunne sammenligne data med andre projekter der har udført systematiske undersøgelser af brugerhenvendelser, ønsker vi først og fremmest at undersøge terminologien på området. Meyer & Abel (2017:737ff.) samt Wiegand et al. (2010:17) pointerer at der længe har været en mangel på fælles terminologi vedr. brugerinteraktion og brugerdeltagelse. Meyer & Abel foreslår en terminologi baseret på Manns distinktion mellem direkte og indirekte bidrag fra brugerne samt

udveksling mellem ordbogens brugere (Mann 2010). Meyer & Abel går videre med en udvidelse af Manns distinktion med mere finkornede kriterier for de enkelte kategorier. Resultatet er en klassificering af brugerinteraktion i elektroniske opslagsværker, og de overordnede kategorier gives der her et kort overblik over (Meyer & Abel 2017:738ff.):

- *Direct user participation*: Denne term betegner alle former for brugerbidrag som har direkte indflydelse på det leksikografiske produkt (Meyer & Abel 2017:738). Ved denne type brugerinteraktion har brugeren mulighed for mere el. mindre direkte at tilføje indhold til ordbogen. Dette ser man fx i Open Dictionary på Macmillans website.
- *Indirect user participation*: Her bidrager brugeren mere indirekte til ordbogen ved at give enten *explicit* eller *implicit feedback* til ordbogen. *Explicit feedback* består i brugerhenvendelser, fx på mail. *Implicit feedback* består i brug af logfiler og registrering af brugernes aktivitet på hjemmesiden.
- *Accessory user participation*: Her bidrager brugerne til ordbogen ved for eksempel at diskutere ordbogens indhold på de sociale medier. *Accessory user participation* vil eksempelvis kunne ses i en Twittertråd omhandlende de nyeste ord der er blevet optaget i ordbogen.

Udover de ovenstående kategorier kan der også være tale om *form-based feedback* og *free-form feedback*. *Form-based feedback* modtages via en formular hvor en række felter skal udfyldes for at henvendelsen kan nå modtageren. *Free-form feedback* svarer til henvendelser som ordbogsredaktionen modtager per mail – der er ingen retningslinjer for hvad henvendelsen skal indeholde for at den når modtageren (Meyer & Abel 2017:742).

I dette lille studie kigger vi på *indirect user participation*, *explicit feedback* baseret på (primært) *free-form feedback*. Spørgsmålet er hvorvidt det er muligt at gå over til mere *form-based feedback* ud fra de emner som brugerne berører i deres henvendelser.

DDO-redaktørernes ønske med denne undersøgelse er blandt andet at sigte mod i fremtiden at modtage mere *form-based feedback* frem for den nuværende store mængde *free-form feedback*.

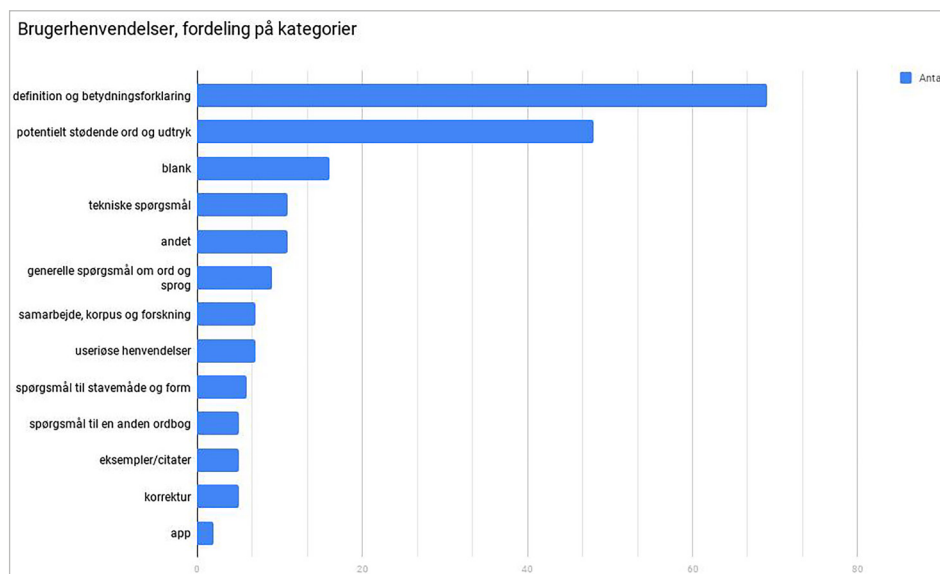
2.3. Metode

Et udtræk fra den fælles indbakke for Det Danske Sprog- og Litteraturselskabs ordbøger blev grundlaget for vores undersøgelse. Mails modtaget i marts måned 2019 blev udvalgt som et repræsentativt udsnit, og heraf blev spam, autosvar og Twiternotifikationer frasortet. Det gav os i alt 125 brugerhenvendelser til vores kvalitative analyse. Den kvalitative analyses kategorier er baseret på et upub-

liceret forstudie udført i foråret 2018 samt en tidligere undersøgelse som blev foretaget i forbindelse med digitaliseringen af Ordbog over det danske Sprog (ODS) (Sørensen 2006). I ODS-undersøgelsen blev der taget udgangspunkt i fem grovkornede kategorier. I denne nye undersøgelse for DDO mente vi at der er behov for en mere nuanceret inddeling.

3. Resultater

Tabel 1 viser de kategorier som vi benyttede i den kvalitative analyse af brugerne henvendelser, samt hvor mange henvendelser der faldt i den enkelte kategori. Da brugerne henvendelser kan berøre flere emneområder, overstiger det samlede antal 125, dvs. det antal brugerhenvendelser fra marts måned som vi tog udgangspunkt i. Tabel 1 viser også fordelingen af brugerhenvendelser i de forskellige kategorier. Her ser vi to klare sværvægtede, nemlig spørgsmål til *definition og betydningsforklaring* samt henvendelser vedr. *potentielt stødende ord og udtryk*. Dog er det også værd at lægge mærke til kategorier såsom *tekniske spørgsmål*, dvs. henvendelser der vedrører tekniske problemer på hjemmesiden for brugeren, *generelle spørgsmål om ord og sprog* (“rimer homofoner?”), samt henvendelser vedr. *samarbejde, korpus og forskning*.



Tabel 1: Fordelingen af henvendelser sorteret efter type og antal.

3.1. Afrunding af resultater

Ud fra vores kvalitative analyse kan vi konstatere at brugerne spørger meget ind til emneområderne *definition og betydningsforklaring* og *potentielt stødende ord og udtryk*. Her er dog en fejlkilde: Netop i marts måned 2019 modtog redaktionen en stor mængde henvendelser vedr. definitionen af ordet *hudfarvet*, og derfor ses der en overrepræsentation af de to nævnte kategorier som dermed kan skævvride resultatet. DDO oplever i perioder en stor interesse fra brugerne vedr. bestemte ord og udtryk i ordbogen, og tidligere har interessen fx berørt ord som *grønlænderstiv*, *muskelhund* og *rocker*, jf. Jensen et al. (2018).

En stor mængde henvendelser som optrådte i undersøgelsen, viste sig at være blanke, og disse henvendelser bør ordbogsredaktionen kunne håndteres spå en mere hensigtsmæssig måde end at de fylder i indbakken.

Den kvalitative analyse viste brugbare resultater i forhold til mindre kategorier såsom *korrektur* og *spørgsmål til stavemåde og form* som også er værd at tage med i de videre overvejelser til en eventuel formular.

Ved en gentagelse eller eventuel udvidelse af studiet kunne en større testmængde af brugerhenvendelser fra indbakken undersøges da vi på den måde ville sikre at enkeltord som *hudfarvet* ikke ville være toneangivende for hele den kvalitative analyse.

4. Brugerinteraktionens fordele

Brugerhenvendelserne er naturligvis ikke kun en gene og en ekstra arbejdsbelastning – mange af dem bidrager allerede i dag til ordbogsplatformen. Det gælder især på de tre områder der behandles i det følgende: 4.1. *Sprogligt*, 4.2. Nye ord, betydninger og udtryk og 4.3. Revision af eksisterende artikler.

4.1. *Sprogligt*

DDO's forside rummer en række faste elementer, bl.a. en oversigt med links til redaktionens seneste optrædener i medierne, *Dagens ord*, der dagligt byder på et tankevækkende ord eller udtryk, og *Sprogligt*, en ugentlig klumme hvor vi publicerer svar på sproglige spørgsmål stillet af ordbogens brugere via e-mail. *Sprogligt* tager altså udgangspunkt i brugerhenvendelser der kun i cirka halvdelen af tilfældene går specifikt på ordbogens artikler: Én bruger spørger fx til oprindelsen af det faste udtryk *have blod på tanden*, en anden vil vide om man bør skrive *der var engang* eller *der var en gang*, og en tredje undrer sig over brugen af historisk præsens i beskrivelsen af fortidige begivenheder. Denne type spørgsmål er tidskrævende at besvare, og emnerne ligger ofte et stykke fra redaktørernes

primære arbejdsområde. Omvendt ligger de generelle spørgsmål mange brugere på sinde, og *Sprogligt* bidrager med ekstra krydderi på ordbogens hjemmeside, ligesom klummen deles på vores Twitterkonto, hvor den er blandt det mest sete indhold.

4.2. Nye ord, betydninger og udtryk

Blandt det mest brugbare input fra brugerne er tip om nye ord, betydninger og udtryk. Indsamlingen af disse er i et vist omfang allerede automatiseret, idet brugere som foretager en resultatløs søgning, får mulighed for at indsende deres søgeord til en database med brugerforslag der deles mellem Det Danske Sprog- og Litteraturselskab og Dansk Sprognævn.



Figur 1: En resultatløs søgning på ordet "supercalifragilisticexpialidocious" i DDO.

En stor del af brugerforslagene resulterer i nye artikler, betydninger eller udtryk – ved opdateringen i marts 2019 var hele 170 ud af de 780 tilføjede ord således baseret på brugerhenvendelser. Brugere kan også vælge at sende en almindelig e-mail hvis de oplever at der mangler bestemte ord, betydninger eller udtryk i DDO.

4.3. Revision af eksisterende artikler

Når det gælder ændringsforslag til eksisterende artikler, har brugerne også mulighed for at kontakte ordbogens redaktion direkte fra den pågældende artikel. Det sker via linket *Rapportér et problem*. Ved mouseover fremgår det at redaktionen ikke besvarer eventuelle spørgsmål tilsendt ad denne kanal.

Henvendelser af denne type er ofte af høj kvalitet, men sjældnere end vi kunne ønske os. Alene det seneste år har brugerne via *Rapportér et problem* hjulpet til med at udrydde korrekturfejl, gøre opmærksom på uheldige citater og

forældede definitioner, forsyne nedsættende ord med passende valøroplysninger, påpege forkerte udtaler og fejlplacerede lydfiler samt bidrage med etymologiske oplysninger.

4.4. Brugerfeedback og formularer

Som nævnt hos Meyer & Abel (2017:742) er der visse risici forbundet med brugen af indberetningsformularer. Navnlig hvis formularerne bliver for lange, udviklede og fintmaskede, kan brugerne have svært ved at diagnosticere og klassificere egne spørgsmål. Det kan føre til at brugeren simpelthen giver op og redaktionen går glip af brugbart input. Omvendt kan en enklere formular med blot et par obligatoriske felter måske afskrække useriøse brugere fra at fremsende usaglig kritik eller spam.

Vi er selvfølgelig ikke interesserede i at indføre en alt for kompliceret formular – snarere i at finde en gylden middelvej. Nedenfor har vi skitseret to løsningsmodeller: én der gør brug af formularer, og én der ikke gør.

5. Effektivisering af arbejdsgangen

Som afrunding kommer her to bud på hvilke konkrete tiltag vi kan sætte i værk i bestræbelsen på at håndtere de mange tilsendte beskeder – og hvilke konsekvenser de hver især vil få for den nuværende arbejdsgang.

5.1. Den nænsomme model

Denne model sigter til at nedbringe arbejdsbelastningen en anelse gennem automatisk frasortering af de mindst brugbare beskeder, men sørger samtidig for at langt de fleste beskeder bliver gennemlæst af en redaktør. Vi vil fortsat modtage stof til *Sprogligt*, men arbejdsbelastningen er til gengæld stadig stor. Modellen består i følgende:

- Automatisk frasortering af tomme beskeder.
- Automatisk fordeling af beskeder på baggrund af modtageradresse (ddo@ordnet.dk, ods@ordnet.dk etc.).
- Førsteprioritet til henvendelser indkommet via *Rapportér en fejl og Send et ord*.

5.2. Den strikse model

Denne model går efter en markant reducere af arbejdsbyrden med henblik på at frigive flest mulige resurser til det ordbogsnære arbejde. Modellen tager udgangspunkt i en simpel indberetningsformular som erstatning for de eksisterende e-mailadresser. Den gør brug af omfattende automatisk sortering, og brugere med generelle sproglige spørgsmål dirigeres alle videre til Dansk Sprog-nævns (DSN) spørgetelefon. Det kan naturligvis resultere i færre relevante input fra brugerne og i slukørede brugere der ikke får svar på deres forespørgsler. Samtidig betyder modellen nødvendigvis et farvel til *Sprogligt* (genudsendelser af eksisterende klummer undtaget). Her er modellen i detaljer:

- Al kontakt foregår via en formular med 1-2 dropdownmenuer og et antal obligatoriske felter. De eksisterende e-mailadresser til en bestemt ordbogs redaktion eller til en specifik redaktør optræder ikke længere på ordnet-plattformen.
- Automatisk sortering og omdirigering af beskeder baseret på brugerens valg fra dropdownmenu med lukket inventar.
- Alle brugere der vælger “generelt spørgsmål om sprog”, dirigeres videre til spørgetelefonen hos DSN.
- Kun beskeder vedr. professionelt samarbejde, tekniske forhold og muligt stødende indhold besvares personligt. Øvrige besvarelser vælges fra et fast inventar af standardsvar. Et eksempel: “Tak for din henvendelse vedr. ordet [drone]. Opslaget er nu udvalgt til revision og vil blive gennemset af en redaktør i den nærmeste fremtid. Der kan gå et stykke tid før eventuelle ændringer kan ses på hjemmesiden. Vi opdaterer normalt ordbogen 2-3 gange årligt. De bedste hilsner fra Den Danske Ordbog”.

6. Konklusion

Gennem nærværende undersøgelse af mulighederne for at optimere Den Danske Ordbogs udbytte af brugernes henvendelser, er redaktionen nu i stand til at handle på et veloplyst grundlag. Den kvalitative undersøgelse af et uddrag af de modtagne henvendelser og den efterfølgende kategorisering af de typiske henvendelsesformer giver mulighed for at automatisere ellers tidskrævende processer, uanset hvilken af de to skitserede modeller for behandling af brugerhenvendelser vi vælger. Selve implementeringen kommer naturligvis til at kræve ekstra resurser, men med disse forstudier er vejen banet for en lettere og mere overskuelig beslutningsproces.

Litteratur

Ordbøger

- DDO = *Den Danske Ordbog*. Trykt udgave 1-6 (2003-05). Onlineversion: <www.ordnet.dk/ddo> (februar 2020).
- Macmillan: Open Dictionary*. Onlineversion: <www.macmillandictionary.com> (februar 2020).
- ODS = *Ordbog over det danske Sprog*. Trykt udgave 1-28 (1918-56) & supplement 1-5 (1992-2005). Onlineversion: <www.ordnet.dk/ods> (februar 2020).
- Wiegand, Herbert Ernst, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer & Werner Wolski (red.) (2010): *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*. Berlin/New York: de Gruyter.

Anden litteratur

- Jensen, Jonas, Henrik Lorentzen, Sanni Nimb, Mette-Marie Møller Svendsen & Lars Trap-Jensen (2018): *Thaipiger, muskelhunde og fulde svenskere: nedsættende ord, stereotyper og ligestilling i Den Danske Ordbog*. I: Ásta Svavarsdóttir, Halldóra Jónsdóttir, Helga Hilmisdóttir & Þórdís Úlfarsdóttir (red.): *Nordiske Studier i Leksikografi* 14, NFL-skrift nr. 15, Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum, 141-151.
- Lew, Robert (2014): *User-Generated Content (UGC) in English Online Dictionaries*. I: Andrea Abel & Annette Klosa (red.): *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikographie"*. OPAL – Online publizierte Arbeiten zur Linguistik. Mannheim: Institut für Deutsche Sprache, 9-30.
- Lorentzen, Henrik (2004): *The Danish Dictionary at large: Presentation, Problems and Perspectives*. I: Geoffrey Williams & Sandra Vessier (red.): *Proceedings of the Eleventh EURALEX International Congress*. Lorient, Frankrig, 285-294. <www.euralex.org/publications/the-danish-dictionary-at-large-presentation-problems-and-perspectives> (februar 2020).
- Lorentzen, Henrik & Lars Trap-Jensen (2006): *Den Danske Ordbog – set i bakspejlet*. I: *Nyt fra Sprogævnnet* nr. 2, 2006. København, 5-11. <www.dsn.dk/nyt/nyt-fra-sprogaevnet/numre/argang-2005-2009/juni-2006-pdf> (februar 2020).

- Mann, Michael (2010): Internet-Wörterbücher am Ende der “Nullerjahre”: Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriterien. I: *Lexicographica* 26, 19-46.
- Meyer, Christian M. & Andrea Abel (2017): User participation in the Internet era. I: Pedro A. Fuertes Olivera (red.): *The Routledge Handbook of Lexicography*. Abingdon/New York: Routledge, 735-753.
- Sørensen, Nicolai Hartvig (2006): Når arvesølvet går online – modtagelsen af ODS på nettet. I: *LEDA-Nyt* 41, 25-28.

Jonas Jensen
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
jj@dsl.dk

Henrik Lorentzen
ledende redaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
hl@dsl.dk

Mette-Marie Møller Svendsen
assisterende redaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
mettemariesvendsen@gmail.com

Brugere og forskellig anvendelse af ONP

Ellert Þór Jóhannsson & Simonetta Battista

In this article we discuss the users and various usage of *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose* (ONP), which is a historical dictionary covering the medieval language of Iceland and Norway. We start by giving a brief history of the project, its editorial principles and what is required of the users of the dictionary. Then we take a closer look at two projects where ONP is used as a standard: *The Medieval Nordic Text Archive (Menota)*, which is a publicly available archive of electronic editions of medieval texts and *Wörterbuch der altnordischen Pflanzennamen*, which is a specialized dictionary of plant terminology in medieval sources. In examining the use of ONP in these projects we consider some of the issues involved and the advantages and disadvantages of using ONP as a standard.

1. Indledning og baggrund

I 1939 startede der i København et ordbogsprojekt over middelaldersproget i Island og Norge. Oprindeligt var det tænkt som en samling af supplerende citat-eksempler til tidligere ordbøger, men det blev senere til en selvstændig ordbog: *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose* (ONP). Formålet med den nye ordbog var en mere fyldestgørende beskrivelse af prosasproget i Island og Norge i middelalderen baseret på mere moderne filologiske principper og nøjagtigere gengivelse af kildematerialet, hvilket adskilte den fra dens forgængere som Cleasby & Vigfusson (1874) og Fritzner (1886-1896). Man besluttede at begrænse ordbogen til prosatekster, idet en ordbog over det poetiske sprog, *Lexicon Poeticum*, allerede var blevet udgivet (Jónsson 1931).

I de første mange år blev der excerperet selektive eksempler, som skulle belyse betydningen for hvert opslagsord og forskellige betydningsnuancer. Ved excerperingsarbejdet blev der lagt særlig vægt på ”ældste forekomst af såvel sproglige som syntaktiske fænomener” og ”kulturhistorisk interessant ord-forråd” (Widding 1964:5).

I 1989 udkom ordbogens registerbind, som er selve nøglen til materialet. Det er et register over tekstkilder og fungerer som et selvstændigt opslagsværk over håndskrifter, videnskabelige tekstudgaver og relevant bibliografi. Registerbindet blev efterfulgt af tre trykte bind med ordbogsartikler, som udkom 1995-2004, og som dækker alfabetet fra *a-em*. Efter 2004 blev udgivelsen af det trykte værk indstillet, og det blev besluttet at gøre ordbogsmaterialet tilgængeligt på nettet, i første omgang ved at scanne det materiale som endnu ikke var behandlet (cf. Jóhannsson 2019).

I 2010 blev onlineudgaven lanceret med både materialet fra de udgivne bind og ikke-udgivet materiale, som primært bestod af lemmalisten samt tilhørende citater i form af scannede sedler fra ONP's arkivsamling. Ordbogsarbejdet fortsætter, og onlineudgaven bliver løbende udvidet med nye ordbogsartikler, samtidig med at anvendelsesmulighederne udvikles (se ONP Online, onp.ku.dk).

2. ONP's principper og brugere

Ifølge ONP's redaktionelle principper skal materialet genspejle originalkilden så nøjagtigt som muligt, hvilket betyder, at citaterne er taget fra videnskabelige udgaver eller direkte fra håndskrifter, og at stavemåden ikke er normaliseret. I de fleste tilfælde refererer ONP til både tekstudgaver og de håndskrifter, udgaverne bygger på. ONP's tilgang til stoffet og redaktionelle principper er diskuteret i detaljer i fx Battista & Jóhannsson (2014) og Jóhannsson & Battista (2016).

Mange brugere af ordbøger over sprog, som kun eksisterer på skrift, bruger dem primært til at forstå specifikke tekster som en del af en videnskabelig undersøgelse i sprog, litteratur, historie osv. På grundlag af de henvendelser ONP har fået i tidens løb kan man konstatere at den typiske bruger er en akademiker, som forsker i norrønt sprog eller litteratur eller i middelalderens historie og kultur.

ONP stiller en del krav til den bruger, som benytter ordbogen som opslagsværk. Den grundlæggende forudsætning er, at brugeren kan læse og forstå sproget og er bekendt med unormaliserede tekstudgaver, dvs. videnskabelige, diplomatariske udgaver af håndskrifter eller håndskriftlæsemåder, hvis et håndskrift bliver citeret direkte. Brugeren skal også have en vis viden om grammatik og grammatisk terminologi og skal kunne orientere sig i morfologiske oplysninger. Brugeren skal kunne "oversætte" eller "konvertere" ordformer. Det betyder, at han eller hun skal kunne generere den normale stavemåde af et bestemt ord og kende dets relevante opslagsform for at kunne slå ordet op. Når brugeren for eksempel støder på en ordform som *nocqvern* i en unormaliseret tekstudgave, skal han eller hun vide, at det ikke er en standardiseret stavemåde og endvidere, at det muligvis er en bøjningsform, for at kunne slå det op i ordbogen. I dette tilfælde findes akkusativ *nocqvern* under nominativformen *nøkkurr*.

3. Brug af ONP som standard

ONP's principper har resulteret i adskillige normer og standarder som kendetegner ordbogens gengivelse af kildematerialet, og som også bliver brugt i andre projekter og publikationer, som ikke er direkte relateret til ordbogen. Her kan vi nævne forskningsnetværk, specialiserede ordbøger, videnskabelige tekstudgaver og artikler. Vi vil nu uddybe hvordan ONP kan bruges som en del af videnska-

beligt arbejde inden for nordisk sprog og litteratur og vise nogle eksempler på, hvordan de normer og standarder som ONP har udviklet er blevet brugt i forskellige sammenhænge.

3.1. ONP som standard i et internationalt netværk

ONP optræder i netværket Menota, en forkortelse for Medieval Nordic Text Archive. Menotas rolle og mål beskrives på dets hjemmeside som: ”et netværk av ledende nordiske arkiver, biblioteker og institutter som arbeider med middelaldertekster og håndskriftsfaksimiler. Målet for Menota er å bevare, formidle og publisere middelaldertekster i digital form, og å utvikle de standarder som kreves for dette”. Menota har stået i spidsen for udviklingen af xml-standarder til opmarkering af oldnordiske tekster og samlet opmærkede elektroniske tekster i et frit tilgængeligt internetarkiv (se www.menota.org).

Menota arbejder på at gøre alle elektroniske udgaver af nordiske middelaldertekster (både vest- og østnordiske) så ensartede som muligt og har udstukket nogle retningslinjer for, hvordan teksterne skal behandles for at kunne udgives i elektronisk form. Menota adresserer forskellige aspekter af behandling af tekstmateriale i en håndbog (Haugen 2019), som er tilgængelig på dets hjemmeside. Her omtales ONP’s standard i forbindelse med tre områder: ortografi, lemma og ordklasse.

3.1.1. Menotahåndbogen

I Menotahåndbogen, del 10.2 er der vejledninger vedrørende normalisering af ortografi i lemmaer. Der står følgende: ”The orthography of lemmata should adhere strictly to established dictionaries in the field. For Old Icelandic and Old Norwegian, we strongly recommend using the orthography of ONP: Dictionary of Old Norse Prose”. Her er anbefalingen at bruge ONP’s ortografi som basis for standardiseret lemmatisering i elektroniske tekstudgaver.

Menotahåndbogen diskuterer også normalisering af tekstudgaver. Her omtales tre normer, som afspejler forskellige traditioner i Island og Norge. ONP’s normaliseringsnorm bliver nævnt som den fjerde mulighed: ”[ONP’s] standard Old Norse normalisation gives students from all over the world a means by which to approach the texts with the help of dictionaries, grammars and other reference tools. In order to provide this shared and open access this dictionary normalises its headwords and Old Norse phraseology.”

Et andet sted i Menotahåndbogen, del 11.6.2 er der vejledninger om valg af lemma: ”We recommend that the lemmatisation is coordinated with the list of lemmata in ONP”, det vil sige, at man som en del af opmærkningsprocessen af en elektronisk tekst skal markere hvert eneste ord med ordklasse og bøjnings-

form. Her anbefales det at man bruger ONP's lemmaliste, når man skal afgøre, hvilket ord en bestemt ordform hører under. Dette er fremhævet i del 11.2: "For Medieval Norwegian and Icelandic texts we suggest that the word-list produced by the Arnamagnæan Commission's Ordbog over det norrøne prosasprog at the University of Copenhagen is used to create the lemma base".¹

Menotahåndbogen henviser også til ONP som standard i forbindelse med ordklasser. Som en del af opmærkingsprocessen af en elektronisk tekst skal man ud over bøjningsform og opslagsform også identificere og markere den ordklasse opslagsordet hører under. I del 11.6.3 står der: "The word classes in ONP should be taken as normative". Her har Menota besluttet, at ONP's evaluering af ordformer og gruppering i ordklasser skal være vejledende ved udgivelse af elektroniske tekster.

3.1.2. Bemærkninger om brug af ONP som norm for tekstudgivelse

Vi har set, hvordan Menota anbefaler brug af ONP's normaliserede ortografi, opslagsord og ordklasse. Det er klart, at et netværk som Menota gerne vil facilitere standardisering af alle elektroniske udgaver af middelaldertekster. Det vil fremme videreudvikling af materialet og gøre det muligt at bruge det i forbindelse med forskellige it-løsninger. ONP's status som en nøjagtig videnskabelig leksikografisk ressource og bred dækning af ordforrådet i middelalder sproget gør det oplagt at bruge den som norm i projekter, hvor standardisering af nogle aspekter af sproget eller materialet er nødvendig.

Brugen af ONP som standard er dog ikke uproblematisk, især når det gælder valg af opslagsord og ordklasse men også som ortografisk standard. Ved brug af ONP's ortografi skal man være opmærksom på, at den afviger fra den traditionelle norrøne ortografi, som er blevet brugt i både Island og Norge i de fleste normaliserede tekstudgaver i de sidste årtier. ONP's ortografi prøver at tage hensyn til både norsk og islandsk sprogudvikling, så den undgår nogle særlandske lydændringer og forsøger at harmonisere tegnene for lange vokaler. Denne tilgang har nogle pædagogiske fordele og udgør et godt kompromis, når man skal udvikle standarder, som kan bruges ved udgaver af såvel norske som islandske tekster. ONP's ortografi har dog kun en begrænset tradition i tekstudgaver. Den er desuden løbende blevet revideret, sidst i 2003, hvilket kunne give det indtryk, at den ikke er så fastlagt som andre ortografiske standarder. Den kan dog i princippet godt bruges i tekstudgaver og det er der nogle eksempler på i Menotas arkiv, fx *Alexanders saga* efter håndskriftet AM 519 a 4to, transskriberet af Andrea de Leeuw van Weenen.

¹ ONP er finansieret af den Arnamagnæanske Kommission i København. Ordbogen er også kendt under sit oprindelige navn den Arnamagnæanske Kommissions Ordbog (AMKO).

Når det gælder brug af opslagsord og ordklasser, kan ONP være lidt vanskelig at bruge som norm. ONP er stadig under udvikling og er ikke et færdigt ordbogsværk. Det sker, at ordbogsredaktører tilføjer nye ord, flytter citater rundt eller ændrer klassifikationen af ordene som en del af redigeringsprocessen. Man skal være opmærksom på, at selv om principperne er fastlagte kan ONP løbende ændre mening og praksis. Det skal dog siges at ordbogens redaktører generelt har et meget konservativt syn på ordklassificering og forsøger at ændre så lidt på data som muligt, især når man ved, at det kan have konsekvenser for andre. Alligevel er det nogle gange nødvendigt at ændre ordenes opslagsform eller ordklasse for at få ordbogsstrukturen til at hænge bedre sammen, fx er adverbiet *sáran* 'smertefuldt, bitterligt' blevet nedlagt som selvstændigt lemma og indlejret i en adverbial afdeling under adjektivet *sárr*.

Menota har i øjeblikket ca. 40 middelaldertekster, som er opmarkeret ifølge deres vejledninger. Ikke alle disse tekster følger alle anbefalingerne, men de fleste har fulgt håndbogen vedrørende lemmatisering og normalisering af lemmaform efter ONP's normer. For nogle af teksternes vedkommende har man desuden valgt at bruge ONP's ortografi i selve udgaven på normaliseret tekstniveau. Man kan derfor sige, at brugerne af Menotahåndbogen ved opmarkering af elektroniske tekster også er en slags brugere af ONP, hvis de vælger at overholde Menotas vejledninger og derved ONP's principper.

3.2. ONP som standard i en specialiseret ordbog

Ud over Menota findes der andre projekter, som viser at ONP kan bruges på flere måder, når det gælder norm og standarder. Her vil vi se nærmere på en leksikografisk undersøgelse af planteterminologi i islandske og norske middelalderkilder: *Wörterbuch der Pflanzennamen im Altwestnordischen* af Wilhelm Heizmann, som udkom i 1993, og som giver en omfattende redegørelse for planter og relateret terminologi i norrøne kilder.

I indledningen til planteordbogen forklarer redaktøren principperne for værket og henviser til ONP vedrørende begrænsning af tidsperioden for materialet:

Der zeitliche Rahmen orientiert sich an dem des arnamagnaeianischen Wörterbuchs [AMKO]. Die Obergrenze wurde demzufolge für isl. Texte auf 1540, dem Erscheinungsjahr des ersten gedruckten Buches in Island, festgelegt. Für norw. Texte gilt als Grenze 1370, da zu diesem Zeitpunkt erhebliche Veränderungen innerhalb der norwegischen Sprache zu einer immer deutlicher ausgeprägten Trennung der beiden Sprachen führen (Heizmann 1993:XIV).

ONP bruger forskellige kriterier for kildemateriale fra henholdsvis Island og Norge. ONP's tidsperiode strækker sig fra det ældste skriftlige materiale på pergament i begge lande, som stammer fra midten af 1100-tallet, men for det norske sprogområdes vedkommende har man valgt at sætte grænsen ved 1370. Efter denne periode er de sproglige ændringer af en sådan art, at man ikke længere kan tale om et fælles sprog for Island og Norge. For Islands vedkommende strækker perioden sig indtil 1540, året for udgivelsen af den første trykte tekst. Disse kriterier har Heizmann valgt at følge ved begrænsning af materialet til sin planteordbog.

Heizmann (1993) har også valgt at bruge henvisningssystemet fra ONP i sin planteordbog. ONP har udviklet et særligt system for signaturer, dvs. unikke henvisninger, som refererer til både en bestemt videnskabelig udgave og det middelalderlige håndskrift, den er baseret på. Dette system er præsenteret i registerbindet af ONP som er udgivet i 1989 og diskuteres bl. a. i Jóhannsson & Battista (2018). Signatursystemet egner sig godt som standard, fordi det består af meget detaljerede henvisninger i forkortet format, som gør det nemmere at sammenligne ord og citater. I Heizmanns indledning står der:

Die Prosasiglen beruhen weitgehend auf denen des arnamagnaeianischen Wörterbuchs. Das Neuartige dieser Siglen besteht darin, daß sie auf die den verwendeten Ausgaben der Texte zugrundeliegenden Handschriften verweisen (Heizmann 1993:XLII).

Det betyder at henvisningerne i planteordbogen også giver oplysninger om værk og håndskrift og nemt kan sammenlignes med ONP.

nøtlufræ n. → Med	oleotré n. → Med
nøtlukál n. → Med	olifa f. <i>Ölbaum</i> (<i>Olea europaea</i> var. <i>europaea</i>); → Poet
óðurt/óðyrt f. → Med	<i>Alex</i> 58.23 <þeir lata bera fire ser fagran vond af tre því er olíva heitir.
ófríðarkorn n. <i>Streitkorn, Korn des Unfriedens</i>	þat var ørugt fríðar tacn íþann tíma> [cum Palladis arbore <i>Alexandris</i> IV, 70]; <i>KlmA1980</i> 168.13var: <i>AM 180 a fol</i>
ólafskorn n. <i>Korn, das am Tag der Olafsmesse (29. Juli) als Almosen verteilt wird</i>	<vnder því tre er olífo heiter> [<i>afr.</i> : desous un oliver <i>Ogier 2012</i>] (→ olífi); <i>Stj</i> ^l
<i>BjarkExcYa</i> ^x 89.25var: <i>Steph</i> 17 ^x (→ leifskorn); <i>Bolt</i> 2.4, 2.7, 2.9, 2.11, 2.13, 2.15, 2.16; <i>FrostKr</i> ^l 142.16; <i>JKr</i> 367.11	60.1 <berandi honum í sínum munnni einn blomgadan kuist af þi tre sem oliua
oleaster lat. <i>Wilder Ölbaum</i> (<i>Olea europaea</i> var. <i>sylvestris</i>)	

Figur 1: Afsnit fra planteordbogen (Heizmann 1993:45). Her ses bl.a. ordet *ólafskorn*.

I figur 1 kan man se eksempler på henvisningssystemet, som er blevet brugt i planteordbogen. Under opslagsordet *ólafskorn* findes der fx en henvisning til side 89, linje 25 i *BjarkExcYa*^x og en variant, *leifskorn*, fra en anden kilde, *Steph*

17^x. Disse forkortelser og værksignaturer er i overensstemmelse med ONP og henviser til side og linje i en udgave og også til et bestemt håndskrift. Figur 2 viser en oversigt fra ONP's registerbind hvor disse signaturer er forklaret sammen med de andre fra samme værk. Registerne er også tilgængelige på ONP Online.

Bjark	Bjarkeyjarrétr	
<i>Bjark</i>	<i>Storm 1885 [NGL 4]</i> 72 ^s -74	AM 315 g fol 2r-2v (c1250)
<i>BjarkExcX^x</i>	ed. cit., 76 ²³ -91 ¹⁶ : venstrespalterne ("Xc")	AM 123 c 4 ^{ox} 1r-8v □ (c1600)
<i>BjarkExcYa^x</i>	ed. cit., 76 ⁷⁻¹⁹ , 79 ⁴⁰⁻⁴³ , 80 ²⁹⁻⁹⁰ : højrespalterne, 91 ³⁴ -95 ²⁵ , 95 ³¹⁻⁹⁷ ("Ya")	Kr 1 ^x 312r(27)-329v(64) (c1700-1725)
<i>BjarkExcYc^x</i>	ed. cit., 76 ²⁰⁻²¹ , 76 ²³ , 80 ²⁸ , højrespalterne, 91 ¹⁷⁻³³ , 95 ²⁵⁻³⁰ ("Yc")	OsloUB 695 4 ^{ox} 125r(19)-126r(21), 130r(29), 135v(40) »124v(18)-137v(44)« (c1675-1725) (cf. var...: OsloUB 695 4 ^{ox})
var...: AM 123 b 4 ^{ox}	ed. cit., var.app. ("Xb")	AM 123 b 4 ^{ox} »2r-7v« □ (c1700)
var...: AM 123 d 4 ^{ox}	ed. cit., var.app. ("Xa")	AM 123 d 4 ^{ox} »1r-13r« □ (c1700)
var...: OsloUB 695 4 ^{ox}	ed. cit., var.app. ("Yc")	OsloUB 695 4 ^{ox} »124v(18)-137v(44)« (c1675-1725) (cf. <i>BjarkExcYa^x</i>)
var...: Steph 17 ^x	ed. cit., var.app. ("Yb")	Steph 17 ^x »12v(24)-28r(55)« (c1700-1750)
<i>BjarkAM^x</i>	<i>Keyser & Munch 1846 [NGL 1]</i> 305-315 ¹³	AM 123 a 4 ^{ox} 2r-10v □ (c1675-1700)

Figur 2: Oplysninger om værket *Bjarkeyjarrétr* fra ONP Register (1989). Her kan man se de forskellige signaturer, og hvilke håndskrifter de referer til.

Figur 2 viser registeroplysninger om værket *Bjarkeyjarrétr*, som er en norsk lovtekst læst efter Storms udgave fra 1885. I artiklen fra planteordbogen citeres håndskrifterne *Kr1^x* og *Steph 17^x*, og man kan i registret se, hvordan de bliver brugt i denne udgave. Håndskriftet *Kr1^x* udgør hovedhåndskriftet for en del af teksten (signaturen *BjarkExcYa^x*) og varianter er taget fra andre håndskrifter, bl.a. *Steph 17^x*. Ydermere kan man slå op i ONP's håndskriftsregister for at få nærmere oplysninger om de enkelte håndskrifter, fx om deres oprindelse og indhold.

I planteordbogen kan man se, hvordan ONP's principper bruges som standard i en anden sammenhæng end i Menotahåndbogen. Her er ONP brugt til at danne tidsrammen om materialet, og endvidere er ONP's henvisningssystem eller signaturer også brugt. I sådan et værk giver det god mening at bruge ONP som norm. Tidsbegrænsningen af kildematerialet har været den samme i hele ordbogens levetid og vil ikke blive ændret. Signatursystemet er også forholdsvis stabilt, selvom det er blevet revideret i tidens løb. I princippet er det til stor gavn for brugerne, når man refererer til en middelalderlig tekst, at henvisningen både er til den foreliggende tekstudgave og til håndskriftskilden eller kilderne, som er blevet lagt til grund for udgaven.

3.3. Andre som bruger ONP som standard

Et par andre vigtige videnskabelige projekter, som bruger normer fra ONP i et vist omfang, kan også nævnes her. Det gælder fx *The Skaldic Project* (www.skaldic.abdn.ac.uk), som er en videnskabelig udgave af middelalderlige skjalde-digte, og det nye *Lexicon Poeticum* (LP) (lexiconpoeticum.org), som er en digital ordbog, som dækker det poetiske sprog i samme tidsperiode som ONP. I disse projekter har man brugt en tidligere version af ONP's lemmaliste og suppleret den med de ord, som kun er overleveret i poesi. Desuden anvender man ONP's henvisningssystem, registre og håndskriftsinformationer. I øvrigt er der en del videnskabelige tekstudgaver (med glossarer) og artikler, hvor ONP's principper er omtalt og brugt, cf. fx de Leeuw van Weenen (2004).

4. Konklusion

Vi har i denne artikel diskuteret ONP's brugere og anvendelse af ordbogen. Vi har gjort rede for ONP's principper og forudsætninger for brug af ordbogen og givet eksempler på, hvordan ONP er blevet brugt som norm eller standard i andre projekter. Vores undersøgelse har vist at ONP er velegnet som standard til afgrænsning af materialet og til kildehenvisning. Når man bruger ONP som norm ved tekstudgivelse skal man være opmærksom på at ONP's lemmaliste og ordklassificering stadig er under bearbejdelse. Vores gennemgang viser endvidere, hvordan ONP har betydning for forskning og arbejde med kildemateriale fra perioden og på den måde når ud til mange flere brugere end blot dem, der henter information på ONP's website eller slår ord direkte op i ordbogen.

Litteratur

Ordbøger

- Cleasby, Richard & Gudbrand Vigfusson (1874): *An Icelandic-English Dictionary*. Oxford: Clarendon Press.
- Fritzner, Johan (1886, 1891, 1896): *Ordbog over Det gamle norske Sprog* 1-3, rev. udg. Kristiania: Den norske forlagsforening.
- Heizmann, Wilhelm. (1993). *Wörterbuch der Pflanzennamen im Altwestnordischen*. Berlin, New York: de Gruyter.
- Jónsson, Finnur (1931): *Lexicon poeticum antiquæ linguæ Septentrionalis / Ordbog over det norsk-islandske skjaldesprog*. Opr. udg. af Sveinbjörn Egilsson. 2. udg. København: Kongelige nordiske oldskriftselskab.

ONP = Helle Degnbol, Bent Chr. Jacobsen, James E. Knirk, Eva Rode, Christopher Sanders & Þorbjörg Helgadóttir (red.): *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose*. ONP Registre (1989). ONP 1: a-bam (1995). ONP 2: ban-da (2000). ONP 3: de-em (2004). København: Den Arnamagnæanske Kommission.

ONP Online = *Ordbog over det norrøne prosasprog Online*. <onp.ku.dk> (september 2019).

LP = *Lexicon Poeticum*. <lexiconpoeticum.org> (september 2019).

Anden litteratur

Battista, Simonetta & Ellert Þór Jóhannsson (2014): Ordbog over det norrøne prosasprog 2004-2014 – fra trykt udgave til netversion. I: *LEDA-Nyt* 57, 6-23.

Haugen, Odd Einar (Gen. ed.) (2019): *The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources*. Version 3.0. Bergen: Medieval Nordic Text Archive <www.menota.org/handbook.xml>.

Jóhannsson, Ellert Þór (2019): Integrating analog citations into an online dictionary. I: C. Navarretta, M. Agirrezabal & B. Maegaard (eds.): *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference*, 250-258.

Jóhannsson, Ellert Þór & Simonetta Battista (2016): Ordbog over det norrøne prosasprog Online: struktur og brug. I: Asgerd Gudiksen & Henrik Hovmark (red.): *Nordiske Studier i Leksikografi* 13: Rapport fra 13. Konference om Leksikografi i Norden, København 19.-22. maj 2015, 165-175.

Jóhannsson, Ellert Þór & Simonetta Battista (2018): Middelaldertekster som sproglig ressource. I: Ásta Svavarsdóttir & Helga Hilmisdóttir (red.): *Nordiske Studier i Leksikografi* 14: Rapport fra 14. Konference om Leksikografi i Norden, Reykjavík 19.-22. maj 2017, 152-161.

de Leeuw van Weenen, Andrea (2004): *Lemmatized Index to The Icelandic Homily Book*. Reykjavík: Stofnun Árna Magnússonar á Íslandi.

Menota: *Medieval Nordic Text Archive*. <www.menota.org> (september 2019).

Widding, Ole (1964): *Den Arnamagnæanske Kommissions Ordbog, 1939-1964: Rapport og plan*. Copenhagen: G.E.C.GADS Forlag.

Ellert Þór Jóhannsson
odbogsredaktør, ph.d.
Københavns Universitet
Njalsgade 136
DK-2300 København S
ellert@hum.ku.dk

Simonetta Battista
odbogsredaktør, ph.d.
Københavns Universitet
Njalsgade 136
DK-2300 København S
sb@hum.ku.dk

Omdannelsen af en flersproget til en monolingval ordbog

Halldóra Jónsdóttir & Þórdís Úlfarsdóttir

This article describes the making of *Íslensk nútímamálsorðabók* (*A Dictionary of Contemporary Icelandic*), a dictionary based on the Icelandic part of ISLEX. The multilingual Scandinavian dictionary ISLEX has for several years been accessible on the web, free of charge (islex.is, islex.dk, islex.no, etc.). The dictionary is medium-sized, having Icelandic as a source language and six Scandinavian target languages. After the completion of ISLEX it became clear that its lexicographic material was well suited for other bilingual dictionaries. Furthermore, it seemed practical to develop an Icelandic monolingual dictionary on the basis of ISLEX. The project was created and named *Íslensk nútímamálsorðabók*.

When ISLEX' database was copied in preparation for the new monolingual dictionary, a large part of the contents had already been completed. What was mainly missing were good quality definitions, as ISLEX only contained rudimentary Icelandic definitions intended as an aid to the editors of the target languages. For the last few years the editorial work of the new dictionary has mainly consisted of revising and rewriting the Icelandic definitions.

1. Indledning

I denne artikel bliver der gjort rede for tilblivelsen af *Íslensk nútímamálsorðabók* (*Ordbog over moderne islandsk*), en ordbog der er baseret på den islandske del af den skandinaviske webordbog ISLEX.

Ordbogen er den første islandske monolingvale ordbog i over halvtreds år der ikke er baseret på en tidligere udgave af en eksisterende trykt ordbog. Den er desuden den anden monolingvale islandske ordbog der er udkommet overhovedet. Hovedvægten i det leksikografiske arbejde lægges på at beskrive islandsk som det tales og skrives omkring og efter året 2000.

Da arbejdet med *Íslensk nútímamálsorðabók* påbegyndtes tidligt i 2014, forelå der omfattende materiale fra ISLEX: en lemmaliste på 50 tusinde ord, kollokationer, faste ordforbindelser, sprogbrugseksempler m.m. Der var oplysninger om verbers og præpositioners kasusstyring samt interne links inde i selve ordbogen. Der fandtes endvidere en del illustrationer og opslagsordenes udtale var til stede i form af lydfiler. Desuden var opslagsordene blevet delt op efter betydning. Den vigtigste og mest tidskrævende del var dog ikke afsluttet, men

det var at skrive islandske ordforklaringer, én eller flere for hvert lemma. De definitioner som forelå i ISLEX var oprindelig tænkt som en hjælpetekst til de skandinaviske redaktører som arbejdede med målsprogene, og teksten var ikke altid af god kvalitet da den ikke var tænkt til offentliggørelse.

Ordforklaringerne skulle revideres eller skrives om hvor det var nødvendigt. Da denne proces var kommet godt i gang blev ordbogen lanceret i 2016 som ”work in progress”, men arbejdet fortsatte indtil værket fik sin endelige form. Denne arbejdsmetode blev valgt gå grund af at webben har ændret folks indstilling til informationsformidling, og på webben finder man ofte materiale som ikke er hundrede procent færdigt. Ikke desto mindre ligger der nyttige oplysninger som er oplagt at gøre tilgængelige til trods for eventuelle mangler.

I de tilfælde hvor en bruger finder en fejl i ordbogen, så har han/hun en mulighed for at indsende en melding, og her kan man tale om at der sker en slags ”crowdsourcing” efter offentliggørelsen. *Íslensk nútímamálsorðabók* er gratis tilgængelig på webben (www.islenskordabok.is).

2. ISLEX-projektet

Den nye monolingvale ordbog henter således sit grundmateriale fra ISLEX, en webordbog med islandsk som kildeprog og seks nordiske målsprog: dansk, bokmål, nynorsk, svensk, færøsk og finsk. ISLEX-projektet blev grundlagt i 2005 som et samarbejdsprojekt mellem seks institutioner:

- Árni Magnússon-instituttet for islandske studier, Reykjavík
- Det Danske Sprog- og Litteraturselskab, København
- Institutt for lingvistiske, litterære og estetiske studier, Universitetet i Bergen
- Institutionen för svenska språket, Göteborgs universitet
- Fróðskaparsetur Føroya, Tórshavn
- Helsingfors universitet

Et stort antal islandske og nordiske medarbejdere bidrog til ISLEX og var med til at forme ordbogens indhold. ISLEX var det første islandske ordbogsværk der udelukkende var designet til digital udgivelse, og var samtidig den eneste eksisterende islandske ordbog som forbandt mange sprog. ISLEX blev åbnet i slutningen af 2011 med målsprogene dansk, bokmål, nynorsk og svensk; senere blev færøsk (2015) og endelig finsk (2018) tilføjet. Figur 1 viser et eksempel fra ordbogen.



Figur 1: Et eksempel fra ISLEX-ordbogen, her med svensk som målsprog.

ISLEX' infrastruktur tager udgangspunkt i en beskrivelse af islandsk som kilde-sprog, og under ordbogsarbejdet blev der indsamlet et righoldigt materiale om islandsk. Det blev hurtigt klart at denne udførlige beskrivelse af moderne islandsk var oplagt at anvende til flere projekter, og ordbogen var struktureret således at det var nemt at omforme materialet til en etsproget, tosproget eller flersproget ordbog.

3. Udformning af den nye ordbog

Det behøver ikke at være en teknisk indviklet proces at omdanne en flersproget ordbog til en etsproget med udgangspunkt i kildesproget, men det forudsætter selvfølgelig en del programmeringsarbejde. Processen sker i nogle trin. Først må databasen kopieres for at danne en ny database for den nye ordbog. I dette tilfælde måtte man smide målsprogene ud samt tilknyttet materiale, så det eneste der var tilbage var det islandske kildesprogsmateriale. Figur 2 illustrerer en ordbogsartikel med det islandske materiale.

Ordbogsredaktionen kunne nu begynde sit arbejde. De midlertidige ordfor-klaringer fra ISLEX skulle revideres og/eller skiftes ud for nye, og nogle gange måtte der tilføjes flere betydninger. Der blev desuden tilføjet nyt materiale, f.eks. flere sprogbrugseksempler og ikke mindst flere lemmaer.

grein subst f

🗣️ **uttale**
→ **BØYING**

1
(trjágrein)
langur, trékenndur útvöxtur á tré, trjágrein

2
(ritsmíð)
ritsmíð, lítill ritgerð, einkum í blaði eða tímariti, sem fjallar um tiltekið efni
grein um bókmenntir

3
(námsgrein)
nám, kennsla og fræði um tiltekið efni, námsgrein, fag
hún var hæst í þremur greinum á lokaprófinu

4
(lagagrein o.þ.h.)
eitt atriði í lista með samfelldum texta, t.d. í reglugerðum og lögum (lagagrein)
20. grein fjallar um tollamál

gera grein fyrir <málinu>
útskýra <málið>

gera sér grein fyrir <ástandinu>
skilja og þekkja ástandið

Figur 2: En ordbogsartikkel hvor målsprogene er forsvundet og kun det islandske indhold vises.

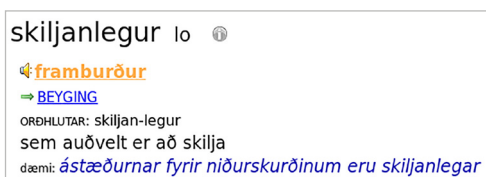
Under redigeringsprocessen er der undervejs blevet afsløret lakuner og mangler i det oprindelige værk, ISLEX-ordbogen, som der måtte rettes op på. Redigeringen har på den måde også haft en tilbagevirkende kraft på ISLEX, som er blevet opdateret samtidig med arbejdet med den monolingvale ordbog.

4. Materialet i den nye ordbog

I dette afsnit vil vi kort gennemgå den nye ordbogs vigtigste komponenter, i makrostrukturen såvel som i mikrostrukturen.

4.1. Lemmaerne

Ordbogens lemmaliste består af 50.000 ord og afspejler sammenhængen i det islandske sprog. Der er således en del norrøne ord, ord fra ældre sprog og poetisk sprog inkluderet, og der findes endvidere fagsprog til en vis grad, f.eks. juridisk sprog, geologi og kemi. Figur 3 viser et tilfældigt lemma fra ordbogen.



Figur 3: Et eksempel fra *Íslensk nútímamálsorðabók*.

Da redigeringsarbejdet indledtes blev det hurtigt klart at der måtte tilføjes nyt ordforråd. De nye ord er især kommet ind ad følgende veje: 1) fra omgivelserne og redaktørernes løbende diskussioner, 2) ved excerpering af tekster i tilfælde af kendte lakuner i ordforrådet, f.eks. i fag som geologi og medicin, og 3) sidst men ikke mindst arbejdes der med frekvenslister fra korpuser. Med alle disse faktorer er lemmalisten vokset med 4.500 nye ord. I de tilfælde hvor der har manglet fraser så er de også blevet tilføjet løbende.

4.2. Øvrige indhold

I de følgende afsnit diskuteres nærmere ordbogens indhold og dens vigtigste bestanddele. Ligesom i de fleste digitale ordbøger består mikrostrukturen af *Íslensk nútímamálsorðabók* af følgende komponenter:

- Islandske definitioner
- Kollokationer og ordforbindelser
- Sprogbrugseksempler
- Orddeling i sammensatte ord
- Bøjning
- Oplysninger om kasusstyring af verber og præpositioner
- Illustrationer
- Udtale som lydfile
- Semantisk inddeling

4.2.1. Definitioner

Definitioner er en grundlæggende komponent i en monolingval ordbog og de skal være af god kvalitet. *Íslensk nútímamálsorðabók* indeholder 62.400 definitioner. Opslagsordene er 50.000, men forskellen beror på at mange ord har flere end én betydning, foruden at de faste udtryk også bliver forklaret. De ord

som har flest betydninger, og dermed de fleste forklaringer, er (efter frekvens) verberne *taka* ('tage'), *koma* ('komme') og *fara* ('gå'), men de har hver for sig dusinvis af betydninger. De navneord som har flest betydninger er *brot* ('brud'), *efni* ('emne'), *hlaup* ('løb'), *mál* ('mål') og *mark* ('mark').

Definitionerne har sin oprindelse i den islandske del af ISLEX-projektet, hvor otte forskellige redaktører havde arbejdet på indholdet og af den grund var de stilmæssigt ret forskellige. Eftersom de først og fremmest blev tænkt som en hjælpetekst for oversætterne var de af meget forskellig kvalitet da de ikke skulle være synlige for brugerne. Der fandtes således definitioner til alle opslagsordene, men det viste sig at mange af disse definitioner var ubrugbare. Opslagsordene måtte gennemgås kritisk, revideres, koordineres og omformuleres inden de blev offentliggjort. Endelig fik alle faste udtryk en forklaring da betydningen tit er ugenomsigtig og ikke fremgår af ordlyden.

I *Íslensk nútímamálsorðabók* bliver der gået ud fra at brugeren er fortrolig med det grundlæggende ordforråd i islandsk eftersom ordene forklares på islandsk. Ordbogens redaktører har haft følgende målgrupper i tankerne som f.eks. islandske skoleelever, udlændinge som lærer islandsk som fremmedsprog, såvel som almene brugere.

4.2.2. Kollokationer og ordforbindelser

En af ordbogens stærkeste sider er det store antal kollokationer og faste udtryk (14.500), herunder 5.000 fraser med verber. Kollokationernes betydning er oftest transparent: *rød i hovedet*, *en udslukt vulkan*, *gå i kloster*.

Betydningen af andre typer faste ordforbindelser kan være helt utransparente da kernen ofte har lidt eller ingenting at gøre med selve ordlyden: *vaxa úr grasi* ('vokse op'), *láta greiðar sópa* ('rage til sig'). Der kan være tale om en metafor og i nogle tilfælde idiomer.

Naveord som indeholder de fleste ordforbindelser er (efter frekvens): *hönd* ('hånd'), *fötur* ('fod'), *mál* ('sag, sprog'), *orð* ('ord'), *auga* ('øje'), *maður* ('mand'), *vegur* ('vej'), *dagur* ('dag'), *líf* ('liv') og *ráð* ('råd'). Sammenlagt findes der i ordbogen omkring 10.000 forskellige slags ordforbindelser.

4.2.3. Sprogbrugseksempler

Ordbogen indeholder mange sprogbrugseksempler og deres rolle er at vise opslagsordet i en typisk sætningssammenhæng og desuden at understrege ordets forskellige betydninger, i de tilfælde hvor ordet har flere end en betydning. Der er ikke tale om autentiske citater med kildehenvisninger, som man kender fra mange andre ordbøger, sprogbrugseksemplerne er derimod specielt konstruerede for ordbogen. En mere udførlig redegørelse for sprogbrugseksemplerne finder man hos Thórdís Úlfarsdóttir (2014).

4.2.4. Orddeling i sammensatte ord

Ligesom de andre nordiske sprog af germanisk oprindelse er islandsk rigt på sammensatte ord. Når opslagsordene er sammensatte vises ordet igen med mindre typer med en bindestreg mellem leddene. F.eks. *samnings-tímabil* ('kontrakt-periode'), *ffallgöngu-maður* ('bjerg-vandrer'), *bak-herbergi* ('bag-lokale'). Dette bliver gjort for at illustrere ordenes led, foruden at det kan fungere som hjælp til rigtig orddeling mellem tekstlinjer.

4.2.5. Bøjning

Opslagsordenes bøjning fremstår med links til bøjningsparadigmerne i BÍN (*Beygingarlýsing íslensks nútímamáls*), en database med bøjninger af flere hundrede tusinde islandske ord. Når man trykker på linken *bøjning* åbnes ordets bøjningsparadigme i sin helhed. Alle ord som bøjes har links: navneord, tillægsord, udsagnsord og stedord. For en del ords vedkommende vises desuden bøjninger i forkortet form i selve ordbogsartiklen.

4.2.6. Kasusstyring

Da islandsk er et bøjningssprog er det i visse tilfælde nødvendigt at give oplysninger om kasusstyring (valens). Dette gælder selvfølgelig for verber og præpositioner, som kan tage med sig forskellige kasus alt efter funktion. Som et eksempel kan nævnes verbet *svara* ('svare') som tager to kasus i dativ: *svara henni* (dat.) *engu* (dat.) ('svare hende ingenting', 'give hende ikke noget svar'). Præpositioner kan tage tre forskellige kasus, akkusativ, dativ eller genitiv, f.eks. *um* + akk. ('om'), *frá* + dat. ('fra') og *til* + gen. ('til'). Disse oplysninger er af stor hjælp for sprogbrugerne, ikke mindst de som lærer islandsk sem fremmedsprog.

4.2.7. Illustrationer

Illustrationer har længe eksisteret i ordbøger, især for unge læsere, og *Ordbog over moderne islandsk* indeholder 3.200 illustrationer. I nogle tilfælde kan et billede forklare et ords indhold bedre end man kan gøre med ord (f.eks. *egern*, *saks*, *tandhjul*, *trådrolle* og *tøjklemme*). Nogle gange kan det også være praktisk at have billeder for at afdække et ords mange betydninger, jfr. ordet *keila* som bl.a. betyder fisken 'brosme' og 'trafikkegle'. Billederne er af forskellig oprindelse, men ca. 700 stammer fra illustratoren Jón Baldur Hlíðberg, hovedsagelig af naturfænomer: planter, fugle, fisk og flere dyr. Desuden findes der i ordbogen en del fotografier og tegninger.

4.2.8. Udtale

Opslagsordenes udtale på lydfiler er en af de komponenter der blev taget i arv fra ISLEX. Det er muligt at lytte til udtalen ved at klikke på et ikon neden for selve lemmaet. Da læseren indlæste ordene i alfabetisk orden viste der sig en tendens til at oplæsningen blev monoton, noget som ikke var ønskværdigt når det drejede sig om udtalen af enkelte ord. Af den grund blev ordene læst i en tilfældig orden og dermed forsvandt monotonien. Foruden de 50.000 opslagsord blev der indlæst godt 700 faste udtryk som forekommer i ordbogsartiklerne.

4.2.9. Semantisk inddeling

Ordbogens samtlige ordforråd er blevet inddelt semantisk og redaktørerne har arbejdet efter den inddeling fremfor den traditionelle alfabetiske orden. Store semantiske grupper er f.eks. *beskæftigelse, indbo, dyr, tøj, vind, vej, geologi* (hovedsagelig navneord); *alder, tilfredshed, tilstand, uduelighed, sundhed, farve, form* (hovedsagelig tillægsord); og *ændring, død, ophold, bevægelse, tale* (hovedsagelig udsagnsord). De fleste lemmaer tilhører flere end et semantisk felt. Den semantiske inddeling er tænkt som et redaktionelt værktøj og er af den grund ikke synlig for brugeren.

Grupperne er af forskellig størrelse, hvor de største indeholder over 1.000 ord, men de mindste kun nogle dusin. Denne arbejdsmetode var en nyhed i islandsk ordbogsarbejde, og har vist sig at være af stort gavn for redaktørerne, bl.a. for at sikre bedre overensstemmelse i behandlingen af semantisk beslægtede ord, hvor der målrettet kan gøres rede for sproglige nuancer.

5. Brugen af korpuser

En af sprogteknologiens vigtige opgaver er dannelsen af tekstkorpuser, en udvikling som ordbogsfolk har høstet frugterne af. Nu er det blevet normen at ordforrådet m.m. i nye ordbøger er baseret på korpuser, og islandske ordbøger er her ingen undtagelse. Man kan endda sige at det er utænkeligt at indlede et ordbogsprojekt uden at anvende korpuser.

Oplysninger som man finder i korpuser er, bl.a.:

- Ordnes frekvens
- Ordnes alder, eller første registrering
- Betydningsbestemmelse, én eller flere

- Sprogligt register (fx slang, formelt)
- Fraser og forskellige faste udtryk

I de seneste to årtier er der sket en hastig udvikling af islandske tekstkorporer, hvoraf de vigtigste er: *Íslenskt textasafn*, *tímarit.is* og *Risamálheildin* (2018). Ovennævnte tre korporer er indbyrdes forskellige. Den førstnævnte indeholder mangeartede teksttyper, mens de to sidstnævnte hovedsagelig består af store mængder nyheds- og dagbladsstof. Alle disse korporer er til gavn for leksikograferne for at udvide deres forståelse for leksikalske enheder.¹

6. Modtagelser og opdateringer

Arbejdet med Ordbog over moderne islandsk påbegyndtes i januar 2014 og der blev åbnet for adgang til ordbogen sent i 2016. På dette tidspunkt var dog visse elementer stadig under bearbejdning, og ordbogens første udgave blev ikke klar før end i marts 2019.

Íslensk nútímamálsorðabók har sin egen hjemmeside (www.islenskordabok.is), og er samtidig tilgængelig på portalen www.málið.is (jfr. Steinþór Steingrímsson et al. 2018a) sammen med andre opslagsværker om det islandske sprog.

Ordbogens modtagelser har været over forventning og i 2018 var brugerne 130.000, som må anses for meget godt hvis man tager hensyn til det antal mennesker som taler islandsk. Brugergrænsefladen har nu fået nyt design som blev åbnet i slutningen af 2019.

I begyndelsen var der 50.000 opslagsord i *Íslensk nútímamálsorðabók*, ligesom i ISLEX, men efter at første version af ordbogen blev åbnet var redaktørerne blevet klare over at der skulle tilføjes en del nye ord. Det nye ordforråd stammer hovedsagelig fra det nyeste islandske tekstkorpor, *Risamálheildin* (2018), og i 2020 bliver der tilføjet 4.500 nye lemmaer som skal redigeres. Det er klart at den omfattende brug af ordbogen giver et fingerpeg om at der er stor efterspørgsel efter et sprogværktøj, der opdateres løbende og er gratis tilgængeligt på webben.

¹ Nærmere oplysninger om *Risamálheildin* er i Steinþór Steingrímsson et al. (2018b).

Litteratur

Ordbøger

- Ordbog over moderne islandsk (Íslensk nútímamálsorðabók)*. Halldóra Jónsdóttir & Þórdís Úlfarsdóttir (red.). Árni Magnússon-instituttet for islandske studier. <www.islenskordabok.is> (november 2019).
- ISLEX-ordbogen*. Þórdís Úlfarsdóttir (red.). Árni Magnússon-instituttet for islandske studier. <www.islex.is> (november 2019).

Anden litteratur

- BÍN = Kristín Bjarnadóttir (red.): *Beygingarlýsing íslensks nútímamáls*. Árni Magnússon-instituttet for islandske studier. <www.bin.arnastofnun.is> (november 2019).
- Risamálheildin* (2018). Árni Magnússon-instituttet for islandske studier. <malheildir.arnastofnun.is> (november 2019).
- Steinþór Steingrímsson, Sigrún Helgadóttir & Eiríkur Rögnvaldsson (2018b): An Icelandic Gigaword Corpus. I: *Nordiske Studier i Leksikografi* 14. Rapport fra 14. Konference om Leksikografi i Norden (NFL) i Reykjavík 30.5-2.6.2018: 246-254.
- Steinþór Steingrímsson, Ari Páll Kristinsson & Halldóra Jónsdóttir (2018a): Málið.is. I: *Nordiske Studier i Leksikografi* 14. Rapport fra 14. Konference om Leksikografi i Norden (NFL) i Reykjavík 30.5-2.6.2018: 255-261.
- Þórdís Úlfarsdóttir (2014): ISLEX—a Multilingual Web Dictionary. I: *Proceedings from LREC 2014, Ninth International Conference on Language Resources and Evaluation*. Reykjavík 26-31 May 2014.
- Timarit.is*. Tímaritavefur Landsbókasafns Íslands – Háskólabókasafns. <www.timarit.is> (november 2019).
- Þórdís Úlfarsdóttir (2013): ISLEX - norræn margmála orðabók. I: Orð og tunga 15. Reykjavík: Árni Magnússon-instituttet for islandske studier.
- Þórdís Úlfarsdóttir (red.): *Íslenskt textasafn*. Árni Magnússon-instituttet for islandske studier. <corpus.arnastofnun.is> (november 2019).

Halldóra Jónsdóttir
ordbogsredaktør
Árni Magnússon-instituttet
for islandske studier
Laugavegur 13
IS-101 Reykjavík
halldora.jonsdottir@arnastofnun.is

Þórdís Úlfarsdóttir
ordbogsredaktør
Árni Magnússon-instituttet
for islandske studier
Laugavegur 13
IS-101 Reykjavík
thordis.ulfarsdottir@arnastofnun.is

Total bruklighetsinskränkning i SAOB – nu och i framtiden

Lennart Larsson

The Swedish Academy Dictionary (SAOB) will be finished in a couple of years, and the editorial staff is already planning a future updated version of the dictionary. The fact that the work with the dictionary has spanned over a century entails that the need for revision is significant. In this article the problems associated with the diafrequent markings to the dictionary are discussed, and it concludes with the proposal of a new order, characterized by a considerable reduction of the number of different markers as well as explicit and precise criteria for their use.

1. Bakgrund

Om några år kommer Svenska Akademiens ordbok (SAOB), efter att i långt över ett sekel ha betat av den ena bokstaven efter den andra, att vara färdig med Ö. Detta betyder emellertid inte att arbetet är över. Förberedelserna för den nödvändiga revideringen är redan på gång, och på redaktionen finns ett antal arbetsgrupper som nagelfar ordboken ur olika aspekter.

Till det som tveklöst pockar på att revideras hör de många bruklighetsangivelserna. I SAOB finns en rik och brokig flora av sådana angivelser, inte bara när det gäller frekvens utan också stilnivå, regionala inskränkningar, fackspråkighet m.m. Det som kommer att diskuteras här är dock bara de angivelser som i SAOB:s språkbruk går under benämningen ”total bruklighetsinskränkning”, alltså användningen eller förekomsten överhuvudtaget av ett visst ord eller en viss betydelse av ett ord.¹

2. Nuläget

I nuläget har SAOB ett helt spektrum av olika totala bruklighetsinskränkningar, alltifrån dödkorset som anger att ordet eller betydelsen är helt obsolet och ner till den lindrigaste inskränkningen ”bl[ott] mera tillf[älligt]”. Oftast är angivelserna

¹ I Mattisson (2013) finns en mera övergripande granskning av bruklighetsstrukturen i SAOB, medan Nilsson & Rosqvist (2018) ger en genomlysning av problematiken kring fackspråksbeteckningarna.

(förutom dödkorset) också försedda med ett inledande ”numera” för att markera att brukligheten gäller i vår tid (eller mera exakt: den tid då artikeln skrevs):

- (†)
- (knappast br.) / (numera knappast br.)
- (föga br.) / (numera föga br.)
- (mindre br.) / (numera mindre br.)
- (ngn gg) / (numera bl. ngn gg)
- (tillf.) / (numera bl. tillf.)
- (mera tillf.) / (numera bl. mera tillf.)

Den här uppsättningen av bruklighetsinskränknings har i stort sett varit konstant ända från ordbokens begynnelse på 1890-talet: de två sista, ”tillf.” och ”mera tillf.”, slog igenom under bokstaven B, medan de övriga är väl belagda redan under A. Dessutom fanns det tidigare ännu en frekvent uppträdande angivelse inom den här kategorin, en angivelse som försvann ca 1975 en bit in på bokstaven S. Det gäller ”(enst.)”, i ordbokens förkortningsordlista förklarad som ”enstaka förekommande”. Ofta var den också – som man kan se redan i det allra första exemplet ADALBOREN (figur 1) – kombinerad med ett dödkors.

1) (enst., †) gm födelsen berättigad till egande af odaljord. Odol betydher ens Egne, frije, Arfuegodz, och Adalboren är then som sådana godz äger, och äga må. **STIERNHIELM Fateb. A 3 a (1643).**

2) (enst.) som har goda anor, ädel. Två hjeltebarn af adalboren ätt. **WIRSÉN Dikter 247 (1872).** — jfr ADEL-BOREN.

Figur 1: Definitionsdelen av artikeln ADALBOREN från saob.se.

För att komplicera saken ytterligare är det här de angivelser som kan stå inledningsvis i definitionerna, men det kan också finnas totala bruklighetsinskränknings inne i definitionerna och i sådana fall är de delvis annorlunda formulerade. Där kan man träffa på ”förr särsk.” eller ”förr äv.” som motsvarar dödkorset och ”i sht förr särsk.” eller ”i sht förr äv.” motsvarande ”föga br.” eller ”mindre br.”. Likaså kan man inne i definitionen möta ”särsk. (o. numera bl.)” eller ”äv. (o. numera bl.)”, vilket då betyder att det som sedan följer är den enda nu levande användningen, och ”särsk. (o. numera nästan bl.)” eller ”äv. (o. numera nästan bl.)” med innebörden att det som tidigare presenterats numera är föga eller mindre brukligt.

2.1. Vad är problemen?

Som allt annat i SAOB har bruklighetsangivelserna sina problematiska sidor. Och om man skulle försöka sammanfatta problemen när det gäller de totala bruklighetsinskränkningarna kan man nog göra det i tre punkter som är nära sammankopplade med varandra, nämligen:

- onödigt många
- inte särskilt precisa
- inte särskilt genomskinliga

Man kan helt klart fråga sig om det verkligen behövs så här många olika angivelser. Och vad står de för egentligen? Är det så självklart att t.ex. ”mindre br.” innebär en större bruklighet än ”föga” eller ”knappast br.”? Eller att ”mera tillf.” faktiskt är mindre tillfälligt än bara ”tillf.”? Och står den här rangordningen klar för användarna? Svaret kan knappast bli annat än nekande.

En annan sak man kan undra över är om man menade detsamma med angivelserna för 100 år sedan som idag. Det finns en manual för oss redaktörer utarbetad av den tidigare ordbokschefen Hans Jonsson från 1993 (med en andra upplaga 1999), och där finns bland mycket annat också mer eller mindre preciserade anvisningar för de olika bruklighetsangivelsernas användning. Mest specifik är den för ”numera knappast br.”, som har följande lydelse:

Inskränkningen ”numera knappast br./knappast br.” ger intrycket av en stegring av ”numera föga br./föga br.”. Men det är bara halva sanningen. Den används när jämförelsevis sena belägg föreligger i en ordbok (i ordböcker) eller liknande källa (källor) trots att ordet eller ordanvändningen bedöms som (praktiskt taget) obrukligt resp. obruklig. Man reserverar sig således för att någon språkbrukare inspirerad just av ordboken osv. i fråga använder ordet i nuspråket. (Jonsson 1999:65)

Så ser alltså kriteriet ut idag. Men det gällde knappast för hundra år sedan, och frågan är nog om inte valet mellan *mindre*, *föga* och *knappast brukligt* tidigare åtminstone till en del handlade om vilket uttrycksätt som den enskilde redaktören föredrog, alltså att det snarast var en fråga om tycke och smak i ordvalet.

Så sammanfattningsvis: bruklighetsangivelserna (och det gäller för övrigt inte bara de totala bruklighetsinskränkningarna utan i mångt och mycket också de andra kategorierna) kunde med fördel bli betydligt färre, mer precisa och mer lättbegripliga.

2.2. Två komplikationer

Innan jag presenterar ett konkret förslag på en nyordning ska jag peka på två speciella egenheter hos just bruklighetsangivelserna.

Den ena egenheten gäller deras tidsberoende. Även om allt i ordboken har en given förankring i tiden är det ändå lite speciellt med bruklighetsangivelserna. Om man tar t.ex. definitionerna är de ju i princip giltiga – och förblir giltiga också i framtiden – även om det ord eller betydelsemoment som definitionen gäller inte längre är brukligt; i en historisk ordbok har även definitioner av äldre betydelser sin plats, de behåller sin språkhistoriska relevans. Detsamma är fallet med språkproven. SAOB har ju som princip att alltid ge det första kända be- lägget för ett visst ord eller en viss betydelse av ett ord. Men även om man vid revideringen med all säkerhet kommer att hitta mängder av nya förstabelägg för ord och betydelsemoment så är ändå de språkprov som finns fortfarande lika giltiga som vittnen om sin tids språkbruk – även om de då förlorat sin speciella status av förstabelägg.

Med bruklighetsangivelserna är det emellertid en annan sak: de berättar om hur brukligheten tedde sig vid den tidpunkt då artikeln skrevs. Och i och med att ordboksarbetet har pågått i drygt 125 år så är det självklart att man hittar otaliga exempel på bruklighetsangivelser som (i alla fall förhoppningsvis) gällde när artikeln skrevs men som helt klart inte gäller idag.

Det här kan man förstås ge hur många exempel som helst på. Jag ska nöja mig med ett, som inte är särskilt iögonfallande eller exceptionellt men som helt enkelt råkar vara det senaste som jag stött på i mitt arbete: det gäller den s.k. särskilda förbindelsen, alltså partikelverbet, BÄRA ÖVER, tryckt 1925 (figur 2).

<p>BÄRA ÖFVER^{10 40}.</p> <p>1) (numera bl. tillf.) till 1: transportera l. flytta öfver; förr äfv.: öfverbringa. STENBOCK (o. OXENSTIERNA) <i>Brefv.</i> 1: 27 (1690). <i>Hamb.</i> (1700).</p> <p>2) [jfr d. <i>bære over sit hjerte</i> o. d., t. <i>über's herz bringen</i>] (†) i uttr. <i>icke kunna bära öfver sitt samvete, icke hafva samvete</i> (att göra ngt). WESTE (1807).</p> <p>3) [af d. <i>bære over med</i>; jfr fsv. <i>ivirbära</i>, mnt. <i>overdragen mit</i>] (enst.) i förb. <i>bära öfver med någon</i>, hafva fördrag l. öfverseende med ngn. En hustru skall ej upphäfvä sig till domare öfver sin man, utan där svaghet finnes, i kärlek bära öfver med hans svaghet. GEIJERSTAM <i>Strömoln</i> 129 (1883).</p>

Figur 2: Artikeln BÄRA ÖVER från saob.se.

Det första betydelsemomentet i BÄRA ÖVER har alltså bruklighetsinskränknin- gen ”numera bl. tillf.”. Lite märkligt kan man tycka, eftersom det handlar om den för oss självklara användningen av typen ”hon bar över kartongerna till gara- get”. En förklaring kunde ha varit att man tidigare, när artikeln kom till i början av förra seklet, föredrog sammansättningen *överbära*. Men det håller knappast; ett av flera indicier på detta är att Dalins ordbok från 1855 förklarar *överbära* med just ”Bära öfver till ett ställe”.

Som sagt: det finns massor av exempel på bruklighetsangivelser som ur våra dagars perspektiv ter sig mer eller mindre felaktiga. I det här fallet var det kanske en missbedömning av redaktören, men i de allra flesta fall handlar det nog helt enkelt om att tiden sprungit ifrån den bruklighet som gällde när artikeln skrevs.

I detta sammanhang kommer man oundvikligen in på en frågeställning som är central för det fortsatta ordboksarbetet efter att den första versionen har nått sitt slut. SAOB har ju genom hela sin historia beskrivit svenskans ordförråd från 1521 och fram till sin nutid, till ett ”numera” som en gång var 1890-talet och i skrivande stund är 2019. Och frågan är om man ska fortsätta som förut, dvs. gå ända fram till en nutid som ständigt förflyttar sig framåt, år efter år och decen- nium efter decennium, eller om man ska ha ett slutårstal, förslagsvis 2020 eller där omkring, kanske 2023 då SAOB:s första upplaga kan antas vara fullbordad. (En sådan tidsgräns skulle ju också ha den tjugusiga poängen att ordboken då kom- mer att sträcka sig över mer eller mindre exakt 500 år.) För min personliga del är jag en avgjord anhängare av ett slutårstal. Och det är bland mycket annat på grund av det här med brukligheterna. Om man inte har något slutårstal kommer bruklighetsangivelserna att fortsatt vara beroende av vilket år artikeln skrevs eller reviderades; och de löper en ständig risk att hamna i otakt med sin tid. Har man däremot ett slutårstal, typ 2023, så är det brukligheten det året som ordboken anger, varken mer eller mindre. (Därmed alls inte sagt att slutårtalet behöver innebära slutpunkten för SAOB. Snarare bör den här modellen betraktas som ett modus operandi, och när ordboken om ett antal decennier föreligger i en uppdaterad, genomarbetad och någorlunda enhetligt utformad andra upplaga finns det förstås god anledning att ta itu med en tredje upplaga.)

Den andra egenheten med bruklighetsangivelserna – som exemplet BÄRA ÖVER (se figur 2) kanske kan sägas vara en illustration av – är deras avsevärda inslag av subjektivitet: det handlar ganska mycket om hur den enskilda redaktö- ren bedömer saken. I övrigt kan ju ordboken göra anspråk på en förhållandevis hög grad av objektivitet och verifierbarhet. De böjnings- och stavningsformer vi anger är de som finns i det bevarade språkmaterialet², de definitioner vi formu- lerar bygger på de autentiska språkproven, och genom att vi återger ett urval av

² Dock finns det undantag. I fråga om stavningen har det ibland hänt att de normerande aspekterna – fastlagda i SAOL – tillåtits fördunkla beskrivningen av det reella språkbruket (se Larsson 2012).

språkproven kan också användaren kritiskt granska definitionerna i förhållande till språkmaterialet. Men när det gäller brukligheten så är det i hög grad redaktörens goda omdöme man får förlita sig på; även om vi har kriterier att rätta oss efter så är de ganska så vaga och lämnar stort utrymme för subjektiva inslag.

3. I framtiden

För att komma tillrätta med de problem som beskrivits ovan är det framför allt två saker som behövs (förutom då ett slutårstal): dels en ordentlig minskning av antalet olika angivelser, dels mer handfasta och explicita kriterier för när de ska användas. Här har vi som ingår i den s.k. bruklighetsgruppen vid redaktionen³ laborerat med diverse olika modeller, modeller som steg för steg gått i riktning mot en allt större förenkling och reducering av antalet angivelser. Och i vårt senaste förslag återstår bara två olika angivelser för inskränkt total bruklighet, nämligen de följande:

- ej belagt i nusvenska (ej efter 1####)
- sällsynt i nusvenska

En grundläggande fråga härvidlag blir förstas hur man avgränsar nusvenskan, något som på senare år diskuterats av Thelander (1987), Hellberg (2005) och Malmgren (2007). Nusvenskan måste ju – som Thelander uttrycker det (1987:13) – vara ett ”rullande” begrepp, något som hela tiden flyttar fram sin bakre gräns. Och en tanke som Malmgren för fram (2007:179) är ”att nusvenskans ålder knappast vid någon tidpunkt är högre än 50 år”.

Går man så 50 år tillbaka i tiden hamnar man kring 1970. Och vårt förslag är att vi ska stanna vid 1973, det år då Svenska Akademiens ordlista, SAOL, utkom i sin tionde upplaga. Grunden för detta är inte bara att de olika upplagorna av SAOL alltmer kommit att te sig som symboliska milstolpar i ordförrådets utveckling. Minst lika viktigt är det faktum att ordböcker har en tendens att halka efter det reella språkbruket: i SAOB finns det mycket gott om fall där sistabeläget för ett bruklighetsinskränkt ord eller en bruklighetsinskränkt betydelse härstammar just från en ordbok. Dessutom skulle den här gränsen också täcka in två andra viktiga ordböcker som ofta uppträder som källa i SAOB, dels Östergrens Nusvensk ordbok vars tionde och sista band avslutades 1972, dels Illustrerad svensk ordbok som kom i sin tredje och sista reviderade version 1964.

³ De övriga medlemmarna i gruppen har varit Bo-A. Wendt och Antti Sivén; det förslag till nyordning som här presenteras är i minst lika hög grad deras förtjänst.

Så den första angivelsen skulle alltså kort och gott betyda att ordet eller en viss betydelse av ordet inte är belagt i källorna senare än 1973 (och med årtalet för det sista belägget angivet inom parentes; självfallet finns språkprovet i fråga också redovisat). Den andra angivelsen däremot talar om att det visserligen finns enstaka belägg efter 1973 men att användningen bedöms vara på väg ut ur språket.⁴ Därtill finns förstås fortfarande behovet av att kunna ange en inskränkt bruklighet inne i definitionen; och här skulle nog det gamla dödkorsets motsvarighet ”förr äv.” också fortsatt kunna användas (med samma årtalsgräns 1973).

Även om inte heller den här modellen är utan komplikationer innebär den ändå att det blir mycket tydligare och handfastare riktlinjer för redaktörernas arbete; det subjektiva inslaget blir kraftigt reducerat. Det blir också tydligt för användaren vad bruklighetsangivelserna står för, de blir genomskinliga, och dessutom kommer de att bli möjliga att utvärdera, att verifiera eller falsifiera. Finns det senare belägg än vad angivelsen ”ej belagt i nusvenska” påstår så är det någonting som inte riktigt stämmer.

Litteratur

Ordböcker

Dalin, A. F. (1850–1855): *Ordbok öfver svenska språket*. Stockholm: Svenska bokförlaget/Norstedts. <spraakdata.gu.se/saolhist>

Illustrerad svensk ordbok, 3 uppl. (1964). Stockholm: Natur och Kultur.

SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien* (1893 ff.). Lund. <www.saob.se>

SAOL = *Svenska Akademiens ordlista över svenska språket*, 1–14 uppl. (1874–2015). Stockholm. <spraakdata.gu.se/saolhist>

Östergren, Olof (1919–1972): *Nusvensk ordbok*, bd 1–10. Stockholm: Wahlström & Widstrand.

Annan litteratur

Hellberg, Staffan (2005): Nusvenskans ålder. I: *Språk i tid. Studier tillägnade Mats Thelander på 60-årsdagen*. Huvudred. Björn Melander. Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet 67. S. 142–151.

⁴De här orden eller betydelseerna blir förstås starka kandidater för att vid nästa revidering ett antal decennier framöver få den kraftfullare bruklighetsinskränkningen ”ej belagt i nusvenska” – då med en ny gränsdragning som förslagsvis kunde utgå från SAOL:s upplaga 2006 eller 2015.

- Jonsson, Hans (1999): *Handbok för redigeringen av Svenska Akademiens ordbok*. 2 uppl. Lund. Otryckt manuskript.
- Larsson, Lennart (2012): Varför inte dub(b)lettformer? Om SAOB som normativ ordbok. I: *Nordiska studier i lexikografi* 11. Rapport från Konferensen om lexikografi i Norden Lund 24–27 maj 2011. Red. av Birgit Eaker, Lennart Larsson & Anki Mattisson. Skrifter utgivna av Nordiska föreningen för lexikografi nr 12. S. 385–395.
- Malmgren, Sven-Göran (2007): Modernt språkligt genombrott, modern nysvenska och nusvenska. I: *Det moderna genombrottet – också en språkfråga?* Studier i svensk språkhistoria 9. Utg. Lars Wollin, Anna Saarukka & Ulla Strohwollin. Skrifter från Svenska institutionen vid Åbo akademi 5/2007. S. 174–180.
- Mattisson, Anki (2013): Om (bristande) bruklighetsstruktur i SAOB. I: *Nordiske Studier i Leksikografi* 12. Rapport fra Konferanse om leksikografi i Norden, Oslo 13.–16. august 2013. Red. av Ruth Vatvedt Fjeld & Marit Hovdenak. Skrifter utgitt av Nordisk forening for leksikografi nr. 13. S. 327–340.
- Nilsson, Pär & Bodil Rosqvist (2018): Inför revideringen av SAOB. Fackspråk och fackspråksbeteckningar i går, i dag och i morgon. I: *LexicoNordica* 25, s. 135–152.
- Thelander, Mats (1987): Nu eller alltid? Perspektiv på 110 års nusvenska. I: *Nysvenska studier* 67, s. 5–14.

Lennart Larsson
docent i nordiska språk, SAOB-redaktör
Svenska Akademiens ordbok, Lund
Institutionen för nordiska språk vid Uppsala universitet
lennart.larsson@svenskaakademien.se;
lennart.larsson@nordiska.uu.se

NAOBs etymologier

Hanne Lauvstad

The Norwegian Academy Dictionary (NAOB) is a modern digital dictionary, documenting Norwegian bokmål (riksmål included), one of the two written standards of Norwegian, from c. 1814 until today. The dictionary is based upon *Norsk Riksmålsordbok* (NRO) (1937–1957) with supplements (1995), but has been expanded, modernized and adapted to a digital format. The few and short etymologies of NRO have been thoroughly revised and augmented considerably. In this article, examples of different types of etymologies (of native words, loan translations, loanwords, simplexes, derivations etc.) in NAOB are presented, to illustrate the editorial choices made. Encyclopedic information is added to the lexical information when considered necessary. Examples are also given of how the etymologies are digitally adapted, especially by means of links between etymons, affixes etc. Editorial templates have been elaborated, in order to secure consistency throughout the dictionary.

1. Innledning

Det Norske Akademis Ordbok (NAOB) er en digital dokumentasjonsordbok for bokmål og riksmål som bygger på *Norsk Riksmålsordbok* (NRO). De fire første bindene av NRO utkom i årene 1937–1957, med to supplementsbind i 1995. NAOB er imidlertid mer enn bare en oppdatering av NRO; det er en moderne ordbok, med et stort, nytt tilfang av lemmer, både ord og uttrykk, språklig- og innholdsmessig oppdaterte definisjoner, og mange flere redaksjons- og (især) sitateksempler. Ordboken retter seg mot et bredt publikum med en viss forkunnskap, f.eks. på nivå med elever i videregående skole (som tilsvarende gymnas). NAOB består i dag av ca. 225 000 lemmer – iberegnet uttrykkslemmene, som inngår som en slags miniartikler i ordboksartiklene. Uttrykkslemmene har egne definisjoner, redaksjons- og sitateksempler, og ofte også egne etymologier.

Som navnet NAOB sier, er ordboken eid av Det Norske Akademi for Språk og Litteratur, en ideell organisasjon. Arbeidet med ordboken har vært utført av en redaksjon i Kunnskapsforlaget, med Tor Guttu som hovedredaktør. Fra 1.1.2020 er ordboken organisert som et eget selskap, Akademiordboken AS. Selv har jeg vært hovedansvarlig bl.a. for etymologiarbeidet siden jeg startet i prosjektet i 2006. Da var det et ønske at etymologiene fra NRO skulle videreutvikles. I NRO hadde det vært svært kortfattede etymologier som oppga lånespråket ved enkelte fremmedord. Fra 2006 skulle også arveordene og ord av nedertysk og tysk opprinnelse få etymologier. NRO var blitt skannet inn og XML-kodet,

men det viste seg snart at det var nødvendig å gjøre noe med hele strukturen. Grunnlagsteksten var innviklet og passet dårlig til XML-strukturen, med dens preg av kinesisk eske, med elementer inni elementer. Dessuten var det påkrevd med koding som viste til funksjoner i teksten, fremfor grafisk utseende. I samarbeid med Tor Guttu og med godkjennelse av Akademiets ordbokskomiteé laget jeg derfor et system for etymologileddene som samtidig ble tilpasset en ny kodestruktur. Dette systemet er blitt finslipt og videreutviklet underveis. Resultatet av arbeidet foreligger nå i NAOB. Det er for øyeblikket noe over 127 700 etymologier i ordboken, dvs. at litt over halvparten av lemmaene har etymologi.

2. Etymologiene i andre store skandinaviske diakrone allmennordbøker

Omfang og utforming av etymologiopplysningene i andre store skandinaviske diakrone ordbøker varierer mye. I den andre store diakrone norske ordboken, *Norsk Ordbok*, som dekker nynorsk og norske dialekter, er det forholdsvis få etymologier, og de som finnes, er kortfattede. Andre store diakrone ordbøker, som ODS og SAOB, har mer omfattende etymologiopplysninger og mer detaljerte redegjørelser for ordhistorien enn vi har i NAOB. Disse ordbøkene er imidlertid også av langt større omfang enn NAOB. Som ordbokstype er NAOB beslektet med disse to verkene, men pga. sin synkron funksjon er NAOB typologisk en kombinasjon av en diakron ordbok (som f.eks. ODS og SAOB) og en synkron ordbok (som f.eks. DDO og SO).

3. Motivasjon og ambisjoner for etymologiene i NAOB

Etymologiopplysninger i ordboksartikler har en viktig funksjon, både for leksikografer og ordboksbrukere. Den etymologiske forståelsen er grunnleggende for lemmatiseringen og betydningsinndelingen, altså både for makro- og mikrostrukturen i en ordbok. Samtidig bidrar etymologiene til brukernes innsikt i ord dannelse og vil forhåpentlig øke deres språklige, især ordhistoriske og semantiske, kompetanse, noe som i sin tur vil kunne bidra til større språkbevissthet og trygghet i omgang med språket, både i lesning og skrivning.

Etymologiopplysningene har derfor vært en selvsagt del av NAOB-arbeidet. Som allmennordbok er NAOB konsentrert om å gi ordbokbrukerne lettfattelige etymologiredegjørelser som samtidig er basert på autoritative kilder.

Ambisjonen har vært en kortfattet redegjørelse for relevant etymologisk informasjon ved hvert oppslagsord hvor det har vært ansett som nødvendig. Arbei-

det har iblant medført at man har måttet sammenstille informasjon fra forskjellige kilder. Iblant har man måttet vurdere hvilke etymologiske hypoteser som har vært mest sannsynlige. I enkelte tilfeller har kildematerialet vært sparsomt eller ikke-eksisterende. Især nyere importord har budt på slike problemer. Med andre ord har arbeidet stilt krav til språklig kompetanse og etymologisk forståelse hos redaktørene. Målsettingen har imidlertid ikke vært å drive selvstendig grunnforskning. I motsetning til f.eks. *Oxford English Dictionary* (OED), som har et eget team på ti personer til å arbeide med etymologiene, har vi i NAOB hatt begrensede ressurser, og etymologiarbeidet har utgjort én blant mange redaksjonelle oppgaver. Det har f.eks. ikke vært et mål i NAOB å utforske opprinnelsen til norrøne former. Her har vi støttet oss på eksisterende kilder. NAOBs etymologier er allikevel langt grundigere enn det man finner i andre norske allmennordbøker, hvor man ofte har nøyd seg med å føre opp f.eks. en opprinnelig form i norrønt eller latin som kan ligge langt unna den endelige formen i moderne norsk. Hvis man ser på bokmålet, kan en moderne ordform ofte være en dansk variant, snarere enn en norrøn arveform, siden bokmålet har en dobbel opprinnelse i dansk og norrønt. Mange latinske lånord har en norsk form som er påvirket av at de er innlånt via f.eks. tysk eller fransk. Ordbokbrukere får derfor altfor snau informasjon dersom man ikke viser disse forholdene eller denne prosessen i etymologiene.

I NAOB har målet vært at etymologiopplysningene skal være korrekte, uten spekulative hypoteser eller desinformasjon. I tilfeller hvor etymologien har vært usikker, har den redaksjonelle føringen vært å angi «av uviss opprinnelse». Der hvor flere troverdige teorier er finnes, er alternativene nevnt. Vi bruker formuleringene «muligens» og «trolig» for å markere større eller mindre grad av usikkerhet.

4. Kildegrunnet for NAOBs etymologier

I utarbeidelsen av etymologiene har NAOB-redaksjonen basert seg på informasjon i relevante oppslagsverk, først og fremst etymologiske ordbøker, men også allmennordbøker og fremmedordbøker med utførlige ordhistoriske redegjørelser. De viktigste kildene våre for arveord og danske former har vært *Våre arveord*, Falk & Torp (1991 [1903–1906]), *Dansk Etymologisk Ordbog* og Torp (1963 [1919]). For lånord og importord vil jeg især fremheve OED, Kluge, Duden *Fremdwörterbuch* og Rey et al. (2006). I tillegg har vi brukt en rekke andre ordbøker, iblant også leksika, det siste for å finne relevant encyklopedisk informasjon. Også idiomer, faste kollokasjoner og ordtak er i stor utstrekning blitt forsynt med etymologiopplysninger i NAOB. Som kilder for disse vil jeg særlig nevne *Bevingede ord* (2006) og tyske Röhrich (1976).

5. Ordtyper med etymologi i NAOB

Etymologitredninger er gjort både for usammensatte ord (simplekser), avledninger, kryptosammensetninger og for en rekke sammensetninger som ikke er direkte gjennomsiktige, f.eks. oversettelseslån og erstatningsord. Dessuten er det gjort rede for etymologien ved en rekke faste uttrykk. Det har vært lagt vekt på å skjelle mellom norrøne arveord og danske former. Ved lå nord og importord har det vært et mål å vise mellomspråket i de tilfeller hvor dette har preget den norske formen eller betydningen. Jeg skal gi eksempler på etymologytypene nedenfor.

6. Etymologytyper og formler

NAOB har vært redigert av ca. 20 medarbeidere. For å oppnå et enhetlig preg i etymologiene har vi brukt en del formellignende uttrykksmåter som «sammen-satt av x og y», «første ledd x; annet ledd y», «avledet av x med suffikset -y». Det er lenket til ledd som er egne oppslagsord i NAOB (her markert med understreking), ellers redegjøres det for leddet i den aktuelle etymologien.

Eksempler på etymologier for arveord, lå nord og oversettelseslån:

- arveord: **hund** av norrønt *hundr*.
- dansk form sammenlignet med norrøn arveform: **aften** dansk form *aften*, av gammeldansk *aftæn*, tilsvarer norrønt *aptann*, *eptann*, **spurv** dansk form *spurv*, av gammeldansk *sparw*, *sporw*, tilsvarer norrønt *sporr*; jf. nynorsk *sporv*.
- lå nord med mellomspråk: **kalender** via tysk *Kalender*, fra senlatin *calendarium*, avledet av latin *calendae* 'første dag i måneden'; jf. *calendarium*.
- arkaiserende nyinnlånt norrøn form: **blotmann** etter norrønt *blótmaðr*; jf. blot.
- oversettelseslån: **brukergrensesnitt** etter engelsk *user interface*.

Ved nydannelser med kjent opphavsmann, er det gjort rede både for orddannelsen og for opphavspersonen, med andre ord er også encyklopedisk informasjon inkludert i etymologien, etter et bestemt mønster:

- **gass** via tysk *Gas*, fra nederlandsk *gas*; jf. engelsk *gas*, fransk *gaz*; betegnelsen dannet av den flamske kjemikeren Jan Baptista van Helmont (1577–1644) etter latin *chaos*, av gresk *khaos* i betydningen 'det tomme rom, det formløse grunnstoff som verden ble skapt av'; jf. kaos”.

Faste kollokasjoner har som nevnt ofte egne etymologiledd. Dette er praktisk, siden uttrykkene gjerne har sin egen spesielle opprinnelse, jf. noen eksempler på uttrykksetymologier i artikkelen for adjektivet **blå**:

- **den blå time**: etter fransk *l'heure bleue*; i senere betydning knyttet til absintens grønnblå farge
- **den blå blomst** | **poesiens blå blomst**: etter tysk *die blaue Blume* 'den blå blomst'; betegnelsen viser til en roman av den tyske forfatteren Novalis (Friedrich von Hardenberg, 1772–1801), hvor den blå blomsten ble symbol på romantikkens evighetslengsler.

7. NAOB-redaksjonens arbeid og ordbokens brukergrupper

I redaksjonen har det vært medarbeidere med kunnskaper om nordiske og en rekke andre germanske språk, dessuten om romanske og slaviske språk, latin og gresk, og vi har også vært så heldig å få hjelp av eksterne eksperter med lånord fra språk som vi selv ikke behersket (f.eks. nordsamisk, arabisk og ungarsk). Men når det gjelder språk som ingen i redaksjonen har kunnet, har vi avholdt oss fra å transkribere ordformer i det aktuelle språket, siden transkripsjonsstandarden krever særlig årvåkenhet, og det lett kan oppstå feil og inkonsekvenser. Her har det vært lagt vekt på pålitelighet og akkuratess fremfor fullstendighet.

Målsettingen med utformingen av etymologiredegjørelsene har vært å lage en oversiktlig og lett forståelig tekst. Preposisjonene «av» ved arveord, «fra» ved lånord og «via» ved (første) mellomspråk og «etter» ved oversettelseslån eller nyinnlånte norrøne former i arkaiserende hensikt tydeliggjør forbindelsene til den moderne norske formen. Dette systemet ble foreslått av hovedredaktør Tor Guttu og den daværende ordbokskomiteen, da artikkelforfatteren startet etymologiarbeidet i 2006. Det er tydelig og systematisk, men i ettertid ville jeg nok ha foretrukket å legge inn denne informasjonen som attributtverdier i taggene e.l., slik at den kunne vises med en spesiell formatering i stilarket. Det ville gitt en mer gjennomført digital systematisering av materialet.

NAOB har ca. 25 000 unike brukere hvert døgn (dvs. det antall personer som slår opp i nettordboken, ev. flere ganger), og interessen er stadig stigende. Redaktørene har kontakt med ordbokbrukerne via mail og sosiale medier (Facebook og Instagram). Ut fra brukerhenvendelsene kan redaksjonen se at ordboken har en sammensatt brukergruppe, med både profesjonelle oversettere, studenter og skoleelever og personer som har norsk som annetspråk. Videre ser man at det er stor interesse for etymologi. Dette betyr at det er stor mulighet for å nå frem til ordbokbrukerne med NAOBs etymologiske informasjon. Håpet er også at etymologiene i NAOB kan være til nytte for andre nordiske lesere. Arveordsetymologiene

gir oversikt over det norske ordstoffet. Redegjørelsen for fordelingen av danske former og norrøne arveformer vil også gi et tydeligere bilde av de østnordiske innslagene i bokmålet. Ikke minst har man i NAOB lagt vekt på å redegjøre for nyere importord, og slike etymologier samlet på ett sted kan også være informative for andre nordiske lesere.

8. Utnyttelse av den digitale ordbokens muligheter

En digital ordbok er helt annerledes strukturert enn en papirordbok. Papirordbokens lineære preg er i digital versjon erstattet med enkeltelementer, som både kan være hele ordbokartikler eller deler av disse, og som kan lenkes sammen til et nettverk. Denne strukturelle endringen er det tatt hensyn til i oppbygningen av NAOB som en digital ordbok. Mer om dette emnet finner man i min kollega Carina Nilstuns artikkel ”Den digitale ordbokens innvirkning på den praktiske leksikografien” (s. 245–252) i denne publikasjonen.

Hver ordbokartikkel har et eget etymogiledde i artikkelhodet. Herfra er det lenket til bestemte betydninger (som vises innrammet) i de tilfeller hvor de krever særskilte etymologiopplysninger, se figur 1.

skole verb

BØYNING skolet, skolet, skoling +

UTTALE [sko:ˈlə] 🔊

ETYMOLOGI til [skole](#); i [denne betydningen](#) etter tysk *schulen*

BETYDNING OG BRUK

1 SKOLEVESEN holde skole; gi skoleundervisning

2 SÆRLIG I ADJEKTIVISK PERFEKTUM PARTISIPP, NÅ SJELDEN [skolere](#)

EKSEMPEL

- *en vel skolet stemme*

SITATER

- *de lite skolede etiopiske tropper (Aftenposten 1940/585/2/2)*
- *jeg vil tage skoling fra dig (Knut Hamsuns brev [🔗] 35)*

Figur 1: Eksempel på lenke fra etymologien til enkeltbetydning i artikkel.

Mulighetene for lenking er utnyttet også ellers i NAOB. Bl.a. har vi et omfattende system med artikler for sammensetningsledd, prefikser og suffikser, hvor mye informasjon er samlet, og som vi kan lenke til ved behov:

- lenke til ledd i sammensetninger: **skravlebøtte** annet ledd bøtte (som viser til en betydning med stilmarkøren «muntlig, dels nedsettende eller spøkefullt, brukt som sisteledd i personbetegnelser»), **riksdaler** sammensatt av riks- og daler.
- lenke til prefiks- og suffiksartikkel: **forflåte** (verb) eldre dansk form *forflade*, etter tysk *verflachen*; jf. prefikset for-, **brennbar** avledet av brenne med suffikset -bar; jf. tysk *brennbar*.
- lenker til prefiks- og suffiksartikkel ved kryptosammensetning: **kryptografi** av krypto- og -grafi. Klikker man på lenkene, kommer man til følgende affiksartikler med egne etymologier: **krypto-** til gresk *kryptos* 'skjult', til *kryptein* 'skjule'; jf. engelsk *crypto-*; beslektet med krypt og **-grafi** fra gresk *grafia*, avledet av *grafein* 'skrive'.

Etymologiene i NAOB inneholder egne tagger for språk (<spraak>), opphavsord/opprinnelige ordformer eller «etymoner» (hos oss <histform>) og betydninger (<grunnbetydning>). Taggen <appellativisering> er også brukt ved «rene» eponymer, dvs. personnavn som uforandret er gått over til å bli appellativer. Taggingen gjør det mulig å søke opp disse funksjonene i redigeringsverktøyet, f.eks. kan man søke på språk. Slik kan man finne ut at det er 20 392 direkte lån fra et annet språk og 2 786 lån via et mellomspråk. Hvis man ser bort fra middelnedertysk og nedertysk, som har gitt norsk en enorm mengde lånord, viser det seg at de hyppigste lånespråkene er

- engelsk, med amerikansk-engelsk 4 110 (uten gammelengelsk og middelengelsk)
- fransk 3 462 (uten gammelfransk og mellomfransk)
- tysk: 2 036, med seks tilfeller av «tysk dialekt» og to tilfeller av sveitsertysk (uten middelhøytysk, middelnedertysk og nedertysk)
- italiensk 632, med tre tilfeller av «italiensk dialekt»

Dette er bare eksempler på de søkemuligheter kodingen gir. Man kan også ved søk på taggen <appellativisering> finne ut hvilke ord som i uforandret form er gått over fra å være proprier til å bli appellativer. Dessuten gjør kodingen det mulig å vise funksjoner på bestemte måter i stilarket, f.eks. viser vi opphavsord kursivert og betydninger mellom enkle anførselstegn, men visningsmåtene kan enkelt endres.

En digital ordbok innebærer ikke minst at man kan oppdatere, korrigere og utfylle informasjonen i ordboken fortløpende. Dette er praktisk. I årene fremover har NAOB-redaksjonen planer for en videre utnyttelse av lenkemuligheter mellom ordboken og andre digitale nettressurser.

9. Oppsummering

En god etymologi er nødvendig for leksikografen og gir ordbokbrukerne en bedre språkhistorisk innsikt. I NAOB er etymologiene også sentrale for hvordan ordbokens makro- og mikrostruktur (dvs. lemmatisering og betydningsinndeling) blir. Arv eller lånevei tillegges vekt ved lemmatiseringen, sammen med andre kriterier som ordklasse og genus. Etymologiene er derfor uunnværlige for ordbokredaktørenes arbeid. For ordbokbrukerne byr etymologiene på muligheter for innsikt i ords og uttrykks betydningsutvikling og slik en bedre kjennskap til språket.

Arbeidet med NAOB har vist hvilke muligheter en digital ordbok byr på. Den digitale strukturen gjør det, som vist ovenfor, mulig å operere med artikler for sammensetningsledd, prefikser og suffikser som man kan lenke til ved behov, med andre ord en økonomisk form for gjenbruk av informasjon i en nettstruktur. Sammen med de gjennomgående formlene i etymologiene er håpet også at dette vil gjøre det lettere for brukerne å orientere seg i ordboken.

I årene fremover har NAOB-redaksjonen planer for en videre utnyttelse av lenkemuligheter mellom ordboken og andre digitale nettressurser. Målsettingen er at NAOB bl.a. skal være en etymologisk nettressurs som når ut til ordboksbrukere i inn- og utland, ikke minst i Norden, at ordboken skal nå ut til stadig større grupper og at den skal være praktisk og brukervennlig. Slik håper vi å bidra til den allmenne kunnskapen om språket og til å styrke norsk språk i en digital tidsalder.

Litteratur

Ordbøker

- Våre arveord = Bjorvand, Harald & Fredrik Otto Lindeman (2007): *Våre arveord. Etymologisk ordbok. Revidert og utvidet utgave*. Oslo: Novus forlag.
 DDO = (2009–) Det Danske Sprog- og Litteraturselskab: *Den Danske Ordbog* (online). <www.ordnet.dk/ddo> (februar 2020).

- Das große Fremdwörterbuch* (2007). 4. opplag. Mannheim – Leipzig – Wien - Zürich [CD-ROM]: Dudenverlag.
- Falk, Hjalmar & Alf Torp (1991 [1903–1906]): *Etymologisk ordbog over det norske og det danske sprog*. Faksimileutgave. Oslo: Bjørn Ringstrøms antikvariat.
- Gundersen, Dag & Snorre Evensberget (2006): *Bevingede ord. Ordtak, sitater og deres opprinnelse*. Oslo: Kunnskapsforlaget.
- Kluge, Friedrich (2002) = *Etymologisches Wörterbuch der deutschen Sprache. Bearbeitet von Elmar Seebold. 24., durchgesehene und erweiterte Auflage*. Berlin/New York: Walter de Gruyter.
- NAOB = Tor Guttu et al. (red.): *Det Norske Akademis ordbok*. Oslo: Kunnskapsforlaget og Det Norske Akademi for Språk og Litteratur. <www.naob.no> (februar 2020).
- Nielsen, Niels Åge (2015): *Dansk Etymologisk Ordbog. Ordenes historie*. 6. utgave, 2. opplag. København: Gyldendal.
- Norsk Ordbok = *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. 1–12. 1930–2016. Oslo: Det Norske Samlaget.
- NRO = *Norsk riksmålsordbok*. Bind I–IV 1937–1957, bind V–VI 1995. Oslo: Kunnskapsforlaget.
- ODS = *Ordbog over det danske Sprog* (1919–1956, bind 1–28 og Supplement (bind 1–5) (1992–2005) <www.ordnet.dk/ods/> (februar 2020).
- OED = *Oxford English Dictionary. The definite record of the English language*. <www.oed.com> (oktober 2019).
- Rey, Alain et al. (red.) (2006): *Dictionnaire historique de la langue française*. Nytt opplag av 2. utgave. Paris: Le Robert.
- Röhrich, Lutz (1976): *Lexicon der sprichwörtlichen Redensarten* (bind 1–2). 4. opplag. Freiburg im Breisgau: Verlag Herder KG.
- SAOB = *Svenska Akademiens ordbok*. A–väva. <www.saob.se> (februar 2020).
- SO = *Svensk ordbok utgiven av Svenska Akademien* (2009). Stockholm: Norstedts. Online: <www.svenska.se> (februar 2020).
- Torp, Alf (1963 [1919]): *Nynorsk etymologisk ordbok*. Uforandret opptrykk. Oslo: Aschehoug & co. (W. Nygaard).

Hanne Lauvstad
leksikograf, dr.art.
Det Norske Akademis ordbok
Grensen 3
NO-0159 Oslo
hanne.lauvstad@naob.no

Multimediale datatyper som leksikografisk merværdi

Patrick Leroyer & Henrik Køhler Simonsen

This article discusses multimedia data types in lexicography. On the basis of a literature review and a review of selected dictionaries and other types of lexicographic resources, we demonstrate how these data are currently used in lexicography. We also demonstrate how the full potential of multimedia data types could be utilized in lexicography and how existing lexicographic data types can be supplemented or replaced by these data and how multimedia data types seem to have a separate role on their own. The theoretical discussion also shows that such data not only support well-known lexicographic functions but also support experiential sub-functions as an added value.

1. Indledning, problem og metode

Man siger som bekendt, at ”et billede siger mere end tusind ord” og vi vil i denne artikel diskutere anvendelsen af forskellige former for multimediale datatyper i leksikografien og hvorfor vi netop bør interessere os mere for dette spørgsmål. Denne artikel er for det første baseret på en bred forståelse af og definition på leksikografi som den informationsvidenskabelige disciplin, der har til formål at designe og udarbejde opslagsværker såsom ordbøger, leksika, encyklopædier, guides, glossarer etc., jf. Leroyer (2011) og Tarp (2018). Artiklen anvender for det andet en lige så bred forståelse og definition af multimediale datatyper, som de mange heterogene datatyper, der ved hjælp af forskellige medialiteter er opstået i takt med den teknologiske udvikling, f.eks. små film og videoer, quizzes, lydfiler, animationer, grafer og interaktive film. Vores diskussion af multimediale datatyper bygger grundlæggende på eksisterende forståelser af f.eks. illustrationer i leksikografi, men udvider forståelsen med de nye medialiteter. Termen *multimedial* forstås her som ”flere forskellige medier”.

Det argumenteres i artiklen, at leksikografien i bred forstand, som naturligt nok interesserer sig mest for ord, til en vis grad er faldet i den klassiske fælde med at blive så opslugt af det lingvistiske materiale, at den måske risikerer at glemme lidt, hvad det er, brugerne efterspørger, først og fremmest andre typer, ikke sproglige tegn. Da leksikografien koncentrerer sig først og fremmest om sproglige tegn, har dette betydet, at billeder som tegn i mange år har levet lidt i skyggen.

Dette er paradoksalt, idet allerede de første hulemennesker som bekendt forsøgte at gengive og berette om verden ved hjælp af hulemalerier for flere tusind år siden, og i dag bruger vi også en lang række former for billeder til at forstå og navigere i den polysemiotiske verden, vi befinder os i.

Samtidig har den teknologiske udvikling betydet, at der i dag er endnu flere former for visuelt indhold end nogensinde før, og enhver kan i dag med deres mobiltelefon tage billeder og videoer af høj kvalitet og poste på internettet. Endvidere gør 360-graders-videoer og AR-videoer det muligt for brugerne næsten af være en del af det visuelle indhold.

Der synes således at være mange grunde til at interessere sig mere for multimediale datatyper i leksikografien, og vi vil i artiklen tage udgangspunkt i følgende tre forskningsspørgsmål:

- 1) Hvordan kan leksikografi med fordel udnytte multimediale datatypers leksikografiske potentiale?
- 2) Hvordan kan eksisterende datatyper suppleres eller erstattes af multimediale datatyper? Hvordan kan multimediale datatyper stå alene?
- 3) Hvilke nye brugssituationer kan multimediale datatyper understøtte og hvilke nye funktioner opstår i denne forbindelse?

Metodisk har vi først og fremmest analyseret den relevante teoretiske litteratur på området og analyseret forskellige forståelser af multimediale datatyper i leksikografien.

Dernæst har vi analyseret udvalgte, praktiske eksempler på leksikografiske ressourcer, som på den ene eller anden måde benytter sig af forskellige former for multimediale datatyper. Endelig har vi på basis af både litteraturgennemgangen og analysen af udvalgte leksikografiske ressourcer udviklet en række overvejelser om anvendelse af multimediale datatyper i leksikografiske ressourcer.

2. Teoretisk diskussion

Der er skrevet ganske meget om illustrationer generelt i den leksikografiske litteratur, herunder også til dels om multimediale datatyper, og i det følgende tager vi bl.a. udgangspunkt i den måde, man kan anvende illustrationer på i fagsproglige ordbøger, jf. Bergenholtz & Tarp (1995:159-166). Et andet eksempel fra Norden er Bergenholtz m.fl. (1997), som i Nordisk leksikografisk ordbok

definerer billeder som ”et fotografi eller en tegning som i det vesentligste gjengir den referenten som skal illustreres”, jf. Bergenholtz m.fl. (1997:85), som senere definerer illustrationer som ”visuel fremstilling af en leksikalsk enhed i en ordbog”, jf. Bergenholtz m.fl. (1997:142).

Hovmark (2010) går et skridt videre og omtaler i artiklen ”Gid man kunne have bragt en tegning!” status quo for og potentielle nye anvendelser af illustrationer i monolingvale ordbøger og diskuterer i artiklen, hvad illustrationer kan bruges til i leksikografien på basis af forfatterens arbejde med inddragelse af illustrationer i redaktionen af *Ømålsordbogen*, jf. Ømo (2019).

Biesaga (2016) undersøger også multimediale datatyper og diskuterer en typologi for illustrationer samt en række forskellige typer illustrationer, hvortil der synes at mangle regler for anvendelse: ”Despite using the narrow (lexicographical) or wider (encyclopedic) perspective, lexicographers should always keep in mind that visual facilities ought to follow certain rules”, jf. Biesaga (2016:106). Biesaga påpeger således, at illustrationer kan tjene to formål, nemlig at bringe information om sproget (sproglig viden) og bringe information om verden uden om sproget (encyklopædisk viden) ved at illustrere sproget og verden uden for sproget. Biesaga (2017:145) fremfører et år senere en kritik af den konventionelle anvendelse af illustrationer og skriver, at “For now visual facility seems only an addition to the exact verbal explanation. In the future, it should be a complementary tool, serving in the process of semantic description”. Sidst men absolut ikke mindst, diskuterer Liu (2017), hvorledes multimediale data med fordel kan anvendes til at udvide meningsunderstøttelsen i elektroniske ordbøger, og det er netop denne udvidede anvendelse og forståelse af multimediale datatyper, som vi her diskuterer.

Det handler således ikke kun om illustrationer i klassisk forstand. Det handler også om det store uudnyttede, illustrative potentiale i multimediale datatyper, som endnu ikke er blevet behandlet dækkende nok hverken teoretisk eller i praksis. Som det vil fremgå af det følgende, argumenterer vi, at multimediale datatyper i høj grad understøtter de veletablerede kommunikative, kognitive og operative funktioner, jf. Tarp (2018). Som bekendt hjælper de brugeren med at læse og forstå og oversætte, med at tilegne sig viden og med at hjælpe brugeren med at følge instruktioner.

Men multimediale datatyper kan også stå alene og understøtte en række nye subfunktioner og dermed skabe øget leksikografisk værdi gennem brugeroplevelser.

3. Analyse

Multimediale datatyper anvendes allerede i en række ordbøger og øvrige opslagsværker, som det fremgår af tabel 1 herunder.

Navn og link	Multimediale elementer
Lexin lexin.udir.no/bildetema	Visuel datatilgang (Mouth) Udtale (Wisdom tooth)
Macmillan Dictionary macmillandictionary.com	Grafik (Word forms) Udtale (Allergy)
Encyclopaedia Britannica britannica.com	Fotos (Allergy) Videoer (Hyperallergenic product, Allergy, Hygiene hypothesis, Skin test, Systemic anaphylactic response)
Medicin.dk min.medicin.dk	Instruktionsvideoer (Brug af inhalator, Hvad er en priktest, Måling af blodtryk) Tabletidentifikation for lægfolk (Foto og identifikation) Tabletidentifikation for læger (Genkend medicin)
Patienthåndbogen sundhed.dk/borger	Animationer (Allergenvacciner, antihistaminer) Fotos (Hudpriktest) Præsentationer (Høfeber)
EcoLexicon ecolexicon.ugr.es	Foto (Earth's energy budget) Illustration (What is a tidal bore?)

Tabel 1: Multimediale datatyper i ordbøger og øvrige opslagsværker.

På hjemmesiden for den franske vinindustri i Bourgogne, Vins de Bourgogne (2020), der indeholder en række leksikografiske ressourcer i form af ordlister (vintermer) og billedforklaringer, kan man f.eks. tilgå små filmsekvenser, der viser vinplantens cyklus eller tage en quiz om smagspræferencer og endda lyde fra den typiske vinproduktion. Disse multimediale datatyper understøtter således et vigtigt, både sprogligt og encyklopædisk læringselement, men måske også et endnu vigtigere oplevelselement, der bidrager til at booste læringen.

I opslagsværket Alt om Hus, jf. Altomhus (2019)¹ er der en visuel datatilgang og en særdeles handlingsorienteret operativ tilgang til rådgivning og tilbud. Man går således fra data til konkret handling i praksis, fordi man vha. opslagsværket kan indhente tilbud fra den nærmeste håndværker. Det argumenteres således, at de multimediale elementer i højre kolonne i tabel 2 har et meget stort, leksikografisk potentiale, ikke alene kommunikativt og kognitivt, men også operativt.

¹ Siden opsætning af artiklen er websitet www.altomhus.dk på grund af flytning til anden server pt. ikke tilgængelig.

Navn og link	Multimediale datatyper
Vins de Bourgogne vins-bourgogne.fr	Film af f.eks. vinplantens cyklus Quizzes om f.eks. smagspræferencer samt genkendelse af lyde ved vinproduktion Vinskole, som er individualiseret efter kompetenceniveau
Alt om Hus altomhus.dk	Visuel datatilgang til byggedele på et hus Operativ tilgang til rådgivning og tilbud
The Right Rhymes therightrhymes.com	Geografisk datatilgang på basis af lokation
LearnMatch learnmatch.net/en	Gamification med fodboldspil Social interaktion (Leg og lær)

Tabel 2: Multimediale datatyper med stort leksikografisk potentiale.

Multimediale kategorier af datatyper i højre kolonne i tabel 3 herunder understøtter således uden tvivl de kommunikative funktioner (hjælp til sprogproduktion), fordi de tilbyder brugeren målrettet hjælp til både reception, produktion, og oversættelse. Men datatyperne i højre kolonne i tabel 3 understøtter også de kognitive funktioner (hjælp til videnstilegnelse), fordi de tilbyder brugeren målrettet hjælp til at tjekke eller tilegne sig viden.

Ligeledes understøtter de multimediale kategorier af datatyper i højre kolonne imidlertid også de operative funktioner, fordi de i lige så høj grad tilbyder brugeren hjælp til at gøre noget i en bestemt situation.

Kommunikative funktioner	Multimediale datatyper
Hjælp til tekstreception, tekstproduktion, oversættelse	Film Animationer Lydklip Udtale Præsentationer
Data indgår i udførelse af igangværende kommunikationsproces	
Kognitive funktioner	Multimediale datatyper
Hjælp til at tjekke eller tilegne sig viden, både sporadisk og systematisk	Film Animationer Lydklip Udtale Præsentationer Fotos Quizzes Interaktive film Stregkoder datatilgang Visuel og geografisk datatilgang
Hjælp til at lære noget i et læringsforløb	
Data er målet for konsultationen	
Operative funktioner	Multimediale datatyper
Hjælp til at handle ved at få vist, hvordan man gør = hvordan skal man bære sig ad i denne situation	Film Animationer Grafik Præsentationer Instruktionsvideoer Visuel datatilgang
Hjælp til at handle ved at få råd og vejledning = hvad skal man gøre i denne situation?	

Tabel 3: Multimediale datatyper, der understøtter hovedfunktioner.

Multimediale datatyper er ydermere i stand til at supplere og erstatte kendte leksikografiske datatyper, som angivet i tabel 4 herunder. En film kan f.eks. både supplere men også erstatte definitioner, eksempler, begreber eller processer, hvilket ligger i tråd med ønsket om at bringe ikke kun en tegning men en helt film som supplement til f.eks. eksemplet, jf. Hovmark (2010).

Multimediale datatyper: supplerer og erstatter	
Datatyper	Supplerer (S) og erstatter (E)
Film	Definitioner, eksempler, begreber, processer (S&E)
Animationer	Definitioner, eksempler, begreber, processer (S&E)
Lydklip, udtale	Udtaleangivelser og IPA (S&E)
Fotos	Definitioner, eksempler, begreber, processer (S&E)
Præsentationer	Definitioner, eksempler, begreber, processer (S&E)
Multimediale datatyper: afløser og står for sig selv	
Datatyper	Afløser – står for sig selv
Quizzes	Lærings- og underholdningsforløb og -evalueringer
Interaktive film	Lærings- og underholdningsforløb og -rum
Eksperimenter	Lærings- og underholdningsforløb og -evalueringer
VR- og AR-videoer	Simulationer og oplevelser

Tabel 4: Multimediale datatyper supplerer, erstatter og afløser.

Multimediale datatyper kan imidlertid også erstatte eller sågar afløse konventionelle leksikografiske datatyper, der normalt bruges i ordbøger og andre leksikografiske informationsværktøjer og stå for sig selv, og dermed udgøre en selvstændig datatype. En velvalgt quiz f.eks. kan stå for sig selv og udgøre et målrettet lærings- og underholdningsforløb, som skaber ny værdi for brugeren. En professionelt produceret VR- eller AR-film vil også kunne stå for sig selv og udgøre en simulation eller en oplevelse, da behovet for at søge information ofte går hånd i hånd med lysten til at få information.

Endelig argumenteres det, at multimediale datatyper i mange tilfælde også understøtter en række vigtige eksperimentielle subfunktioner, som anført i venstre kolonne i tabel 5.

Der er altså tale om, at de multimediale datatyper i sig selv skaber leksikografisk merværdi ved at skabe leksikografiske oplevelser, f.eks. underholdningsværdi, læringsudbytte, social oplevelse eller imageoplevelse. Disse subfunktioner er særligt interessante i læringssituationer og i forbindelse med virksomheders eller organisationer, hvor informationer anvendes i marketing og brandingøjemed på online platforme, der er struktureret leksikografisk.

Multimediale datatyper: understøtter subfunktioner	
Subfunktioner	Skaber værdi gennem oplevelser
Genererer underholdning	Underholdningsværdi og oplevelse
Supporterer læringsforløb	Læringsudbytte og læringsoplevelse
Fremmer socialisering	Social oplevelse og socialt udbytte
Fremmer fællesskabsopbygning	Social oplevelse og socialt udbytte
Genererer salg	Salg og service oplevelse
Virker som kanal for marketing	Reklameoplevelse og kognitiv og affektiv konsonans
Medvirker til branding	Imageoplevelse og kognitiv og affektiv konsonans

Tabel 5: Multimediale datatyper understøtter subfunktioner.

Illustrationsmæssigt set er leksikografi mere end ordrepræsentationer og udlægning af det sproglige materiale. Leksikografi er også multimediale datatyper og dataarkitektur, der kan opfylde en række nye informationsbehov og dermed skabe leksikografisk merværdi.

4. Konklusion

Litteraturgennemgangen indikerer, at potentialet i multimediale datatyper endnu ikke er behandlet dækkende nok og diskussionen her viser, at de multimediale datatypers potentiale endnu ikke i tilstrækkeligt omfang er blevet realiseret i leksikografien i praksis.

Som svar på det første forskningsspørgsmål om nytteværdien af brugen af multimediale datatyper argumenteres det, at vi ved at sætte de leksikografiske hoved- og subfunktioner endnu mere i centrum, så vil leksikografi med fordel kunne anskues som bruger- og situationstilpasset dataarkitektur. En sådan arkitektur bør nødvendigvis opbygges på basis af funktionelle og subfunktionelle valg. Leksikografens opgave er så at selekttere multimediale datatyper, der i særlig grad kan bidrage til at realisere de valgte hoved- og subfunktioner.

Som svar på det andet forskningsspørgsmål om multimediale datatypers rolle argumenteres det, at multimediale datatyper både kan supplere og erstatte de konventionelle datatyper. Men de multimediale datatyper kan løsrive sig, hvorved de hverken skal supplere eller erstatte, men skabe ny værdi og åbne op for nye leksikografiske dimensioner gennem brugeroplevelser.

Som svar på det tredje og sidste forskningsspørgsmål om nye brugssituationer argumenterer vi, at multimediale datatyper er særdeles velegnet til at understøtte

leksikografiske brugssituationer som underholdning, digital læring, socialisering og fællesskabsopbygning, salg og marketing samt branding.

Vi har nu set, at leksikografi forstået som opslagsværker og informationssystemer, som benytter sig af og leverer værdifuld information, som kan understøtte underholdning, branding etc. kan komme til at spille en væsentlig rolle i videns- og informationssamfundet. Dermed understøtter multimediale data ikke kun de veletablerede kommunikative, kognitive og især operative funktioner, men også en række eksperientielle funktioner, og subfunktioner i tilknytning til de nye leksikografiske brugssituationer, som vi har identificeret.

Der er sammenfattende et stort leksikografisk potentiale i de multimediale datatyper og leksikografien anskuet som funktionel dataarkitektur synes at være et godt udgangspunkt til at realisere dette potentiale. Multimedialitet i sin leksikografiske, digitale udformning kan komme til at spille en stigende rolle i videns- og informationssamfundet, fordi den både faciliterer og optimerer vores tilgang til og udnyttelse af komplekse informationer i vores stærkt polysemiotiske virkelighed, og er med til at skabe leksikografisk merværdi.

Litteratur

Ordbøger og andre leksikografiske ressourcer

Altomhus (2019) = *Alt om hus*. <altomhus.dk> (februar 2020). Siden opsætning af artiklen er websitet altomhus.dk på grund af flytning til anden server pt. ikke tilgængelig.

Bergenholtz, Henning m.fl. (1997): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.

Vins de Bourgogne (2020) = *Vins de Bourgogne*. Éditeur: BIVB (Bureau Interprofessionnel des Vins de Bourgogne). <www.vins-bourgogne.fr> (februar 2020).

Ømo (2019) = *Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer 1*. København 1992–

Anden litteratur

Bergenholtz, Henning & Sven Tarp (red.) (1995): *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins Publishing Company.

Biesaga, Monika (2016): Pictorial Illustration in Dictionaries. The State of Theoretical Art. I: Margalitadze, Tinatin & Meladze, George (eds.): *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity*. Tbilisi: Ivane Javakhishvili Tbilisi State University, 99-108.

- Biesaga, Monika (2017): Dictionary Tradition vs. Pictorial Corpora: Which Vocabulary Thematic Fields Should Be Illustrated? I: *Lexikos* 27, 132-151.
- Hovmark, Henrik (2010): Gid man kunne have bragt en tegning! Om definitioner og illustrationer i ensproglige ordbøger. I: *Nordiska Studier i Lexikografi* 10. Rapport från Konferens om lexikografi i Norden, Tammerfors 3.-5. juni 2009, 192-208.
- Hupka, Werner (1989): *Wort und Bild: Die Illustrationen in Wörterbüchern und Enzyklopädiën*. Tübingen: Max Niemeyer Verlag 1989.
- Leroyer, Patrick (2011): Change of Paradigm in Lexicography. From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools. I: Fuertes Olivera, Pedro A. and Bergenholtz, Henning (eds.): *E-lexicography: Internet, Digital Initiatives and Lexicography*. London, New York: Continuum. 121-140.
- Liu, Xiqin (2017): Multimodal Exemplification: The Expansion of Meaning in Electronic Dictionaries. I: *Lexikos* 27, 287-309.
- Tarp, Sven (2018): Lexicography as an Independent Science? I: Fuertes-Olivera, Pedro A. (ed.): *The Routledge Handbook of Lexicography*. London and New York: Routledge, 19-33.

Patrick Leroyer
 lektor, PhD
 Aarhus Universitet
 School of Communication & Culture
 Jens Chr. Skous Vej 4
 DK-8000 Aarhus C
 pl@cc.au.dk

Henrik Køhler Simonsen
 ekstern lektor, PhD, MA, MBA
 Copenhagen Business School
 Department of Management, Society
 & Communication
 Dalgas Have 15
 DK-2000 Frederiksberg
 hks.msc@cbs.dk

Ut med *adamsslekt* og inn med *arveprinsesse*? Leksikografiske metodar i revisjonen av *Bokmålsordboka* og *Nynorskordboka*

Gunn Inger Lyse

In the project revising the two Norwegian dictionaries *Bokmålsordboka* and *Nynorskordboka* (www.uib.no/revisjonsprosjektet), the lexicographers are to revise these two existing dictionaries in tandem. This article presents the source material, tools, and methods that are utilised to enable the lexicographers to identify a modernised and relevant vocabulary in both dictionaries in an efficient and empirical manner.

1. Innleiing

Brukarvenlege søkjeverktøy og eit godt kjeldegrunnlag er avgjerande for at leksikografen kan arbeide effektivt og med solid empirisk forankring. Denne artikkelen presenterer kjeldematerialet, verktøy og metodar som blir brukte og utvikla i prosjektet Revisjon av *Bokmålsordboka* og *Nynorskordboka* (Revisjonsprosjektet). Dette prosjektet skal modernisere og forbetre dei digitale versjonane av ordbøkene.

Sentralt står dei vitskaplege språksamlingane ved Universitetet i Bergen (UiB), som delvis ligg opne. I tillegg har prosjektet teke i bruk verktøy for språkkanalyse som er tilgjengelege via språkinfrastrukturen CLARINO (2019). Det eine verktøyet er Korpuskel, som let oss gjere avanserte søk på ord og frasar. Gjennom Revisjonsprosjektet er Korpuskel utvida slik at ein kan søkje i fleire korpus samstundes. Utvalet av korpus gjev til saman eit materiale på om lag 2,4 milliardar ord. Det andre verktøyet er INESS (2019), som let oss søkje etter syntaktiske konstruksjonar som ikkje er så lett å søkje etter i eit tradisjonelt tekstkorpus. Denne artikkelen gjev døme på når tekstkorpus er til god hjelp – men òg døme på at korpus kan vere misvisande.

Del 2 av artikkelen gjev ein kort introduksjon til Revisjonsprosjektet. I del 3 følgjer eit overblikk over kjeldegrunnlaget. I del 4 blir søkjesystema Korpuskel og INESS presenterte.

2. Revisjon av *Bokmålsordboka* og *Nynorskordboka*

Revisjonen av *Bokmålsordboka* og *Nynorskordboka* starta opp ved UiB i 2018 og er finansiert av Kulturdepartementet (sjå www.uib.no/revisjonsprosjektet). Ordbøkene inngår i Språksamlingane ved UiB, er eigd og drifta av UiB og Språkrådet, og dei ligg ope tilgjengeleg gjennom nettsøk og som nedlastbar app.¹

I Noreg står *Bokmålsordboka* og *Nynorskordboka* i ein særskild posisjon som Språkrådet sin viktigaste kanal for å formidle informasjon om gjeldande offisiell rettskriving til publikum (Språkrådet 2014:9). Desse er dei einaste fritt tilgjengelege digitale ordbøkene som gjer greie for heile breidda av normerte ord- og bøyingsformer, og dei blir løpande oppdaterte i tråd med rettskrivingsvedtak i Språkrådet. Ordbøkene er mykje brukte, med rundt 125 000 søk per døgn i 2018 (Rauset 2019:162).

Revisjonsprosjektet skal revidere dei to eksisterande ordbøkene parallelt, med ei tidsramme på fem og eit halvt år. Leksikografane har tre hovudoppgåver: sikre at definisjonane er i tråd med dagens språkbruk, peike ut eit modernisert og relevant ordtilfang i begge målformene, og gjere utvalet av lemma mellom ordbøkene likare (Selback 2020). Dette er omfattande produksjonsmål på avgrensa tid, og det er avgjerande med gode søkjeverktøy som let leksikografane jobbe etterretteleg og samstundes effektivt.

3. Kjeldegrunnlag

Bokmålsordboka og *Nynorskordboka* skal gje kvalitetssikra og oppdatert informasjon om normert rettskriving og om vanleg språkbruk for skriftspråka nynorsk og bokmål (Språkrådet & Universitetet i Bergen 2016). Som mellomstore ordbøker skal dei ha eit moderne ordtilfang. Den empiriske forankringa er viktig i ordbokarbeid både for å sikre kvaliteten på, og å etterprøve, arbeidet.

Eit tekstkorpus kan definerast som ei søkbar samling av digital tekst, der treffa kan bli analyserte på ulike måtar. Tekstkorpus er ei god empirisk kjelde både for å vurdere kva oppslagsord som bør vere med i ordbøkene (lemmautval), kva orda betyr, korleis orda oppfører seg syntaktisk og for å finne gode døme på typisk bruk av eit ord. Det er då vanleg å basere seg på ordfrekvens. Det er naturleg å ta omsyn til frekvensdistribusjonen over tid, altså om ordet er stabilt i bruk over tid, og korleis ordet er distribuert på tvers av domene, altså om ordet er del av allmennspråket. Sidan ordbøkene skal reflektere moderne skriftspråk, ser redaksjonen i hovudsak på ordfrekvensen innanfor dei siste 50 åra.

¹ Ordbøkene på nett: ordbok.uib.no; appen heiter Ordbøkene.

Ordbokarbeidet kviler likevel ikkje åleine på søk i tekstkorpus. For det første tek vi omsyn til normeringsvedtak i Språkrådet. Ordbokredaksjonen ved UiB vel ut kva slags ord og tydingar som skal vere med i ordbøkene, medan Språkrådet bestemmer korleis orda skal skrivast og bøyast. For det andre bruker vi dei vitenskaplege Språksamlingane ved UiB, som ligg ope tilgjengeleg (www.usd.uib.no). Mellom anna sjekkar vi i Setelarkivet og dei andre kjeldesamlingane til *Norsk Ordbok* det som vi ikkje finn belegg for i tekstkorpus. For det tredje konsulterer ordbokredaktøren òg språkkjensla si og eigen grammatikkunnskap i vurderinga av korleis ei leksikografisk eining blir brukt.

4. Søkjeverktøy for språkanalyse

Generelt er det vanskeleg å finne korpus som let deg gjere søk både ut frå frekvens, tid og domene, samstundes som materialet er stort nok (Gjesdal & Lyse 2016). Forskingsinfrastrukturar som CLARIN er interessante, fordi dei jobbar for å gjere ulike korpus tilgjengelege og søkbare gjennom felles grensesnitt (CLARIN 2019). Gode døme på dette er Glossa (Nøklestad et al. 2017), Sketch Engine (Kilgarriff et al. 2014), Korp (Borin et al. 2012), Korpuskel (Meurer 2012) og Weblicht (Hinrichs et al. 2010). Nasjonalbiblioteket i Noreg utviklar eit programmeringsgrensesnitt (ein API) for å gjere språkanalysar på alt det digitaliserte materialet som ligg hos Nasjonalbiblioteket (Johnsen 2019).

Ordbokredaktørane i Revisjonsprosjektet brukar i dag hovudsakleg to verktøy for språkanalyse som er tilgjengelige gjennom den språkteknologiske infrastrukturen CLARINO (2019). Desse to verktøya blir presenterte nedanfor.

4.1. Korpuskel-Lex

Korpuskel er eit grensesnitt for å tilgjengeleggjere, søkje i, og analysere, tekstkorpus (Meurer 2012). Korpuskel er eit generisk verktøy med over 60 korpus på ulike språk, format og mediatypar. Korpuskel-Lex er ei vidareutvikling av Korpuskel som vart implementert i 2018 for Revisjonsprosjektet, og let brukaren søkje på tvers av mange korpus samstundes. Som del av CLARINO er Korpuskel-Lex ope tilgjengeleg, innanfor rammene av brukslisensane som gjeld for dei enkelte tekstkorpusa.

Korpusa i Korpuskel-Lex utgjer om lag 2,4 milliardar ord, og eit oversyn over korpusa ligg føre i tabell 1. Dette utvalet svarar på fleire behov. For det første er dette korpus der vi kan vente at teksten har vore gjennom ein redaksjonell prosess før publisering (i motsetnad til t.d. bloggtekst). Dette er viktig for

å ha eit kjeldegrunnlag av ein viss skriftleg kvalitet. Revisjonsprosjektet treng vidare å skilje mellom bokmål ('bm') og nynorsk ('nn') (jf. kolonna *målform*). Det er nyttig med eit stort materiale frå ulike domene (jf. kolonnene *storleik* og *sjanger*) og som dekkjer eit breitt tidsaspekt (jf. kolonna *tidsrom*) – dette fordi vi rett nok fokuserer på moderne språk, men samstundes treng vi i revisjonen òg å sjå bruken av orda i eit større tidsperspektiv. Vi ønskjer å kunne søkje på språkleg analysert materiale (lemma og ordklasse) der det er mogleg (jf. kolonna *søkjeattributt*).

Korpusnamn	Målform	Storleik (mill.)	Tidsrom	Sjanger	Søkjeattributt
Aviskorpus (bokmål)	bm	1880,62	1998–2019	sakprosa (avis)	ordform
Aviskorpus (nynorsk)	nn	19,6	1998–2015	sakprosa (avis)	ordform
Aviskorpus annotert	bm	28,97	2001–2009	sakprosa (avis)	ordform, lemma, ordklasse
Forskning.no (2017)	bm	21,47	1998–2017	sakprosa (avis)	ordform, lemma, ordklasse
Leksikografisk bokmålskorpus	bm	102,29	1985–2013	blanda (sakprosa, skjønnlitteratur)	ordform, lemma, ordklasse
NBs frie tekster (bokmål)	bm	516,39	1765–2013	blanda (sakprosa, skjønnlitteratur)	ordform
NBs frie tekstar (nynorsk)	nn	46,02	1850–2010	blanda (sakprosa, skjønnlitteratur)	ordform
Norsk Ordboks nynorskkorpus	nn	107,8	1866–2012	blanda (sakprosa, skjønnlitteratur)	ordform, lemma, ordklasse
Talk Of Norway	bm, nn	63,8	1999–2016	sakprosa	ordform, lemma, ordklasse

Tabell 1: Tekstkorpusa som er del av Korpuskel-Lex.

Korpuskel-Lex har alle funksjonar som òg finst i Korpuskel, i tillegg til to funksjonar som har blitt utvikla særskilt for behova til ordbokredaktørane. Den første er at Korpuskel-Lex, til skilnad frå Korpuskel, let brukaren søkje i meir enn eitt korpus om gongen. Treffa kan bli sorterte ut frå ulike attributt som korpuskjelde eller tidsrommet treffa er fordelt over (jf. kolonnane i figur 1). Dermed får redaktøren raskt overblikk over bruken av eit ord, utan å måtte gå frå kjelde til kjelde og sjekke.

corpus	treff	source	year		
avis-nno	s og Medkila i Prestegardsmoen i dag. Ballen	curla	inn i lengste hjørnet utan sjanse for målvakt	BT	2008
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	AP	2011
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	SA	2011
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	AA	2011
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	FV	2011
avis-nno	r kjøpt OL-rettar. Altså risikerer vi i framtida	curling	med reklameavbrot. Samtidig har NRK vist a	AP	2011
avis-nno	all, tennis, basketball, bordtennis, volleyball,	curling	, golf og synkronsymjng. Kvifor akkurat des	BT	2015
avis-nno	YTO: ARNE S. GJONE ¶ Les også En skole for	curlingbarna	- 06.02.2005 Svikt i lærebøkene, ¶ svikt hos	AP	2005
avis-nno	ar det, går det ut på eitt. Nokre har kalla dei	curlingbarna	, barna som får kosta banen sin fri av foreldr	BT	2006
avis-nno	nge menneske i dag er ikkje kalla glasur- og	curlinggenerasjon	for ingenting, men det er det vaksne som ha	SA	2015
avis-nno	lange kallenamn, som «lydig», «flinkis» og «	curlinggenerasjonen	». Angrepa kjem gjerne når nokon går ut i rr	BT	2014
avis-nno	lt; F ¶ ~ Norsk seier over Japan ¶ De norske	curlinggutta	slo Japan 8-5 i torsdagens første VM-kamp. '	SA	2006
avis-nno	det på havets botn ligg ein tusen år gammal	curlingstein	. -  Nei, nei - han er ikkje tusen - ha	BT	2015
avis-plain	Paramount går snart i gang med " Girl in the	curl	", og Warner Bros skal lage " Lisa Andersen :	SA	2001

Figur 1: Skjermbilete frå Korpuskel-Lex: konkordansar for søk etter ord som byrjar på *curl*.

Den andre funksjonen er at det er mogleg å avgrense søkjetreffa ut frå ord i konteksten. I *Nynorskordboka* er t.d. ordet *appelsinbåt* oppført med ei generell tyding ('bit av ein skrelt appelsin') og den meir spesifikke tydinga 'båt som fraktar appelsin', som stammar frå tida då appelsinar vart frakta til Noreg med båt. For å finne korpusdøme på denne meir spesifikke tydinga, kan ein raffinere søket til at ordet *kome* må finnast innanfor ei ramme på t.d. 10 ord før og etter *appelsinbåt* (som i: *appelsinbåten kom*).

Korpuskel-Lex har eit avansert søkjespråk med regulære uttrykk, og har òg eit forenkla, grafisk grensesnitt for søk (Meurer 2019). Det er mogleg å lagre ofte brukte søk, enten privat for den innlogga brukaren eller ope, slik at alle i ordbokredaksjonen kan dele søk. Ein kan òg spare søket som ein URL, som let deg repetere same søk i dei same korpusa i etterkant.

Der er fleksible moglegheiter for å vise resultatet etter eit søk. Treffa kan bli viste som konkordansar (vising av søkjeordet i kontekst), som i figur 1, eller som frekvensordlister, som i figur 2. Frekvensordlister, som i figur 2, er praktiske for å peike ut gode kandidatar til oppslagsformer som bør vurderast med i ordboka. Døme på dette er ordet *curlingforeldre* (nedst i figur 2), som ikkje låg i ordboka allereie. Lister som dette viser òg i kva grad eit ord er produktivt i samansetjingar, som i enkelte høve kan gjere at eit grunnord vert teke med, sjølv om ordet åleine ikkje er spesielt frekvent.

Antall treff: 6693, unique values: 429. Attributt: word | ignorer storskriving | sorter: etter frekve

Side 1 av 2. Previous Next

1514 (22,62%)	curling	21 (0,31%)	curlingpresident	7 (0,10%)	curling-
913 (13,64%)	curling-VM	20 (0,30%)	curlinglandslag	7 (0,10%)	curling-gull
439 (6,56%)	curlinggutta	20 (0,30%)	curlingspilleren	7 (0,10%)	curlingbarn
301 (4,50%)	curlet	19 (0,28%)	curlingguttene	7 (0,10%)	curlingbarna
204 (3,05%)	curling-EM	18 (0,27%)	curlingdamene	7 (0,10%)	curlingdamer
159 (2,38%)	curlingjentene	18 (0,27%)	curlingsteiner	7 (0,10%)	curlingforelder
150 (2,24%)	curlinglaget	16 (0,24%)	curlingforbundets	7 (0,10%)	curlingkarrieren
129 (1,93%)	curlinglandslaget	16 (0,24%)	curlingklovnene	7 (0,10%)	curlingkolleger
121 (1,81%)	curlinghallen	15 (0,22%)	curling-gutta	7 (0,10%)	curlinglagene
116 (1,73%)	curlingherrene	15 (0,22%)	curlingklubb	7 (0,10%)	curlingnasjon
113 (1,69%)	curlinglag	14 (0,21%)	curlingbukse	6 (0,09%)	curling-finalen
112 (1,67%)	curle	14 (0,21%)	curlinggull	6 (0,09%)	curlingdrønning
111 (1,66%)	curler	14 (0,21%)	curlingkvinner	6 (0,09%)	curlingfans
99 (1,48%)	curlingmillioner	14 (0,21%)	curlingmiljøet	6 (0,09%)	curlingfeberen
71 (1,06%)	curlingforbundet	14 (0,21%)	curlingseier	6 (0,09%)	curlingfolket
65 (0,97%)	curlingbanen	13 (0,19%)	curlere	6 (0,09%)	curlinghistorie
62 (0,93%)	curlingguttas	12 (0,18%)	curla	6 (0,09%)	curlinginteressen
60 (0,90%)	curlingforeldre	12 (0,18%)	curlingbukser	6 (0,09%)	curlingjenter
54 (0,81%)	curling-seiser	12 (0,18%)	curlingfeber	6 (0,09%)	curlingkampene

Figur 2: Skjermbilete frå Korpuskel-Lex: frekvensordliste, som resultat på søk etter ord som byrjar på *curl*.

Treffa kan òg bli visualiserte med distribusjonen over tid og på tvers av korpus (trendlinjer), som er nyttig for å vurdere kor etablerte orda er i språket. Som døme kan vi ta orda *adamsslekt* ('menneska'), som ligg i begge ordbøkene i dag, og *arveprinsesse*, som ikkje låg i ordbøkene då Revisjonsprosjektet tok til. I våre kjelder fann vi berre tre treff på *adamsslekt* som er nyare enn 1940 (eitt frå 1971 og to frå 1981). Til samanlikning har *arveprinsesse* ein ganske stabil frekvens ut over 2000-talet. Korpuskel-Lex viser òg enkle statistiske kollokasjonar, slik at eit søk på *hatt* kan returnere statistisk signifikante kollokasjonar som *fjør i hatten*.

Sjølv om korpus og frekvensar er nyttige leksikografiske reiskaper, må ein hugse på at frekvensar ikkje fortel alt. Eit ord kan tilsynelatande ha høg frekvens, samstundes som frekvensen kan vere ujamnt distribuert over tid eller på tvers av ulike korpus eller domene. Korpuskel-Lex har nok dessutan, per i dag, ei uønskt overvekt av avistekst.

Eit døme på ujamn distribusjon er ordet *aortastenose* ('hjerterklaffproblem'). Dette ordet kom opp medan redaktørane jobbe med bokstaven *a* og søkte på ord som byrja på *aor-*. Ordet *aortastenose* stod for 51,33 prosent av alle treffa i resultatlista (totalt 113 treff), medan det intuitivt meir kjende ordet *aorta* berre var halvparten så frekvent. Men ser ein på distribusjonen til treffa på *aortastenose* (jf. figur 3), ser ein at nesten alle treffa er frå 2005 – som var då norske kong Harald vart innlagd på sjukehus med hjerterklaffproblem. Dette ordet fekk såleis ein relativt høg frekvens i eit veldig kort tidsrom, og hovudsakleg i avistekst. Liknande døme er ord som vart svært frekvente i ein kort periode i samband med vulkanutbrotet på Island i 2010 (t.d. *askesky*).

	(sum)	<input checked="" type="checkbox"/> aortastenose
(sum)	58 (100,0)	58 (100,0)
	58 (100,0)	58 (100,0)
2005	57 (100,0)	57 (100,0)
2014	1 (100,0)	1 (100,0)

Figur 3: Skjermbilete frå Korpuskel-Lex: distribusjonen av ordet *aortastenose* relativt til år.

På same måte kan eit ord vise seg å dominere i eit bestemt korpus eller domene – t.d. har ordet *assistenttrener* over 13 000 treff i dei samla korpusa som er tilgjengeleg i Korpuskel-Lex. Men ser ein på distribusjonen per korpus, ser ein at dette ordet er overrepresentert i Aviskorpuset (12 977 av treffa), som har veldig mykje tekst frå domenet sport. Ordet vart likevel inkludert som eit nytt ord i ordbøkene, ettersom ein kan argumentere for at høgfrekvente idrettsomgrep kan reknast som ein del av allmennspråket. Redaksjonen må difor aktivt undersøkje kvar treffa er funne, og òg konsultere andre kjelder enn dei primære tekstkorpusa som vi bruker.

4.2. NorGramBank i INESS

Det andre verktøyet vi har teke i bruk er den norske trebanken NorGramBank, som er søkbar gjennom verktøyet INESS (Rosén et al. 2012). Ein trebank kan definerast som eit tekstkorpus der kvar setning har ein detaljert syntaktisk analyse. INESS er ein infrastruktur for å bevare og gjere trebankar tilgjengelege for brukarar, og er del av CLARINO. NorGramBank er ein trebank for bokmål og nynorsk (Dyvik et al. 2016). Analysene i denne trebanken er produserte automatisk med bruk av den maskinelle grammatikken NorGram, som er skriva innanfor rammeverket leksikalsk-funksjonell grammatikk (LFG). NorGramBank inneheld om lag 80 millionar ord, fordelt på skjønnlitteratur, sakprosa og stortingstekst. Fordelinga mellom bokmål og nynorsk i NorGramBank er som i materialet i Korpuskel-Lex, med om lag 90 prosent bokmål og 10 prosent nynorsk.

NorGramBank let ordbokredaktøren søkje opp WordSketch-liknande informasjon om ord, som dei vanlegaste verbrammene til eit ord (*gifte seg, gifte bort, gifte seg til*) eller typiske døme på korleis at-setningar blir styrte av preposisjonar (*etter at, utan at*). NorGramBank har vist seg særleg nyttig for funksjonsord (fordi ein kan gjere søk baserte på dei ulike funksjonane til ordet), ord med systematiske skilnader knytt til syntaktiske relasjonar (t.d. partisippformer og adjektiv) og fleirordsuttrykk (t.d. *auge: ha auge for, få auge på, falle i auga*). Fi-

gur 4 viser døme på søk etter fleirordsuttrykk som inneheld bokmålssubstantivet *øye*. Eit fleirordsuttrykk treng berre ein forekomst i INESS for å bli registrert (sjå Dyvik et al. 2019). Leksikografen må altså framleis vurdere om fleirordsuttrykka er tilstrekkeleg etablerte i allmennspråket til å bli tekne med i ordboka.

Som dømet i figur 4 viser, vart det funne tre ulike fleirordsuttrykk i NorGramBank som inneheld *øye* (*få øye på*, *holde øye med*, *ha øye til*), sorterte etter frekvens (kolonna lengst til venstre). Ved å klikke på eit fleirordsuttrykk, t.d. på *holde#øye*med*, kan ein gjennomgå setningane som fleirordsuttrykket vart funne i.

Click on a row to see the matching sentences. |
Copy format: plain BRO

Count	#p: <i>atom</i>	globals: <i>language</i>
76	få#øye*på	nob
30	holde#øye*med	nob
1	ha#øye*til	nob

Figur 4: Skjermbilete frå INESS: fleirordsuttrykk som er registrerte i den norske trebanken NorGramBank for substantivet *øye* på bokmål, sorterte etter frekvens.

5. Konklusjon

Korpus er til god hjelp for leksikografen når det gjeld lemmatilfang, fleirordsuttrykk, ordtydingar, syntaktisk åtferd og bruksdøme. Samstundes må korpus brukast klokt, og dei kan ikkje erstatte leksikografiske vurderingar. I Revisjonsprosjektet ser vi ein stor gevinst i å knytte IT- og datalingvistisk kompetanse til leksikografiske prosjekt som driv korpusutvikling og -tilpassing etter behova som ulike prosjekt har. Vi finn stor synergieffekt i å nytte aktivt andre forskingsprosjekt og infrastrukturar ved UiB – fordi aktiv bruk frå vår side stimulerer til betre løysingar som òg kjem nye brukarar til gode. T.d. har korpusverktøyet Korpuskel blitt utvida til å få nye tekniske løysingar som er tilgjengeleg for fleire, grunna våre behov. Samstundes ser vi behovet for å vidareutvikle både tekniske løysingar og statistiske hjelpemiddel som kan forenkle og effektivisere ordbokarbeidet ytterlegare.

Litteratur

Ordbøker

Bokmålsordboka og Nynorskordboka = Universitetet i Bergen og Språkrådet.
<ordbok.uib.no> (oktober 2019)

Annan litteratur

Borin, Lars, Markus Forsberg & Johan Roxendal (2012): Korp – the corpus infrastructure of Språkbanken. I: *Proceedings of LREC 2012*. Istanbul: ELRA, 474–478.

CLARIN (2019): Nettside for den europeiske infrastrukturen CLARIN. <www.clarin.eu> (oktober 2019).

CLARINO (2019): Nettside for CLARINO, den norske noden i CLARIN. <clarin.w.uib.no> (oktober 2019).

Dyvik, Helge, Gyri Smørdal Losnegaard & Victoria Rosén (2019): Multiword expressions in an LFG grammar for Norwegian. I: *Representation and parsing of multiword expressions: Current trends*. Language Science Press, 69–108.

Dyvik, Helge, Paul Meurer, Victoria Rosén, Koenraad De Smedt, Petter Haugereid, Gyri Smørdal Losnegaard, Gunn Inger Lyse & Martha Thunes (2016): NorGramBank: A ‘Deep’ Treebank for Norwegian. I: Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk & Stelios Piperidis (red.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*. Portorož, Slovenia: ELRA, 3555–3562.

Gjesdal, Anje Müller & Gunn Inger Lyse (2016): Exploring an environmental neologism in Norwegian across corpora. I: *Neologica* 10, 39–57.

Hinrichs, Erhard W, Marie Hinrichs & Thomas Zastrow (2010): WebLicht: Web-Based LRT Services for German. I: *Proceedings of the ACL 2010 System Demonstrations*, 25–29.

INESS (2019): Nettside for infrastrukturen INESS. <clarino.uib.no/iness> (oktober 2019).

Johnsen, Lars G. (2019): *Mining digital libraries*. 23. april 2019. <www.github.com/Yoonsen/Tutorial-DHN2019> (oktober 2019).

Kilgarrieff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel (2014): The Sketch Engine: ten years on. I: *Lexicography*. Springer, 7–36.

- Meurer, Paul (2012): Corpuscle – a new corpus management platform for annotated corpora. I: Gisle Andersen (red.): *Exploring Newspaper Language: Using the Web to Create and Investigate a large corpus of modern Norwegian*. Studies in Corpus Linguistics 49, John Benjamins.
- Meurer, Paul (2019): *Corpuscle-Lex: The query language*. <clarino.uib.no/lex/page?page-id=The_query_language> (oktober 2019).
- Nøklestad, Anders, Kristin Hagen, Janne Bondi Johannessen, Michal Kosek & Joel Priestley (2017): A modernised version of the Glossa corpus search system. I: Jörg Tiedemann (red.): *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, 251–254.
- Rauset, Margunn (2019): Bokmålsordboka og Nynorskordboka – einegga, toegga eller siamesiske tvillinger? I: *LexicoNordica* 26, 155–175.
- Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik (2012): An open infrastructure for advanced treebanking. I: Jan Hajič, Koenraad De Smedt, Marko Tadić & António Branco (red.): *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*. Istanbul, Turkey: ELRA, 22–29.
- Selback, Bente (2020): «Å nei, det ordet er ikkje lov på nynorsk! » Eller ...? Om parallell redigering av to norske ordbøker. I: *Nordiska studier i lexikografi* 15, 297–305.
- Språkrådet (2014): *Norsk ordbokpolitikk*. Notat fra Språkrådet til Kulturdepartementet 3. mars 2014. <www.sprakradet.no/upload/Aktuelt/Notat%20om%20norsk%20ordbokpolitikk.pdf> (mars 2019).
- Språkrådet & Universitetet i Bergen (2016): *Samarbeidsavtale*. Upublisert.

Gunn Inger Lyse
 ordbokredaktør, ph.d.
 Universitetet i Bergen
 Sydneplass 7
 NO-5007 Bergen
 gunn.lyse@uib.no

Redigeringsmiljön för Stora finsk-svenska ordboken – återbruk, samarbete och vidareutveckling

Monica Martens

This article describes the process of developing a web based editorial platform and temporary public web interface for the bilingual Finnish-Swedish dictionary *Suomi–ruotsi-suursanakirja* published by the Institute for the languages of Finland. The printed dictionary has been out of stock for many years and as Swedish is one of the official languages in Finland, there is a constant demand for an up-to-date and reliable dictionary. The new editorial platform was implemented in co-operation between the Institute and the Dept. of Swedish, Gothenburg university, which as one of the pioneers in computational linguistics has decades of experience with lexical databases. The process of development and continuous enhancement of the editorial platform in close co-operation between lexicographers and the software developer has proved to be of mutual benefit for the bilingual Finnish-Swedish dictionary update project and an ongoing university project concerned with maintenance and development of a monolingual lexical database owned by the Swedish Academy. Both projects use a similar architecture and technical platform. Data is stored in a relational database and php/html scripts are used for both editing and presentational purposes, which gives the editors immediate access to the finalized layout.

1. Bakgrund och presentation av deltagande institutioner

Stora finsk-svenska ordboken gavs ursprungligen ut i samarbete mellan Institutet för de inhemska språken (dåvarande Forskningscentralen för de inhemska språken) och förlaget WSOY (SFSO 1997). Den senaste upplagan kom ut i tryck 2004 (SFSO 2004) och i elektronisk form 2006. Webbversionen tillhandahölls senare av förlaget Talentum. År 2016 inleddes ett samarbete mellan Institutionen för svenska språket vid Göteborgs universitet och Institutet för de inhemska språken (nedan kallat Språkinstitutet) med främsta mål att tillhandahålla en redigeringsmiljö för att möjliggöra en uppdatering av ordboken. I uppdraget kom också att ingå att hålla en utvecklingsversion av ordboken tillgänglig för allmänheten då det vid tillfället saknades resurser för en anpassning av ordboken till befintlig infrastruktur vid Språkinstitutet.

Finskans ställning som officiellt minoritetsspråk i Sverige (Lag2009:724) har stärkt språkets ställning och förväntningarna på att ordboken även ska tillgodose de sverigefinska behoven avseende bl.a. administrativa termer. Den finska enheten vid Institutet för språk och folkminnen (ISOF) är därmed en viktig intressent i uppdateringsprojektet.

Korpusbaserad lexikografi och språkteknologi har under varierande namn och organisationsformer haft en etablerad position vid Göteborgs universitet alltsedan 1960-talet. Arbetet inleddes av forskningsgruppen för modern svenska som under Sture Alléns ledning bl.a. gav ut Nusvensk frekvensordbok och fortsatte sedan inom Institutionen för språkvetenskaplig databehandling bildad 1977 (se Allén 1979). Numera bedrivs verksamheten inom institutionen för svenska språket. Det fortsatta arbetet med den lexikaliska databas som fick sin form under 1980-talet (se Sjögreen 1988) garanterades 2010 då ett 50-årigt överlåtelse- och samarbetsavtal slöts mellan Svenska Akademien och Göteborgs universitet. Under senare år har arbetet breddats mot nya utgivningsformer (se Holmer et al. 2015).

Av avgörande betydelse för projektet har varit den samlade kunskap och erfarenhet som tillförts dels genom direkt medverkan av Stora finsk-svenska ordbokens huvudredaktör Nina Martola, dels genom tillgång till de redigeringsystem som framförallt tidigare systemansvarige Christian Sjögreen byggt upp för liknande projekt vid Göteborgs universitet (Sjögreen 2013). Projektet som jag beskriver i denna artikel har i sin tur i väsentlig grad berikat vidareutvecklingen av redaktionsmiljön för Svenska Akademiens lexikaliska databas vid Göteborgs universitet. Förhoppningsvis kan denna beskrivning av den valda datastrukturen och programmiljön och det täta samarbetet mellan lexikografer och systemutvecklare vara av intresse för andra projekt som står i färd med att utvärdera sina redaktionssystem och publiceringsprocesser.

2. Uppdraget

Efter en inledande kontakt hösten 2015 inleddes arbetet med ett månadslångt pilotprojekt 2016 under vilket Nina Martola och jag tittade närmare på vilka indata vi hade att utgå ifrån och undersökte lämplig systemmiljö och systemarkitektur. Vi gjorde även en proof-of-concept (konceptverifiering) för att verifiera att det tänkta upplägget fungerade för redigering i Helsingfors mot en server i Göteborg. Efter ett framgångsrikt genomfört pilotprojekt skrevs ett avtal för 2017–2018 med option på årsvis förlängning, vilket har utnyttjats två gånger. Det nuvarande avtalet om drift och underhåll av redigerings- och publiceringsmiljön gäller alltså till och med år 2020.

Stora finsk-svenska ordboken (SFSO 2019) har legat tillgänglig för allmänheten på svenska institutionens server sedan april 2017 med länk från Språk-

institutets webbplats, där också omtexterna finns (SFSO-omtexter). Den publika installationen uppdateras några gånger per år med nyinlagt och redigerat material. En viktig källa till nya uppslagsord är nyorden i ordboken över det finska allmänspråket, *Kielitoimiston sanakirja*, som regelbundet tillhandahålls i digitalt format för inläsning i ordboksdatan. Redaktör och representant för uppdragsgivaren är lexikograf Nina Martola, som även samordnar kontakterna med finska enheten vid Institutet för språk och folkminnen, som i sin tur ansvarar för språkrådgivning för de nationella minoritetsspråken i Sverige och bl.a. utarbetar sverigefinska ordlistor.

Sporadiska förbättringar av användargränssnittet och utökade sökmöjligheter har införts även utifrån önskemål som inkommit via den kontaktmejlfunktion som finns på söksidan. Förslag till förbättringar av ordboksartiklar inkommer också den vägen och kanaliseras till redaktören.

Det nuvarande redigeringsystemet är avsett att användas endast under ett begränsat antal år, varefter den uppdaterade ordboken skall flyttas hem till Språkinstitutet och integreras med där befintliga system. Fokus har därför legat på att snabbt och resurseffektivt få till stånd ett användbart redigeringsystem för en eller ett fåtal expertanvändare. Ett krav var också att det skulle finnas en snabb och smärtfri process för att publicera nya versioner på nätet.

3. Genomförande

3.1. Verktyg

Efter en snabb inventering av tillgängliga verktyg valdes en beprövad lösning i form av webbgränssnitt skrivet i programmeringsspråket php och datalagring i relationsdatabashanteraren MariaDB på en linuxserver. I ett inledningsskede hade vi hoppats på att kunna använda Språkbankens KARP-system (KARP), men den redigeringsmiljön bedömdes inte vara tillräckligt utvecklad för att snabbt kunna tas i drift för en ny tillämpning.

I stort sett samma systemarkitektur, med egenutvecklade webbskript som kommunicerade med en relationsdatabas via en Apache-webbserver på operativsystemet Linux, hade tidigare använts på Göteborgs universitet bl.a. för Svensk Morfologisk Databas (SMDB), se Berg & Cederholm (2001). Även den av Svenska Akademien ägda lexikaliska databasen låg i en relationsdatabas, men inte MariaDB utan Ingres, och redigeringsystemet var byggt med 4GL-verktyget OpenRoad som inte lämpade sig för en modern webblösning. Dessa system kunde alltså i viss utsträckning fungera som förlagor för arbetet med redigeringsystem för den finsk-svenska ordboken, men det var ändå ett helt nytt system med egna åtkomstlösningar, egen datamodell och egna gränssnitt som behövde byggas.

3.2. Datamodell

Utgångspunkten för vårt arbete var den version av Stora finsk-svenska ordboken som förlaget WSOY färdigställde år 2004. Ordboksartiklarna levererades i form av en XML-fil (se figur 1) och uppdraget var att skapa ett redigeringsystem som möjliggjorde uppdatering av ordboken.

```
<ENTRY ID="SURUSSK:aakkonen" HWD="aakkonen"><HEAD>aakkonen<CODE>38</CODE>
</HEAD><TARG>bok|stav</TARG><INFL>-staven -stäver; </INFL><EXPL>(aakkoset)
</EXPL><TARG>alfabet </TARG><INFL>-et =; </INFL><USG>(kuv </USG><EXPL>alkeet)
</EXPL><TARG>abc </TARG><INFL>--t --n </INFL><SYM>▶ </SYM><EX>latinalaiset
aakkoset </EX><TEX>latinska alfabetet; </TEX><EX>Morsen aakkoset
</EX><TARG>morsealfabetet; </TARG><EX>merenkulun aakkoset </EX><TEX>abc i
navigation; </TEX><EX>keittotaidon aakkoset </EX><TEX>kokkonstens
abc</TEX></ENTRY><ENTRY HWD="aakkosellinen"><HEAD>aakkosellinen<CODE>38</CODE>
</HEAD><TARG>alfabetisk</TARG></ENTRY>
```

Figur 1: Utdrag ur XML-filen som utgjorde indata till det nya redigeringsystemet.

Tabell	Kolumn	Beskrivning
heads	Inr	lemmanr, unikt löpnr, nyckel
	homonr	homografnr
	head	uppslagsform (inklusive ev. paradighänvisning inom {})
	head2	variantform
	dis2	bruklighet o.dyl.
	visas	0=ej klar, 1=klar för visning
	kommentar	redaktörens interna noteringar
	entry	id för referens, ex <ENTRY ID="SURUSSK:aatami" HWD="aatami">
	date_inserted	datum
	user_inserted	användarsignum
senses	Inr	lemmanr, gemensam nyckel med heads-tabellen
	nnr	betydelsegruppnr (intern numrering)
	num	momentnummer (för visning)
	explnr	betydelsenr (intern numrering)
	translnr	översättningsnr (intern numrering)
	pre_expl	taggning som föregår finskspråkig förklaring
	expl	finskspråkig förklaring
	post_expl	taggning som efterföljer finskspråkig förklaring
	pre_svexpl	taggning som föregår svensk översättning
	svexpl	svensk översättning
	infl	böjningsinformation av det svenska ordet
	anm	efterföljande anmärkningar och hänvisningar till andra ord
	kommentar	interna noteringar
	refid	refererbart momentid
	exempel	0=ej exempel, 1= exempel (påverkar visningsformatering)
	date_inserted	datum
	user_inserted	användarsignum

Figur 2: Beskrivning av tabeller och kolumner efter inläsning i databasen, namngivningen ansluter till taggnamnen i infilen.

Tyvär visade det sig att det mesta av den omsorgsfullt definierade struktur som funnits i det ursprungliga SGML¹-baserade editeringssystemet Author Editor hade gått förlorad på vägen. Vi blev därmed tvungna att arbeta med en förenklad datamodell som delade upp och grupperade ihop den utplattade informationen i XML-filen till två huvudtabeller, en för artikelhuvuden och en för betydelser, där den senare innehöll en rad per översättning/ekvivalent (se figur 2 och 3).

lnr	explnr	translnr	pre_expl	expl	post_expl	pre_svexpl	svexpl	infl	ann
2	1	1					bok stav	-staven -stäver;	
2	2	1		(aakkoset)			alfabet	-et =;	
2	3	1	<USG> (kuv </USG>	alkeet)			abc	--t --n	
2	4	1	<SVN> </SVN>	latinalaiset aakkoset			latinska alfabetet;		
2	5	1		Morsen aakkoset			morsealfabetet;		
2	6	1		merenkulun aakkoset			abc i navigation;		
2	7	1		keittotaidon aakkoset			kokkonstens abc		
24	1	1					alfabetisk		
29	1	1					alfabetiskt;		
29	1	2					i bokstavsordning		
35	1	1					alfabetisk ordning		
35	1	2					bokstavsordning	-en	
42	1	1					alfanumerisk		
47	1	1					ordna(1) ngt		
47	1	2					alfabetiskt;		
47	1	3					ordna(1) ngt i		
47	1	4					bokstavsföljdd;		
47	1	5					alfabetisera(1);		
47	1	6			<SPEC>(atk myös)		sortera(1) ngt		
47	1	7			</SPEC>		alfabetiskt;		
47	1	8					sortera(1)		
57	1	1							<RSEP>ks. </RSEP><REF IDREF="SUURUSSK:aakkonen">aakkonen</REF>
63	1	1		(aakkosjärjestys)			bokstavsordning	-en;	
63	1	2		(aakkosjärjestys)			alfabetisk ordning;		
63	2	1		(aakkostaminen)			alfabetisering	-en -är;	
63	3	1	<RSEP>ks. myös </RSEP>käännösmalli 2						

Figur 3: Början av senses-tabellen, som innehåller en rad per ekvivalent eller referens.

Problemet beskrivs så här av Nina Martola (mejlkorrespondens februari 2020):

Det jag upplever som grundläggande var att WSOY hade frångått den ursprungliga tanken att ett systematiskt uppmärkt material kan ges olika utformning. På 90-talet kunde vi få ut utskrifter i tre olika layouter, varav två inte alls såg ut som traditionella ordboksartiklar. De användes för olika redaktionella kontroller. För att det här skulle lyckas var strukturen hierarkisk och inte lineär, så att räckvidden för varje angivelse var klar, och så att exempelvis efterställda angivelser, typ myös lääk ’även medicinskt’, strukturellt kopplades samman med rätt ekvivalent. På WSOY utformades uppmärkningen senare enbart med tanke på sidlayout och gjordes gemensam för flera ordböcker. Man tog bort en mängd ”mellantaggar” så att bara taggen närmast varje enskilt innehållsligt element blev kvar. På så vis gick ju information om vad som hör vart förlorad. I WSOY-versionen lade redaktörerna själva ut skiljetecknen medan de hade genererats utifrån

¹ SGML (Standard Generalized Markup Language) var en föregångare till XML (Extensible Markup Language) och HTML (Hypertext Markup Language).

taggningen i 1990-talsversionen. De skrev också ut förkortningarna och tog bort det vi kallade ”entity”, dvs. att alla förkortningar var uppmärka element, så att man i slutändan kunde välja hur man ville realisera dem, som förkortningar med eller utan punkt, eller utskrivna.

I den fil vi hade att arbeta med saknades alltså explicit hierarkisk information inom artikeln. Analysen komplicerades ytterligare av att för strukturen betydelsebärande skiljetecken, såsom semikolon, låg inbäddade i texten. Vid konverteringen blev fokus därför på att i artikeltexten identifiera och återskapa de grundläggande enheterna: betydelsemoment, finskspråkig förklaring och svenskspråkig översättning. Övriga attribut såsom bruklighetsangivelser och metaspråklig information grupperades utifrån deras placering. Räckvidden för framförallt framförställda förklaringar var många gånger svår att avgöra entydigt, så implementationen fick styras av en pragmatisk hållning som tillät en ursprungstrogen formatering av artikeln även när analysen var osäker.

3.3. Redaktionsgränssnittet

Redaktionsgränssnittet har genomgått en stegvis vidareutveckling, men består i grunden av en söksida (se figur 4) och en radorienterad redigeringsvy (se figur 5) där man kan lägga till och ändra i tabellraderna i databasen.

Om ordboken [Ta kontakt](#) [Grammatik](#) [Suomeksi](#) [KS](#)

Välkommen till Stora finsk-svenska ordboken

(sidan senast uppdaterad 16:54:10 finsk tid)

Sökord: NM

Sökalternativ

- bland uppslagsorden/hakusanosta
- i texten/tekstista
- i kommentarerna
- i finska exempel och förklaringar (expl)
- i svenska översättningar (svexpl)
- pre_expl
- post_expl
- pre_svexpl
- inflection
- anm

Avgränsa till uppslagsord mellan och ösa

Urval: endast nya/udet. ej klara/epåvalmt alla/kaikki

[Sök i svenska/finska listorna/Haku ruotsinsuomalaisista listoista \(73\)](#) Håll sökalternativen öppna/Pida auki

Modifierad sökning: hela ord ord början ordstut

med jokertecken (_ och %)

Visa som:

Klicka på ett fetstilt ord för att se hela artikeln

kaltistus Ej klar att visa [Redigera](#) z1938/NMz NYTT

-aaltainen Ej klar att visa [KS Redigera](#) z1851/NMz zUTGÄRz under ssg

-aiheinen [KS Redigera](#) z1851/NMz

[Redigera](#) z1851/NMz

-aiheinen [KS](#)

tasa-aiheinen seos en jämn (homogen) blandning

© Institutet för de inhemska språken, Helsingfors

© Institutet för de inhemska språken, Helsingfors (artikelinnehåll) Göteborgs universitet, Inst. för svenska språket (teknikplattform)

Figur 4: Redaktörernas söksida: sökresultat i listform nere till vänster, vid klick i listan öppnas komplett artikel till höger.

Utöver information ur ordbokstabellerna har man från redaktörssidan även åtkomst till tvåspråkiga terminologordlistor från det svenska Institutet för språk och folkminnen, för att underlätta arbetet med den specifikt sverigefinska vokabulären.

Under menyvalet Grammatik återfinns delar av inledningen till den tryckta ordboken, bl.a. en komplett listning av använda förkortningar.

Beteckningarna *pre_expl*, *post_expl* och *pre_svexpl* på redaktörernas söksida avser bruklighetsinformation och annan specialtaggad tilläggsinformation som visas före respektive efter den finskspråkiga beskrivningen respektive före den svenskspråkiga översättningen. De ursprungliga taggarna har bevarats för korrekt formatering i artikelvisningen.

Fältet *pre_expl* kan t.ex. innehålla informationen *kuv*, förkortning av *kuva-kielessä*, *kuvallisesti*, *kielikuva* (bildligt) eller *sot.slg*, förkortning av *sotilaslangissa* (soldatslang), som inledare till den finskspråkiga betydelsebeskrivningen. Fältet *post_expl* innehåller information som ska presenteras efter eventuell finskspråkig betydelsebeskrivning, t.ex. latinska namn på växter.

Fältet *pre_svexpl* innehåller information som visas framför den svenskspråkiga översättningen, så som stilvärde, geografisk tillämpning (Sverige/Finland) och annan metainformation avseende den svenska översättningen.

Skriv in ett (finskt) ord eller id
OBS om bryt+omnumrering önskas måste sökning göras på id, ej ord (för att undvika homografproblem)

ord: alternativt, id: 1033376 [Visa/Uppdatera](#) (sidan senast uppdaterad 16:42:27)

Omnumrering: nnr (+5) och expl (+1) för alla rader med nnr >= 0 (skriv in nnr för den rad fr.o.m. vilken omnumrering ska ske) [Numrera om](#)

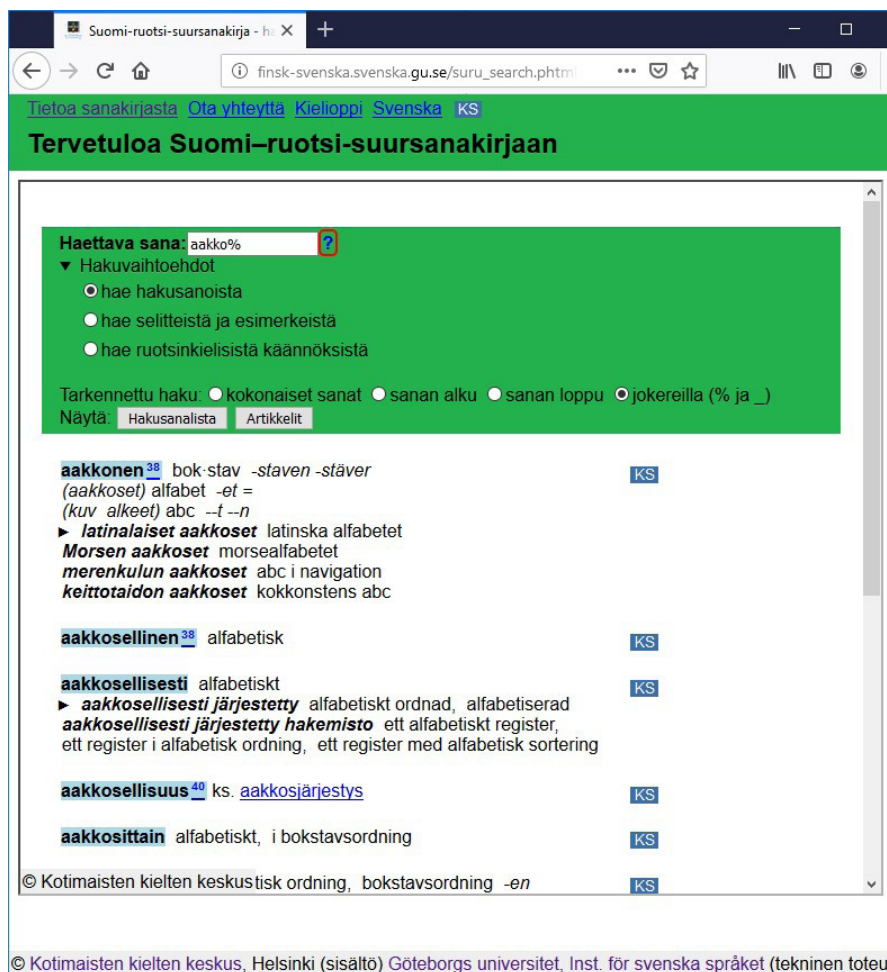
Visa formaterad artikel		vildakko ^{4A}		id: 1033376 head2: dis2:		komm: Redigera	
num	pre-expl	finska (expl)	post-expl	pre-svenska	svenska	infil	övrigt
1	1			1	djungel	- <i>eln -ler</i>	Redigera
Lägg till ny betydelse/nytt exempel							
4	1	2	(<i>sademetsä</i>)	1	regnskog	- <i>en -ar</i>	Redigera
Lägg till ny betydelse/nytt exempel							
7	1	3	(<i>aarniometsä</i>)	1	urskog		Redigera
Lägg till ny betydelse/nytt exempel							
--							
9	2	1	(<i>tiheikko</i>)	1	snår	- <i>et =</i>	Redigera
9	2	1	(<i>tiheikko</i>)	2	snårskog		Redigera
Lägg till ny betydelse/nytt exempel							

Figur 5: Redigeringsvyn.

Som framgår av figur 4 och 5 förväntas redaktören vara svenskspråkig och väl förtrogen med datamodellen. I databasen har alla kolumner engelskspråkiga namn, i anslutning till den taggning som fanns i infilen (se figur 1 och 2). I redaktionsgränssnittet har dessa namn endast i undantagsfall getts en kompletterande förklaring.

3.4. Det publika gränssnittet

Det publika gränssnittet är en kopia av redaktörernas sökgränssnitt, dock så att endast de ord som är markerade som klara att visas är åtkomliga. Sökfunktionaliteten är reducerad på så sätt att klassificeringskolumner och interna kommentarer inte är sökbara (se figur 6). Varken javascript eller cookies krävs för tillgång till ordboken. I databasen sparas avidentifierad information om sökta ord så att återkoppling till redaktörerna kan ges i form av statistik över gjorda sökningar inklusive information om sökningar på ord som ej fått träff i databasen.



Figur 6: Den publika söksidan, här den finskspråkiga versionen.

Den publika sidan är fritt tillgänglig utan begränsningar och utan krav på någon form av registrering och riktar sig till alla typer av användare. Tack vare återkoppling från användarna vet vi dock att en viktig användargrupp är professionella översättare. Angående målgruppen för ordboken se vidare Martola, N. i denna volym.

4. Framtida utveckling

Det är i skrivande stund inte klart vad som kommer att hända med Stora finsk-svenska ordboken efter att nuvarande avtal går ut år 2020. Avsikten är att systemet skall inkorporeras med Språkinstitutets övriga lexikonresurser vilket förmodligen innebär en migrering till en annan typ av system men hur det kommer att ske rent praktiskt återstår att se. Under åren 2018–2019 har den lexikaliska databasen vid Göteborgs universitet, ur vilken Svensk ordbok utgiven av Svenska Akademien tas fram, genomgått en migrering från databashanteraren Ingres till MariaDB och från 4GL-redigeringsystemet OpenRoad till webbaserad redigering. I den processen har man haft god nytta av erfarenheterna från arbetet med den tvåspråkiga ordboken både vad gäller praktiska systemerings- och programarkitekturlösningar och användarvänlig redigeringsmiljö.

Allt tyder på att den tvåspråkiga finsk-svenska ordboken och den enspråkiga svenska inom de närmaste åren, förhoppningsvis ömsesidigt stärkta och mer livskraftiga, kommer att gå skilda vägar efter att under några år härbärgerats och växt till sig i samma systemmiljö.

Litteratur

- Allén, Sture (1979): Språkvetenskaplig databehandling. I: *Proceedings of NODALIDA 1979*, s. 171–179.
- Berg, Sture & Yvonne Cederholm (2001): *Att hålla på formerna. Om framväxten av Svensk morfologisk databas*. <hdl.handle.net/2077/24079>
- Holmer, Louise, Monica von Martens & Emma Sköldberg (2015): Making a dictionary app from a lexical database: the case of the Contemporary Dictionary of the Swedish Academy I: I. Kosem, M. Jakubiček, J. Kallas & S. Krek (eds.): *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, 11-13 Aug-ust 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., pp. 32–50 <www.elex.link/elex2015/proceedings/eLex_2015_03_Holmer+vonMartens+Skoldberg.pdf> (oktober 2019).
- ISOF = Institutet för språk och folkminnen <www.isof.se/sprak/minoritetsprak.html>

- KARP = <www.spraakbanken.gu.se/swe/forskning/infrastruktur/karp>
 Lag2009:724 = *Lag om nationella minoriteter och minoritetsspråk* <www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-2009724-om-nationella-minoriteter-och_sfs-2009-724>
- Martola, Nina (2020): Den tvåspråkiga ordboken – ett absurt åtagande. I: *Nordiska studier i lexikografi* 15, s. 25–41.
- Martola, Nina (2020). Mejlkorrespondens februari 2020.
- SFSO 1997 = *Suuri suomi–ruotsi sanakirja/Stora finsk-svenska ordboken*. (1997). Redaktion: Birgitta Romppanen (huv.red), Ilse Cantell, Nina Martola & Mats-Peter Sundström. Forskningscentralen för de inhemska språken & WSOY. Helsingfors.
- SFSO 2004 = *Suomi–ruotsi-suursanakirja/Stora finsk-svenska ordboken*. (2004). Forskningscentralen för de inhemska språken & WSOY: Helsingfors.
- SFSO 2019 = *Suomi–ruotsi-suursanakirja/Stora finsk-svenska ordboken* <finsk-svenska.svenska.gu.se> (oktober 2019).
- SFSO-omtexter = *Stora finsk-svenska ordboken* <www.sprakinstitutet.fi/sv/vara_ordbocker/flersprakiga_ordbocker/stora_finsk-svenska_ordboken> (oktober 2019).
- Sjögreen, Christian (1988): *Creating a Dictionary from a Lexical Database*. I: *Studies in Computer-Aided Lexicology*. Stockholm 1988
- Sjögreen, Christian (2013): Svenska Akademiens Ordlista – sedd ur en tekniker synvinkel. I: *Language, Football and All That Jazz: a Festschrift for Sölve Ohlander* / (eds. G. Bergh, R. Bowen & M. Möbärg), Göteborg, Acta Universitatis Gothoburgensis, Göteborg 2013, s. 469–480.

Monica Martens
 systemutvecklare, fil. kand.
 Göteborgs universitet
 Inst. för svenska språket
 Box 200
 SE-405 30 Göteborg
 monica.martens@gu.se

Nettbrukaren i fokus: digitalisering og revisjon av Norsk Ordbok *a–h*

Johan Myking, Oddrun Grønvik & Stig Jarle Helset

The dictionary Norsk Ordbok (Norwegian Dictionary) was initiated in 1929 and accomplished in 2015 with the publication of its twelfth and final volume. The Dictionary covers the Nynorsk written standard as well as the vocabulary of Norwegian dialects in all their variety. Two thirds of the 330 000 lemmas (letters *i–å*) are open to digital search. Work has now been initiated to revise the contents of the first third (letters *a–h*) as well as making this part available for digital searches. The revision will have to adapt to the needs of digital-minded user groups, as no updated paper version is planned. Such adaptive measures include the extensive use and development of electronic corpora.

1. Innleiing

Norsk Ordbok er den einaste ordboka som startar i tidleg nynorsk periode (rundt 1600) og dekkjer heile tidsspennet fram til i dag. Den er den einaste norske ordboka som kan gjera vitskapleg bruk av den komparative språkvitskapen til å identifisera norsk ordtilfang diakront og synkront. Meir enn halvparten av ordtilfanget i Norsk Ordbok er uttrykt der i normert form for første gong.

Arbeidet med Norsk Ordbok vart sett i gang i 1929 ved Universitetet i Oslo og vart avslutta i 2015. Det Norske Samlaget har gjeve ut verket i 12 band, fullført gjennom prosjektet Norsk Ordbok 2014 (NO2014) i åra 2002–2015. Verket inneheld om lag 330 000 oppslagsord. Dei siste to tredelane, bokstavstrekket *i–å*, er digitalt tilgjengelege.

Ordboksverket inkluderer eldre kjelder og siktar mot å dokumentera leksikalsk talemålsvariasjon i eit større omfang enn vanleg i ordbøker der litterære kjelder dominerer. Norsk Ordbok fører vidare Ivar Aasens prinsipp for skriftstandardisering på talemålsgrunn:

Den gjennomgåande dekninga av norsk talemål sikrar at alle brukarar av norsk skal kunna finna att noko av sitt språk i verket, systematisert, vitskapleg fortolka og løfta inn i ein samanheng som viser både dialektale særdrag og språkleg samanheng mellom ulike område av landet. (Wetås & Aa 2017:291)

Ordboksarbeidet ved Universitetet i Oslo vart fasa ut og avslutta då papirutgåva av Norsk Ordbok var ferdigstilt. Materialet og det faglege ansvaret vart flytta til Universitetet i Bergen i 2016 og etablert som Språksamlingane ved Universitetsbiblioteket der, saman med andre delar av dei samlingane av språkleg materiale som Universitetet i Oslo sa frå seg ansvaret for. Tre typar infrastruktur blir drøfta nedanfor: Ordbokshotellet, Metaordboka og Nynorskorpuset, sjå avsnitta 4 og 5. Dessutan er Setelarkivet viktig, der ligg 3,2 millionar ordsetlar samla inn frå 1930-talet og til i dag lagra i digital form. Denne infrastrukturen utgjer hjelperessursane som var utvikla for Norsk Ordbok, men blir altså vidareført i Språksamlingane.

2. Ideen om Norsk Ordbok som nettordbok – NO-AH

Universitetet i Bergen (UiB) har i samarbeid med Høgskulen i Volda (HVO) teke ansvaret for å gjera ferdig den digitale utgåva av Norsk Ordbok og revidera alfabetstrekket *a–h*. Prosjektet har fått akronymet NO-AH, og har etter søknad til Kulturdepartementet fått midlar over statsbudsjettet, i første omgang for 2019 og 2020. Mange av dei faglege føresetnadene peikar utover alfabetstrekket *a–h*, men det er altså dette strekket prosjektet må konsentrera seg om.

NO-AH blir eit nytt leksikografisk prosjekt som må retta seg mot nettbrukaren og derfor byggjast på til dels nye faglege prinsipp, av di det ikkje kjem noka ny papirutgåve. Ideen om overgang til digital publisering kom som del av den omfattande digitaliseringa av produksjonsprosessen etter 2002:

Norsk Ordbok fremstår som en tradisjonell, trykt ordbok. Den er planlagt i 12 bind og skal være ferdig i 2014. Under den synsvinkelen har Norsk Ordbok svært lite med nettordbøker å gjøre. Men bak denne trykte fasaden foregår redigeringen av Norsk Ordbok i et heldigitalt redigeringsverksted. De trykte ordboksbindene er en presentasjon i tradisjonelt trykkformat av bare en del av informasjonen i databasene. (Ore 2012:489)

I søknaden til Kulturdepartementet for 2019 er målet for prosjektet formulert slik: «fullføra Norsk Ordbok på nett ved å revidera innhald og form på alfabetstrekket *a–h* (NO-AH), og slik skapa ei komplett elektronisk Norsk Ordbok». Det planlagde resultatet er då eit elektronisk system av ordboksdatabasar i Språksamlingane som skal gje ein eksplisitt, lettbrukt og vitskapleg fundert presentasjon av norsk ordtilfang.

2.1. Ein brukaranalyse i fem drag

Utviklinga av ei slik moderne digital ordbok må byggjast på ein brukaranalyse, fordi ordbokbrukarane i dag kan ha utvikla andre føresetnader for bruken av ordboka enn dei som galdt då første band kom ut i 1966. Vi ser såleis fem grunnleggjande drag ved det norske publikumet som må få følgjer for arbeidet:

- 1) Det norske publikumet har svært stor interesse for språk, både for tale, skrift og etymologi

Informasjonsbehovet må tilpassast det allmenne språkinteresserte publikumet, som er stort, og derfor må materialet òg vera så heildekkjande som mogeleg. I og med at dei digitale kunnskapane varierer, må ein då ha ei vurdering av brukarkompetanse: Kor stor språkkunnskap, digital kunnskap og kunnskap om ordboksstruktur kan ein venta av publikum, og kva konsekvensar skal det få? Dette er av dei spørsmåla som vil bli greidde ut tidleg i prosjektperioden, i samråd med representantar for viktige brukargrupper.

- 2) Det norske språksamfunnet er karakterisert av stor variasjon, både i tale og skrift

Ordboka må derfor førast vidare ut frå det etablerte prinsippet om å inkludera både talemål/målføre («det norske folkemålet») og skriftspråket, med eit dokumenterande meir enn normerande formål:

Eit ordbokverk med eit eksplisitt dokumenterande føremål og med ei diakron kjeldedekning som strekkjer seg over ein periode på 500 år, har eit anna føremål og eit anna potensielt publikum enn ei preskriptiv produksjonsordbok (Wetås & Aa 2017:291).

Den normerande effekten kan ein dels oppnå med koplingar til dei to offisielle standardordbøkene Nynorskordboka og Bokmålsordboka, dels ved at oppslagsorda er kopla til gjeldande rettskrivingsformer. Eit grensesnitt mot forskning i norsk talemål kan sikra at variasjonen i talemålet blir synleggjord (jf. målføresynopsisen, nedanfor).

3) Det norske samfunnslivet er prega av aukande og intensivert digitalisering og «snakkande» teknologi

I søknaden om prosjektet *Norsk ordbok a–h* heiter det at resultatet av arbeidet ikkje berre skal vera ei ordbok, men eit system av databasar og språkressursar. Desse ressursane må ha gjenbruks- og overføringsverdi, dei må vera utvikla med tanke på integrasjon i språkteknologi som del av dagens satsingar på digital infrastruktur.

4) Papir er «ute»

Dette er eit trivielt faktum: Det digitale mediet overtek, ingen nye papirutgåver er p.t. planlagde av nokon av dei større norske ordboksverka, som NAOB og standardordbøkene. Samstundes veit ein at brukargruppa for Norsk Ordbok ikkje er homogent datakompetent – ein skal leggja til rette både for dei som er mest vane med papirmediet, samstundes som praksisendringar gjev høve til å tilfredsstilla andre og meir avanserte brukarbehov enn i dag.

5) Eldre band treng revisjon!

Den siste føresetnaden gjeld ikkje mediet, men innhaldet. Band I av ordboksverket kom ut i 1966, deretter kom det tre band i 1978, 1994 og 2002 før ein gjekk over til digitale produksjonsformer i gjennom prosjektet NO2014 frå 2002. Dei fire første banda og første halvdel av band V er ikkje digitalt tilgjengelege. Verda har endra seg i dei meir enn 50 åra sidan band I kom ut, dels ved at nye omgrep er komne til, dels ved at eldre fenomen blir sedde gjennom nye briller. I band I finn ein følgjande artiklar om dei opphavlege innbyggjarane i Australia:

australiemann m. [-ra'] (He.Fo, Vo.Ty), australiar m [-ra'-; til *Australia*, avl. t lat *australis* 'sørleg' av *auster* 'sønnavind'] (kvit) person som eig heime i el er frå Australia. [...] **australneger** m [-ra'] ein av det opphavlege folket i Australia.

Nemninga *aborigin* kom i bruk sist på 60-talet, og vil bli redigert i den reviderte Norsk Ordbok, men dei eksisterande artiklane kan ikkje sløyfast. I den digitaliserte versjonen vil denne typen oppslagsord bli merkte som forelda, medan definisjonane vil bli justerte.

2.2. Praksisendring og endra vektlegging

Som nemnt kan overgangen til det digitale mediet krevja endra leksikografisk praksis på nokre område, som skal presenterast nedanfor. Samstundes skal ein ta vare på heilskapen i verket, og må dermed respektera dei etablerte konvensjonane så langt råd er. Verket skal helst framstå nokolunde samla og homogent i form og struktur.

Når det gjeld innhaldssida, er det rimeleg å ta omsyn til eit primærbehov hos moderne nettbrukarar, som er å finna svaret på spørsmålet med ein gong. Dette gjeld særleg definisjonar og bruksdøme. *Norsk Ordbok i-å* (frå band V) vart redigert både for papir og nett, og redigering for papir inneber plassparing. Norsk Ordbok har derfor mange snautt formulerte definisjonar. I NO-AH vil det vera meir rom for fullskrivne definisjonar. Der *Norsk Ordbok i-å* måtte nøya seg med synonymdefinisjonar, vil NO-AH kunna kosta på seg både ein fleirordig definisjon og ei synonymtilvising eller to.

I kva grad desse ulike omsyna kan harmoniserast, vil visa seg eit stykke inn i framtida. Fokuset no ligg på alfabetstrekket *a-h*, men det vil opplagt bli behov for innhaldsrevisjonar av strekket *i-å* òg, sjølv om dette ved dags dato ikkje ligg inne i søknadene om finansiering.

3. Korleis presentera talespråkleg variasjon?

Som alt nemnt, har den grunnleggjande ideen bak Norsk Ordbok vore å gje ein dokumentasjon både av det nynorske skriftspråket og av det talemålstilfanget som ligg til grunn. Dermed har Norsk Ordbok ei større utfordring enn mange andre ordboksverk: Korleis bør informasjon om målførevariasjon presenterast i ei vitskapleg nettordbok, slik at utbyttet blir optimalt for heile spennet av ordboksbrukarar?

Den behovsanalysen – altså kva ein går ut frå at brukarane treng å få vita – som ligg til grunn for målføreinformasjonen i Norsk Ordbok, er ikkje direkte uttrykt, men framgår av opplysning om at føreseielege (sjølvgeivne) målføreformer er sløyfa. Norsk Ordbok på papir vender seg primært til lesarar som kan systematikken i det norske målførekartet. Omtalen i brukarrettleiinga (Vikør 2018:78ff.) syner at 1) oppstillinga av former (med kjelde) føreset kunnskap om konvensjonane for bøyingsmorfologi i norsk, og dermed litt meir enn grunnleggjande grammatisk terminologi; 2) dei viktigaste isoglossane for norsk talemål er føreset kjende, og likeins det områdekartet som desse isoglossane gjev opphav til. Såleis er omtalen av jamvekt, «tjukk l», infinitivsending og apokope

ofte forenkla til det som blir kalla «samleformer». Det gjer at forma «skjere-» i dømet nedanfor også blir brukt til å representere jamningsformer i tilfanget, som /sjyrru-/ og /sjurru-/:

I **skjere-** i sms [målf òg (stundom berre i einskilde sms) *skjer-* (Furnes, Folldal, Gbr sumst, Hall, Uvdal, Ri, Modum, Andebu, Tinn, Rog sumst, Selje, Bud, Ndm og Trl og No sumst), *skjera-* (SAustl sumst, Jr, Vik, Saltal, Beiarn, Gildeskål), *skjerar-* (YSandsvær, Andebu, Rødøy, Meløy, Kjerringøy, Vstrål, Vadsø); jfr òg *skjerarfolk* og *skjerarlår*] til *II skjera*. 1 om handlinga å skjera, såleis t d: skjere-fart, -måte, -resultat, -rørsle. 2 om (del av) reiskap til å skjera med, såleis t d: skjere-kant, -maskin, -reiskap, -verkty. 3 om underlag til å skjera noko på, såleis t d: skjere-brett, -fjøl. (Transkribert versjon, redigeringshandboka (Grønvik m.fl. (red.). 2016: 162)

I dømet ovanfor kan ein røynd brukar av Norsk Ordbok sjå at poenget med målføreoppstillinga er å visa fugevariasjonen for førsteleddet, fordelt geografisk. For nye brukarar av Norsk Ordbok vil dette ikkje vera opplagt. I nettutgåva kan ein også få vist dokumentert utbreiing i kart, noko som raskt vart populært då tenesta vart lansert i 2014.

Ei revidert Norsk Ordbok vil som før sagt vera bunden av eksisterande konvensjonar. Men dagens ordboksbrukarar skal også kunna ha nytte og glede av målføreopplysningane utan så omfattande førehandskunnskapar som tidlegare føresett. Det står att å finna ut kva slags form på talemålsinformasjonen eit breiare publikum ser seg tent med. Det må uansett bli eit system som kan gjera bruk av eksisterande oppsett i redigerte ordartiklar.

4. Det digitale etterslepet etter NO2014

4.1. Korleis handtera eldre kjelder?

Noverande band I er tynt når det gjeld dei eldste kjeldene, men denne mangelen kan no supplerast gjennom bruk av dei digitale verktøya som har vore utvikla for Norsk Ordbok, særleg den såkalla Metaordboka, men òg andre. Viktig digitalisering er gjennomført i tida etter at NO2014 vart avslutta – særleg er det viktig at ordbøkene til Aasen (1873) og Ross (1895) er digitaliserte og no ligg i det såkalla Ordbokshotellet (jf. Grønvik & Ore 2018), slik at dei er søkbare i seg sjølve og via Metaordboka. At NO-AH blir heildigital, vil innebera at redaktørane får tilgjenge til det digitale etterslepet etter NO2014, som t.d. fulldigitalisering av talemålskjeldene Norsk målføresynopsis og Norsk dialektatlas, og nettilgjenge til dei eldre ordsamlingane (1600–1850).

I Oslo hadde redaksjonen tilgjenge til fleire unike papirarkivaliar, men det må det no bli slutt på – risikoen og slitasjen er for stor. Heildigitalisering må innebera at Setelarkivet berre blir digitalt tilgjengeleg, slik at ein ikkje lenger sorterer lappar på skrivebordet. Dermed må ein gjera noko med dei 18 000 setlane som manglar registrert oppslagsord og dei ca. 120 000 setlane som manglar registrert kjelde (synleg på foto, men ikkje registrert i metadata).

4.2. Metaordboka

Metaordboka er eit register over alt tilfang som er knytt til ein gjeven «leksikalsk gjenstand» ('lexical item', jf. Atkins & Rundell 2008), dvs. ein teiknstreng som krev eigen ordboksartikkel. Metaordboka knyter (knippe av) daterte oppslagsformer til kvar artikkel, slik at den leksikalske gjenstanden får kopla saman sine normerte uttrykk over tid med tilfanget som utgjer grunnlaget for innhaldsanalysen:

Metaordboka er et register der hver innførsel eller «artikkel» har et hode som tillater indeksering av belegg med (1) språkform (bokmål og nynorsk), (2) flere oppslag for hver språkform, (3) normeringsstatus for hver oppslagsform, (4) datering av hver status, (5) ordklasse, (6) leddanalyse. (Grønvik & Ore 2018:89)

Det er meininga at ordboka skal bli ein felles indeks for Språksamlingane. I høve til Norsk Ordbok er Metaordboka reiskapen som skal gjera det mogleg å søkja i Norsk Ordbok i moderne rettskriving. Bakgrunnstilfanget for normering og datering ligg i ordlistene i Ordbokshotellet. Språksamlingane samarbeider med Språkrådet om den såkalla Ordbanken, som inneheld alle normerte norske ord med bøyingsformer. Ordbanken er ein fullformsgenerator med lenkjer til alle oppslagsformer i dei to standardordbøkene Bokmålsordboka og Nynorskordboka (Grønvik & Ore 2013). Om Metaordboka blir kopla til Ordbanken, vil ein også kunna opplysa om gjeldande bøyning for oppslagsformene i Norsk Ordbok (utover 1938-rettskrivinga, som ordboka opplyser om). Metaordboka kan på denne måten utviklast til eit omfattande register over historisk normering.

5. Korpus – kjeldekritikk og kvalitet

5.1. Korpusgrunnlaget for Norsk Ordbok

For nynorsk er det bygd opp eit tekstkorpus med ein diakron dimensjon – 1866 til i dag, det såkalla Nynorskkorpuset. Eit tilsvarande for bokmål finst ikkje. Det er enno ikkje avklåra kort stort omfang korpusarbeidet kan få i NO-AH, men det er ein intensjon om sterk utvikling i regi av Språksamlingane. Norsk Ordbok og prosjektet NO-AH vil stå sentralt, men arbeidet må peika lenger – ut over prosjektperioden for NO-AH.

I tida før prosjektet NO2014 som starta i 2002, var parolen å utnytta alt tilfang til siste komma, men under NO2014 vart ordsamlingane brukte både som redigeringsgrunnlag og til å proporsjonera ordboka. No er overvekta av nytt avistilfang eit problem på grunn av mengda. Ein balansert korpuspolitikk peikar seg dermed ut som avgjerande metodisk grunnlag for det vidare redigeringsarbeidet.

5.2. Nynorskkorpuset

Nynorskkorpuset omfattar om lag 106 millionar ord frå tekstar som høyrer til ulike sjangrar, skjønnlitteratur, avistekstar, og sakprosa frå 1870-åra og fram til i dag. Etter ein pause på nokre år er arbeidet med utvikling av Nynorskkorpuset sett i gang att, med juridisk og fagleg grunnlagsarbeid. Korpuset har ein sjølvstendig bruksverdi i tillegg til den leksikografiske. Særleg er Språkbanken ved Nasjonalbiblioteket som leverandør av strukturert tekstmateriale til språkteknologiske formål ein stor interessent og mottakar av denne typen tekst.

Ved universitetet i Oslo vart det òg utvikla eit leksikografisk korpus over bokmål, og begge desse korpusa høyrer til Språksamlingane ved UiB i dag. I samarbeid mellom fleire universitet og fagmiljø er det òg utvikla eit stort Norsk aviskorpus. Potensialet er altså stort for å driva korpusbasert leksikografi på norsk.

Oppsummert vil dette seia at korpusdelen av arbeidet med NO-AH må ha tre mål: For det første må korpuset aukast og utvida så det blir meir representativt, eventuelt supplert med nye spesialkorpus. For det andre er leveranse til språkindustriell bruk eit politisk ønske som det er viktig å ta omsyn til. Og for det tredje, for å få dette til, må det leksikografiske fagmiljøet arbeida med auka vekt på kjeldekritikk og brukarkompetanse.

6. Norsk Ordbok a–h som løysinga på ein politisk situasjon

Norsk Ordbok har sidan den første starten i 1930 vore ei politisk sak på den måten at finansieringa har skjedd over statsbudsjettet. Ikkje minst var den radikale omlegginga til NO2014 i 2002 avgjerande for at verket låg føre i tolv band i 2015. Også spørsmålet om digitalisering og revisjon av strekket a–h er eit spørsmål som krev politiske løysingar. *Norsk Ordbok a–h* (NO-AH) er såleis resultatet av ein politisk situasjon og måten den vart løyst på.

Då UiB i samarbeid med HVO i 2016 gjekk inn og overtok Språksamlingane og ansvaret for digitalisering av *Norsk Ordbok a–h*, følgde ansvaret for dei to standardordbøkene Bokmålsordboka og Nynorskordboka med i porteføljen. Ein ser no at overtakinga føreset ei strategisk satsing på leksikografi som forskingsfelt. Ei slik satsing representerer eit potensial for utvida aktivitet og «gjødsling» av den nasjonale språkvitskaplege satsinga på norskspråkleg infrastruktur og språkteknologi.

Litteratur

Ordbøker

NAOB = *Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur og Kunnskapsforlaget. Digital utgåve: <www.naob.no>

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet. Band I–XII. 1966–2015. Oslo: Det Norske Samlaget. Digital utgåve i-å: <no2014.uib.no>

Ross, Hans (1895): *Norsk Ordbog. Tillæg til 'Norsk Ordbog' af Ivar Aasen*. (1971). Fotografisk opptrykk. Oslo: Universitetsforlaget.

Aasen, Ivar (1873): *Norsk Ordbog med dansk Forklaring. Fjerde uforandrede udgave, (1918)*. Kristiania: Vestmannalaget og Cammermeyers forlag.

Annan litteratur

Atkins, Beryl T. (Sue) & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Grønvik, Oddrun m.fl. (red.) (2016): *Redigeringshandbok for Norsk Ordbok 2014*. <no2014.uib.no/eNo/tekst/lenkjer.html>

Grønvik, Oddrun & Christian-Emil Smith Ore (2014): Samvirket mellom ordbank og ordbok. I: Fjeld, Ruth Vatvedt & Marit Hovdenak (red.): *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.-16. august 2016*. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 13. Oslo. Novus forlag, 1–22.

- Grønvik, Oddrun & Christian-Emil Smith Ore (2018): Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforsking av ordtilfang. I: Ásta Svavarsdóttir; Halldóra Jónsdóttir; Helga Hilmisdóttir & Þórdís Úlfarsdóttir (red.): *Nordiske Studier i Leksikografi 14, Rapport fra 14. Konference om Leksikografi i Norden Reykjavík 30. maj - 2. juni 2017*. Nordisk Forening for Leksikografi, 87–95.
- Ore, Christian-Emil Smith & Oddrun Grønvik (2018): Comparing Orthographies in Space and Time through Lexicographic Resources. I: Jaka Čibej, Vojko Gorjanc, Iztok Kosem Simon & Krek (eds.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, Faculty of Arts, 159–172. <e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1.pdf>
- Ore, Christian-Emil Smith (2012): Nettordbøker og Norsk ordbok – hvordan etablere en vitenskapelig nettordbok. I: Eaker, Birgot, Lennart Larsson & Anki Masttisson (red.): *Nordiska studier i Lexikografi 11. Rapport från Konferensen om lexikografi i Norden Lund 24.-27. mai 2011*. Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr. 12. I samarbeite med Språkrådet i Norge og Svenska Akademiens ordboksredaktion, 488–499.
- Vikør, Lars S. (2018): *Inn i Norsk Ordbok. Brukarrettleiing og dokumentasjon*. <no2014.uib.no/eNo/tekst/lenkjer.html>
- Wetås, Åse & Leiv Inge Aa (2017): Museale betraktningar av eit moderne ordbokverk – tilsvar til John Ole Askedal si melding av Norsk Ordbok. I: *LexicoNordica*, 289–300.

Johan Myking
 professor
 Universitetet i Bergen
 Boks 7805
 NO-5020 Bergen
 johan.myking@uib.no

Oddrun Grønvik
 forskar
 Universitetet i Bergen
 Boks 7805
 NO-5020 Bergen
 oddrun.gronvik@uib.no

Stig Jarle Helset
 førsteamanuensis
 Høgskulen i Volda
 Boks 500
 NO-6101 Volda
 stig.helset@hivolda.no

Den digitale ordbokens innvirkning på den praktiske leksikografien

Carina Nilstun

The advent of computers revolutionized lexicography. Editing tools were developed along with digital corpora, yet the format of the output remained the same: a book with columns of strongly condensed text. Then came the digital dictionary, and many paper dictionaries were supplemented by a digital version. These digital versions clearly reveal their paper format origin and the constraints associated with the latter medium. Now, the indigenously digital dictionaries are here. This might seem as a small step (or even a setback, as one medium is lost), but is actually a marked shift in our field as it allows lexicographers to elaborate how the dictionary entries are built and presented. The *Norwegian Academy Dictionary* (NAOB) was published on the internet December 2017 (www.naob.no) as a solely digital dictionary. This article is about the advantages of creating and publishing a dictionary only digitally – discussing both implemented features from NAOB, and thoughts for future work.

1. Innledning

Datamaskinen revolusjonerte leksikografien da den ble tatt i bruk i ordboks-redaksjonene. Det ble utviklet digitale redigeringsverktøy og elektroniske tekst-korpus. Forandringene gikk i hovedsak på metodikken; produktet var fortsatt en bok, og ordboksartiklene så ganske like ut.

Med ordboken som bare finnes digitalt, åpner muligheten seg til å la teknologien utvikle også ordboksartiklene. Jeg vil i denne artikkelen trekke frem temae ne plassbruk, brukermedvirkning, nye muligheter for betydningsbeskrivelse og fremfor alt lenking.

2. Fra *Norsk Riksmålsordbok* til *Det Norske Akademis ordbok*

Utgangspunktet for dette bidraget er arbeidet med *Det Norske Akademis ordbok* (NAOB). Samtlige eksempler som gis er hentet fra den. Siden det ikke er plass til å gjengi alle ordboksartiklene det refereres til, bes leseren om å søke opp disse selv på naob.no.

Først vil jeg kort gjengi veien frem til dagens NAOB. NAOB bygger på firebindsverket *Norsk Riksmålsordbok* (NRO) som utkom i perioden 1937–1957, med to supplementsbind i 1995. Disse to alfabetstrekke ble slått sammen da innholdet ble digitalisert for rundt 20 år siden. På dette grunnlaget har man bygget videre i et slikt omfang at NAOB ikke kan regnes som en ny utgave av NRO, men som et selvstendig verk. Prosjektets historikk er presentert på nettstedet (www.naob.no/om). Lenge var tanken at NAOB skulle finnes både på nett og i bokform, men i årene frem mot publiseringen i 2017 ble det klart at man ikke skulle forsøke å få finansiert en papirutgave, og at man – uten å utelukke at det kunne komme – ikke ville ta noen hensyn til en eventuell papirutgave i det videre arbeidet.

Mange ordbøker finnes både på papir og digitalt, og det varierer nok i hvilken grad den digitale ordbokens muligheter er utnyttet. Man kan for eksempel ha egne spesifiseringer for hver av utgivelsene, blant annet med tanke på ekspansjon av forkortelser, men svært ofte gjenkjenner vi papirordboken og gjenfinner dens begrensninger i den digitale ordboken. Petzell (2017) behandler dette i større detalj i sin anmeldelse av SAOB på nett. Han betoner den enorme forskjellen det er på en statisk papirordbok som også finnes digitalt tilgjengelig og en digital ressurs som gjennom lenking og avanserte søkemuligheter har potensiale til langt å overgå papirordboken (Petzell 2017:274). Det er i alle tilfeller et interessant spørsmål om det er papirordboken eller nettordboken redaktørene av ordbøkene som utgis i begge medier har i bevisstheten når de redigerer.

Papirordboken er nok ikke død, men for de store prosjektene synes den å være i livets høst. Heller ikke *Den Danske Ordbog* (DDO) har planer om noen ny utgivelse av papirutgaven. Det betyr at vi i arbeidet ene og alene har den digitale ordboken for øye; det er den vi skriver for. Når man er kommet dit, ser man plutselig hvilket stort skritt dette er for den praktiske leksikografien: Ordboksartikkelens utforming kan nå utvikles videre. Papirformatets begrensninger er tilbakelagt.

Nedenfor ser vi først et bilde fra NRO (figur 1) som starter på artikkelen *konferanse*. Det er tettskrevne spalter med sterkt kondensert tekst. Neste bilde (figur 2) viser artikkelen *konferanse* i NAOB. Det visuelle uttrykket er langt luftigere¹.

¹ Ordboksartikkelen er gjengitt slik den vises på smarttelefon.

konferanse [-fɛra'ngsɛl, en, -r (fr. *conference*, mlat. *conferentia* til *conferre* 'bringe sammen'), mest litt., adm. **1**] **a**) (politisk) forhandling ell. rådslagning; forberedende møte av representanter for stater; jvf. *fredskonferansø* og *kongress*. **b**) møte, samtale som en person har (med en annen) ell. som finner sted (mellem to ell. flere) for å drøfte en (viktig) sak: *ha konferanse med en om noget* | *embedsmæssige konferanser* (B. Lie, *Overm.*, 164) | (*de holdt*) *raadslagning og konference* (J. Lie, *Trold*, 73). **2**) fag., det å konferere (2), sammenholde (skriftstykker, dokumenter): *de to stenografer er optatt med konferanse med å konferere sine stenogrammer*.

konferencier [-rangsie'], en, -er (fr. *conférencier*, egtl. 'foredragsholder'), person som i en revy presenterer de opptredende og binder de enkelte nummer sammen, oftest med humoristiske bemerkninger.

konferere, -te, -t, -ing (ty. *konferieren*, fr. *conférer* til lat. *conferre*), mest litt., adm. **1**] **a**) forhandle (med en om noget): (*direksjonen*) *havde gjort dette uden at konferere med ham* (Kiell., *Fort.*, 154). **b**) (om to ell. flere) rådslå i fellesskap: *De er vel ikke kommet hit over bare for at konferere om klokken?* (Elster d. y., *Skyg.*, 237). **2**) sammenholde, sammenligne (bøker ell. dokumenter for å se om innholdet stemmer).

Figur 1: Artikkelen *konferanse* (med mer) i NRO.

konferanse substantiv

BØYNING en; konferansen, konferanser

+

UTTALE [kånfɛra'ŋsɛ], [konfɛra'ŋsɛ] ?

ETYMOLOGI fra fransk *conférence*, av middelalderlatin *conferentia*, til *conferre* 'bringe sammen'

BETYDNING OG BRUK

1 større arrangement med foredrag og diskusjoner, hvor deltagerne utveksler kunnskap om, erfaringer med og synspunkter på bestemte (faglige) emner

EKSEMPLER

- *en internasjonal konferanse*
- *arrangere konferanse*
- *det var over 300 deltagere på konferansen*

Figur 2: Starten på artikkelen *konferanse* i NAOB.

3. Momentene

Jeg vil etter tur tematisere de ulike sidene som berøres i overgangen fra papirformatet og til den rendyrkede digitale ordboken.

3.1. Plassen

Det man merker umiddelbart når man skriver for den digitale ordboken, er all plassen som plutselig byr seg. Det betyr på ingen måte at definisjonsspråket skal miste sin stramhet, men mye annet har godt av å få breie seg litt.

I NAOB finnes det nå ingen andre forkortelser enn de som er alminnelige i løpende tekst, for eksempel *e.l.* (*eller lignende*). I tillegg har vi innført det vi kan kalle ledetekster, eller overskrifter, for de ulike elementene, så som uttale, sitater osv. Den type informasjon lå tidligere underforstått i oppsettet. Og tenk bare på en ting som linjeskift: på bildet av *konferanse*-artikkelen i NRO (figur 1) ser man at det bare var ved nytt lemma at man spanderte linjeskift.

Sist, men ikke minst gir det digitale formatet mulighet til å ha luft i og rundt ordboksartiklene. De massive tekstspaltene fra papirordbøker kan lett få en ordboksbruker til å miste motet.

3.2. Illustrasjoner, lyd og video

Illustrasjoner finner man i noen papirordbøker, men bruken begrenses av plassmangel. Med økt plass kan man i større grad ta i bruk illustrasjoner i betydningsbeskrivelsen – som supplement til definisjonene. Noen ord, for eksempel *sprengskisse*, er vanskelige å beskrive bare med ord, mens en illustrasjon vekker umiddelbar gjenkjennelse. Det finnes foreløpig ikke illustrasjoner i NAOB, men det er noe vi ønsker oss.

Også lyd og/eller video kan være nyttige som del av betydningsbeskrivelsen. Definisjonen av *beatboxe* kunne ha tjent på å utstyres med lyd, mens *dab* og *dabbe*, kunne ha vært ledsaget av en videosnutt.

Mulighetene for å formidle hva et ord betyr er altså blitt flere, og man kan sette inn ulike måter i kombinasjon, alt etter hva slags ord det er man skal definere.

3.3. Lenking

Lenking er kanskje den største nyvinningen som har kommet med den digitale ordboken. Særsilt vil jeg nevne lenkingens fortrefelighet ved beskrivelse av orddanning. Vi har i NAOB lagt vekt på systematisk beskrivelse av suffikser, prefikser og sammensetningsledd, og på å lenke så presist som mulig inn i be-

tydningsstrukturen. I dataene får man da også oversikt over frekvensen for de ulike orddannelsene ved at man kan se hvor mange lenker det går til de ulike artikler/betydninger.

Eksempel: Suffikset **-ete** (med variantform **-et**) danner adjektiver og er delt inn i tre betydninger etter hvilken ordklasse suffikset kobles til, alle med eksempler. Adjektivet **møkkete** har som etymologi «avledet av møkk med suffikset -ete», hvor -ete lenker til riktig betydning i suffiksets artikkel. Lauvstad (dette nummer) går dypere i detalj om NAOBs etymologier.

Man kan i papirordbøker naturligvis henvide hit og dit og spesifisere ved å sette betydningsnummer i parentes, men det krever sitt vedlikehold og det er ikke særlig brukervennlig.

I NAOB lenker vi også en del internt i en ordboksartikkel. Lange artikler har både innholdsfortegnelse og uttrykksfortegnelse i starten av artikkelkroppen. Disse kan ekspanderes og minimeres, og hver betydning og hvert uttrykk er en lenke til plasseringen lenger nede i artikkelen.

3.4. Henvisning og henvisningslenking

Henvisning mellom uttrykk har vært vanlig i papirordbøkene. Av plasshensyn har det vært uaktuelt å føre samme uttrykk flere steder, og henvisningen har således vært en ren service til leserne. Når man søker på «ugler i mosen» i NAOB og får to treff, hvorav ett er en henvisning, er det rester etter analog praksis. I en digital ordbok må slikt regnes som støy. I et rent digitalt miljø ønsker vi at uttrykket skal vises i de to artiklene **ugle** og **mose** – og naturligvis kunne søkes opp og vises separat. I praksis vil det si at de faste uttrykkene er egne elementer, som lemmaene, men at de vises flere steder. Her er DDO forbilledlige: Uttrykket «der er ugler i mosen» kan søkes opp direkte, i tillegg til at det finnes på alfabetisk plass blant uttrykkene under både **ugle** og **mose**.

En annen type henvisning eller henvisningslenking er fra unormerte, foreldede eller feilstavede former og til rett artikkel. I papirordboken kan man ikke gjøre så mye av dette, dels av plasshensyn og dels fordi det vil oppleves forstyrrende. I den digitale ordboken kan slike former legges inn skjult, som en ren brukerhjelp til dem som trenger det. Dette gjøres både i NAOB og i DDO.

3.5. Lenking 2.0

Med den rene digitale ordboken kan lenkingen tas et steg videre: Man kan lenke til eksternt innhold.

En relevant ekstern lenking er lenking til selve litteraturen som de autentiske beleggene er hentet fra, noe vi har gjennomført i NAOB. For samtlige verk som finnes allment tilgjengelig i digital versjon, har nå verksnavnet fått understrek-

ning (som er den etablerte formen for å vise lenker) og i tillegg en liten firkant med pil, som indikerer at lenken leder til eksternt innhold. Flesteparten av disse lenkene går til Nasjonalbiblioteket² (www.nb.no), men vi lenker også til for eksempel Henrik Ibsens skrifter (www.ibsen.uio.no), og etter hvert skal vi også lenke til Bibelen (www.bibel.no/nettbibelen). Hvis man klikker på en hyperkoblet verkstittel, åpnes boken i en egen fane. For verkene til Nasjonalbiblioteket legges et utsnitt av det aktuelle sitatet automatisk inn i søkefeltet, og hvis teksten er godt optisk lest, kommer man til riktig side og med deler av sitatet markert med gul bunn. Man skal naturligvis alltid være varsom med eksternt lenking. Vi lenker derfor ikke til for eksempel nettaviser, hvor vi vet at lenkene er flyktige. Lenkene må gå til institusjoner som har en bevissthet rundt dette med tilgjengelighet og dokumentasjon, slik som Nasjonalbiblioteket, som bruker varige lenker.

Mange språkbrukere har ikke noe innarbeidet forhold til skillet mellom *proprier* og *appellativer*, og særlig ikke til det at *proprie*ne tradisjonelt ikke beskrives i ordbøkene. En ordbok har altså oppslag for *skandinavisk* og *skandinav*, men ikke for *Skandinavia*. Dette skillet trenger ikke å være så absolutt lenger. *Skandinavia* nevnes i etymologien til *skandinav* og *skandinavisk*, og nå har vi mulighet til å lenke til informasjon om *proprier* eksternt. Det samme gjelder *eponymer*, altså ord som er dannet til personnavn. Grunnstoffet *curium*, kan vi lese i etymologien, har fått sitt navn etter ekteparet Curie. Også her vil ordboksartikkelens bruksverdi øke hvis man fra disse navnene lenker til eksternt informasjon om personene. Nettopp dette med navn i ordbøker er tema i LexicoNordica 24. Trap-Jensen & Lorentzen (2017) tematiserer dette inngående for DDO og peker på at argumentet knyttet til plassmangel ikke lenger er gyldig, gitt den digitale ordboken. De sier videre at selv om *proprie*ne mangler betydning (og derfor tradisjonelt har vært utelukket fra allmennordbøker), har de andre sider som passer inn i en ordbok (normert form, uttale, etymologi), og de nevner at grensen mellom *appellativer* og *proprier* ofte er vanskelig å trekke – både for brukerne og for leksikografene (Trap-Jensen & Lorentzen 2017). I NAOB har vi allerede ordklassen *substantiv (egennavn)*, men med få forekomster. Et eksempel er *Kristus*. Det planlegges ikke å åpne for dette som en stor klasse, men lenking til eksternt innhold kan være en spennende viderutvikling av ordbøkene på dette feltet.

NAOB avviklet sitt samarbeid med Kunnskapsforlaget ved overgangen til 2020. Prosjektets eier, *Det Norske Akademi for Språk og Litteratur*, har opprettet et aksjeselskap til å administrere ordboken, og fra november 2019 deler vi lokaler med *Store norske leksikon* i Oslo sentrum. Vi vil sammen se nærmere på

² Dessverre er ikke alle Nasjonalbibliotekets digitaliserte bøker tilgjengelige fra utenlandske IP-adresser.

hvordan vi gjensidig kan forbedre våre tjenester. Leksikonet finnes nå også bare digitalt, og ordboken og leksikonet som tidligere stod side om side i bokhyllene, kan nå gripe inn i hverandre gjennom det digitale formatet. Leksikografisk må dette anses som nybrottsarbeid.

4. Brukerne

Avstanden mellom brukerne og redaksjonen kan minskes betydelig ved overgangen fra papirordbok til nettordbok. Brukerne er en viktig ressurs for ordboksredaksjonen, og terskelen for å ta kontakt er mye lavere for brukerne av en digital ordbok enn for brukerne av en papirordbok. På naob.no er hver artikkel utstyrt med en knapp med teksten «send oss en kommentar til denne artikkelen». Foretar du derimot et søk som ikke gir treff, får du opp knappen «send inn forslag til et oppslagsord». Begge leder til et skjema med send-knapp nederst.

Brukerne kan enkelt melde fra om feil og mangler, og om ord, betydninger og uttrykk de savner.

Feil kan rettes av redaktørene og eksporteres til naob.no med én gang. Gode forslag til nye ord redigeres og publiseres i månedlige bolker. Forslagene vi får er overveiende gode, og vi har etter hvert en del brukere som bidrar jevnlig. Interaksjonen mellom redaksjon og brukere er høyst reell, og dermed får vi en levende ordbok. Det at vi responderer hurtig og også siterer fra de helt nyeste bøker og aviser, imponerer og motiverer brukerne.

Vi har også stor nytte av søkeloggene. Her kan vi se hva som søkes på, men som ikke gir treff, og vi kan se hva det søkes mye på. På topplisten gjenkjenner man ofte saker fra nyhetene, og iblant er det aktuelt å føye til nyere sitater og kanskje nye betydninger eller nye uttrykk i disse artiklene.

5. Avslutning

Det er en veldig spennende og inspirerende fase NAOB nå er inne i. Basen er gjennomarbeidet fra A til Å og bragt ajour frem til nåtiden. Det gir mulighet for videreutvikling, og vi ser at mulighetene er store og nye siden vi ikke har noen papirordbok å ta hensyn til. Å skrive for den rene digitale ordboken oppleves som en helt annen disiplin enn å skrive papirordbøker, og denne nye disiplinen er ikke ferdig utformet.

Litteratur

Ordbøker

DDO = *Den Danske Ordbog* (2003–2005). København: Det Danske Sprog- og Litteraturselskab & Gyldendal. <www.ordnet.dk/ddo> (oktober 2019).

NAOB = Tor Guttu et al. (red.): *Det Norske Akademis ordbok*. Oslo: Kunnskapsforlaget og Det Norske Akademi for Språk og Litteratur. <www.naob.no> (oktober 2019).

NRO = *Norsk Riksmålsordbok*. Bind I–IV 1937–1957, bind V–VI 1995. Oslo: Kunnskapsforlaget.

Annen litteratur

Lauvstad, Hanne (2020): NAOBs etymologier. I: *Nordiska studier i lexikografi* 15, 195–203.

Petzell, Erik M. (2017): Svenska Akademiens ordbok på nätet. I: *LexicoNordica* 24, 267–276.

Trap-Jensen, Lars & Henrik Lorentzen (2017): Navnestof i ordbøger – nye tider, nye måder. I: *LexicoNordica* 24, 109–130.

Carina Nilstun
leksikograf MA
Det Norske Akademis ordbok
Grensen 3
NO-0159 OSLO
carina.nilstun@naob.no

Å ta Hans Ross på ordet: Ross' ordbok i relasjon til Aasens med Metaordboka som verktøy

Christian-Emil Smith Ore

Ivar Aasen's and Hans Ross' Norwegian dictionaries are pivotal in the study of Norwegian language in the 19th century. Ross presented his dictionary as an addendum although it is larger than Aasen's. The paper analyzes the relation between the two dictionaries based on the set of lexical items defined by the *Metadictionary*, a repository for source material developed for the project *Norsk Ordbok 2014*. The paper contains a brief introduction to the research program of Aasen and Ross and a description of the Metadictionary and the repository for dictionaries, *Ordbokshotellet*. Finally, we discuss a solution unifying the dictionaries.

1. Innledning

I denne artikkelen vil jeg se på hvordan en kan bruke *Ordbokshotellet* og de leksikalske gjenstandene i *Metaordboka* som et verktøy for å sammenlikne ordinventaret i de beslektete ordbøkene til Ivar Aasen (Aasen 1873) og Hans Ross (Ross 1895). Ordbokshotellet og *Metaordboka* ble lagd i forbindelse med prosjektet *Norsk Ordbok 2014* (NO2014) for å gi en systematisk inngang til retrodigitaliserte ordbøker og kildematerialet generelt. Jeg vil også kort vise hvordan en kan bruke *Metaordboka* og *Ordbokshotellet* til å lage en enhetlig digital utgave av Aasens og Ross' ordbøker. Ordbøkene er helt sentrale i studiet av norsk språk i det nittende århundret, men Aasen og Ross hadde svært ulik bakgrunn, og deres språkvitenskapelige program er også forskjellig. Jeg vil derfor gi en liten introduksjon til Aasen og Ross som målreisere, lingvister og leksikografer.

2. Ivar Aasen (1813–1896)

Ivar Aasen var en norsk språkforsker og dikter. Han var en selvlært bondesønn fra Sunnmøre. I 1836, som 22 år gammel omgangsskolelærer, skrev han det korte språkpolitiske manifestet «Om vort Skriftsprog» (Aasen 1836). Manifestet ble først publisert i 1909. Hovedpåstanden var at de norske dialektene er en direkte videreføring av norrønt språk. På grunnlag av en systematisk dokumentasjon av de enkelte dialektene mente han at et «Selskab, opprettet af sprogkyndige Mænd» kunne lage et felles skriftspråk», og skriver at «Vi ønske os just et folkesprog, et

som enhver Landsmand uden Møie kan tage Deel i» (Aasen 1836). I dette forsiktig formulerte manifestet understreker han at dette bare skulle være et tilbud og ikke pådyttes folket. Det ble aldri et slikt selskap og heller ikke noen separate beskrivelser av de enkelte dialektene.

I 1842 fikk Aasen gjennom anbefaling fra biskopen i Bergen et toårs stipend med mulig forlengelse fra DKNVS (Det Kongelige Norske Videnskabers Selskab) i Trondheim med oppdrag å utarbeide en felles eller flere ordbøker og grammatikk(er) for et utvalg dialekter. Det har vært diskutert om Aasen fikk stipendet fordi DKNVS ønsket å dokumentere at norsk (folke)språk var en videføring av norrønt og ikke bare var et slags forkvaklet dansk eller svensk som mange hevdet på den tiden. Aasens støttespillere i DKNVS tenkte seg neppe at en skulle erstatte dansk med et nytt norsk skriftspråk. Som resultat av Aasens arbeid kom først *Det norske Folkesprogs Grammatik* (Aasen 1848) og så *Ordbog over det norske Folkesprog* (Aasen 1850). Begge fikk svært positive anmeldelser av den norske historikeren P.A. Munch (Munch 1848, 1850). Men Munch var mer skeptisk til forsøkene på en moderne standardortografi og foretrakk en tilnærmet norrøn stavemåte.

En kan vel si at så lenge det etablerte miljøet så Aasens målreising som akademisk dokumentasjonsvirksomhet, var den uproblematisk. Det var først da en oppfattet at Aasens forslag til skriftspråk var ment som en konkurrent til dansk at den til dels svært usaklige målstriden begynte. I følge Walton (1996) var språkstriden på 1850-tallet medvirkende til at Aasen i større grad fokuserte på det internasjonale forskersamfunnet som målgruppe både for *Norsk Grammatik* (Aasen 1864) og *Norsk Ordbog* (Aasen 1873).

I forordet til *Norsk Grammatik* (Aasen 1864) tar Aasen opp igjen argumentasjonen fra manifestet (Aasen 1836). Fra forordet kan en lese at grammatikken skulle være en hjelp til å skrive landsmål og en språkdokumentasjon som kunne brukes i internasjonal (sammenliknende) språkforskning der norsk språk til da var uteglemt.

Ivar Aasens grammatikk generelt og orddanningskapittelet spesielt var en nødvendig forutsetning for å kunne definere en rettskrivningsnormal og for en systematisk behandling av ordforrådet. I tillegg til grammatikker og ordbøkene satte Aasen opp en norsk tesaurus, *Norsk Maalbund* (1876). Den var inspirert av *Roget's Thesaurus* (Roget 1852). Norsk Maalbund ble publisert som manuskript i 1925 (Aasen 1925).

Med moderne språkbruk kan en si at i ungdomsmanifestet beskriver Aasens forskningsprosjektet sitt og de viktigste forskningsspørsmålene knyttet til dette. Aasen ønsket å undersøke om det var mulig å lage et godt begrunnet og solid utkast til et moderne og standardisert norsk skriftspråk på grunnlag av systematiske undersøkelser av de norske dialektene. Gjennom grammatikkene og ordbøkene demonstrerte han at det var mulig.

3. Hans Ross (1833–1914)

Hans Ross ble født inn i embedsborgerskapet; faren og farfaren var prester. Selv tok Ross artium og ble cand. theol. i 1855. Men han ville ikke bli prest og arbeidet som lærer ved flere høyere skoler inntil han fikk statsstipend fra 1877 og livet ut. Historien skal ha det til at Ross ble landsmålsforkjemper etter å ha blitt introdusert til Aasens grammatikk (Aasen 1848) av sin gymnaslærer. Ross traff Aasen på slutten av 1850-tallet og var en aktiv nynorskforkjemper siden. Han arbeidet også som publisist, organisasjonsbygger og ikke minst som innsamler av språkdokumentasjon gjennom stort sett årlig feltarbeid til omlag 1890, se også Venås (2009).

Ross var ikke alltid strateg på egne vegne. Han var åpenhjertig og tenkte kanskje ikke så mye på hva som lønte seg for ham selv. Han hevdet at dansk ikke hadde noen fremtid i Norge fordi den ikke hadde rot i det talte språket i Norge (Grønvik 1989). Om nordmenns bruk av dansk hevdet han:

Telemarkinger har sagt mig, at de leste lettere bøger lavede i Danmark end de i Kristiania lavede. Dansken kan nemlig sit sprog, bevæger seg i det som Fisken i havet; mens vi ikke kan Dansk skikkeligt, men ligger og spræller og gisper i vasspytterne, som står igjen i fjæren efterat ebben har ført de store vande ut. (Ross 1878:7)

Dette og tilsvarende utsagn kan være medvirkende årsak til at Universitetet (i Christiania) ikke ville ha Ross som professor i 1881. Men Ross fikk siden full anerkjennelse, ble medlem av Vitenskapsakademiet i 1891, og ordboka (Ross 1895) ble svært godt mottatt.

I motsetning til hva Aasen gjorde i sitt ungdomsmanifest, formulerte Hans Ross aldri noe livslangt forskingsprogram. Men en kan likevel si at i Ross' implisitte språkvitenskapelige program står studiet av talemålet sentralt. Talemålet må læres og studeres der det brukes. Synkron dokumentasjon er viktigere enn den diakrone dokumentasjonen basert på historiske røtter. En må transkribere uttalevarianter, og alle observerte varianter må tas med. Både sikre og mindre sikre metodevalg og funn må dokumenteres. Videre mente Ross at en kan ikke ta hensyn til (lande)grenser når en dokumenterer språk. Ross baserte sine analyser på Aasens systematiske beskrivelse av norsk fonologi og leksikon.

4. Metaordboka som analyseverktøy

Metaordboka ble lagd i 1999 som et verktøy for å gi en felles inngang til det digitaliserte kildematerialet til prosjektet Norsk Ordbok 2014. Metaordboka er ingen statisk database og vokser etter hvert som nytt kildemateriale kommer til.

Hver innførsel eller «artikkel» har et hode og ett eller flere oppslag. For hvert oppslag kan en registrere grunnform, ordklasse, leddanalyse, språkform og normeringsstatus med tidsrom. Oppslagene beskriver ord som har eller har hatt status som normerte former i rettskrivningen. Artikkeltroppen består av referanser eller ”hyperlenker” til ulike kildedatabaser som speiler den variasjonen som foreligger i skrift og tale over lang tid. Informasjonen som kan registreres for hvert oppslagsord i Metaordboka gjør at den ligger tett opp til fullformsystemet Norsk ordbank (se Ore & Grønvik 2018), men Metaordboka har ikke bøyninginformasjon og regler for å generere fulle bøyningsskjema.

Metaordboka samordner normert og unormert tilfang, tekst og transkribert tale. Materialet som er lenket sammen gjennom Metaordboka kan grovt grupperes som primærkilder eller sekundærkilder:

Primærkildene omfatter (1) ordsedler, med metadataene oppslagsord, ordklasse, stedfesting eller litterær kilde og medarbeidersignatur, beleggene kan være tekst, lyd, bilde og video, og (2) korpuskonkordanser.

Sekundærkildene omfatter utvalgte ordbøker, for eksempel grunnmanuskriptet til Norsk Ordbok, A. Torp *Nynorsk etymologisk ordbok*, (Torp 1919) og tidligere utgaver av *Nynorskordboka* og *Bokmålsordboka*. En særskilt gruppe er de 84 ordbøkene, ordlistene og ordsamlingene som ligger i Ordbokshotellet. Det er her vi finner ordbøkene til Aasen og Ross.

Det generelle ordningsprinsippet i Metaordboka er at hver artikkel skal representere en leksikalsk gjenstand (*lexical item*, Atkins & Rundell 2008; Ore & Grønvik 2018). Men det har ikke vært et krav at Metaordboka skal være homografseparert. Så en artikkel kan representere flere leksikalske gjenstander ved homografi. Fra 2015 har Metaordboka blitt utviklet til å bli et verktøy for å beskrive norsk normeringshistorie både for bokmål og nynorsk og begge målformene under ett.

5. Ordbokshotellet

Ordbokshotellet er lagd for å være et felles rammeverk for rasjonell lagring og tilgang til retrodigitaliserte ordbøker, ordsamlinger og ordlister, se også Ore & Grønvik (2018).

En ordbok som skal lastes opp i Ordbokshotellet skal være kodet i henhold til TEI P5 Dictionary-format, se TEI (2020) for en nærmere beskrivelse af dette xml-formatet for å kode ordbøker. I hotellet vil hele ordboka lagres som en enhet. Den vil også bli splittet opp i de enkelte TEI-kodete artiklene indeksert etter originalt oppslag og etter et normalisert oppslag.

Koplingen til Metaordboka er på artikkelnivå. Under innlastningen foretas en maskinell oppkopling der de normaliserte oppslagene brukes. Lenkingen

til Metaordboka kan endres ved senere korrekturarbeid. Det generelle samordningsprinsippet i Metaordboka gjør at artiklene i de ulike ordbøkene og ord-samlingene i Ordbokshotellet vil være koplet på leksikalsk gjenstandsnivå og i utgangspunktet uten hensyn til homografseparering.

6. Ordinventaret og metaordboksmetoden

Ivar Aasen bruker norsk i grunnformer, bøyde former og brukseksempler. Dialektformer er transkribert til det norske alfabetet. Forklaringer og metatekst er på dansk. Ordboka er slik sett tospråklig. Aasen skriver at han vurderte å bruke tysk eller engelsk for å lette tilgangen for et internasjonalt forskerpublikum, men slo det fra seg.

Ross følger i hovedsak Aasens oppsett, på tross av den faglige utviklingen innenfor synkron talemålsgransking. Men han bruker en mer detaljert lydtranskripsjon og har en mer detaljert stedfesting. Ross tar med lemma Aasen kanskje ville ha sløffet, og tar med tilfang med usikker identifikasjon.

	Ordbøker	Antall artikler i ordbøkene	Antall artikler i MO (=leksikalsk gjenstand)
1	A2	38 711	36 439
2	R, R1-2	41 872	35 533
3	R, R1-6	49 220	39 542
4	A2 og/eller R, R1,2	79 870	60 850
5	A2 og/eller R, R1-6	87 931	64 536
6	A2 og R, R1-2	-	10 616
7	A2 og R, R1-6	-	11 303

Figur 1: Oversikt over artikler og leksikalske enheter. A2 er Aasen (1873) mens R, R1,2 er Ross (1895). R, R1-6 er ordbok og alle tillegg 1895–1913.

Når en skal sammenligne ordforrådet i to eller flere ordbøker i Ordbokshotellet, kan en enten bruke de originale oppslagene (ordform og ordklasse) eller Metaordbokas gruppering av artiklene på leksikalsk gjenstandsnivå. Siden både stavemåte og ikke minst ordklasseangivelse kan variere sterkt, vil den første metoden ofte være vanskelig å bruke for annet enn for standardiserte ordbøker. Den andre, metaordboksmetoden, fjerner støy fra ulik staving og ikke-standard angivelse av ordklasser, men kan gruppere hardere enn en kanskje ønsker. En må være oppmerksom på slike feilkilder, og analyse av materialet i Ordbokshotellet krever god databasekunnskap og leksikografisk eller leksikalsk innsikt. I

sammenlikningen av Aasens og Ross' ordbøker har jeg brukt metaordboksmetoden. Jeg har også gjort kort rede for hvordan analysen ville bli om en brukt hotell-oppslagene i stedet.

Den høyre kolonnen i figur 1 viser antall metaordboksartikler for ordbøkene. En metaordboksartikkel svarer til en leksikalsk gjenstand og grupperer oppslagsord som bare er ortografiske eller usignifikante uttalevarianter. Den venstre kolonnen er antall artikler. Forholdet mellom de to kan si noe om grad av standardisering. Legg merke til at den relative avstanden mellom disse tallene er mindre hos Aasen enn hos Ross. Det skyldes nok den mer lyd nære skrivemåten til Ross og alle variantene som står på oppslagsplass. Aasens formål var å lage en standardisert rettskrivning og har dermed mindre varians. I tillegg til ord på oppslagsplass fører både Aasen og Ross opp en mengde målføreformere inne i artikkelteksten, henholdsvis 14 000 og 24 000 målformer. Det store flertallet av disse har ikke egen artikkel. Men disse ekstra formene gjør det uklart hva Ross mener når han i forordet til ordboka (Ross 1895) skriver at det er mellom 35 000 og 40 000 nye ord i hans ordbok.

I sammenlikningen over bruker jeg metaordboksmetoden og sammenlikner leksikalske gjenstander. Dersom vi ønsker å sammenlikne oppslag i Ordbokshotellet, er det ikke så lett å få dette meningsfullt siden det er flere måter å si at to oppslag er like. Jeg har satt opp noen av resultatene en kan få i figur 2. Med oppslag menes parene oppslagsform og ordklasse slik de står i ordbøkene. En forstyrrende faktor er at verken Aasen eller Ross har standardisert ordklassebetegnelsene fullt ut, og at de, spesielt i henvisningsartiklene, ikke oppgir ordklasse for oppslaget. Samlet gjelder dette forhenholdsvis 22 % og 10 % av artiklene. Å beregne antall oppslag som forekommer hos begge er derfor vanskelig og krever mye av den samme innsatsen som å kople dem til Metaordboka. Figur 2 viser en tilsvarende varians som figur 1.

	Ordbøker	Antall oppslag	Antall oppslag uten duplikater
1	A2	43 235	42 194
2	R, R1,2	43 825	41 462
3	R, R1-6	51 390	46 681
4	A2 og/eller R, R1,2	87 060	76 756
5	A2 og/eller R, R1-6	94 625	81 863

Figur 2: Oversikt over antall oppslagsord på oppslagsplass. A2 er Aasen (1873) mens R, R1,2 er Ross (1895). R, R1-6 er ordboka og alle tilleggene 1895–1913.

Av Aasens oppslag (leksikalske gjenstander) er det om lag 73 % (25 826) som ikke er beskrevet eller referert til i R1,2. Motsatt er det om lag 70 % (24 920) av oppslagene i R1,2 som ikke finnes i Aasen. Overlappet er 10 616 oppslag/leksikalske gjenstander (henholdsvis 27 % og 30 %). Når Ross skriver at det sentrale ordforrådet finnes i Aasen (1873) og at de to ordbøkene derfor må leses under ett, er det rimelig å ta Ross på ordet og kalle Ross' ordbøker et tillegg til Aasen (1873). Det til tross for at Ross' ordbok med tilleggene er 15 % større enn Aasens ordbok.

7. Utvalget av ord i ordbøkene

Det har i mange beskrivelser av ordutvalget vært hevdet at Ross, i motsetning til Aasen, tok med ord som gir verdiladete karakteristikk om folk og fe og da spesielt om kvinner (se for eksempel artikkelen om Ross i Norsk biografisk leksikon (NBL 2020)). Jeg har foretatt en liten nærlesning for å se om det er noe hold i det.

Aasen (1873) har 152 artikler der 'Kvinde...' forekommer. Litt under 40 (24 %) av disse er negativt ladete og betegner åndsevner, arbeidsevne og kroppsføring. Ross (1895) med alle tilleggene har 443 slike artikler. Drøye 200 (47 %) er negative. De relaterer seg typisk til grad av fedme, orden, hygiene, åndsevner, arbeidsevne og til ubeskjeden adferd. Disse negative artiklene utgjør henholdsvis en og fire promille av artiklene i de to bøkene. Det viser at Aasen og Ross hadde en signifikant forskjellig innstilling til hva slags ord en skal ta med. I en moderne ordbok som Nynorskordboka (NOB 2019) er det 730 artikler der 'kvinne...' forekommer, men bare 36 (5 %) av disse er negative. En kan vel konkludere med at Ross' ordbøker gir et visst sexistisk inntrykk i forhold til Aasens og spesielt i forhold til Nynorskordboka. Men jeg vil peke på at dette er sammenliknbart med hva en kan observere i norske synonymordbøker fra tidlig på 1970-tallet. Så det avspeiler vel tiden og kanskje bruk av mest mannlige informanter. Om det er riktig, så er en mer alvorlig bivirkning at ord relatert til kvinnehverdagen og kvinnes arbeidsområde ikke har kommet med i så stor grad som de burde.

En kan også fundere på hvor mye av hele ordutvalget i ordbøkene som er tatt med videre til moderne ordbøker. Her kan Metaordboka gi et raskt svar. For alfabetstrekket I-Å er det 1 352 metaordboksartikler med Aasen-belegg som ikke er lenket til en artikkel i Norsk Ordbok (2016). Ut fra redaksjonsreglene til Norsk Ordbok (2016) skulle det vært nærmest ingen. Her er det imidlertid en usikkerhetsfaktor. Den manglende lenkingen kan også skyldes litt mangelfull manuell kontroll av lenking. Tallet er uansett lite. For Ross (1895) er det tilsvarende tallet 2 802. For Nynorskordboka er det 51 % av oppslagene i Aasen (1873) som er lenket, mens for Ross (1895) er det 30 %. For Bokmålsordboka er de tilsvarende

tallene 27 % og 15 %. Ser en litt nærmere på det, viser det seg at så godt som alle oppslagene (leksikalske gjenstandene) som er lenket til Bokmålsordboka hører til fellesvokabularet for bokmål og nynorsk.

8. Ordbøkene som en digital enhet

I forordet til ordboka skriver Ross: «Benyttelsen af denne Bog sammen med Ivar Aasens, hvortil den er ”Tillæg”, vil blive noget møisom. Man vil have fire Steder at slaa efter paa for at faa den nu bogførte Rede paa Ordet, idet begge Ordbøger har ”Tillæg og Rettelser»» (Ross 1895:XIX). Med de fem ekstra tilleggene er det faktisk ni steder å slå opp. Mye kan løses gjennom å bruke Norsk Ordbok i stedet, men da får en ikke direkte tilgang til ordbøkene til Aasen og Ross. For å bøte på dette og som en øvelse i å se hvor anvendelig Ordbokshotellet og Metaordboka er, har jeg prøvd å realisere en felles digital Aasen-Ross ordbok. Som beskrevet tidligere, er det relativt lite overlapp mellom de to ordbøkene, om lag 10 000 oppslag (leksikalske enheter). I Ross (1895) og tilleggene, er det rundt 7 500 artikler som er ført som utvidelser av artikler i Aasen (1873). Ross fører disse «tilleggsartiklene» ganske uensartet, og tilleggsopplysningene kan ikke flettes mekanisk inn i den refererte artikkelen i Aasen (1873). Eksempelene nedenfor er tatt fra de digitale transkripsjonene av Aasen (1873) og Ross (1895). Henvisningen til Aasen (1873) er markert i anførselstegn og avsluttes med en 'A'. Eksempelet viser at det krever en faglært leksikograf til å flette de to artiklene sammen til et hele:

tormen (ò), adj. vantreven, seent voxende, saftløs; om Træer. Østerd. (Aamot). – **Tormeskog**, m. vantreven Skov. Tormetre, n. Træ som ikke vil voxer.

”**tormen** (o’) adj. vantreven”. A. NØsterd. (Tynset). Törmen Gul. ”Tormen Skog, Føsk”. Og: **Tormskog** m. **Tormfisk** m. slunken vantreven Fisk (“Føsk”). Se tolmen.

I den digitale Aasen-Ross utgaven¹ er det derfor ikke gjort noe forsøk på å endre originaltekstene. Det er i stedet lagd hyperlenker mellom slike artikkelpar. Igjen viste Metaordboka seg som et nyttig verktøy. Dersom en Ross-artikkel hører til en metaordboksartikkel som har en og bare en Aasen-artikkel med samme oppslag var det enkelt å gjennomføre lenkingen (5 200 tilfeller). Dersom det er flere Aasen-artikler knyttet til en metaordboksartikkel, ble problemet løst med en rask manuell gjennomgang (800 tilfeller). I tillegg finnes det en del tilfeller (850) der Ross’ henvisning ikke svarer til noe oppslagsord i Aasen (1873). Der er det ofte slik at Ross viser til et ord inne i en artikkeltekst i Aasen (1873). Her krevdes det mer detektivarbeid.

¹ Se www.edd.uio.no/aasen_ross/aasen_ross.html.

Dette knippet av digitale ordbøker er lagd over samme mal som beskrevet i Ore (2016). Det er mulig å velge å søke i en eller flere av ordbøkene. Når det er flere, er både tilslagene, de alfabetiske oppslagslistene og artiklene samsorterte. I tillegg er det selvfølgelig lenker til faksimileutgavene.

9. Oppsummering

I denne artikkelen har jeg sett på hvordan en kan bruke Ordbokshotellet og de leksikalske gjenstandene i Metaordboka som et verktøy for å sammenlikne ordinventaret i de beslektete ordbøkene til Ivar Aasen og Hans Ross. Men det er også vist graden av videreføring av ordinventaret til moderne ordbøker, i dette tilfellet *Bokmåls-* og *Nynorskordboka*.

Den digitale knippeutgaven gjør det mulig å sammenstille Aasens ordbok med alle tilleggene til Ross til et hele. I Ore (2016) foreslår jeg at en kan krysslenke dialektinformasjonen i de ulike nordiske ordbøkene til et nettverk. Det er ikke så arbeidskrevende som en skulle tro. Det er mest snakk om organisering og samarbeid om programmeringsgrensesnitt (API).

Når Ross skriver tilslutt i forordet «En Trediemand vil i en Fremtid vide at sammenarbeide de to Værker, rette paa de Mangler som maatte være der, og indarbeide i det Sammenstøbte, de nye Oplysninger der, som før antydet ved Omtalen af dette Materiales Tilveiekomst, visselig imidlertid vil have strømmet ind» (Ross 1895:XIX), så er dette realisert ikke av en person, men gjennom *Norsk Ordbok* (2016).

Litteratur

Ordbøker

Bokmålsordboka (2019) <ordbok.uib.no> (november 2019).

Norsk Ordbok (2016) = *Norsk ordbok : Ordbok over det norske folkemålet og det nynorske skriftmålet*, Oslo, Samlaget 1950–2016.

Nynorskordboka (2019) <ordbok.uib.no> (november 2019).

Roget, Peter Mark (1852): *Thesaurus of English Words and Phrases*.

Ross, Hans (1895): *Norsk ordbog : Tillæg til "Norsk ordbog"* af Ivar Aasen. I:

Ross, Hans: *Norsk ordbog*. Universitetsforlaget. Oslo. 1971.

Torp, Alf (1919): *Nynorsk etymologisk ordbok*. Kristiania: Aschehoug, 1919.

Aasen, Ivar (1850): *Ordbog over det norske Folkesprog*. Christiania: Feilberg & Landmark.

Aasen, Ivar (1873): *Norsk Ordbog med dansk forklaring*. Christiania: Mallings Boghandel, 1873.

Annen litteratur

- Atkins, Sue, & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford. Oxford University Press.
- Grønvik, Oddrun (1989): Meir enn eit tillegg til annan manns verk? Om Hans Ross og ordboka hans. I: *Ord og Mål. Festskrift til Magne Rommetveit*. Oslo, Kringkastingsringen 1989. ss. 91–111.
- Munch, Peter Andreas (1848): *Samlede Afhandlingar I-II*. Christiania. Cammermeyer. 1873–1874, I ss. 360–374.
- Munch, Peter Andreas (1850): *Samlede Afhandlingar I-II*. Christiania. Cammermeyer. 1873–1874, II ss.433–443.
- NBL 2020: *Norsk bibliografisk leksikon* <nbl.snl.no/Hans_Ross> (mai 2020)
- Ore, Christian-Emil Smith (2016): Gamle ordbøker og digitale utgaver. I: *Nordiske Studier i Leksikografi 13 Rapport fra 13. Konferanse om Leksikografi i Norden*. København 19.-22. mai 2015. Nordisk forening for leksikografi 2016 ISBN 978-87-992447-6-8. ss. 203–216.
- Ore, Christian-Emil Smith & Oddrun Grønvik (2018): Comparing Orthographies in Space and Time through Lexicographic Resources. I: *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Znanstvena založba Filozofske fakultete 2018 ISBN 978-961-06-0097-8. ss. 159–172.
- Ross, Hans (1878): *Maalreisningen: her hjemme, belyst fra udlandet*, Stavanger Carl M. Olafsons Bogtrykkeri.
- TEI (2020): *The Text Encoding Initiative* <www.tei-c.org/guidelines/p5/> (mai 2020).
- Venås, Kjell (2009): *Hans Ross; Målføregranskar, Ordboksskrivar Og Grammatikar*. Oslo: Novus forlag.
- Walton, Stephen (1996): *Ivar Aasens kropp*. Oslo: Samlaget, 1996.
- Aasen, Ivar (1836): *Om vort Skriftsprog*, trykt i Syn og Segn 1909, Det norsk samlaget, Kristiania, ss. 1–5.
- Aasen, Ivar (1848): *Det norske Folkesprogs Grammatik*, Christiania.
- Aasen, Ivar (1864): *Norsk Grammatik*, Christiania Mallings Boghandel.
- Aasen, Ivar (1925): *Norsk Maalbunad*, Oslo: Samlaget 1925.

Christian-Emil Smith Ore
 førsteamanuensis, dr.scient.
 Institutt for lingvistiske og nordiske studier
 Universitetet i Oslo
 Boks 1102, Blindern
 NO-0317 Oslo
 c.e.s.ore@iln.uio.no

The roles of Rasmus Rask and Gustaf Renvall in the publishing and editing of Christfrid Ganander's *Nytt Finskt Lexicon*

Elina Palola & Petri Lauerma

After his visit to Finland in 1818, Rasmus Rask organized the publication of Christfrid Ganander's *Nytt Finskt Lexicon* (MS 1786–1787) with the financial help of Russian patron Nikolay Rumyantsev. The dictionary was edited by Rask's Finnish teacher Gustaf Renvall. This article shows that despite Renvall's abridgements, the dictionary retains archaic expressions.

1. Rasmus Rask and Finland

The famous Danish scholar Rasmus Rask (1787–1832) was one of the founders of comparative linguistics. His studies in Germanic and Nordic languages formed the basis for the understanding of sound correspondencies between related languages. Rask also studied several other languages, among them Finnish and Lappish, finding them unrelated to Indo-European languages. Wanting more material for his comparisons he visited Finland as a part of the long trip that eventually led him through Russia to places as distant as Caucasus and Ceylon. (On Rasmus Rask e.g. Hovdhaugen, Karlsson, Henriksen & Sigurd 2000:159–164.)

Rask came to Turku in March 1818. He was in Finland for only a month but managed to meet many eminent Finnish scholars and inspire them to study Finnish language and mythology. At first Rask started to take Finnish lessons from Gustaf Renvall (1781–1841), who at that time was adjunct (associate professor) of history and docent of Finnish language. Rask's notes from these lessons formed the basis of his own unpublished Finnish grammar (Lauerma 2019). When Rask expressed interest in Finnish vocabulary, he was shown the unpublished manuscript of Christfrid Ganander's *Nytt Finskt Lexicon* (1786–1787) as the only available major source. (On Rask's trip to Finland see Ekman 1899:21–23; Korhonen 1986:24–25; Lauerma 2005:125; Häkli 2017:10–12.)

2. Christfrid Ganander and his dictionary

Christfrid Ganander (1741–1790) was, alongside Henrik Gabriel Porthan (1739–1804), one of the most significant Finnish representatives of the humanistic sciences in the end of the 18th century. He was born in Haapajärvi and worked as a chaplain in Rantsila from 1775 to 1790. He published many works, such as scientific articles, a collection of riddles and fairy tales, a book of pharmacy and also one of animal diseases. Ganander is also called the father of comparative religious studies in Finland on the basis of his *Mythologia Fennica* (1789), a reference book of Finnish mythology. Above all, he is remembered as a dictionary writer, an investigator of the Finnish mythology and a collector of traditional poems. (Nuutinen 1997:IX, XII–XIII; Häkkinen 1998:634–635; Hovdhaugen, Karlsson, Henriksen & Sigurd 2000:41.)

The most important of Ganander's works is his comprehensive dictionary *Nytt Finskt Lexicon*, which was the most complete dictionary up to that day. The dictionary is very wide with 35 122 entries and over 180 000 occurrences of Finnish words. In addition to the Finnish entry, it usually includes a note of the part of speech and the Swedish or Latin equivalent or definition. Example sentences are presented also in Swedish, and often in Latin and German, too. Furthermore, Ganander gives the entries clarifying synonyms as well as equivalents in dialects and other languages. (Nuutinen 1997:IX–X, 2000:351.)

The dictionary is not only linguistically significant, but it is irreplaceable as an information source about the literature, tradition and mythology of its time. It includes references to old Finnish literature, of which a great deal was destroyed in the Great Fire of Turku in 1827 or otherwise disappeared. (Häkkinen 1998:635; Hovdhaugen, Karlsson, Henriksen & Sigurd 2000:41.)

Ganander systematically gathered vocabulary from the literature of his time. He also had the material gathered by his grandfather Henrik Hiden with whom Ganander lived after his father's death in 1752. Ganander describes the use of the words with an abundance of examples and gives a great amount of vernacular vocabulary in his dictionary. The dictionary is also said to be the first etymological Finnish dictionary because of its etymological explanations. (Häkkinen 1998:635; Hovdhaugen, Karlsson, Henriksen & Sigurd 2000:40–41; Kilpiö 2011:131.)

The first impetus for Ganander's dictionary was the need to complement the dictionary of Daniel Juslenius (1676–1752), printed in 1745. Porthan and Ganander started the work together, but Porthan soon handed over his materials for Ganander to use and Ganander continued the work on his own. After finishing his work, Ganander sent his thick pile of manuscripts to Porthan to check in 1789, a year before his death. Porthan was completely surprised by the magnitude of the work. He thought that Ganander's work was not yet ready for publication

because of the old etymological views and inadequate dialect materials, among other reasons. Porthan was planning to complement and correct the dictionary, but the work was never finished. Thus, the manuscript was not printed, but it ended up in the university library, where Rask came across this work. (Ekman 1899:19–20; Nuutinen 1997:IX–X; Häkkinen 1997:25, 1998:635; Hovdhaugen, Karlsson, Henriksen & Sigurd 2000:40–41; Romppanen 2001:133–137; Lauerma 2005:124.)

3. Financing and editing plan emerges in St. Petersburg

Rask had not forgotten Ganander's dictionary when he continued his journey from Finland to St. Petersburg, where he spent more than a year. The most important person Rask met in St. Petersburg was count Nikolay Petrovitsh Rumyantsev (1754–1826), a great patron of sciences, who had been in eminent positions as diplomat and still had political interests. Rask suggested to Rumyantsev that a Finnish dictionary based on Ganander's manuscript should be compiled and published, and proposed Renvall as the editor for this work. Rumyantsev accepted the plan, because he thought that the publishing of such a work would strengthen the position of Finnish and its affinity to other Finno-Ugrian languages which were all mainly spoken in Russia. At the same time this would lessen the impact of Swedish, which was the language of Finland's previous rulers. (On Renvall's plan and Rumyantsev's role in it see Elmgren 1874:27; Ekman 1899:33–37; Korhonen 1986:25; Lauerma 2005:125; Häkli 2017:12–14.)

From Renvall's letters we can follow the process of Rask persuading him to accept the post of editor, and see some of Rask's ideas concerning the dictionary, like the alphabetical order of words. But it was Renvall who turned down Rask's ideas of writing Finnish front vowels with Danish and German letters and marking vowel quantity with diacritics instead of doubling the letters. (Ekman 1899:37–39; Lauerma 2005:126; Häkli 2017:21–22.) The choice of languages used in the dictionary became a bigger problem. Ganander had used Swedish and Latin when explaining the meanings of Finnish words, but Rask thought that Latin would be adequate for the use of foreign scholars. Renvall emphasized that Swedish would be essential for the use of dictionary in Finland. As a compromise German was accepted as third language in the dictionary. (Ekman 1899:38; Lauerma 2005:126; Häkli 2017:21.)

Renvall started his editorial work already in 1818, practically alone, as he had wished, but getting relatively little new lexical material from other scholars (Elmgren 1874:28; Korhonen 1986:25; Lauerma 2005:126–127). The project took quite a long time, because he had other occupations and there were many omissions and changes to make in the contents.

4. From Ganander's manuscript to Renvall's dictionary

Though Renvall's dictionary became trilingual, it was not possible to translate all the nuances of Ganander's Swedish definitions and expressions into German, especially since Renvall had to shorten the German definitions. Therefore, the main language for the word articles from Ganander became Latin. (Hakulinen 1967:99.)

The new dictionary was meant to be a compact, cleaned-up version of Ganander's manuscript. Renvall omitted a large amount of Ganander's lengthy explanations from the articles as well as examples regarding word use and literature quotes. It is also known that Renvall removed a large number of Ganander's entries and phraseology during his editorial work. Renvall left out especially unusual derived words, expressions he considered dialect and descriptive words and loan words. (Hakulinen 1967:99–100.) The following observations are based on the material which Palola – one of the authors of this article – collected for her dissertation (2009).

4.1. The omission of certain poetical expressions

Ganander has often included in his dictionary so called coinages, poetisms, temporary constructs, only because of a single line of a folklore poem. One example is *pilkooma*, which Ganander (1997 s.v. *pilkooma*) has accepted as an entry into his dictionary, and for which he has given the meaning 'gash' without any mark of poetism. He provides with the entry *pilkooma* examples *pikomata*, *pilkkomata* 'without doing any harm, any damage' and explains that these words are used in poems. I have not found any examples of the word *pilkooma* 'gash' in the folklore poems or in other old texts including the old Finnish dictionaries. It would seem that Ganander has made the entry *pilkooma* solely because of these very rare examples *pikomata*, *pilkkomata* 'without damage' in folklore poems, without any evidence of the noun *pilkooma* 'gash'. (See Palola 2009:185.) It seems that Renvall has recognized such items and omitted them.

Other coinages from Ganander that Renvall has not continued to use include *herankarva* 'otter' and *sirakainen* 'bear', for example. Both of them are very rare, even in folklore poems. Ganander has noted that these coinages are used in poems, yet still he has chosen to accept such kinds of words as entries into his dictionary. There are also some coinages that both Renvall and Lönnrot (in his dictionary compiled 1866–1880) have accepted from Ganander's material, even when such coinages have been marked as being poetisms.

Ganander provides the poetic noun *torikainen* 'bear' together with *sirakainen* in his dictionary. *Sirakainen* is not included as an entry in any of the dictionaries apart from the one by Ganander, and neither Renvall (1826) nor Lönnrot (1958)

mentions it in connection with *torikainen*. Their opinions were probably based on Ganander's dictionary. No other traditional poems use *sirakainen* or *torikainen* meaning 'bear'; only a poem in the Arwidsson-Bergh collection (SKVR), which is a parallel text to Ganander's poem, includes the verses *siwun käyfwyk sirakainen, kaufwas karfwas tårikainen* 'go away you Sirakainen, go far away you Torikainen', which Ganander (1997 s.v. *sirakainen*) gives in his dictionary.

The nouns written down by Ganander may be based on genuine traditional poems even though other examples are difficult to find. It is possible that Ganander had some information that other people did not write down and did not possess after him. In any case, how Ganander, Renvall and Lönnrot mention, *sirakainen* and *torikainen* are both poetisms, temporary constructs, and not actually names for a bear, at least not in language other than poems. They are nevertheless semantically justified and genuine; they have simply not been retained in any other literary monuments of the Finnish language. It is interesting that Renvall, and subsequently also Lönnrot, have accepted *torikainen* (also *tarikainen*) but not *sirakainen*.

Renvall was critical towards the eastern dialects, although later he did accept some features of these dialects (Renvall 1837:24–26; Hovdhaugen, Karlsson, Henriksen & Sigurd 2000:124, 144; Lauerma 2005:137–141). This may have affected especially the omission of expressions from the folklore poems, as most of these were from the areas of the eastern dialects.

Ganander's dictionary has examples of very rare and complicated reflexive forms, which Renvall seems to have omitted completely from his own version (see Palola 2009:139–141). In the reflexive forms there is a specific suffix that shows that the action is directed to the subject (Penttilä 1963:214). Renvall shunned the reflexive forms in other than poetic language use, too, and even publically objected to their use (Renvall 1837:25–26; Lauerma 2005:137). It is particularly interesting that his own attitude towards language and language use is also strongly reflected in the editing work of Ganander's manuscript.

4.2. The preservation of rare expressions and false interpretations

Renvall has been praised for how well and critically he has chosen and organized the most essential word articles for his collection from Ganander's work, which has been described as illogical (see Rapola 1962:69; Hakulinen 1967:99–101; Häkkinen 1998:635–637). However, it is possible to find such rare words and expressions in Renvall's dictionary from Ganander's materials that have not been found in other sources, as well as some of Ganander's questionable interpretations.

One of the most valuable legacies of Ganander's dictionary for the posterity is the fact that he happened to record such rare language use that has not survived

anywhere else (see Palola 2009: 84, 140, 191, 236). Renvall has copied this kind of materials into his dictionary. For instance, he has taken the very rare form of the Finnish negation clitic *kAAAn* ‘neither’, namely *nak*, and included it in his dictionary and called it a Savonian equivalent of clitic *kAAAn* (see Renvall 1826 s.v. *Nak, Näk*; Palola 2009:80–84). Renvall’s example *ei minullanak* ‘neither have I’ cannot be found in any other sources, as it is probably based on a quote from Ganander’s dictionary (Palola 2009:80–84).

Regarding vocabulary, Renvall has kept some rare expressions and their special meanings that Ganander has used. These include for example *kaiha, kaihi* as in ‘evil’ and *kävyn kanta* as in ‘a poor person, beggar’. (See Palola 2009:141–143.)

Renvall has in some cases continued to use Ganander’s interpretation in an edited form, though it seems to be uncertain. In his entries Ganander often cites old folklore examples, which he explains. In particular, he interprets the world of folklore: what the line meant in its place (or context) in the poem and the world older than his own. On the other hand, he sometimes explains a line in a very concrete manner even if it has a metaphorical meaning regarding spells, for instance. (Palola 2009:240–242.)

It would seem that Ganander has accepted some entries into his dictionary, even if the only evidence we have found of this kind of words is from the language of the old poems. For instance, Ganander (1997) has an entry *käyteliä* which has the forms *käyteliäk, käyteliäs*, for which he gives the meaning ‘someone which is creeping’. He provides an example of using this noun in a poem and notes that within its context it has a meaning ‘lizard’. It is remarkable that Ganander has not marked the noun *käyteliäk* as a poetism, even if he gives an example where this word is used in a poem. There might be a difference. Renvall (1826 s.v. *käyteliäs*) explains that *käyteliäs* in poems means ‘something which is creeping, for example a lizard or a snake’. It is significant also, that Renvall has not accepted the form with a final *k*, which seems to be a mistake by Ganander. (See Palola 2009:201–203.)

There are also incorrect materials that both Ganander and Renvall seem to have accepted. For example, Ganander (1995:79) notes in his *Mythologia Fennica* that in old poems a mythic oculist called Munnu occurs. Renvall (1826) has an entry *munnu*, with a meaning ‘healer’. However, *munnu* is the incorrect interpretation of incorrectly copied lines of a poem, when *Nunnus Ilman tyttäriä* ‘Nunnus the daughter of the Air’ has been copied as follows: *Munnu silmän tyttäriä* ‘Munnu the daughter of the Eye’, which Ganander gives in his dictionary as an example of the entry *munnu* ‘deastri’ (see SKVR XV:120–121; Palola 2009:110–111). It seems that Renvall’s critical approach here, regarding Ganander’s materials, has failed, and future sources have an incorrect citation when describing the Finnish language. That is, even if Renvall has edited Ganander’s interpretation somewhat.

By continuing the use of that Ganander's very rare material, Renvall has made it possible that these rare words and expressions and their meanings have appeared in future works and thus influenced the description of the Finnish language in dictionaries and grammatical works.

5. Publication and later history

Renvall published his edition – which was named *Lexicon Lingua Fennicae, Suomalainen Sanakirja* – in two parts. Some copies of the first part were printed in 1823, but its preface and the second part of the dictionary were not completed until 1826 (Toivonen 1938). The book got a fairly good reception, only the German descriptions of meaning were criticized. Unfortunately, in the following year 1827 the main part of the edition was destroyed in the fire of Turku. (Elmgren 1874:28–29.) By a lucky coincidence the manuscript of Ganander's dictionary was saved, because Renvall had forgotten to return it to the library (Hormia 1961:112). Though Ganander's original dictionary was not published until the 20th century (as a facsimile edition in 1937–1940, and as a thoroughly edited printed edition 1997), Renvall's publication nevertheless guaranteed that a significant part of Ganander's work continued its life as material for later dictionaries.

Literature

Dictionaries

- Ganander, Christfrid (1786–1787): *Nytt Finskt Lexicon*. Manuscript. Helsinki University Library.
- Ganander, Christfrid (1937–1940): *Uusi suomen sanakirja I–III*. [Facsimile edition of *Nytt Finskt Lexicon*]. Helsinki: WSOY.
- Ganander, Christfrid (1997): *Sanakirja Nytt Finskt Lexicon hakemistoineen*. Ed. by Liisa Nuutinen. Helsinki: SKS.
- Lönnrot, Elias (1958 [1866–1880]): *Suomalais-Ruotsalainen Sanakirja I–II* [Finnish-Swedish Dictionary I–II]. Helsinki: WSOY.
- Renvall, Gustavus ([1823–]1826): *Suomalainen Sana-Kirja – Lexicon Linguae Fennicae, cum interpretatione duplici, copiosiore latina, brevior germanica, I–II*. Aboae: Frenckell.

Other Literature

- Ekman, E. A. (1899): Rasmus Raskin osallisuudesta G. Renvallin sanakirjatyön alkuunpanoon. In: *Virittäjä* 3, 17–21, 33–43.
- Elmgren, S. G. (1874): Lefnadsteckning af Gustav Renvall. In: *Lännetär* V, 21–29.
- Ganander, Kristfrid (1995 [1789]): *Mythologia Fennica*. Ed. by Juha Pentikäinen. Klaukkala: Recallmed.
- Hakulinen, Lauri (1967): Suomen sanakirjoista. In: *Kirjoituksia suomen kielestä*. Tietolipas 51. Helsinki: SKS, 83–107.
- Hormia, Osmo (1961): *Gananderin sanakirjan lähteet*. Helsinki: SKS.
- Hovdhaugen, Even, Fred Karlsson, Carol Henriksen & Bengt Sigurd (2000): *The History of Linguistics in Nordic Countries*. Jyväskylä: Societas Scientiarum Fennica.
- Häkkinen, Kaisa (1997): Ordböcker med finska och svenska från Schroderus till Karlsson. In: Mauno Koski, Eeva Lähdemäki and Kaisa Häkkinen (eds.): *Fennistica festiva in honorem Göran Karlsson septuagenarii*. Fennistica 9. Turku: Finska institutionen vid Åbo Akademi, 21–36.
- Häkkinen, Kaisa (1998): Retki menneeseen maailmaan. In: *Virittäjä* 102, 634–640.
- Häkli, Esko (2017): *Rasmus Rasks korrespondes med språkforskare i Finland*. Helsingfors. <www.urn.fi/URN:NBN:fi-fe201706287562> (September 2019).
- Korhonen, Mikko (1986): *Finno-Ugrian language studies in Finland 1828–1918*. The History of Learning and Science in Finland 11. Helsinki: Societas Scientiarum Fennica.
- Kilpiö, Matti (2011): Old English etymologies in Christfrid Ganander's *Nytt Finskt Lexicon* (1787). In: Olga Timofeeva and Tanja Säily (eds.): *Words in Dictionaries and History*. Terminology and Lexicography Research and Practice Vol. 14. Amsterdam: John Benjamins, 131–152.
- Lauerma, Petri (2005): Gustaf Renvall suomen kirjakielen standardoijana [Abstract: Gustaf Renvall and the standardisation of the Finnish Literary Language in the 19th century]. In: *Sananjalka* 47, 119–157.
- Lauerma, Petri (2019): Observations on the Finnish grammar of Rasmus Rask. In: *Folia Uralica Debreceniensia* 26, 193–208. <finnugor.arts.unideb.hu/fud/fud26/11a_lauermapetri.pdf> (March 2020).
- Nuutinen, Liisa (1997): Christfrid Ganander ja hänen elämäntyönsä. In: Christfrid Ganander 1997: *Sanakirja Nytt Finskt Lexicon hakemistoineen*. Helsinki: SKS, IX–XVII.
- Nuutinen, Liisa (2000): Gananderin sanakirjan uusi tuleminen. – *Piipiakielestä kirjakieleksi*. Helsinki: Kotimaisten kielten tutkimuskeskus, 346–351.

- Palola, Elina (2009): *Arkaismi, konteksti ja kirjuri. Sananloppuisen k:n merkintä Christfrid Gananderin kansanrunokokoelmassa*. [Abstract: Archaism, context, and the scrivener. The marking of the word-final k in Christfrid Ganander's collection of old traditional poems.] Acta Universitatis Ouluensis, Series B, Humaniora, n:o 89. Oulu: Oulun yliopisto.
- Penttilä, Aarni (1963): *Suomen kielioppi*. 2., tarkistettu painos. Helsinki: WSOY.
- Rapola, Martti (1962): Suomalaisen sanakirjan historiaa. In: *Virittäjä* 66, 69–73.
- Renvall, Gustaf (1837): *Grammatiska uppsatser*. Åbo: Christ. Ludv. Hjelt.
- Romppanen, Birgitta (2001): *Från målspråk till källspråk. Utvecklingen av den finsk-svenska och den enspråkiga finska ordboken*. Acta Wasaensia n:o 91. Wasa: Wasa universitet.
- SKVR [online publication]. = *Suomen Kansan Vanhat Runot* [The Ancient Poems of the Finnish People]. <www.skvr.fi>
- SKVR XV = *Suomen Kansan Vanhat Runot XV 1997: Runoja Henrik Florinuksen, Kristfrid Gananderin, Elias Lönnrotin ja Volmari Porkan kokoelmista*. – Matti Kuusi ja Senni Timonen (toim.). Helsinki: SKS.
- Toivonen, Y. H. (1938): Renvallin sanakirjan ilmestymisvuodet [Referat: Die Druckjahre von Renvalls Lexicon]. In: *Virittäjä* 42, 100–105.

Elina Palola (chapters 2 and 4)
Ph. D., researcher
Finnish language, University of Oulu
elina.palola@oulu.fi

Petri Lauerma (chapters 1, 3 and 5)
Ph. D., senior specialist,
docent in Finnish language
Institute for the Languages of Finland
petri.lauerma@kotus.fi

Suffixavledda verb av finskt ursprung i de svenska dialekterna i Finland

Caroline Sandström & Harriet Eriksson

In this article we give a systematic overview of verbs ending with *-isa*, *-ista*, *-itta*, *-otta* and *-utta*, based on searches in the A–P section of the online edition of *Ordbok över Finlands svenska folkmål* ‘the dictionary of Swedish dialects in Finland’ (kaino.kotus.fi/fo). Etymologically, most of these verbs appear to have an equivalent in Finnish dialects. The aim of our study is to examine if there are regional differences in the distribution of verbs with different derivative suffixes. From a contrastive language contact perspective we also examine if the suffixes are productive as word formation elements in the Swedish dialects.

1. Inledning

Finland är ett språkkontaktområde där lån mellan svenska och finska, inte minst mellan dessa språks olika dialekter, varit utbrett. I denna artikel undersöker vi ett tämligen outforskat språkligt fenomen, nämligen suffixavledda verb av finskt ursprung i Finlands svenska dialekter. Materialet för undersökningen härstammar från *Ordbok över Finlands svenska folkmål*, som beskriver de svenska dialekter som talas i Österbotten, på Åland, i Åboland och i Nyland. Ordboken har sedan 2013 getts ut som nätordbok¹ och hittills har ca 74 500 artiklar inom avsnittet A–P publicerats.

Under arbetet med redigering och utgivning av artiklar på p- framträdde särskilt en grupp verb med finalt *-isa* (ex. *popisa* ’mumla; pladdra’, *prättisa* ’gräla, käbbla emot’) och *-ista* (ex. *prottista* ’gräla, käbbla emot; mumla’, *pullista* ’svälla’) i ordboksmaterialet. Genom mer utvidgade sökningar kunde vi snart konstatera att verb av denna typ även finns inom tidigare avsnitt av den publicerade ordboken. Vi lade också märke till tre andra grupper av flerstaviga verb, nämligen verb på *-itta*, *-otta* och *-utta*. Gemensamt för dessa fem grupper är deras finska ursprung. Majoriteten är direkta lån från finska, men i samtliga grupper

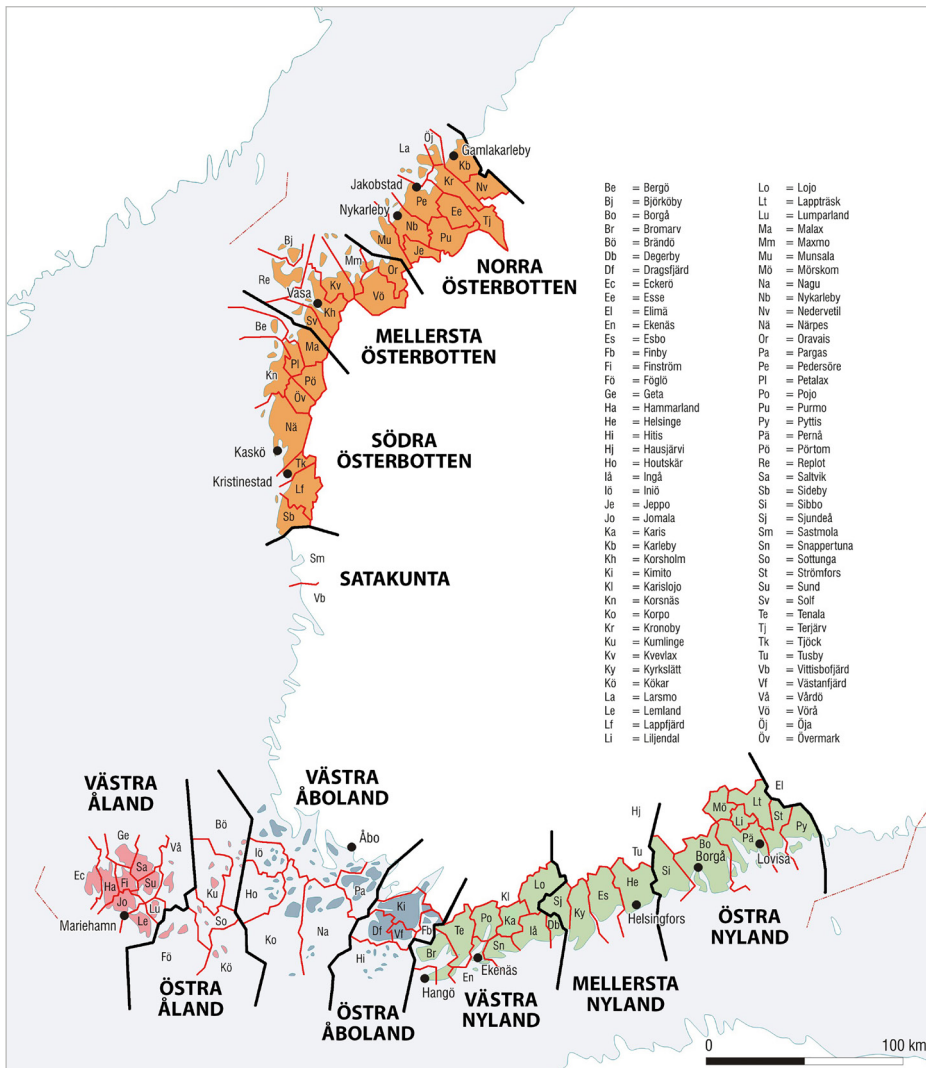
¹ Nätordbokens adress är kaino.kotus.fi/fo. På webbplatsen finns ytterligare bakgrunds-uppgifter om ordboken.

förekommer också verb som inte har finska motsvarigheter². Detta tyder på att suffixen kan ha blivit produktiva ordbildningselement som gett upphov till nybildningar i de svenska dialekterna i Finland.

Svenska dialekter talas i områdena längs den finländska väst- och sydkusten; se karta 1 som illustrerar de svenska dialektområdena i Finland. Tvåspråkighet och direkt kontakt mellan människor med olika modersmål har förekommit, vilket är en viktig förutsättning för språkkontakt och lån; se t.ex. (Thors 1981; Martola, Mattfolk & Sandström 2014a). I de områden där kontakten mellan svenska och finska varit speciellt tät syns detta i såväl de finska som i de svenska dialekterna. Svenska och finska är dock inte besläktade, utan typologiskt olika språk. Bl.a. Weinreich (1953) har konstaterat att språk som är typologiskt lika lättare lånar språkligt material av varandra, medan lånen är färre mellan typologiskt olika språk. Eftersom finska och svenska under lång tid levt tätt inpå varandra under samma sociala och kulturella omständigheter har det emellertid skett en hel del lån mellan språken. (Se diskussion i Thomason & Kaufman 1988:14–19 och Gardani m.fl. 2014.)

Ivars (2004:40–42) visar att tidskriften *Folkmålsstudier* innehållit bidrag om lånord och språkkontakt mellan finska och svenska i dialektal miljö. I synnerhet under 1930-talet, då det ingick ett stort antal bidrag av T.E. Karsten, och under 1950-talet, då V.E.V. Wessman flitigt dokumenterade finska lånord i de svenska dialekterna. Sandström (2013:94ff) konstaterar ändå att det med tanke på den språkliga situation som rått i Finland är förvånansvärt få finlandssvenska dialektforskare som mer ingående fokuserat på språkkontakt. Det banbrytande verket var Saxéns (1895–1898) studier av finska lånord i östsvenska. I Sandström (2011) diskuteras en mindre grupp etablerade och tillfälliga lånord i de svenska dialekterna utifrån *Ordbok över Finlands svenska folkmål*, se även Martola, Mattfolk & Sandström (2014b). I samband med denna studie vill vi särskilt nämna Thors (1981) iakttagelse att de finska lånordens antal i de svenska dialekterna varierar, men är störst i Nyland, i synnerhet i östra Nyland, samt i mellersta och norra Österbotten. Hans studie bygger huvudsakligen på frekvensuppgifter som ges i Wessman (1957:275–359).

² I ordboken ges explicit etymologiska uppgifter för finska lån i de fall ordens ursprung kunnat fastslås.



Karta 1: De svenska dialektområdena i Finland.

2. Teoretiska utgångspunkter

Denna studie utgår från Thomason & Kaufmans (1988:21, 37) definition av lån, dvs. att lån är inkorporering av främmande element i en talares modersmål. Vid sidan av inomspråkliga ljudförändringar och analogi har lån traditionellt ansetts vara en av de viktigaste drivkrafterna för språkförändring. Inom forskningen har lexikon i allmänhet stått i fokus, men i viss mån har även grammatiska och strukturella lån diskuterats. Morfologiska lån av såväl böjningssuffix som avledningssuffix har uppmärksamats av bl.a. Weinreich (1953). Han hävdar att avledningssuffix i större utsträckning än böjningssuffix tenderar att lånas mellan språk. Inom språkkontaktforskning har man ofta ställt upp lånehierarkier som tar fasta på vilka språkliga områden som påverkas och i vilken ordning påverkan sker. Även här har vi utgått från den modell som presenteras av Thomason & Kaufman (1988), nedan återgiven i vår översättning.

Enligt den lånehierarki (eng. *borrowing scale*) som Thomason & Kaufman (1988:74ff) presenterar kan man på en första nivå (i) med tillfällig kontakt bara iakta förändringar i lexikon med lån av innehållsord, ofta lån som sker av kulturella orsaker. På en andra nivå (ii) med något tätare kontakt påverkas lexikon med både innehålls- och funktionsord, men i viss mån även smärre fonologiska, syntaktiska och lexikaliskt-semantiska drag i strukturen. Fonologiska lån förekommer på denna nivå enbart i lånord. Den tredje nivån (iii) kännetecknas av tät kontakt som inverkar på både struktur och på lexikon med lån av såväl innehållsord som funktionsord (t.ex. prepositioner, pronomen, låga räkneord). Påverkan på lexikon och struktur är betydande och det är inte ovanligt att avledningssuffix abstraheras ur lånord och sedan kommer in som ordbildningselement i modersmålet. I fråga om strukturen kan fonologiska lån ge upphov till helt nya fonem, och det kan ske förändringar i prosodi och stavelsestruktur. På de följande nivåerna (iv) och (v) handlar det om starkt kulturellt tryck och kraftig påverkan som leder till ökat antal strukturella lån av t.ex. böjningssuffix och omfattande syntaktiska förändringar som kan medföra genomgripande typologiska förändringar.

I den aktuella finländska situationen med kontakt mellan svenska och finska dialekter rör vi oss på den andra eller tredje nivån. Vi kan på dessa nivåer tala om en tät kontakt där innehållsord samt vissa funktionsord och tydligen även avledningssuffix kan lånas, vilket vi närmare studerar i denna artikel. När det gäller delar av det svenska språkområdet – som norra svenska Österbotten vid språkgränsen, Karleby och Nedervetil, men även Pyttis i östra Nyland (se karta 1 ovan) – finns det evidens för påverkan från finska inom prosodi och fonologi, men också förenklingar av genussystemet som sannolikt beror på språkkontakt. Generellt verkar kontakterna med finska dock inte ha gett upphov till omfattande syntaktiska eller typologiska förändringar i de svenska dialekterna.

3. Undersökningen: regional fördelning och fördelningen mellan lån och nybildningar

3.1. Syfte, metod och material

Syftet med denna undersökning är tredelat. För det första vill vi kartlägga verben på *-isa*, *-ista*, *-itta*, *-otta* och *-utta* i det hittills publicerade avsnittet A–P i *Ordbok över Finland svenska folkmål*. För det andra vill vi undersöka om det finns regionala skillnader i fördelningen mellan olika suffix. Är t.ex. verb på *-isa* särskilt vanliga i någon viss region och verb på *-utta* i någon annan? För det tredje vill vi ur ett språkkontaktperspektiv undersöka suffixens produktivitet i de svenska dialekterna i Finland.

För att få fram vårt material använde vi oss av nätordbokens avancerade sökfunktion som tillåter avgränsningar av olika slag. Vi begränsade sökningen till huvudord, verb och **isa*, **ista*, **itta*, **otta* respektive **utta*. Härfter gjorde vi manuella kontroller för att utesluta icke-relevanta träffar, dvs. verb som inte har något samband med finska som t.ex. *bevista* och *förlisa*.

Våra sökningar resulterade i totalt 104 trestaviga, suffixavledda verb fördelade på fem olika grupper (se tabell 1), där verben på *-isa* och *-ista* dominerar. Vi kunde snabbt konstatera att betydelsen hos merparten av dem hänger samman med någon form av ljud eller ljudåtergivning (ex. *kolisa* 'bullra, slamra, dåna', *firrista* 'brinna med sprakande ljud', *kotkotta* 'skrocka, kackla (om tjäderhöna)' samt *marista* respektive *jamsutta* 'gnälla; tjata').

Under sammanställningen av materialet kunde vi därtill kategorisera verben etymologiskt i tre grupper, vilket här exemplifieras med *kolisa*, *krakista* och *jamsutta*. *Kolisa*, 'bullra, slamra, dåna' är välbelagt i hela Nyland. Det är ett direkt lån från presensformen *kolisee* av finskans *kolista* med samma betydelse. Även parallellformen *kolista*, lånad från den finska infinitivformen och belagd i norra Österbotten, finns med i ordboken. *Krakista* 'hosta, harkla sig, klara strupen', likaså belagt i norra Österbotten, är i sin tur ett lån från dialektal finska, nämligen *krakistaa*, där förleden med stor sannolikhet har en svensk förlaga i det dialektala svenska *kraka/krakas*, se **kraka** v.1 i ordboken. Här rör det sig alltså om ett återlån, ett s.k. pingpong-lån. *Jamsutta* är slutligen en nybildning, bildat av svenskans *jamsa* 'kälta, tjata' och suffixet *-utta*.

3.2. Resultat

I det följande presenterar vi våra resultat i två tabeller, där den första visar den regionala fördelningen av suffixavledda verb av finskt ursprung i vårt material:

Suffix Σ 104 Region	<i>-isa</i> , n 26	<i>-ista</i> , n 49	<i>-itta</i> , n 6	<i>-otta</i> , n 6	<i>-utta</i> , n 17
nÖB	2	27	-	3	8
mÖB	4	1	-	-	-
sÖB	-	1	-	-	-
SK	-	1	-	-	-
ÅL	-	-	-	-	-
ÅB	-	9	-	1	-
vNL	12	5	-	2	2
mNL	22	10	1	3	6
öNL	17	16	6	3	8

Tabell 1: Fördelning av suffixavledda verb av finskt ursprung enligt region.

Som tabellen visar fördelar sig de totalt 104 verben ganska ojämnt inom de fem olika grupperna. De klart största grupperna är verb på *-isa* (26) och *-ista* (49), där de förra dominerar i hela Nyland (NL) och de senare i norra Österbotten (nÖB). Av de 26 olika verben på *-isa* förekommer belägg på 12 i västra, 22 i mellersta och 17 i östra Nyland, medan belägg på 27 av de 49 olika verben på *-ista* finns i norra Österbotten, 10 i mellersta och 16 i östra Nyland. Verben på *-isa* respektive *-ista* uppvisar alltså tydliga regionala skillnader i fördelningen. Verben på *-ista* framträder också i Åboland (ÅB), enligt materialet mer specifikt i de östra delarna närmare språkgränsen (Pargas och Kimito). Verben på *-itta*, *-otta* och *-utta* är vidare de tre minsta grupperna i vårt material, men också de dominerar i Nyland och i norra Österbotten. På Åland (ÅL) finns inga träffar och i södra Österbotten (sÖB) endast en, vilket är föga förvånande då dessa områden inte haft så mycket språkkontakt (jfr Thors 1981:96). I Satakunta (SK), ett finskdominerat område i närheten av Sydösterbotten, kunde man ha väntat sig något fler träffar, men så var inte fallet i vårt material.

Som sammanfattning kan sägas att tabellen ger en helhetsbild. Verben på *-ista* tycks vara vanligast i såväl norra Österbotten som i Nyland, medan verben på *-isa* dominerar i Nyland, med en jämn fördelning mellan västra, östra och mellersta Nyland. Att vissa regioner och i synnerhet vissa orter uppvisar flera fall

av finska lån kan hänga samman med att insamlarna här varit synnerligen aktiva och fokuserade på just finska lån i de svenska dialekterna. Man ska alltså inte dra alltför långtgående slutsatser, men resultaten stämmer till övervägande del väl överens med Thors (1981) iakttagelser då de visar tecken på större språkkontakt i Nyland och i norra Österbotten än i de övriga regionerna.

Vår andra tabell visar härnäst fördelningen mellan finska lån och nybildningar bland de undersökta verben:

Suffix	-isa	-ista	-itta	-otta	-utta	Alla
Lån från fi. el. dial. fi.	23	35	3	5	8	74
Nybildningar	2	12	2	1	7	24
Osäkra fall	1	2	1	-	2	6
Sammanlagt	26	49	6	6	17	104

Tabell 2: Fördelning av finska lån och nybildningar bland de undersökta verben.

Ovanstående tabell baserar sig på de etymologiska uppgifter som ordboken tillhandahåller om varje enskilt verb. Utifrån siffrorna kan vi snabbt konstatera att de flesta av de 104 verben är direkta lån från finskan (74 stycken). I den största gruppen, verb på *-ista* (49 stycken), är knappt en fjärdedel nybildningar (ex. *bjänista* 'brumma, svärja', *präppista* 'kåbbla emot, opponera sig', *pöllista* 'visa sig dum och enfaldig; fåna sig'). Ett par fall är osäkra medan resten är direkta lån. I fråga om verb på *-isa* men också *-itta* och *-otta* är andelen nybildningar liten. Av de undersökta typerna är det bara verben på *-utta* som urskiljer sig genom att bestå av ungefär hälften lån och hälften nybildningar.

Vid en jämförelse mellan verb på *-isa* och verb på *-ista* framträder liksom gällande *kolisa* och *kolista* ett mönster där bägge varianterna förekommer, t.ex. *kilisa/kilista*, *marisa/marista*, *mutisa/mutista* och *potisa/potista*. Det finns en tendens att verben på *-isa* (som lånats från de finska verbens presensform) förekommer i Nyland, och typen *-ista* med ett synonymt verb i norra Österbotten, men också i mellersta och östra Nyland (se även tabell 1 ovan).

Vid lån överförs alltså såväl de finska förlagornas uttal som deras betydelse till de svenska dialekterna. Som ovan nämndes syns det finska böjningsmönstret t.ex. i att finska presensformer står som förlaga till många verb på *-isa*. Men i de svenska dialekterna förekommer också en annan grupp av verb med suffixet *-ista*. Det handlar om verb som *krakista* 'hosta, harkla sig', lånat av dialektala finska *krakistaa*, som diskuterades ovan. Andra exempel är *kopista* 'bulta, knacka; slamra', ett lån av finska *kopistaa* och *mörista* 'råma, böla, morra', som är lånat

av motsvarande dialektala finska *möyristää*. Det är påfallande att det i fråga om dessa verb saknas parallella bildningar på *-isa* i det svenska dialektmaterialet. Dessa mönster för hur verbsuffixen *-ista* respektive *-isa* beroende på sitt finska ursprung används i svenska dialekter tyder på att de svensk- och tvåspråkiga i områden som östra Nyland och norra Österbotten behärskat finska tämligen väl.

Som tabell 2 visar är nybildningarna en mycket mindre grupp än de direkta lånen i vårt material. Enligt andra undersökningar av kontaktsituationer mellan olika språk verkar samma tendens finnas. Lån av morfologiska drag som avledningssuffix kan förekomma också mellan språk som inte är besläktade med varandra, men i de flesta fall dominerar ändå lexikaliska lån (se t.ex. Gardani m.fl. 2014; Thomason 2014).

4. Diskussion

Syftet med vår undersökning, som bör ses som en pilotstudie för vidare forskning, var tredelat. För det första ville vi kartlägga och överskådligt sammanställa de suffixavledda verben med finskt ursprung i den hittills publicerade ordboken. Denna kartläggning kan bl.a. ligga till grund för det fortsatta arbetet att fastslå finska etymologier och tjäna som stöd för lemmatisering och uppdelning i homografer i ordboken.

För det andra ville vi undersöka om regionala skillnader i antalet lån och i fördelningen mellan suffixen *-isa*, *-ista*, *-itta*, *-otta* och *-utta* kunde iakttas. Som ovan konstaterats framträder tydliga skillnader i förekomsten av dessa verb, vilket går hand i hand med den forskning som tidigare gjorts på området (jfr Wessman 1957; Thors 1981).

För det tredje hade vi en tes om att de finska suffixen kan ha blivit produktiva ordbildningselement i våra svenska dialekter. Som vi har sett ger materialet även belägg för detta, men nybildningarna är trots allt färre än de direkta lånen. Sägans bör också att en del av materialet är insamlat för flera decennier sedan, vilket gör det omöjligt för oss att utifrån denna undersökning slå fast hur produktiva suffixen är i de dialekter som i dag talas i Svenskfinland.

Litteratur

Ordböcker

Ordbok över Finlands svenska folkmål. 2013–. Institutet för de inhemska språkens webbpublikationer 33. Helsingfors: Institutet för de inhemska språken. Nätpublikation HTML. URN:NBN:fi:kotus-201330 – ISSN 2323-3370. Nätordboken kaino.kotus.fi/fo uppdateras fortlöpande. Uppdatering 04.06.2019.

Annan litteratur

- Gardani, Francesco m.fl. (2014): Borrowed morphology: an introduction. I: Gardani, Francesco (red.): *Borrowed Morphology*. De Gruyter Mouton.
- Ivars, Ann-Marie. (2004): Folkmålsstudier 1–40: författare, ämnesområden, beskrivningsnivåer. I: *Folkmålsstudier* 43, 33–55.
- Martola, Nina, Leila Mattfolk & Caroline Sandström (2014a): När språk möts. (Språkkontakt). I: *Språkbruk* 1/2014, <www.sprakbruk.fi/-/nar-sprak-mots> (citerad 12.6.2020).
- Martola, Nina, Leila Mattfolk & Caroline Sandström (2014b): Lån i svenska dialekter i Finland. (Språkkontakt). I: *Språkbruk* 4/2014, <www.sprakbruk.fi/-/lan-i-svenska-dialekter-i-finland> (citerad 12.6.2020).
- Sandström, Caroline (2011): Finska lån och hur de anpassas i de svenska dialekterna i Finland. I: *Studier i dialektologi och sociolingvistik. Föredrag vid Nionde nordiska dialektologkonferensen i Uppsala 18–20 augusti 2010*. Uppsala, 249–257.
- Sandström, Caroline (2013): Från folkmålsstudier till interaktionell dialektologi. I: *Folkmålsstudier* 51, 87–113.
- Saxén, Ralf (1895–1898): *Finska lånord i östsvenska dialekter. Språkhistoriska studier I och II*. Svenska landsmål och svenskt folkliv 11 och 11:3. Uppsala.
- Thomason, Sarah G. (2014): When is the diffusion of inflectional morphology not dispreferred? I: Gardani, Francesco (red.): *Borrowed Morphology*. De Gruyter Mouton.
- Thomason, Sarah G. & Terrence Kaufman (1988): *Language Contact, Creolization and Genetic Linguistics*. University of California Press.
- Thors, Carl-Eric (1981): Finska påverkningar på de finlandssvenska folkmålen. I: Loman, Bengt (red.): *De finlandssvenska dialekterna i forskning och funktion*. Åbo: Åbo Akademi, 95–99.
- Weinreich, Uriel (1953): *Languages in Contact*. The Hague: Mouton.
- Wessman, V.E.V. (1957): Främmande inflytelser i de finlandssvenska folkmålen. *Folkmålsstudier* 17–18.

Caroline Sandström
fil.dr, huvudredaktör
Ordbok över Finlands svenska folkmål
Institutet för de inhemska språken
Hagnäskajen 6
FI-00530 Helsingfors
caroline.sandstrom@sprakinstitutet.fi

Harriet Eriksson
fil.mag., redaktör 04/2018–07/2019
tidigare anställd vid Institutet för de
inhemska språken
harriet.c.eriksson@gmail.com

Finnic **litna* ‘town, castle’, a possible alternative origin

Johan Schalin

Etymologin för det urfinska ordet **litna*, som har betydelseorna ’stad’ och/eller ’borg’, är omstridd. Bara en etymologi har föreslagits, nämligen ett lån från urformen för sv. *fläck*, dvs. ← *+flikna-* el. hellre **flekna-*. Här bedöms det förslagets gångbarhet och vägs mot en ny möjlig etymologi, nämligen ett lån från urformen för svenskans *klippa*, dvs. ← **klibn-*.

1. The Finnic data and the Proto-Finnic reconstruction

Words with the meanings ‘town, borough’ or/and ‘castle, fortification’ occur in all branches of Finnic. The ancestor is reconstructed as Late Proto-Finnic (LPF) **litna*. The cognates that have the meaning ‘town’ occur in all daughters except Livonian and modern Finnish, while the cognates with the meaning ‘castle’ occur only in Finnish, Karelian and Livonian (EES, LägLoS, SKES). The aim here is to advance the discussion on the origin of this etymon.

Finnish	linna	‘castle, fortification’
Karelian	linna	‘castle, etc.’; ‘town, etc.’
Ingrian	linna	‘town, borough’
Lydic	l’inn, lidn(ε)	‘town, borough’
Veps	l’idn	‘town, borough’
Votic	lidna (dial.: linna)	‘town, borough’
Estonian	linn	‘town, borough’
Livonian [†]	nīn	‘castle, fortification’
South-Estonian	liin	‘town, borough’

Table 1: Formally equivalent cognate lexemes meaning ‘town’ or ‘castle’.

[†] At least the initial nasal in Livonian is phonologically irregular.

The Livonian form *nīn* is not phonologically expected, but considering that a similar sporadic remote assimilation to another nasal occurs in *nī'em* ‘cow’ (cf. LPF **lehmā*) the word must be considered inherited, and as such admissible

testimony of its semantics. Finnish and Karelian are diachronically close mutual relatives while Livonian is more remote (Kallio 2014), and not one with much direct contact with Finnish. Hence, unless we accept that the same semantic drift has occurred twice, the meanings ‘castle etc.’ and ‘town etc.’ will both have to be reconstructed in early dialects of Late Proto-Finnic.

For the purpose of reconstructing the medial sequence **-tn-* there are not many parallel cases to rely on. Sequences with an obstruent followed by a resonant do not occur at all in Pre-Finnic Uralic etymologies. Accordingly, such words with known etymologies are all loanwords, typically from Proto-Germanic (PGM) or Proto-Baltic (PBlT). A few examples include:

- 1) Fi. *neula* ‘needle’ < **nekla* ← PGM **nēpla-* > Sw. *nål* (LägLoS)
- 2) Fi. *sauna* ‘sauna’ < **sakna* ← PGM **stagna-* > Sw. *stack* (Kallio 2008)
- 3) Fi. *teeri* ‘capercaillie’ < **tetri* ← PBlT **tetr-* > Latvian *teteris* (SSA)
- 4) Fi. *vuona* ‘lamb’ (cf. Ingr. *voonna*) < **vootna* ← PBlT **ōgna-s* (SSA)

In addition to **vootna* above there is yet another example with the sequence **-tn-* (or **-ðn-*) with a wide distribution in Finnic, as illustrated in Table 2.

Proto-Finnic	<i>*litna</i>	<i>*vatnas</i> (<i>*vaðnas</i>)	‘coultter’ (prow)
Finnish	<i>linna</i>	<i>vannas, vaana, vaarnas</i>	‘coultter’; ‘prow’
Karelian	<i>linna</i>	<i>vannas, voarnas</i>	‘coultter’
Ingrian	<i>linna</i>	<i>vaarnas</i>	‘coultter’
Lydic	<i>l’inn, lidn(e)</i>	<i>vannaz (vadnaz)</i>	‘coultter’
Veps	<i>l’idn</i>	<i>vadnaz</i>	‘coultter’
Votic	<i>lidna</i> (dial.: <i>linna</i>)	<i>-vadnaz</i> (dial.: <i>vaarna-</i>)	‘coultter’
Estonian	<i>linn</i>	<i>vannas (vadnas, varnas)</i>	‘prow’; ‘coultter’
Livonian	<i>nīn</i>	<i>va’dmōl</i> (derived form)	‘prow’
South-Estonian	<i>liin</i>	–	

Table 2: Comparison of etymons to support the reconstruction of **-tn-*.

For the purpose of reconstructing vowel length there is a potential issue involved, with secondary pre-cluster lengthening in the word **vatnas/vaðnas* in some northern daughters and in the word **litna* in the southernmost daughters. Neither of these distributions of length matches the descendants of the word

**vootna*, which has a long vowel in a wide-ranging distribution, the only exceptions being secondary pre-cluster shortening in Veps and Votic as well as a data gap in South Estonian, where the word is not attested (EES: s.v. *voon*). The difference is best explained by a substitution of a long vowel in the loan original of **vootna*, namely PBlT **ōgna-s*.¹

So, considering the data, the reconstruction **litna* is solid enough and its two meanings are conceivably both ancient. Not least owing to the phonotactic structure of the medial sequence *-*tn*-, we know that the noun should be a loanword. The same conclusion is supported by the occurrence of -*i*- in the initial syllable when occurring in a back-vocalic stem (Häkkinen 2019:34f). The wide distribution of the etymon indicates a quite ancient LPF borrowing. The same is suggested by the fact that no transparent loan original is available in Iatish Iron Age Proto-Norse or Proto-Slavic. Accordingly, the only loan original proposed so far is of greater age. It has sparked controversy, as discussed in the next section.

2. The merits of assuming a borrowing from PGm +*flikna*-²

The loan original that has been proposed is a supposed PGm +*flikna*- with descendants in Sw. *fläck* ‘spot, fleck; stain’, in German *fleck* ‘spot, fleck; stain; place’ and in the other Scandinavian and continental West Germanic languages. The proposal goes back to T.E. Karsten (1915, 1943–1944) but is quite explicitly rejected in AEW (s.v. *flekkr*) and LägLoS (s.v. *linna*). It is however argued anew by Jorma Koivulehto (1999:121) in his review of LägLoS.

2.1. The semantics of the assumed loan original

As for the semantics Koivulehto (1999:121) points out that the cognates in many Germanic languages have a meaning of ‘place’ whereas *vleck(e)* in Middle High German also meant ‘marketplace’ (cf. derived German *flecken* ‘borough’). A meaning of ‘village’ has developed in Dutch *vlek*. Also, in Old High German

¹ In addition to these etymons there is the verb **vatno-dak* where the reconstruction of the cluster relies on a possible Sámi cognate *vuordnut* (Koivulehto 1999:121). Note that the validity of the Sámi testimony would be no less if the word had been borrowed early enough from Proto-Finnic. Apart from the Sámi word, there are no attestations in Finnic itself to the presence of a cluster in Proto-Finnic. This could be explained by greater age, if the earlier cluster had developed into *-*nn*- before a reintroduction of a LPF cluster *-*tn*- through later loanwords. I owe this observation to Petri Kallio (personal communication), who has also contributed with some other useful comments, for which I express my gratitude.

² Together these are the reasons why in this paper the reconstruction +*flikna*- is prefixed with a ‘+’-sign to indicate that it is here understood to be erroneous.

a meaning of ‘village’ is reported in AEW (s.v. *flekkr*). A chain of development ‘spot’ > ‘place’ > ‘marketplace’ > ‘village’ or ‘town’ must be deemed probable against this background.

Nevertheless, while the conclusion may be drawn from this data that a semantic development from ‘spot’ to ‘town’ in general could take place, and indeed that it has taken place in some continental West Germanic daughters neighbouring each other, it is not verified that it had done so already in early Proto-Germanic. Certainly, the meaning of ‘town’ is absent in the North Germanic cognates, for example Old Swedish (OSw) *flæcker*, Old Norse (ON) *flekkr*, Old Danish *flæk*, Övdalian *flekk* etc. And indeed, based on the medial cluster, which developed into *-kk-* already in Proto-Germanic, it is at the PGm level where the assumed loan original of this etymon must be sought.

In conclusion the chain of semantic changes assumed for the loan original is longish and hypothetical but not impossible. For the etymology to work, however, all these changes should have been concluded already in early Proto-Germanic, for which the argument is based on conjecture.

2.2. The phonology of the assumed loan original

To recall again, the form that must be reconstructed in Late Proto-Finnic is **litna*. Matching this form with a reconstruction of the assumed loan original raises a couple of problems. These are invoked in LägLoS (s.v. *linna*) to underpin a rejection of the etymology.

Based on the examples of Finnic **vootna* ‘lamb’, **vatnas* ‘coultter (prow)’ and **vatno-dak* ‘to swear’ Koivulehto has, contrary to the doubt expressed in LägLoS, argued well that a substitution of Finnic **-tn-* for **-gn-* (or **-g^wn-*, or by implication **-kn-*) could have happened, given that the borrowing occurred early, before Finnic **-kn-* had become phonotactically permitted. His argument is valid, but indeed it requires a quite early date of borrowing: the etymology would have to be older than that of *sauna* < **sakna* in example 2 above, which in itself is a remarkably early Germanic borrowing owing to the substitution of word-initial **s-* (instead of **t-*) for PGm **st-*.

Once this problem has been cleared there is another one, which proves to be more difficult, namely the reconstruction of a high vowel in *+flikna-* (or *+fligna-*). The high vowel relies on one fatal and one dubious assumption.

First and foremost, a fatal problem for the suggested reconstruction follow from the Germanic data, where all descendants display a non-high vowel *-e-* or *-æ-*. The viability of the reconstruction is preconditioned on the hypothesis that a lowering umlaut would have occurred in all branches. Yet this is not supported by a check of sub-minimal pairs in Scandinavian, in which the same high vowel consistently remains non-lowered in phonological environments where equally

active triggers for the lowering umlaut are present, such as in Sw. *sticka* ‘splinter, sliver; knitting needle’ with its Scandinavian cognates, each without any lowering from < *stikkōn-.³

Secondly, the etymology invoked to account for the high vowel is at best uncertain. It would be a zero-grade cognate of ON f. *flik(a)* ‘rag’ (EWDS: s.v. *flekk, flecken*) < PreGm *plei-g- ‘torn off piece’ (SEO: s.v. *flik*), in eastern Scandinavian also meaning ‘patch; flap, tab’. This possibility is explicitly dismissed in SEO (s.v. *fläck*) and no longer even discussed for example in AEW (s.v. *flekkr*) or DEO (s.v. *flæk*). Koivulehto (1999:121) addresses this problem with a reference to EWDS: “Das schon traditionell angesetzte urgerm. -i- (vgl. Streitberg [...] Pokorny [...]) ist auch heute durchaus vertretbar (siehe Kluge – Seebold s.v. *flekk, flecken*)”.⁴ However, Koivulehto’s source itself (EWDS:ibid.), only presents this etymology as its secondary option and by a quite cautious and conditional formulation. A preferred etymology, being semantically superior, is a full-grade (or “e-grade”) cognate of English *flake* < PGm *flaka- (SEO:ibid.). The latter requires a reconstruction with the ablauting mid-vowel, as in *flekna- and is thus best compatible with the etymological discussion in EDPG (s.v. *flaka).⁵

Finally, no one has suggested yet that *-i- in *litna could have been substituted for *-e- in *flekna-. It goes without saying that in this particular case, where also the semantics are difficult, there should be no compelling reasons to resort to this remote and quite unparalleled possibility.

³ There are numerous other examples, but many might be borrowed from Low German like Sw. *vicka* ‘to wiggle, jiggle’ ← .. < *wignan- (or *wikkan-) and other may equally well be intensive or iterative derivatives, the formation of which are difficult to prove earlier than the lowering umlaut, such as Sw. *vricka* ‘to sprain, twist’ < *wrikkan- or *wrignan-. Since no parallels for lowering exist an *ad hoc* auxiliary assumption would be called for, such as the Nordic word being a borrowing from West Germanic.

⁴ There is no coherent treatment of Grimm’s law in either Koivulehto (1999:121) or in LägLoS. Koivulehto denotes Indo-European *pliknó, which renders a PGm *-g- and makes the etymology incompatible with the very explanation in EWDS (s.v. *fleck, flecken*) that he invokes for the vowel. LägLoS (s.v. *litna*) on the other hand denotes “frühurg. *flign- > urgerm. *flikna-” with -g- retained after the operation of Grimm’s law on the initial f-, thus apparently displaying an internal chronology of the different stages of Grimm’s law and representing a different consonant than Koivulehto.

⁵ The appearance of the full-grade in the shape of the PGm vowel */e/ in the PGm ablaut *flaka-/*flekna- excludes all roots with a final laryngeal *h2, as in *pelh2-/pleh2-. This means that none of the Latin and Greek parallels invoked to support the etymology in SEO (s.v. *flak*) can be accepted. For the same reason the etymon does not compare to the theme discussed in EWDS (s.v. *flach, fladen, flur*) and VAEO (s.v. *flade, flor*), nor does it benefit from the Latin and Greek parallels that DEO (s.v. *flagre, flæk*) invokes in support of its preferred etymology for the Danish cognate *flæk*.

3. The merits of assuming a borrowing from PGm **klibna*

Given the severe problems with the one proposal for an etymology of Late Proto-Finnic **litna*, there is scope to seek other possibilities to explain it. A potential original that seems to have escaped attention is the PGm ancestor of Middle Low German *klippe* ‘cliff, rock’ (on land or by the sea). The etymon has correlates in Old Danish (DEO: s.v. *klippe*) and OSw *klippa* ‘rock (off the seashore)’ *biærgh klippa* ‘cliff (on land)’ and *biærgh klippa* ‘ibid.’ (FSLDB: s.v. *biærgh*, s.v. *biærghklippa*, s.v. *klippa*).

Just as German *Klippe* ‘cliff, reef, rock (by the sea)’ (EWDS: s.v. *Klippe*) and Dutch *klip* ‘ibid.’ (EWN: s.v. *klip*), the Scandinavian word is thought to be borrowed from Middle Low German (SEO: s.v. *klippa*). In New Norwegian, the word is considered a loan over Bokmål from Danish (BMO-NNO: s.v. *klippe*). The direction of borrowing conforms to the fact that the Scandinavian weak declension must be secondary, since a reconstruction *+klibnōn-*, based on it, would pose a problem since it contains a double reflex of the nasal stem formant. Despite that the oldest occurrence thus must be considered limited to Old Saxon, the supposed PGm etymology **klibn-* still presupposes high age, i.e. that the word existed already in early PGm when **-bn-* developed into *-pp-*.

3.1. The assumed phonological sound substitution

The etymology discussed here implies a unique sound substitution of **-tn-* for **-bn-*. At first glance a hypothetical but non-existent Proto-Finnic cluster *+pn-* would have been a most faithful candidate to substitute any of the clusters **-bn-*, **-fn-* or **-pn-* in a PGm loan original. Against this background a prerequisite for the viability of this etymology is a further assumption that Proto-Finnic phonotactic constraints would have discarded this candidate in favour of **-tn-*. This assumption is plausible, considering that for example in Finnish no cluster **-pn-* has appeared until modern times, not even on morpheme boundaries. Consider also that precisely **-tn-* was a preferred substitute at an early Proto-Finnic time when **-kn-* was still discarded by constraints. This is shown by equally early loanwords, as presented with reference to Koivulehto (1999:121) in subsection 2.2 above.⁶ The same may thus well have been the case for a disallowed *+pn-*.

⁶ This is not contradicted by the fact that clusters of ascending sonority with the labial followed by a liquid have been introduced through loanwords, exemplified by **-pr-* in *sepra* ‘company’ ← PBlT **sēbra-* (SSA: s.v. *seura*) and (perhaps only later) **-pl-* as in *kepla* ‘bow, prow’ ← PGm **skeula-* ‘shelter’ (LägLoS: s.v. *keula*). A substitution such as in **revna* ‘edge’ ← PBlT **breuna-* (SSA: s.v. *reuna*) is not equivalent either owing to the different sonority contour.

Nonetheless, in no other Finnic word has a sound substitution of **-tn-* for **-bn-* been identified. This may be an accidental data gap, since the set of borrowed Finnic lexical items is limited, and the clusters **-bn-* **-fn-* **-pn-* are not overly common in early Proto-Germanic anyway. In lack of a direct parallel for this cluster substitution, a piece of evidence could be invoked to testify that the very same phonotactic constraint seems to have operated in a somewhat younger Finnic etymology, where *+pn-* was also not used to substitute **-bn-* or **-fn-*. Thus Early Finnish/Carelian **aineh* (perhaps alongside a variant **aun-*), which developed into Finnish *aine* ‘substance, material; cause, reason’, has been presumed to be a borrowing from Proto-Norse **afnia-/abnia-* > ON *efni* ~ OSw *æmne* ‘ibid.’ (LägLoS: s.v. *aine* with references). If this cluster was discarded as a valid sound substitution in Early Finnish/Carelian, this must so much more have been the case in the phonotactically more rigid Proto-Finnic, which was contemporary with the Proto-Germanic original.⁷ Unfortunately, also the sound substitution in this etymology lacks a parallel case to support it, which for the reasons already quoted may again be a gap in the data. In any event, this etymology must be deemed rather uncertain despite the near-perfect semantic match.

In conclusion, the sound substitution may be argued to be plausible, especially considering that better substitutions may not have been available. The lack of better parallels may be accidental gaps in the data but will nonetheless impede the explanation based on **klibn-* from being conclusively ascertained.

3.2. The semantics of the assumed loan original

The descendants of the word **klibn-* have two wide-spread meanings, the better known meaning being ‘cliff or rock in/by a lake or a sea(shore)’. In West Germanic, this meaning is the only one that has survived. Yet, originally a *klip(pe)* may well have been on land as stated in EWN (s.v. *klip*):

Until the 16th century, the word *klip(pe)*, both in Dutch and in Low and High German, could also indicate a rock on land, in particular as a shelter or dwelling place for animals. Nowadays, only rocks in the sea or on the coast are still called *klippen*.⁸

⁷ The substitution would obviously be much younger, because the Proto-Norse cluster must have become composed after the sound law **-bn-/fn-* > **-pp-* had ceased to be productive in Proto-Germanic (if not, the modern Scandinavian forms would contain a geminate, such as *+eppe*).

⁸ Tot in de 16e eeuw kon *klip(pe)*, zowel in het Nederlands als in het Neder- en Hoogduits, ook een rots op het land aanduiden, met name als schuilplaats of woonplaats voor dieren. Tegenwoordig noemt men alleen rotsen in zee of aan de kust nog *klippen*.

In fact, many of the oldest attestations in Swedish has nothing to do with the sea or a lake. In Old Swedish, for example, the compounds *biærghklippa* or *biærghs(-)klippa* occur four times in a medieval version of the book of Judges (FSVLDB: s.v. *biærgh*, s.v. *biærghklippa*) and in all of them the narrative stays firmly on land, for example the mentioning on page 89:

Oc vppa thentidh waro madyans tiel j dalenom widh fwdhra fydhona widh høgha bærgx klippona. – And at that time the Midyanites' tents were in the valley on the south side by the high mountain cliff.

These “terrestrial” meanings are diverse in the richer corpus of literary Swedish of the 16th century. One of them is at times close to ‘fortification’. In SAOB under the entry of *klippa* (sbst.) the following meaning is given and exemplified, among others, by the two subsequent examples:

(more or less) stand-alone, steep, often bare (woodless) cliff in particular when situated in or near water (usually sea or lake); sometimes about such rock under the water surface; in earlier times occasionally approaching the meaning ‘rock fortress’; often in comparisons to denote firmness or immobility.⁹

The säya berghomen och clippomen, Faller vthoffuer oss och skyler oss. Upp. 6:16 (NT 1526). – They tell the mountains and the rocks, Fall on us, and shelter us. (Apocal. 6:16, New Testament from 1526)

Han tractade honom efter lif ..., för thet han icke ville late sättie en hop uthaf hans krigsfolk in upå Bahus klippa. RA I. 2:326 (1569). – He wished him dead..., for the sake that he would not let a horde of his troops be allotted to the Bohus castle rock. (Parliam. acts from 1569)

Further, under the entry *klippa* (sbst.) another meaning is given and exemplified, as follows:

(especially in religious style) to designate someone or something as the solid foundation or the reliable support or as a safe haven and a safe shelter for someone or something; especially about God or Christ.¹⁰

Herren är mijn klippa och mijn borgh. 2 Sam. 22:2 – The Lord is my rock and my fortress. 2 Sam. 22:2 (Bib. 1541)

⁹ 1) (mer eller mindre) fristående, brant, ofta kal (trädlös) bergmassa; i synnerhet om en sådan i eller vid vatten (vanligen hav eller sjö); stundom om dylik under vattenytan; förr någon gång närmande sig betydelsen: klippfästning; ofta i jämförelser för att beteckna fasthet eller orubblighet.

¹⁰ 3) ... a) (i synnerhet i religiös stil) för att beteckna någon eller något såsom den fasta grunden eller det pålitliga stödet eller såsom en trygg tillflykt och ett säkert beskydd för någon eller något; i synnerhet om Gud eller Kristus.

These last meanings in 16th century Swedish are somewhat different from both meanings in the West Germanic daughters, including from the terrestrial meaning of 'rock shelter (for animals)', which has gone extinct. The question is where the meanings of a 'firm and solid foundation', 'safe haven' and 'fortress rock' come from. They appear to defy the general tendency, according to which innovations in 16th century religious Swedish mainly draw on Luther's High German. Could the new meanings really have developed independently of that tendency, spontaneously, in just 150 years after borrowing from 14th century Low German? Or could they reflect an innovative calque of the German word *Fels* 'rock'? Or would the meanings continue a semantic inheritance from Proto-Germanic, more or less invisible in the earlier corpus of Old Saxon? Notwithstanding the uncertainty regarding the more specific meaning of 'fortification', this does not challenge the reconstruction of a slightly more general PGM meaning such as 'rock shelter' or 'safe haven'.

3.3. The assumed semantic development of the loan word

A meaning of 'town' may easily develop from the meaning 'fortification' in a context where villages or towns were protected by fortified safe havens. This is for example the assumption made for the PGM word **burgz* and its descendants in SEO (s.v. *borg*) and VAE0 (s.v. *borg*). EWDS (s.v. *burg*) argues at length in favour of the reverse direction of semantic development, which must be considered equally possible.

The question here is which direction of development is the one that occurred in the Finnic languages. This is also a matter for archaeology. Koivulehto (1999:121) argues in favour of a semantic development from 'town, borough' to 'fortification' and 'castle' by the argument that the earliest Estonian **litna*-communities were no refuge fortifications but protected settlements often situated on promontories or islands. This part of his text comes without references. While there are no immediate reasons to doubt his position, it cannot be evaluated further here.

4. Comparing the merits of the two possible etymologies

For reasons stated in Section 1, the etymology is surely old. In the context of the Early Metal Age a reasonable expectation is that the word **litna*, which reflects a settled and partly urbanised society would originate from a southern rather than a northern zone of language contact. Moreover, many words in the field of construction are Proto-Baltic, such as **tilta* 'bridge', **talo* 'house' and **širte* 'log'. Against this background, it may be asked why the etymology should be

Proto-Germanic in the first place. Yet given the alternatives at hand, we have undertaken to weigh two available PGM proposals for etymologies against each other and leave a future quest for a PBlT etymology to others.

At this point, we may state that neither of the etymologies is impossible from a semantic point of view, but both are problematic. There are gaps to be filled in, which makes the reasoning hypothetical. In both cases, the data raise misgivings about reconstructing the required meaning as far back as in Proto-Germanic. Admittedly these gaps seem no less severe for the new etymology based on **klibn-*, compared to Koivulehto's etymology based on **flign-/flikn-*.

As far as the sound substitution is concerned, both etymologies are relatively vulnerable to criticism. The assumed sound substitution in **klibn-* → **litna* can be argued to be plausible but unlike Koivulehto's etymology it does not benefit from attested parallel cases. In both cases, the reconstruction of the consonant cluster in the loan original is open to controversy, but also in both cases, the required reconstruction is the mainstream majority view. When it comes to the vocalism, the new etymology is far superior, as the reconstruction of a high vowel in *+flign-/flikn-* goes against the best available etymological and phonological evidence.

References

Dictionaries

- AEW (1977) = Jan de Vries: *Altnordisches etymologisches Wörterbuch* (2. verbesserte Aufl [1. ed. from 1957-60]). Leiden: Brill.
- BMO-NNO = *Bokmålsordboka og Nynorskordboka*. Universitetet i Bergen & Språkrådet <ordbok.uib.no> (November 2019).
- DEO (1985) = Niels Åge Nielsen: *Dansk etymologisk ordbog: ordenes historie* (4. opl. [1. ed. from 1966]). København: Gyldendal.
- EDPG (2013) = Guus Kroonen: *Etymological Dictionary of Proto-Germanic*. Leiden: Brill.
- EES (2012) = I. Metsmägi, M. Sedrik, & S.-E. Soosaar (Eds.): *Eesti etimoloogia-sõnaraamat*. <portaal.eki.ee> (November 2019).
- EWDS (2002) = F. Kluge & E. Seebold: *Etymologisches Wörterbuch der deutschen Sprache* (24., durchges. und erw. Aufl). Berlin: de Gruyter.
- EWN (2003–2009) = Philippa, M., F. Debrabandere, A. Quak, T. Schoonheim & N. van der Sijs (Eds.): *Etymologisch woordenboek van het Nederlands*. Amsterdam: Amsterdam University Press.
- FSVLDB = *Fornsvensk lexikalisk databas*. Göteborg: Språkbanken, Göteborgs universitet. <www.spraakbanken.gu.se/fsvldb> (November 2019).

- LägLoS (1991–2012) = A. D. Kylstra, S.-L. Hahmo, T. Hofstra & O. Nikkilä: *Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen* (1991st–2012th ed., Vols 1–3). Amsterdam; Atlanta; New York: Rodopi.
- NAOB (2017) = Tor Guttu et.al. (Eds.): *Det Norske Akademis ordbok*. <www.naob.no>
- SAOB = *Ordbok över svenska språket, utgiven av Svenska Akademien*. <www.saob.se>
- SEO (1980) = Elof Hellquist: *Svensk etymologisk ordbok* (3. korrigerade upplagan [1. ed. from 1922], Vols 1–2). Lund: Liber.
- SKES (1955–1981) = Toivonen, Y. H., E. Itkonen & A. J. Joki (Eds.): *Suomen kielen etymologinen sanakirja I–VII*. Helsinki: Suomalais-ugrilainen seura.
- SSA (1992–2000) = Itkonen, E., & U.-M. Kulonen (Eds.): *Suomen sanojen alkuperä: etymologinen sanakirja* (Vols 1–3). Helsinki: Suomalaisen Kirjallisuuden Seura : Kotimaisten kielten tutkimuskeskus.
- VAOE (2007) = Harald Bjorvand & Fredrik Otto Lindeman: *Våre arveord: etymologisk ordbok* (rev. og utv. utg [1. ed. 2000]). Oslo: Novus forlag.

Other Literature

- Häkkinen, Jaakko (2019): Kantasuomen keskivokaalit: paluu. In: S. Junttila & J. Kuokkala (Eds.): *Petri Kallio rocks : Liber semisaecularis 7.2.2019*. Helsinki: Kallion Etymologiseura, 24–40.
- Kallio, Petri (2008): The Etymology of Finnish sauna ‘Sauna’. In C. Dekker, A. A. MacDonald, & H. Niebaum (Eds.): *Northern Voices: Essays on Old Germanic and Related Topics, Offered to Professor Tette Hofstra. Germania Latina VI*. Leuven, Dudley, Mass: Peeters, 313–320.
- Kallio, Petri (2014): The Diversification of Proto-Finnic. In J. Ahola, Frog, & C. Tolley (Eds.): *Fibula, Fabula, Fact: The Viking Age in Finland*, 155–168. <www.doi.org/10.21435/sfh.18> (November 2019).
- Karsten, Torsten Evert (1915): *Germanisch-Finnische Lehnwortstudien. Ein Beitrag zu der ältesten Sprach- und Kulturgeschichte der Germanen*. <www.archive.org/stream/actasocietatissc45suom#page/n38> (November 2019).
- Karsten, Torsten Evert (1943–1944): Finnar och germaner: Folk- och kulturberöringar från tre årtusenden. *Folkmålsstudier*, IX–X (1943–44).
- Koivulehto, Jorma (1999): Review of LÄGLOS, Bd. II. *Kratylos*, 44, 119–122.

Johan Schalin
 PhD
 University of Helsinki
 FI-06100 Borgå
 johan.schalin@alumni.helsinki.fi

«Å nei, det ordet er ikkje lov på nynorsk!» Eller ...? Om parallell redigering av to norske ordbøker

Bente Selback

The dictionaries for the two written varieties of Norwegian, *Bokmålsordboka* and *Nynorskordboka*, have previously been edited independently by different people. While the dictionaries used to be two separate books, they now come in a digital version at ordbok.uib.no, where they can be viewed side by side on the screen. This way the differences between them are more visible. This article shows that users feel uncertain when a word can only be found in one of the dictionaries, as is the case with the word *animasjon* ('animation'), for instance. Both dictionaries are now being revised simultaneously at The University of Bergen. An important goal is to limit the number of unmotivated differences between them.

1. Revisjon av *Bokmålsordboka* og *Nynorskordboka*

I 1986 kom førsteutgåvene av standardordbøkene *Bokmålsordboka* (BOB) og *Nynorskordboka* (NOB) ut. Dei var meinte å vera to likeverdige eittbands allmennordbøker for dei to norske målformene. Ordbøkene er digitaliserte og samla i eit felles søkjegrensesnitt på ordbok.uib.no.

Begge ordbøkene er bygde opp på same måten. Dei inneheld relativt korte artiklar som gjev opplysningar om alle godkjende skrivemåtar av oppslagsordet, ordklasse og bøyning, og i ein del tilfelle finn ein òg informasjon om uttale og etymologi. Dei fleste artiklane har dessutan ein relativt kortfatta definisjon og gjerne eitt eller fleire bruksdøme.¹ BOB og NOB skal no reviderast, og tre sentrale oppgaver er å oppdatere lemmautvalet, sikra at tydingane er i tråd med språkbruken i dag, og jamna ut umotiverte skilnader mellom dei to ordbøkene.

Denne artikkelen dreier seg om arbeidet med det siste punktet, altså utjamninga mellom dei to ordbøkene. Eg vil kort komma inn på den norske skriftspråksituasjonen og korleis ordbøkene blei utarbeidde i si tid, før eg seier noko om korleis dei blir brukte og oppfatta av brukarane i dag, korleis me arbeider med utjamning i revisjonen, og utfordringane med at skilnadene mellom dei viser seg å vera større enn venta.

¹ Når nokre artiklar manglar definisjon, kjem det av overføringa frå papir til digitalt format der ein del samansetjingar automatisk blei gjorde om til eigne artiklar som viser skrive- og bøyingsmåte, men utan at det blei utarbeidd definisjonar.

2. Er BOB og NOB like som to tvillingar?

BOB og NOB har ofte blitt omtalte som tvillingordbøker, men sjølv om det er mange likskapar, er dei likevel ikkje heilt like, og av og til kan skilnadene mellom dei føra til feilslutningar hjå brukarane.

2.1. Då Harald skulle skriva stil – ei lita brukarhistorie

Ein ettermiddag var eg på besøk hjå familien til Harald. Han går første året på vidaregåande skule og sat og jobba med leksene sine. Det var ei stiloppgåve om film, som skulle skrivast på nynorsk, og han lurte på korleis ein stavar *animasjon*. Eg tenkte at dette var eit godt høve til å reklamera for ordbøkene våre, og viste han korleis han kunne søkja opp ordet. Resultatet blei ikkje heilt som venta (sjå figur 1).

Bokmålsordboka	Nynorskordboka
<p>animasjon substantiv m BØYING</p> <p>REF animasjonsfilm, animert, tegnefilm,</p> <p>OPPHAV fra engelsk</p> <p>BETYDNING filmteknikk der en med en rekke enkeltbilder får f.eks. tegnede figurer eller dokker til å virke levende</p>	<p>Vi har dessverre ingen informasjon om ordet 'animasjon' i nynorskdbasen.</p> <p>Vanlege feilkjelder:</p> <ul style="list-style-type: none"> • Skrivefeil: Sjekk skrivemåten! • Søk i feil ordbok: Bruk Begge-knappen!

Figur 1: Oppslag på *animasjon* i BOB og NOB før revisjonen.

Det viste seg at *animasjon* ikkje var eit oppslagsord i NOB, og det fekk Harald til å konkludera med at ordet ikkje var «lov» å bruka på nynorsk. Eg prøvde å forklara at ordbøker ikkje har med alle moglege ord, noko han kunne vera med på, men han meinte at det nok ikkje var tilfellet her sidan han jo fekk treff i BOB. Det måtte bety at dette var eit typisk bokmålsord som ikkje kunne brukast på nynorsk.

Denne vesle historia illustrerer at brukarane ofte legg meir i skilnadene mellom ordbøkene enn det det faktisk er grunnlag for. Dette må sjåast i lys av den norske språksituasjonen med to offisielle skriftspråk der store delar av ordtilfanget er felles, men der målformene likevel har hatt kvart sitt særpreg. Det kan ha medverka til at ein del språkbrukarar er usikre på kva som blir rekna som «rette» og «gode» ordval. I tilfellet *animasjon* er det snakk om ein tilfeldig skilnad i lemmautvalet i ordbøkene, og han speglar ikkje ein reell skilnad mellom målformene. Likevel blir den manglande oppføringa i den eine ordboka tolka som eit forbod. Harald er slett ikkje åleine om denne tolkinga, noko mange av spørsmåla ordbokbrukarane sender til redaksjonen, viser (Rauset 2019).

2.2. Bakgrunnen for ordbøkene

For å forstå litt meir av kvifor det er ein del ulikskapar mellom dei to ordbøkene, kan det vera nyttig å sjå kort på korleis dei blei til.

Etter ein lang periode med språkstrid blei det i 1964 sett ned ein komité med mandatet å «ta opp til drøfting heile språksituasjonen i landet og gjera framlegg om tiltak som komitéen meiner kan tene til å ta vare på og utvikle vår norske språkarv» (Innstilling 1966:3). Eitt av framlegga deira var å etablere eit leksikografisk institutt, og dei peiker på behovet for gode ordbøker og «nevner som en mulighet at ordbøker [...] kunne lages for begge målformer under ett» (Gundersen 1967:120).

I 1974 var det nyoppretta Norsk leksikografisk institutt ved Universitetet i Oslo i samarbeid med Norsk språkråd i gang med arbeidet med to nye norske allmennordbøker, ei for kvar målform. Dei to ordbøkene hadde kvar sin redaksjon, men i starten samarbeidde redaksjonane tett om opplegget, og dei starta på kvar sin del av alfabetet, for så å byta manus.

Når ordbøkene likevel ikkje blei likare kvarandre, hadde det samanheng både med ulike tradisjonar og ideologiar knytte til dei to målformene og med ulikt kjeldetilfang. BOB-redaksjonen kunne mellom anna stø seg på *Norsk riksmålsordbok* og *Tanums store rettskrivningsordbok*, som inneheldt mange samansetjingar. For nynorsk og målføra var ein i gang med det store verket *Norsk Ordbok*, men mykje arbeid stod att, og ein hadde heller ikkje noka ordbok som svarte til *Tanum*. Det var difor trong for å ha med fleire samansetjingar i NOB. På nynorsk-sida ønskte ein dessutan å halda ute ein del av dei danske og tyske lånorda som var vanlege i bokmålet (Nes 1984; Gerdener 1986), og i staden ta inn fleire ord frå dei norske målføra (Worren 1998).

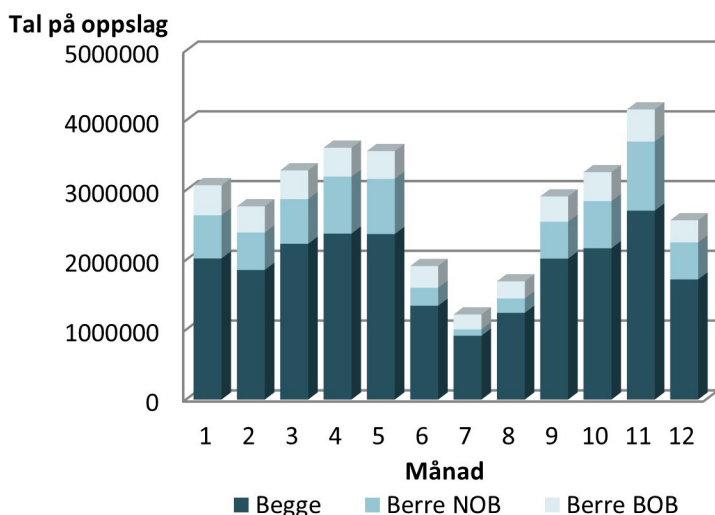
Førsteutgåvene på papir låg føre i 1986, og omfanget viste seg å vera på om lag 65 000 oppslagsord i BOB og 90 000 i NOB. Sidan den gongen er begge ordbøkene komne ut i fleire nye utgåver delvis uavhengig av kvarandre. I revisjonane er det helst rettskrivinga og bøyingane som er oppdaterte i samsvar med nye reformer, i tillegg til at nyord er lagde til, men det har ikkje vore gjort store revisjonar av innhaldet (Hovdenak 2013).

2.3. Bruken av ordbøkene

Den gongen ordbøkene var nye, var det kan henda ikkje så farleg om det var ein del skilnader mellom dei. Då var ordbøker noko som fanst på papir. Lurte ein på noko, slo ein opp i ordboka for den målforma ein skulle skriva på. Det var nok dei færraste som kom på å samstundes slå opp i ordboka for den andre målforma for å samanlikna artikkane.

I dag er situasjonen ein annan. Ordbøkene finst i elektroniske utgåver, på nett sidan 1994 og som app sidan 2017. Her er begge ordbøkene samla på éin stad, og dei blir oppfatta som éin ressurs. Ein kan søkja på eit ord og få opp treffa frå begge ordbøkene ved sida av kvarandre.

Figur 2 viser oppslaga i nettordbøkene i 2018, fordelte på månad og ordbok. Me ser at det er litt fleire som slår opp i NOB (20 %) enn i BOB (13 %), men mest slåande er den store delen som bruker standardsøket som blir gjort parallelt i begge ordbøkene (67 %).² Det fører til at det òg er langt fleire enn tidlegare som legg merke til at dei to tvillingane ikkje alltid er så like, og utjamning av skilnadene mellom dei blir ei viktigare oppgåve enn det har vore tidlegare.



Figur 2: Bruk av nettsida ordbok.uib.no i 2018. Kjelde: Ordbokstatistikk (2019).

3. Parallellredigering i praksis

Som nemnt er det no i gang ein revisjon av dei to ordbøkene. Prosjektet er femårig (2018–2023), finansiert av Kulturdepartementet, og det er eit samarbeid mellom Universitetet i Bergen og Språkrådet. I prosjektsøknaden blei det lagt vekt på at målet var å få til to mest mogleg parallelle ordbøker: «BOB og NOB er og skal være tvillingordbøker, dvs. at de skal være parallelle i innhold bortsett fra når forhold som er særegne for den ene eller den andre målformen, tilsier noe annet» (Språkrådet 2017).

² Talet på oppslag i NOB er høgt med tanke på at berre om lag 12 prosent av elevane i grunnskulen i dag har nynorsk som sitt primære skriftspråk (SSB 2019).

Det som er heilt nytt samanlikna med tidlegare arbeid med ordbøkene, er at det er sett ned éin redaksjon som skal arbeida parallelt med både BOB og NOB. I redaksjonen jobbar me slik at ein redaktør får tildelt eit alfabetstrekk som han eller ho så redigerer parallelt i begge ordbøkene.

Det første som blir sjekka, er om oppslagsordet er med i begge ordbøkene, noko det jo ikkje var i dømet med *animasjon*. Dersom det ikkje er det, må redaktøren ta stilling til om det er rimeleg å berre ha det med i den eine ordboka, eller om det bør opprettast ein tilsvarande artikkel i den andre. Det er òg mogleg at oppslagsordet i visse tilfelle er så utdatert eller perifert at det ikkje treng å vera med vidare i nokon av ordbøkene. For å prøva å avgjera dette bruker me tilgjengelege tekstkorpus, men det er ei utfordring at tekstomfanget er mykje større på bokmål enn på nynorsk (Lyse 2020, i dette nummeret).

Dersom oppslagsordet alt er med i både BOB og NOB, vurderer redaktøren om artiklane er rimeleg parallelle, eller om det er trong for å gjera dei likare kvarandre når det gjeld anten etymologi, uttaleopplysningar, definisjonstekst, bruksdøme eller kan henda heile artikkelstrukturen.

Figur 3 viser eit døme på korleis ein artikkel, *glamour*, kan sjå ut i dei to ordbøkene. Me ser at begge har med dette oppslagsordet, og at artikkelstrukturen er mykje den same, med unnatak av at BOB har med eit bruksdøme til slutt, og det burde vore med i NOB òg. Bøyingsparadigma er sette opp på same måten, men uttaleopplysningane sprikjer. I BOB får me vita at ordet skal uttalast /glamo´r/, medan NOB hevdar at uttalen skal vera /glæm´mer/. Det er ingen grunn til at uttalen skal vera ulik i nynorsk og bokmål, så her trengst det ei utjamning, og truleg er uttalen i BOB den vanlegaste i dag. Når det gjeld etymologien og definisjonen, står ikkje opplysningane i direkte motsetnad til kvarandre. Essensen er mykje den same, men formuleringane er ulike. Her ser me òg ein av føremonene med parallellredigeringa, nemleg at ein kan la seg inspirera frå det beste i begge dei to artiklane i forsøket på laga ein ny og likare tekst i begge ordbøkene.

Bokmålsordboka	Nynorskordboka
<p>glamour substantiv m BØYING</p> <p>UTTALE <i>glamo´r</i></p> <p>OPPHAV engelsk egentlig skotsk variant av <i>grammar</i> 'grammatikk', fordi lærdom ble forbundet med magi</p> <p>BETYDNING magisk skjønnet, sjarm (som ofte er utvendig) <i>hun har glamour</i></p>	<p>glamour substantiv m BØYING</p> <p>UTTALE <i>glæm´mer</i></p> <p>OPPHAV engelsk 'fortrolling, synkverving', omlaga av <i>grammar</i> 'grammatikk'</p> <p>TYDING strålende (men ofte tom og utvendig) venleik og sjarm ytre glans</p>

Figur 3: Oppslag på *glamour* i BOB og NOB.

Bokmålsordboka	Nynorskordboka
<p>av rette verb BØYING</p> <p>OPPHAV etter tysk</p> <p>BETYDNING</p> <ol style="list-style-type: none"> dressere <i>avrette hunder</i> gjøre rett <i>avrette platekanter</i> <i>avrette en mur</i> henrette <i>bli avrettet i den elektriske stolen</i> 	<p>av rette, av retta verb BØYING</p> <p>OPPHAV etter tysk av rette (III)</p> <p>TYDING avlive (ein dødsdømd) <i>bli avretta i den elektriske stolen</i></p>

Figur 4: Oppslag på *avrette* i BOB og NOB.

Figur 4 viser oppslaga for verbet *avrette* der sjølve artikkelstrukturane skil seg frå kvarandre. Det er ein liten skilnad i etymologiane, men den største skilnaden er at BOB har tre ulike tydingar, medan NOB berre har éi, som svarer til tyding 3 i BOB. Tydingsstrukturen i artikkelen i BOB er henta meir eller mindre direkte frå førsteutgåva av *Norsk riksmålsordbok*. Det er naudsynt å undersøkje dei tilgjengelege tekstkorpusa nærmare for å finna ut om alle desse tydingane framleis er i bruk i moderne bokmål, og om dei òg finst i nynorsk. Når det gjeld tyding 2 i BOB, er det relativt lett å finna henne i moderne bokmål, men for nynorsk må ho undersøkjast nærmare. Tradisjonelt har nynorsk gjerne hatt laust samansette verb der bokmål har faste samansetjingar, slik at det har vore vanleg å til dømes skriva at ein «rettar av eit golv», heller enn at ein «avrettar» det, men her er målforma i endring (Almenningen 2006:90–91). Det er ofte ei utfordring å finna den rette balansen mellom for mykje og for lite utjamning.

4. Røyslene med utjamninga så langt

Kor stort er eigentleg omfanget av umotiverte skilnader mellom ordbøkene? For å få eit bilete av situasjonen har eg gjort ei lita undersøking av dei 300 første oppslagsorda i alfabetbolken som startar med *aviarium*, og sett på korleis artiklane fordelte seg på dei to ordbøkene i utgangspunktet, og kva endringar me har gjort framlegg om.³

Av dei 300 undersøkte oppslagsorda var det 14 prosent som i utgangspunktet berre var med i BOB, og 33 prosent som berre var med i NOB. Resten (53 %) hadde artiklar i begge ordbøkene (sjå tabell 1).

	Før revisjonen (N = 300)	Etter revisjonen (N = 237 ⁴)
Berre i BOB	14 %	7 %
Berre i NOB	33 %	11 %
I begge	53 %	82 %
Til saman	100 %	100 %

Tabell 1: Fordeling av dei undersøkte artiklane på ordbøkene.

Etter den første revisjonsrunden er særtilfanget i dei to ordbøkene redusert slik at 82 prosent av dei oppslagsorda som framleis er med, no har artikkel i begge ordbøkene, 7 prosent har berre artikkel i BOB, og 11 prosent har berre artikkel i NOB.

I dei artiklane som alt fanst i begge ordbøkene, er det òg gjort ein del tilpassingar for å gjera ordbøkene likare kvarandre. Opphavleg var 158 av dei undersøkte artiklane med i begge ordbøkene, og i revisjonen har me gjort 100 av dei likare kvarandre, det vil seia om lag to tredelar. Det kan dreia seg om endringar i uttaleopplysningar, etymologi, definisjonstekst eller heile artikkelstrukturen. Samla sett var det altså ein stor del av dei undersøkte artiklane som trong ei eller anna form for samkøyring.

³ Oppslagsorda frå BOB og NOB slegne saman i ei felles liste. Oppslagsord som finst i begge ordbøkene, utgjer éi oppføring på lista.

⁴ Ein del av dei opphavlege artiklane er tekne ut av ordbøkene i revisjonen (anten fordi me finn få belegg på bruk, eller fordi det dreier seg om lite sentrale samansetjingar), og difor er N redusert her. Det er i staden lagt til heilt nye oppslagsord i alfabetstrekket, men dei er haldne utanfor denne opteljinga. Så langt ser det ut til at BOB veks mest, slik at omfanget av dei to ordbøkene vil bli likare.

5. Konklusjon

Med overgangen frå papir til digitale ordbøker oppdagar ordbokbrukarane ulikskapar dei ikkje såg tidlegare, og ei av hovudoppgåvene i revisjonsarbeidet er difor å gjera BOB og NOB likare. Om me skal samanfatta røynslene så langt, har arbeidet avdekt at ulikskapane er større enn venta, noko som har ført med seg at denne delen av prosjektet har blitt meir omfemnande enn me hadde sett føre oss. For å kunna fullføra prosjektet innanfor dei rammene me har, kan det bli naudsynt å prioritera dei viktigaste oppgåvene, og me trur det vil bety meir for brukarane at lemmautvalet og definisjonane er likast mogleg, enn at etymologiar, uttaleopplysningar og bruksdøme er det. Me håper uansett at revisjonsarbeidet skal gjera ordbøkene meir brukarvennlege i framtida slik at verken Harald eller andre blir villeidde av tilfeldige skilnader.

Litteratur

Ordbøker

BOB (1986–) = *Bokmålsordboka*. Oslo: Kunnskapsforlaget. <ordbok.uib.no> (oktober 2019).

NOB (1986–) = *Nynorskordboka*. Oslo: Det Norske Samlaget. <ordbok.uib.no> (oktober 2019).

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet (2016). Oslo: Det Norske Samlaget. <no2014.uib.no> (oktober 2019).

Norsk riksmålsordbok (1937–1957). Oslo: Aschehoug.

Tanums store rettskrivningsordbok (1964–). Oslo: Kunnskapsforlaget.

Annan litteratur

Almenningen, Olaf (2006): *Innføring i nynorsk for høgare utdanning*. Oslo: Det Norske Samlaget.

Gerdener, Wilhelm (1986): *Der Purismus im Nynorsk. Historische Entwicklung und heutige Sprachgebrauch*. Münster: Hans Hölker Verlag.

Gundersen, Dag (1967): *Fra Wergeland til Vogt-komiteen. Et utvalg av hovedtrekk og detaljer fra norsk språknormering*. Oslo: Universitetsforlaget.

Hovdenak, Marit (2013): Bokmålsordboka og Nynorskordboka gjennom ein generasjon. I: Ruth Vatvedt Fjeld & Marit Hovdenak (red.): *Nordiske studier i leksikografi* 12. Oslo: Nordisk forening for leksikografi, 229–246.

- Innstilling (1966): *Innstilling om språksaken fra komitéen til å vurdere språk-situasjonen m.v. oppnevnt ved kongelig resolusjon 31. januar 1964*. Oslo: Kirke- og undervisningsdepartementet.
- Lyse, Gunn Inger (2020): Ut med adamsslekt og inn med arveprinsesse? Leksikografiske metodar i revisjonen av Bokmålsordboka og Nynorskordboka. I: *Nordiska studier i lexicografi* 15, 215–224.
- Nes, Oddvar (1984): *Om nokre affiks i nynorsk – mest om -heit*. Bergen: Nordisk institutt, Universitetet i Bergen.
- Ordbokstatistikk* (2019). Universitetet i Bergen. <ordbok.uib.no/perl/ordbokstatistikk.cgi> (oktober 2019).
- Rauset, Margunn (2019): Bokmålsordboka og Nynorskordboka – einægga, toegga eller siamesiske tvillinger? I: *LexicoNordica* 26, 155–175.
- Språkrådet (2017): *Prosjektsøknad – revisjon av Bokmålsordboka og Nynorskordboka*. Upublisert.
- SSB (2019): *Elevar i grunnskolen*. <www.ssb.no/utdanning/statistikker/utgrs/aar> (oktober 2019).
- Worren, Dagfinn (1998): Om å avgrense eit ordtilfang – saga om målføreorda i Norsk handordbok. I: Ruth Vatvedt Fjeld & Boye Wangenstein (red.): *Normer og regler. Festskrift til Dag Gundersen 15. januar 1998*. Oslo: Nordisk forening for lexicografi, 59–70.

Bente Selback
redaktør for revisjonen av
Bokmålsordboka og Nynorskordboka
Universitetet i Bergen
Postboks 7805
NO-5020 Bergen
bente.selback@uib.no

Augmented Writing: nye muligheder og nye teorier

Henrik Køhler Simonsen

This article discusses Augmented Writing (AW) in lexicography. The article is based on an analysis of sixteen selected AW services and the empirical data collected seem to indicate that this technology challenges lexicography, but that it also needs lexicographic theory and practice. The insights from the analysis led to a number of theoretical considerations on what AW is and on how it may affect and challenge lexicography. The article also offers theoretical reflections on what lexicography can contribute to AW. Finally, the article offers six theses on how lexicography can make the technology better by providing curated data, curated algorithms and lexicographic world knowledge.

1. Indledning og problem

Vi lever i disruptionens tidsalder og er lige nu vidne til en lang række store omvæltninger og fundamentale skift i bl.a. banksektoren, forsikringsbranchen, advokatbranchen og sundhedssektoren. De store omvæltninger skyldes bl.a. en stigende udbredelse af automatisering samt kunstig intelligens (AI).

Kunstig intelligens i form af *Augmented Writing* (AW) er nu også kommet til journalist- og kommunikationsbranchen, hvor der i stadigt stigende omfang anvendes forskellige former for ny teknologi til at understøtte f.eks. journalisters og tekstproducenters research og arbejde. Mange virksomheder har allerede taget AW til sig og anvender den nye teknologi i mange situationer. McDonalds har f.eks. indgået en global aftale med AW-tjenesten Textio, jf. Textio (2019), som specialiserer sig i AI-understøttet rekruttering og virksomhederne Chase, Sky, Vodafone og American Express har indgået aftaler med AW-tjenesten *Persado*, som specialiserer sig i AI-understøttet produktion af marketingtekster. Og mange andre virksomheder og organisationer bruger allerede AW til produktion af juridiske tekster samt generering af jobopslag etc.

Formålet med denne artikel er at diskutere følgende tre forskningsspørgsmål:

- 1) Hvad er AW?
- 2) På hvilke områder synes leksikografien at blive udfordret af AW?
- 3) Hvad kan leksikografi bidrage med og hvad er leksikografiforbedret AW?

2. Metode og data

Bidraget er baseret på en empirisk analyse af seksten specielt udvalgte, og meget forskellige, AW-tjenester, som i varierende grad anvender forskellige typer af data og AI-algoritmer til på den ene eller anden måde at hjælpe brugeren med at optimere tekstproduktionen, enten autonomt (hvor der automatisk genereres en tekst på basis af nogle få stikord) eller med menneskelig mellemkomst (hvor teksten løbende forbedres på basis af valg foretaget af brugeren). De seksten AW-tjenester er blevet selekteret for yderligere test og analyse på basis af hvor relevant den pågældende AW-tjeneste synes at være for tekstproduktion.

Testen indebar, at jeg analyserede hver AW-tjenestes opgavefokus, anvendelsesområde, grad af autonomi, grad af integration med brugerens arbejdsområde og den vurderede output-kvalitet. Tabel 1 viser en oversigt over de seksten AW-tjenester og deres opgave- og anvendelsesfokus.

Navn	Opgave	Anvendelse
Articoolo	Tekstproduktion	Alle domæner
Essaybot	Tekstproduktion	Universitære opgaver
Grammarly	Tekstanalyse	Alle domæner
Grammarly	Grammatik & Stavning	Alle domæner
Linguix	Grammatik & Stavning	Alle domæner
MessagePath	Tekstanalyse	Marketing og jura
Persado	Tekstanalyse	Marketing
Phrasee	Tekstanalyse	Marketing
ProWritingAid	Tekstproduktion	Alle domæner
TalktoTransformer	Tekstproduktion	Alle domæner
TapRecruit	Tekstproduktion	Jobopslag og HR
Textio	Tekstproduktion	Jobopslag og HR
Whitesmoke	Grammatik & Stavning	Alle domæner
WordRake	Tekstanalyse	Alle domæner og jura
Write Assistant	Oversættelse	Oversættelse
WritingAssistant	Grammatik & Stavning	Sociale medier

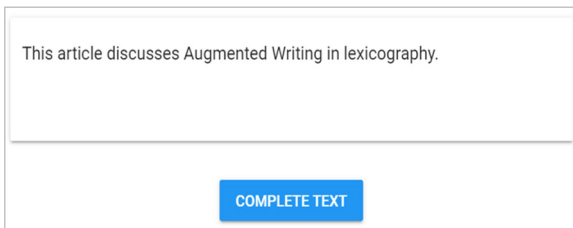
Tabel 1: Oversigt over de 16 AW-tjenester.

Analysen påviste meget store forskelle på de seksten AW-tjenester og herunder gennemgås af pladshensyn fire udvalgte eksempler.

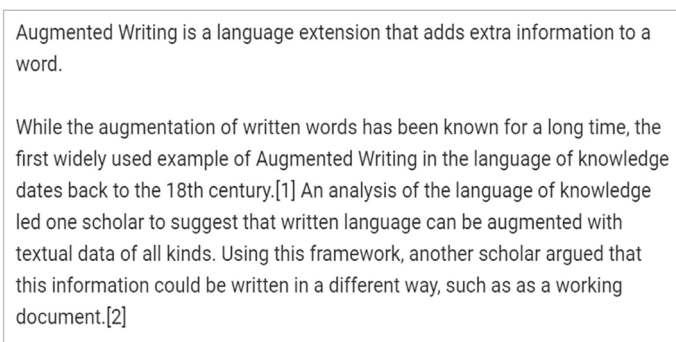
To eksempler på AW-tjenester, som har en høj grad af autonomi og som ikke kræver nogen menneskelig mellemkomst er *Articoolo* og *TalktoTransformer*. Graden af integration med brugerens arbejdsområde er lille, idet begge AW-tjene-

ster afvikles på separate websites. Begge tjenester genererer automatisk tekster på basis af input fra brugeren, men den vurderede output-kvalitet er lav.

Figur 1 herunder viser to udsnit fra *TalktoTransformer*, hvor jeg først har indsat en sætning på syv ord, som vist i figur 1 og derefter klikket på knappen *Complete text*, hvorefter *TalktoTransformer* automatisk har genereret en tekst som vist i figur 2. *TalktoTransformer* anvender bl.a. et neuralt netværk med sprogmodellen GPT2, som er baseret på 1,5 milliarder parametre.



Figur 1: Indtastningsfelt i TalktoTransformer (2019).

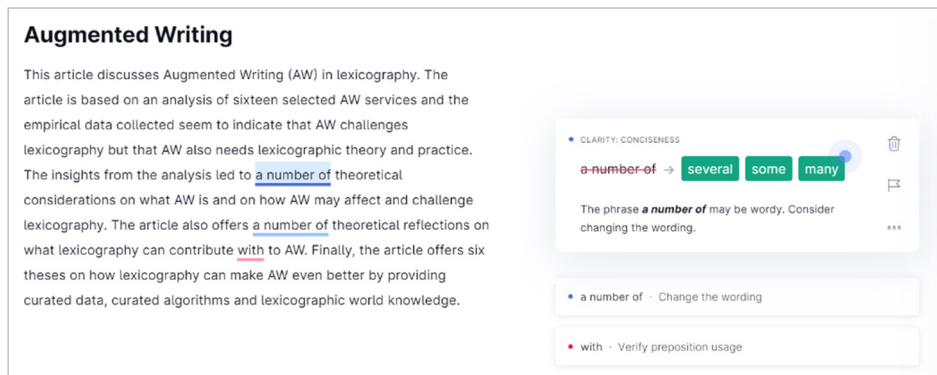


Figur 2: Artikelfelt i TalktoTransformer (2019).

Som det fremgår af den automatisk genererede tekst i figur 2, så lader tekstens kvalitet meget tilbage at ønske, og den bør i høj grad efterredigeres og forbedres med relevant omverdensviden og faktatjekkes.

To eksempler på AW-tjenester, som har en meget lille grad af autonomi og dermed hele tiden kræver menneskelig stillingtagen er f.eks. *Grammarly* og *ProwritingAid*. Disse AW-tjenester har en meget høj grad af integration, idet man både i de fleste Microsoft Office-produkter og Google Chrome har sømløs adgang til begge AW-tjenester, som løbende tjekker stavning, grammatik, stil og tekstens opsætning. Output-kvaliteten af *Grammarly* og *ProwritingAid* afhænger naturligvis af brugerens løbende valg, men det argumenteres, at output-kvaliteten, med den hjælp de to AW-tjenester tilbyder, er høj.

Figur 3 herunder viser et abstract efter, at jeg har bedt AW-tjenesten *Grammarly* om at tjekke grammatik og tone i denne artikels abstract. Som det fremgår er der værdifuld hjælp at hente i *Grammarly*.



Figur 3: Grammtik- og stiltjekker i Grammarly (2019).

Til forskel fra de to første mere autonome AW-tjenester er der således for f.eks. *Grammarly* og *ProwritingAid* ikke behov for, at teksten skal efterredigeres og forbedres med leksikografisk omverdensviden, idet dette er sket løbende.

3. Definition og teoretisk diskussion

Da Augmented Writing er et helt nyt felt, er det først nødvendigt at etablere en dækkende definition af AW med afsæt i f.eks. Marconi (2017). På basis af eksisterende litteratur samt analysen og testen af de seksten AW-tjenester var det muligt at udarbejde følgende definition, som besvarer det første forskningsspørgsmål om, hvad AW er:

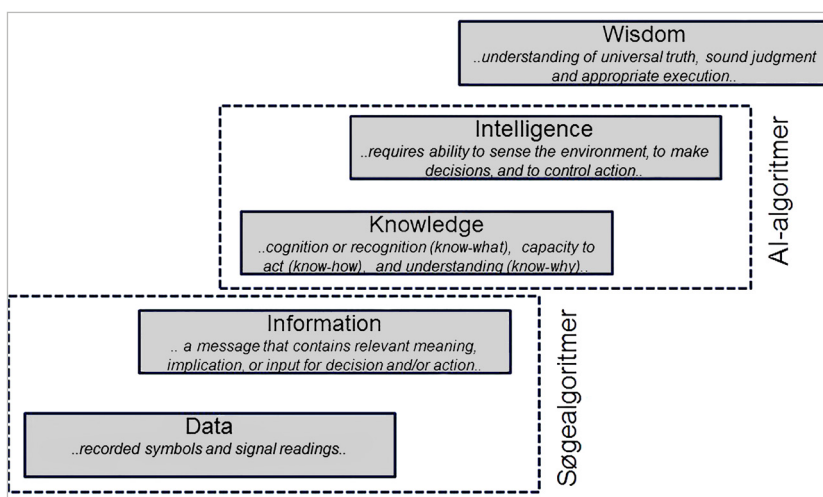
Augmented Writing er et AI-baseret værktøj, der med eller uden hjælp fra et menneske på forskellige niveauer hjælper tekstproducenter med at producere, revidere, analysere eller oversætte tekster, og som automatisk kontrollerer grammatisk, syntaktisk og stilistisk korrekthed og/eller som foreslår forskellige typer forbedringer.

En anden og helt afgørende teoretisk overvejelse, før vi kan behandle det andet forskningsspørgsmål om, på hvilke områder leksikografien synes at kunne blive udfordret af AW, er spørgsmålet om, hvordan vi mennesker egentlig skriver en tekst. Ifølge Abas et al. (2018:20) kan skriveprocessen opdeles i ”Drafting, Planning, Prewriting, Revising and Editing og Pausing and Reading”, men disse trin

er på ingen måde lineære. Denne forståelse ses også i f.eks. Liferichpublishing (2019), som anbefaler den såkaldte ARRR-metode, som dækker over en dynamisk opstående tilgang til faserne ”Add, Rearrange, Remove, Replace”.

En tredje relevant gruppe af teoretiske overvejelser er at forstå leksikografiens konsultationsprocesser. Dette er på ingen måde et enkelt spørgsmål, og der findes da også allerede en lang række teoretiske overvejelser på dette felt. Af særlig relevans for denne artikel er Tarp et al. (2017:495), som blandt andet diskuterer ”L2 writing assistants and context-aware dictionaries” samt foreslår syv typiske konsultationstrin. Som det blev påvist i figur 1 og 2 synes både skrive- og konsultationsprocesser at blive udfordret af AW.

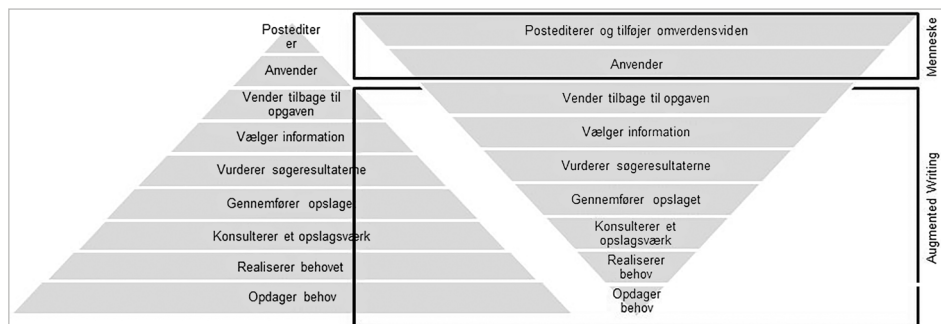
Et fjerde og helt afgørende spørgsmål er, hvordan opdelingen mellem menneske og maskine allerede er ved anvendelse af visse AW-tjenester. Det er nødvendigt at se på, hvilken rolle søgealgoritmer og AI-algoritmer spiller for skrive- og konsultationsprocesserne og se på de fundamentale forskelle mellem hhv. data, information og viden. For at behandle dette spørgsmål har jeg videreudviklet Liews DIKIW-model, jf. Liew (2013:56-60), og angiver vha. stiplede linjer hvor søgealgoritmer og AI-algoritmer udfolder deres potentiale. Figur 4 herunder hjælper os således med at forstå sammenhængen mellem data, information, viden og intelligens og viser, at søgealgoritmer især udfolder sit potentiale ved fremfindning af hhv. data og information, og at AI-algoritmer især udfolder sit potentiale ved fremfindning og skabelse af viden og intelligens. Det kan selvfølgelig hævdes, at viden og intelligens skabes i hjernen hos mennesket, men som analysen af de autonome AW-tjenester viste, så skabes viden og intelligens rent faktisk også i AW-tjenesten selv, jf. f.eks. TalktoTransformers (2019), som vha. sprogmodellen GPT2 producerer en sammenhængende tekst et ord af gangen.



Figur 4: Algoritmer og DIKIW-modellen.

En femte teoretisk overvejelse, som ligeledes synes at være vigtig for forståelsen af, på hvilken måde AW udfordrer leksikografien, er en erkendelse af hvad hhv. algoritmer og mennesker kan. Almindelige søgealgoritmer synes allerede at være bedre end vi mennesker til at fremfinde f.eks. data og information og AI-algoritmer bliver snart bedre end mennesker til at fremfinde og skabe viden og intelligens. Derfor må vi se på de leksikografiske konsultationsprocesser, her forstået som de processer, som brugeren går igennem i forbindelse med opslag og søgning. Der kan nemlig med fordel ske en arbejdsdeling mellem menneske og maskine og lade algoritmerne gøre det, de er bedst til og mennesker gøre det de er bedst til, jf. f.eks. Fadel et al. (2015:76), som anvender en lignende opdeling.

Hvis det nu antages, at de syv processer fra Tarp et al. (2017:495) udbygges med to ekstra processer nemlig *Anvende valgt information* og *Posteditere information*, vil den konventionelle leksikografiske procespyramide se ud som i venstre side af figur 5. Men hvis vi nu opdeler indsatsen, således at AI gør det, den er bedst til og mennesket gør det vi er bedst til, så må vi ”vende leksikografien på hovedet” som vist i højre side af figur 5. Argumentet er, at vi bør se leksikografien og især de leksikografiske processer i et helt nyt lys, da mange af de eksisterende processer formentlig vil blive erstattet af AW og at vi mennesker i stedet for bør bruge tid på at efterredigere og tilføje omverdens- og relationsviden til teksten.



Figur 5: Leksikografien i en AW-tidsalder.

Ovenstående teoretiske diskussion gør det nu muligt at svare på det andet af artiklens forskningsspørgsmål, nemlig spørgsmålet om, på hvilke områder leksikografien synes at blive udfordret af AW. For det første synes AW at ville kunne erstatte en lang række opslags- og søgehandlinger ved automatisk at foreslå kontekstrelevant hjælp, hvilket er gældende for langt de fleste AW-tjenester, herunder f.eks. Grammarly (2019). For det andet vil der formentlig også være en række leksikografiske produktionsprocesser, som i en AW-tidsalder vil blive

erstattet eller kraftigt suppleret af AI og avancerede sprogmodeller, f.eks. korpusarbejde, lemmaselektion og ækvivalentselektion. Men enhver AI har brug for hjælp, idet mennesker føder, renses og tagger de data, som en AW skal bruge for at kunne generere kontekstafhængig hjælp.

Dette fører til besvarelsen af det tredje forskningsspørgsmål om, hvad leksikografien kan bidrage med, og hvad der menes med leksikografiforbedret AW, herefter kaldet LAW. Som det argumenteres ovenfor, har leksikografien en række muligheder for at gøre AW endnu bedre. F.eks. har AW brug for leksikografien for at kunne håndtere forskellige leksikalske betydninger og for at kunne tilføje den nødvendige omverdens- og relationsviden, som kun vi mennesker (endnu) kan tilføje. Samtidig kan det argumenteres, at AW de facto stiller endnu højere krav til sine brugeres sproglige kompetencer, idet der med mere AW formentlig vil blive brug for endnu mere efterredigering, hvilket i øvrigt blev påvist i figur 1 og 2.

Jeg argumenterer således, at fokus i højere grad bør være at hjælpe brugeren med efterredigering, dvs. der hvor brugerens udfordringer nu ligger. AW er kommet for at blive og vil formentlig kun blive bedre, men samtidig lader kvaliteten meget tilbage at ønske. Man kan måske endda tale om ”leksikografiforbedret AW” (LAW), som med fordel kan hvile på følgende seks teser:

- 1) LAW skal anvende kuraterede leksikografiske data
- 2) LAW skal anvende kuraterede algoritmer
- 3) LAW skal hjælpe brugeren med efterredigeringen
- 4) LAW skal tilføre omverdensviden i efterredigeringen
- 5) LAW skal tilføre relationel viden i efterredigeringen
- 6) LAW skal understøtte domænespecifikke brugssituationer.

For det første bør LAW hvile på kuraterede data. AI bygges af mennesker, og hvis maskinen fodres med data af tvivlsom kvalitet, bliver outputkvaliteten tilsvarende tvivlsom. Det samme er i høj grad gældende for de algoritmer, der anvendes. De skal i særlig grad hvile på kuraterede og alment anerkendte etiske retningslinjer, således outputtet ikke på den eller anden måde bliver biased. For det tredje skal LAW understøtte den essentielle efterredigering, som i mange tilfælde vil være nødvendig. Det gælder altså om at sætte brugeren i stand til at efterredigere tekster, således de får et højt sprogligt niveau. For det fjerde og femte skal LAW netop øge output-kvaliteten ved at tilføre den helt nødvendige omverdensviden og relationelle viden, som kun vi mennesker (endnu) har. Især her vil leksikografer og sprogteknologer kunne spille en afgørende rolle. For det sjette argumenterer jeg, at LAW bør fokusere på domænespecifikke brugssituationer for at sikre output-kvaliteten.

5. Konklusion

AW udfordrer leksikografien, men AW bygges og trænes af mennesker og leksikografien har i særlig grad meget at tilbyde AW. Som det er fremgået af de empiriske data og den teoretiske diskussion, så udfordres leksikografien af AW, fordi især søge- og AI-algoritmer er bedre og hurtigere til at fremfinde data og information og til automatisk at generere viden end vi mennesker.

Men der mangler omverdens- og relationsviden. Her er mennesket (endnu) AW overlegent, og især leksikografiens saglige, kuraterede og etisk afbalance-rede tilgang til data, information og viden er nødvendig, ligesom også leksiko-grafiens bruger- og behovsorienterede tilgang til udvælgelse og præsentation af data, information og viden må være central for AW.

Leksikografien er på ingen måde lagt i graven endnu. Leksikografien vil på flere punkter kunne bidrage til udviklingen af AW-tjenester og bidrage til udviklingen af nye leksikografiforbedrede AW-tjenester (LAW).

Litteratur

AW-tjenester

- Articoolo (2019) = *Articoolo*. <www.articoolo.com> (marts 2019).
 Essaybot (2019) = *Essaybot*. <www.essaybot.com> (marts 2019).
 Grammark (2019) = *Grammark*. <www.grammark.org/dist/#/> (marts 2019).
 Grammarly (2019) = *Grammarly*. <www.grammarly.com> (marts 2019).
 Linguix (2019) = *Linguix*. <www.linguix.com> (marts 2019).
 MessagePath (2019) = *MessagePath*. <www.messagepath.com> (marts 2019).
 Persado (2019) = *Persado*. <www.persado.com> (marts 2019).
 Phrasee (2019) = *Phrasee*. <www.phrasee.co> (marts 2019).
 ProWritingAid (2019) = *ProWritingAid*. <www.prowritingaid.com> (marts 2019).
 TalktoTransformer (2019) = *TalktoTransformer*. <www.talktotransformer.com> (marts 2019).
 TapRecruit (2019) = *TapRecruit*. <www.taprecruit.co> (marts 2019).
 Textio (2019) = *Textio*. <www.textio.com> (marts 2019).
 Whitesmoke (2019) = *Whitesmoke*. <www.whitesmoke.com/writing-assistant> (marts 2019).
 WordRake (2019) = *WordRake*. <www.wordrake.com> (marts 2019).
 Write Assistant (2019) = *Write Assistant*. <www.writeassistant.com/da> (marts 2019).
 WritingAssistant (2019) = *WritingAssistant*. <www.writing-assistant.com> (marts 2019).

Anden litteratur

- Abas, Imelda, Abd Aziz & Noor Hashima (2018): Model of the Writing Process and Strategies of EFL Proficient Student Writers: A Case Study of Indonesian Learners. I: *Pertanika* 26, 1815-1842.
- Fadel, C., B. Trilling & M. Bialik (2015): *Four-Dimensional Education: The Competencies Learners Need to Succeed*. Center for Curriculum Redesign. ISBN-13: 978-1518642562.
- Liew, Anthony (2013): DIKIW: Data, Information, Knowledge, Intelligence, Wisdom and their Interrelationships. I: *Business Management Dynamics*. Vol. 2, Issue 10, April 2013, 49-62.
- LifeRichPublishing (2019): *The 5-Step Writing Process: From Brainstorming to Publishing*. <www.liferichpublishing.com> (marts 2019).
- Marconi, Francesco (2017): *The Year of Augmented Writing*. <www.niemanlab.org/2016/12/the-year-of-augmented-writing> (marts 2019).
- Tarp, Sven, Kasper Fisker, & Preben Sepstrup (2017): L2 writing assistants and context-aware dictionaries: New challenges to lexicography. I: *Lexikos*, 27(1), 494-521.

Henrik Køhler Simonsen
 ekstern lektor, PhD, MA, MBA
 Copenhagen Business School
 Department of Management, Society & Communication
 Dalgas Have 15
 DK-2000 Frederiksberg
 hks.msc@cbs.dk

Islandske ordbøger i 1800-tallet og deres rolle i sprogstandardiseringen

Ásta Svavarsdóttir

This article gives an overview of Icelandic 19th century dictionaries and discusses their role in the development of the national language standard emerging in the same period. It addresses two main questions: Who compiled the dictionaries, for whom and for what purposes? Did the lexicographical work have a direct or indirect effect on the standardisation process, and if so, how?

1. Indledning: Ordbøger og sprogstandardisering

I indlægget diskuteres 1800-tallets islandske ordbøger med henblik på deres stilling i sprogstandardiseringsprocessen. Ifølge Einar Haugens model for udviklingen af standardsprog, består sprogstandardisering af fire faser: (1) valg (af norm), (2) kodificering, (3) implementering og (4) elaborering, hvoraf de to første gælder sprogets form, mens de to sidste angår dets funktion (Haugen 1966:933). Ordbøger – samt grammatikker, retskrivningsordbøger og stilistiske vejledninger – spiller en vigtig rolle i kodificeringsfasen og de kan også være af betydning for implementering af normen, dvs. ”the gradual diffusion and acceptance of the newly created norm across speakers as well as across functions” (Deumert & Vandenbussche 2003:7). Standardisering af det moderne nationalsprog i Island – i almindelighed anset og omtalt som ”sprogreform” – foregik samtidigt med den politiske selvstændighedskamp i det 19. og tidlige 20. århundrede. I samme periode udkom der flere islandske ordbøger: Hvem skrev disse ordbøger, for hvem og hvorfor? I hvilken grad har de haft direkte eller indirekte indflydelse på den nationale sprogstandard som var under udvikling?

Kapitel 2 fokuserer på det første spørgsmål. Indledningsvis (2.1) gives der en oversigt over islandske ordbøger udgivet i 1800-tallet og derefter (2.2) følger en analyse af deres redaktører, primære målgruppe og formål, først og fremmest baseret på ordbøgernes forord og til dels på lemmaudvalg og struktur. I kapitel 3 diskuteres ordbøgernes rolle i sprogstandardiseringen og i kapitel 4 er der kortfattede konklusioner.

2. Islandske ordbøger i 1800-tallet

2.1. Oversigt

1800-tallet var en frugtbar periode mht. ordbøger over islandsk, mens der tidligere kun var blevet udgivet en håndfuld værker (jf. SOOÍ). Tabel 1 viser en oversigt over ordbøger udgivet i det 19. århundrede.

År	Islandsk som kildesprog	År	Islandsk som målsprog
1814	Björn Halldórsson: <i>Lexicon Islandico-Latino-Danicum</i>	1819	Gunnlaugur Oddsson: <i>Orðabók sem inniheldur ...</i> (dansk-islandsk)
1860	Sveinbjörn Egilsson: <i>Lexicon poëticum</i>	1851	Konráð Gíslason: <i>Dönsk orðabók ...</i> (dansk-islandsk)
1863	Eiríkur Jónsson: <i>Oldnordisk Ordbog</i>	1896	Jónas Jónasson: <i>Ný dönsk orðabók ...</i> (dansk-islandsk)
1874	Cleasby/Guðbrandur Vigfússon: <i>An Icelandic-English Dictionary</i>	1896	Geir T. Zoëga: <i>Ensk-íslenzk orðabók</i> (engelsk-islandsk)
1876/1895	Jón Þorkelsson: <i>Supplement til islandske Ordbøger</i> (1 og 2)		
1890-1897	Jón Þorkelsson: <i>Supplement til islandske Ordbøger</i> (3, I-II)		
1900	Björn Jónsson: <i>Íslenzk stafsetningarorðabók</i>		

Tabel 1: Ordbøger fra 1800-tallet med islandsk som kildesprog eller målsprog.

Tabellen viser kun værker som blev (delvis) udarbejdet af islandske leksikografer og udgivet i sin helhed, fx ikke ordbøger som kun er overleveret i manuskript. Glossarier og ordlister inkluderes ikke.

De fleste tidlige værker var udarbejdet og udgivet i København. Cleasby & Guðbrandur Vigfússons ordbog blev færdiggjort i England og udgivet i Oxford i 1870-erne, men det oprindelige manuskript var udarbejdet af islandske leksikografer i København i perioden 1840-1854 (jf. ordbogens forord; 1874:v-vi). Det var først i 1800-tallets sidste fjerdedel som udgivelse af islandske ordbøger flyttedes for det meste til Island.

Alle ordbøgerne, undtagen Björn Jónssons (1900) retskrivningsordbog, er tosprogede og i de fleste er islandsk kildesprog. Det øvrige sprog er oftest dansk, men engelsk præsenteres også i den sidste del af århundredet. Den ældste ordbog (Björn Halldórsson 1814) har to målsprog, latin og dansk. Det latinske målsprog i ordbogen er en arv fra ældre tid – det var reglen i 1600- og 1700-tallets ordbøger og Björn skrev sin ordbog sent i 1700-tallet selv om den først blev udgivet posthumt i 1814. Rasmus Rask fik ansvaret for udgivelsen og han og hans medarbejdere tilføjede det danske målsprog i takt med voksende vægt på de moderne sprog.

En uklar grænse mellem oldislandsk og samtidssproget er et fælles træk ved især ordbøger med islandsk kildesprog og det genspejles gerne i titlerne. I *Oldnordisk ordbog* findes der fx fra det samtidige talesprog ”Ord, Ordbetydninger og Talemaader, hvis oldnordiske Charakter syntes deels aabenbar, deels temmelig utvivlsom ifølge de senere Afændringer i de scandinaviske Dialecter” (Eiríkur Jónsson 1863:xxxviii). I indledningen til *An Icelandic-English Dictionary* står: ”This work is a Dictionary of the Old Icelandic Language” (Cleasby & Vigfusson 1874:iii). Alligevel er det klart at der er blevet inkluderet væsentligt materiale fra yngre sprog, inklusive samtidssproget, men i indledningen siges jo også at ”though the Dictionary is mainly intended for the old authors, both in prose and poetry, it endeavours to embrace an account of the whole language, old and new” (1874:v). Værket var oprindelig planlagt og udarbejdet som en ordbog over det oldislandske prosasprog og materiale fra samtidskilder blev kun tilføjet på et senere trin, ifølge forordet til ordbogen (1874:v-vii). I ordbøger hvor islandsk er målsprog er lemmavalget og artiklernes struktur betinget af kildesproget (jf. Konráð Gíslason 1851:III; Jón Hilmar Jónsson 1991:xiv-xvi), dansk eller engelsk, men også i dem fremtræder oldsproget. Konráð Gíslason (1851:iv) anså fx henvisninger til og citater fra oldislandsk litteratur som en af de største fordele ved sin dansk-islandske ordbog og han undgik ækvivalenter af fremmed oprindelse, medmindre de var dokumenteret fra oldsproget. I indledning til genudgivelsen af Gunnlaugur Oddsons ordbog over (især) danske fremmedord siger Jón Hilmar Jónsson (1991:xix) at blandt de islandske forklaringer og ækvivalenter findes der en blanding af ord som tydeligt røber en fremmed oprindelse og ord som stammer fra oldislandsk litteratur.

2.2. Af hvem, for hvem og hvorfor?

2.2.1. Leksikografer

De som var ansvarlige for 1800-tallets leksikografiske arbejde var en forholdsvis lille gruppe islandske akademikere og studerende i København. Nogle leksikografer havde andel i flere en ét værk: Konráð Gíslason var hovedredaktør ved udarbejdelsen af den islandsk-engelske ordbog initieret af Cleasby i København i 1840-erne og han udgav også en dansk-islandsk ordbog i eget navn (Konráð Gíslason 1851). Eiríkur Jónsson, forfatter til *Oldnordisk ordbog* (1863), deltog ligeledes i Cleasbys redigeringsgruppe i København (jf. Cleasby & Vigfusson 1874:v-vi). Guðbrandur Vigfússon, som afsluttede den islandsk-engelske ordbog, og Jón Þorkelsson, forfatter af supplementer til ældre islandske ordbøger, bidrog begge til udgivelsen af *Lexicon Poëticum* (Sveinbjörn Egilsson 1860; jf. Jón Ólafsson 1904). Og Björn Jónsson, forfatter til retskrivningsordbogen (1900), var også udgiver og medredaktør af Jónas Jónassons (1896) dansk-islandske ordbog og skrev bl.a. forordet til værket.

I 1800-tallet var der i Danmark og Nordeuropa efterspørgsel efter mænd med sproglige og filologiske kundskaber af oldislandsk og de klassiske oldnordiske tekster. Mange leksikografer var uddannede filologer eller filologistuderende og i tillæg til ordbogsredigeringen arbejdede de på at udgive og oversætte oldislandske tekster. Det var således en stærk fokus på oldislandsk i deres virksomhed alment.

Til sidst bør det nævnes at nogle af leksikograferne var aktive i tidens sprogreform, fx som redaktører af tidsskrifter og aviser eller som lærere. Konráð Gíslason var en af redaktørerne til det navnkundige tidsskrift, *Fjölnir*, hvor han også skrev artikler om sprog og sprogbrug; Jón Þorkelsson var lærer og senere rektor i latinskolen i Reykjavík, skrev lærebøger og deltog i tidens sprogdebat; og Björn Jónsson var bl.a. udgiver og redaktør af et af landets vigtigste aviser, *Ísafold*, i den sidste del af 1800-tallet.

2.2.2. Brugere

Hvilken målgruppe havde udgivere og redaktører af 1800-tallets islandske ordbøger i tankerne? Der er få direkte udsagn om eventuelle brugere i selve værkerne, men (næsten) alle ordbøger med islandsk som kildesprog var tosprogede og havde forord og metaoplysninger på målsproget og det tyder på at de primært var tiltænkt udenlandske brugere. Vægten som lægges på oldislandsk indikerer

også at målgruppen var især brugere med interesse for oldnordisk litteratur – filologer, historikere og andre entusiaster – snarere end dem som ville tilegne sig samtidssproget.

Forordet til *Oldnordisk ordbog* viser en interessant synsvinkel mht. målgrupper og behov for ordbøger. Efter en omtale om tilvækst i udgivne oldnordiske tekster og deres oversættelser siges:

derved fik enhver Ynder af denne Literatur sine Önsker nogenlunde tilfredsstillede, i det Folket i Island fik Grundtexten, Folket i Danmark og de andre scandinaviske Lande danske Oversættelser, og de Lærde Oversættelser paa Latin. (Eiríkur Jónsson 1863:xvi)

Dette tyder på at efter forfatterens mening blev behovet for ordbøger mindre desto flere oldislandske tekster, gerne med glossarier, blev udgivet og oversat. Forfatteren mente tilsyneladende også at samtidige islændinge ikke behøvede forklaringer for at forstå de gamle tekster og dermed at de ikke havde brug for ordbøger over (old)islandsk.

I modsætning til de ovenfor beskrevne værker var ordbøger med islandsk som målsprog primært tiltænkt islandske brugere. Målgruppen genspejles bl.a. i at i alle ordbøger fra dansk eller engelsk til islandsk, som publiceredes i perioden, var forord og metaoplysninger på islandsk. Gunnlaugur Oddsson (1819) fortæller fx i forordet til sin ordbog at han har sendt et ”Subscriptionsblad” til Island for at tilbyde sine landsmænd at købe værket. Retskrivningsordbogen af Björn Jónsson (1900) var også tiltænkt islandske brugere.

2.2.3. Formål og funktion

Hovedformålet med de fleste af 1800-tallets ordbøger med islandsk som kilde-sprog var åbenbart at skaffe hjælpemidler for udenlandske læsere og forskere for at forstå, undersøge og oversætte den oldnordiske litteratur og det stemmer overens med brugernes forventninger. Samtidssproglige oplysninger blev set som supplement til ord og eksempler som stammede direkte fra oldislandske kilder og som et middel til at kaste nærmere lys over de gamle tekster. Kun to af ordbøgerne fokuserede hovedsagelig på samtidssproget, dvs. Björn Halldórssons ordbog (1814) og Jón Þorkelssons supplement fra 1890-erne. Björn (1724-1794) var en islandsk præst og oplysningsmand og man mener at han byggede sin ordbog først og fremmest på daglig tale, som fremgår fx af forordet til Eiríkur Jónssons ordbog (1863:xv). Mange brugere blev skuffede over ordbogen fordi de ikke fandt hvad de søgte, dvs. gavnlige oplysninger for at forstå oldislandske

tekster, og værket blev kritiseret i danske og tyske tidsskrifter. Rask, som var hovedredaktør ved udgivelsen af ordbogen, besvarede kritikken som han syntes at genspejlede en misforståelse af forfatterens hensigt med værket:

Ordbogens Anmelder beretter: ”at man [...] var betænkt paa, at udvide Bogen til det nyere og omgangssproget.” Han har altsaa ganske taget feil af Ordbogens Indretning og Forfatterens Plan og Hensigt. [...] Bjørn Haldorsen, som [...] talte og skrev Islandsk som Modersmaal, har ikke trukket nogen Grænse imellem det ældre og nyere Sprog [...]. Han har derfor samlet af Sagaerne, hvad der var brugeligt og forstaaeligt, uden ængstelig at opsøge ethvert kunstord eller forældet Udtryk, som ingen Islænder siden Reformationens Indførelse har brugt, eller enhver ubetydelig Sammensætning, som en Forfatter har brugt en enkelt Gang, men som ikke har faaet nogen Fasthed i Sproget. (Rask 1816:267-268)

Rask var lingvist snarere end filolog, dvs. ikke mindre interesseret i selve sproget end i oldislandsk litteratur, og det har uden tvivl påvirket hans syn.

Ordbøger med islandsk som målsprog var først og fremmest tænkt som hjælpemidler for islandske brugere og hovedformålet var at fremme forståelse af tekster på kildesproget (jf. Jónas Jónasson 1896:iv). Dette genspejles i Gunnlaugur Oddsons ordbog som fokuserer på ”fremmede og sjældne ord, som danske forfattere har lånt fra andre sprog, og de danske ord, som mindst ligner islandsk” (1819:[iv]), som det udtrykkes i forordet. Som kilde til sin lemmaliste har han således først og fremmest anvendt tidens danske fremmedordbøger (Jón Hilmar Jónsson 1991:xii-xvi). Man har også til dels tænkt på produktion, ikke mindst oversættelse fra kildesproget, og i enkelte tilfælde er der taget hensyn til talesproget ved at vise udtale af enkelte ord (Gunnlaugur Oddsson 1819) eller opslagsordene alment (Zoëga 1896).

Björn Jónssons retskrivningsordbog fra 1900 havde et andet formål og en anden karakter end de øvrige værker, dvs. at sætte ortografiske retningslinjer og til dels også at give anden sproglig vejledning til modernmålsbrugere.

3. Ordbøgernes rolle i sprogstandardiseringen

Ordbøger fokuserer på ordforrådet og deres eventuelle påvirkning på den moderne islandske sprogstandard er derfor knyttet til ordniveaue. De kan være med til at kodificere bestemte ortografiske og strukturelle varianter af ordene og de kan spille en rolle i implementering af disse varianter hvis de når en tilstrækkelig distribution blandt almenheden. I mange tilfælde er kodificering af bestemte varianter begrænset til opslagsordet (hvis islandsk er kildesprog) eller til ækvi-valenternes grundform (hvis islandsk er målsprog). Den kan desuden dække op-

slagsordets ordklassetilhørighed, kanoniske bøjningsformer, køn af substantiver og kasusstyring af verber hvis ordbogsartiklerne indeholder sådanne oplysninger. Hvis der vises fx ordforbindelser og sprogbrugseksempler giver de i tilæg indirekte morfologiske og syntaktiske oplysninger som kan bidrage til kodificering af bestemte varianter.

Med hensyn til sprogstandardisering er ordbøger, som fokuserer på samtidssproget og på modernmålsbrugere, af størst interesse. Kun to ordbøger med islandsk som kildesprog opfylder mindst én af disse forudsætninger, dvs. Björn Halldórssons ordbog (1814), baseret på omgangssproget i den sidste halvdel af 1700-tallet, og Jón Þorkelssons supplement (1890-1897), baseret bl.a. på skriftsprogskilder fra 1800-tallet. Derimod gælder det alle ordbøgerne med islandsk som målsprog, samt Björn Jónssons retskrivningsordbog (1900).

Ortografien er stort set normaliseret i hvert enkelt værk. Rask har tydeligvis valgt den ortografiske norm i udgaven af Björn Halldórssons ordbog, fordi den stort set svarer til fx ortografien i den islandske litteraturforenings skrifter, hvor Rask var et indflydelsesrigt medlem i de første år efter dens stiftelse 1816 (Jón Aðalsteinn Jónsson 1959:77). Ifølge forordet til genudgivelsen af ordbogen viser en sammenligning med forfatterens manuskript at Rask og hans medarbejdere i mange tilfælde har ændret hans ortografi, fx brugte han den moderne *-ur* endelse i nominativ af maskuliner, mens den i udgaven er blevet ændret til varianten *-r* som i oldsproget (1992:xxiv). Jón Þorkelsson følger derimod en anden norm, som står nærmere den moderne officielle ortografi, fx brugte han endelsen *-ur*. I modsætning til de andre ordbøger havde retskrivningsordbogen (Björn Jónsson 1900) et klart formål og målgruppe som en præskriptiv vejledning for modernmålsbrugere. Ordbogen var udgivet af journalistforeningen og præsenterede den ortografiske norm som var blevet udviklet og accepteret af foreningens medlemmer, selv om den ikke var alment vedtaget. 1800-tallets debat om ortografi reflekteres således i ordbøgerne, hvor leksikograferne kodificerer hver sin udvalgte norm, men ingen af dem svarer præcis til den standard som til sidst blev vedtaget tidligt i 1900-tallet.

Det er vanskeligere at se klare linjer mht. ordforråd og ordformer. Lemmabestanden bestemmedes hovedsagelig af de kilder som brugtes, men der viser sig en tendens til at undgå fremmedord, især fra senere århundreder. I forordet til et af Jón Þorkelssons ordbogssupplementer (om ældre ordforråd, ca. 1200-1700) siger han bl.a.:

det [har] ikke kunnet være min Hensigt at optage lutter klassiske eller mønstergyldige og efterlignelsesværdige Ord. Dog har jeg ikke optaget alle de Ord, som findes i de Bøger fra det 16de og 17de Aarhundrede, som jeg har brugt ved Udarbejdelsen af Ordsamlingen. I disse Bøger findes nemlig en Del Ord, som paa Grund af deres Oprindelse og Dannelse ere

saa fremmede for Sproget, at de ikke kunne betragtes som islandske og desuden aldrig have vundet Borgerret i Islandsk. (Jón Þorkelsson 1879-1885:vii)

Det samme gælder valg af ækvivalenter i ordbøger med islandsk som målsprog, ikke mindst i Konráð Gíslasons dansk-islandske ordbog, hvor han i forordet (1851:iv) undskylder brugen af ord som kan anses som ikke ”full-íslenzk” (’helt islandske’), fx *planta* som han rettfærdigør med at stammen allerede findes i oldsproget. Han ser også henvisninger til og eksempler fra oldislandske kilder som en stor fordel ved sin ordbog. Selv i retskrivningsordbogen gør purismen sig gældende i et appendiks, ”Nokkur mállýti” (’Nogle sprogfejl’; 3 sider), med eksempler på sprogbrug som burde undgås. Det drejer sig hovedsagelig om formodede danismer, med forslag om en anden formulering, fx *mögulegur* (d. *mulig*; = *fær*, *tækur*...), *æfa* (d. *øve*; = *iðka*, *temja sér*) og *teikning* (d. *tegning*; = *dráttlist*), som alle er almindelige i nutidssproget mens andre, fx *eiginlegleiki* og *tilsvara*, er forældede. I lemmalisten og i enkelte artikler opføres der endvidere ord, ordformer og ordforbindelser som anses som ”ukorrekte eller mindre rigtige” markerede som sådanne med efterfølgende forslag om det rigtige form, fx ortografien *atla* (tilsvarende udtale med monoftong) i stedet for det korrekte *ætla* (med diftongen [ai]).

Ordbøgerne genspejler således vigtige træk af den sprogdebat og standardisering som var i gang i 1800-tallet, nemlig forskellige synspunkter og forslag til ortografisk standardisering, sproglig purisme og vægt på det oldislandske forbillede.

4. Konklusioner

Min konklusion er at 1800-tallets ordbøger over samtidssproget har været vigtige for udvikling og kodificering af den moderne sprogstandard, især ortografien selv om en officiel retskrivningsstandard først blev vedtaget i 1918. Vægten på oldislandsk i periodens ordbøger, både på grund af målgruppens forventninger og leksikografernes baggrund som filologer og udgivere af oldislandske tekster, har sikkert bidraget til at oldsproget blev en så vigtig model for den moderne sprogstandard som var under udvikling.

Det er mindre sikkert hvor stor påvirkning ordbøgerne har haft på udbredelse og acceptering af standarden blandt almene islandske sprogbrugere, eftersom den primære målgruppe for de fleste ordbøger med islandsk som kildeprog var udlændinge. Værkerne var dog kendt og brugt i Island, men sandsynligvis var brugen for det meste begrænset til akademikere.

Det leksikografiske arbejde kan alligevel have haft en indirekte påvirkning på udbredelsen af den standard som ordbøgernes redaktører var med til at udvikle og kodificere i værkerne. Mange af dem som deltog i arbejdet var aktive i tidens sprogdebat og har givetvis selv praktiseret deres standardiseringsforslag i deres egne skrifter. Debatten førtes ikke mindst i tidsskrifter og aviser som havde stor udbredelse blandt almenheden. Den kan således have påvirket den almene sprogbrug, både debattørernes synspunkter og argumenter om ideal sprogbrug og deres egen sprogbrug som en eventuel model for andre.

Litteratur

Ordbøger

Björn Halldórsson (1814 [1992]): *Lexicon Islandico-Latino-Danicum Björnonis Haldorsonii*, I-II. Rasmus K. Rask (red.). Havniæ. [Ny udg. (1992): *Orðabók. Íslensk, latnesk, dönsk*. Red.: Jón Aðalsteinn Jónsson. (Orðfræðirit fyrir alda 2.) Reykjavík: Orðabók Háskólans.]

Björn Jónsson (1900): *Íslenzk stafsetningarorðabók*. Reykjavík.

Cleasby, Richard & Guðbrandur Vigfusson (1874): *An Icelandic-English Dictionary*. Based on the ms. collections of the late Richard Cleasby; enlarged and completed by Gudbrand Vigfusson. Oxford: The Clarendon Press.

Eiríkur Jónsson (1863): *Oldnordisk Ordbog*. København: Det kongelige nordiske Oldskrift-Selskab.

Gunnlaugur Oddsson (1819 [1991]): *Orðabók, sem inniheldr flest fágæt og vandskilinn ord, er verða fyrir í dönskum bókum*. København. [Ny udg. (1991): *Orðabók sem inniheldur flest fágæt, framandi og vandskilin orð er verða fyrir í dönskum bókum*. Red.: Jón Hilmar Jónsson & Þórdís Úlfarsdóttir. (Orðfræðirit fyrir alda 1.) Reykjavík: Orðabók Háskólans.]

Jón Þorkelsson (1876-1899): *Supplement til islandske Ordbøger*. Første og Andet Samling. Reykjavík 1876; Anden Samling. Ny Udgave. København 1895; Tredje Samling, I-II. Reykjavík 1890-1897; Fjerde Samling. Reykjavík 1899.

Jónas Jónasson (1896): *Ný dönsk orðabók með íslenzkum þýðingum*. Reykjavík.

Konráð Gíslason (1851): *Dönsk orðabók með íslenzkum þýðingum*. København.

Sveinbjörn Egilsson (1860): *Lexicon poeticum antiquæ linguæ septentrionalis*.

København: Societas regia antiquariorum septentrionalium.

Zoëga, Geir T. (1896): *Ensk-íslenzk orðabók*. Reykjavík.

Anden litteratur

- Deumert, Ana & Wim Vandenbussche (2003): Standard languages: Taxonomies and histories. I: Deumert & Vandenbussche (red.): *Germanic Standardizations. Past to present*. Amsterdam/Philadelphia: John Benjamins, 1-14.
- Haugen, Einar (1966): Dialect, Language, Nation. I: *American Anthropologist* 68(4), 922-935.
- Jón Aðalsteinn Jónsson (1959): Ágrip af sögu íslenzkrar stafsetningar. I: *Íslenzk tunga* 1, 71-119.
- Jón Hilmar Jónsson (1991): Um orðabók Gunnlaugs Oddssonar. I: Gunnlaugur Oddsson, *Orðabók sem inniheldur flest fágæt, framandi og vandskilin orð er verða fyrir í dönskum bókum* [Ny udg.] Reykjavík: Orðabók Háskólans, ix-xxxvii.
- Jón Ólafsson (1904): Jón Þorkelsson Dr. phil., R. Dbr. I: *Andvari* 29, 1-13.
- Rask, R.K. (1816): Om Bjørn Haldorsens Islandske Ordbog. I: *Dansk Litteratur-Tidende for Aaret 1816*. No. 17, 265-270.
- SOOÍ = *Skrá yfir orðabækur og orðasöfn sem varða íslensku*. Stofnun Árna Magnússonar í íslenskum fræðum. <www.arnastofnun.is/is/skra-yfir-ordabækur-og-ordasofn-sem-varda-islensku> (16.7.2019).

Ásta Svavarsdóttir
 forskningsdocent
 Árni Magnússon-instituttet for islandske studier
 Laugavegi 13
 IS-101 Reykjavík
 asta.svavarsdottir@arnastofnun.is

Superordbog og salmesang: ordbogslinkning i praksis

Mette-Marie Møller Svendsen, Nicolai Hartvig Sørensen & Thomas Troelsgård

Music and Language in Danish Hymns from the Reformation is a cross-disciplinary project based at the Society for Danish Language and Literature. Part of the project is the compilation of a “super-dictionary” consisting of a total of 12 digitised and interlinked dictionaries and glossaries relevant to the language used in the hymnbooks from the Reformation. This article provides a brief outline of the practical steps taken in the process of linking our lexical data, such as how the lemma equivalents are chosen, a few examples of how the linking is carried out, as well as a presentation of the tools used in the process.

1. Om projektet

Projektet *Musik og sprog i reformationstidens danske salmesang* er et tværvindskabeligt projekt hjemmehørende på Det Danske Sprog- og Litteraturselskab (DSL). Som en del af projektet udvikles en ”superordbog” for renæssancens sprog, som bliver tilgængelig som en ordbog i DSL’s ordbogsplatform der i forvejen inkluderer ni moderne og historiske ordbøger for dansk. Denne ordbog vil blive benyttet som ressource af den internetportal som udvikles i projektet, og som skal præsentere salmebøgerne (se nedenfor). Superordbogen kommer til at rumme i alt 12 digitaliserede ordbøger og gloselister som er relevante i forhold til sproget i reformationstidens salmebøger og generelt for perioden. Ordbogsværkerne omfatter dansk-dansk, dansk-latin, latin-dansk og latin-dansk-græsk. Projektet er støttet af Carlsbergfondet og Velux Fonden og forventes afsluttet i 2021.

Superordbogen udnytter at DSL er i gang med at linke alle Selskabets ordbøger sammen på lemmaniveau i en overordnet ”metaordbog” og alle superordbogens links udgør i praksis en delmængde af denne metaordbog.

På samme måde som den norske metaordbog giver et overblik over samtlige opslag på et ord fra en række forskellige materialer (Ore 2001), så giver superordbogen også mulighed for at skabe et indblik i sprogets udvikling ved hjælp af en mængde danske resurser. I øjeblikket er der stort internationalt fokus på de muligheder som linkning af forskellige leksikalske resurser tilbyder, både for leksikografer og brugere, som det fx kan ses hos Hudeček & Mihaljević (2019:74).

Internetportalen danner udgangspunkt for tre tværfaglige forskningsprojekter:

- Kontinuitet og nybrud i reformationstidens salmesang: I dette projekt fokuseres der på forholdet mellem middelalderligt og nyt melodistof, mellem latinske og danske tekster og mellem det gammeldanske ordforråd ift. nyskabelserne i de lutherske salmer.
- Varians i melodier og tekster: I dette projekt undersøges en eventuel korrelation mellem musikalske og litterære variationer, der kan tolkes som udtryk for forskellige opførelsespraksisser.
- Spredning af kirkens melodier og tekster: I dette projekt undersøges spredningen til den verdslige sang, specifikt genbrug af salmemelodier og salmernes indflydelse på visernes ordforråd og fraseologi.

Den avancerede internetportal kommer til at bidrage med nye værktøjer og materialer til forskningsprojekterne, og de tre tværfaglige projekter forventes at blive påbegyndt i 2020. Superordbogen bliver tilgængelig som en selvstændig ordbogsdel på portalen, hvor man kan slå ord op samlet i de relevante ordbøger. Desuden bliver den integreret i portalens tekstpræsentation, så man kan slå ord op direkte fra centrale tekster.

Tabel 1 viser de ordbøger der direkte eller indirekte inddrages ved opbygningen af superordbogen. I denne artikel gennemgås arbejdet med linkningen i praksis.

Udgiver	Titel	Årstal	Omfang
Christiern Pedersen	<i>Vocabularium ad usum dacorum</i>	1510	13.000
Henrik Smith	<i>Hortulus synonymorum</i>	1520	1.000
Jon Tursen	<i>Vocabularius rerum</i>	1561	8.000
Henrik Smith	<i>Libellus Vocum Latinarum</i>	1563	6.200
Poul Nielsen Hingelberg	<i>Vocabulorum variorum expositio</i>	1576	3.500
Mads Pors	<i>De nomenclaturis Romanis</i>	1594	3.200
Poul Jensen Colding	<i>Dictionarium Herlovianum</i>	1626	30.000
Otto Kalkar	<i>Ordbog til det ældre danske sprog (2. udg. 1976)</i>	1881-1918	76.000
Jensen & Goldschmidt	<i>Latinsk-Dansk Ordbog (2. udg. 1920)</i>	1886	19.000
DSL	<i>Ordbog over det danske Sprog (ODS)</i>	1918-56	181.000
Kristian Sandfeld	Ordbog til <i>Hans Mogensens Oversættelse af Philippe de Commines Memoirer</i>	1919	5.500
Kristian Sandfeld	Ordbog til <i>En Ræffue Bog, Herman Weigeres Oversættelse af Reinke de Vos</i>	1923	6.100
Marius Kristensen	Ordbog til <i>Danske Viser fra Adelsvisebøger og Flyveblade 1530-1630</i>	1931	7.100
DSL	<i>Supplement til ODS (ODS-S)</i>	1992-2005	40.000
D. Duncker m.fl. (Københavns Univ.)	Ordlister udarbejdet i forb. med projektet <i>Dansk Folkevisekultur 1550-1700</i>	2001	4.500
DSL	<i>Den Danske Ordbog (DDO)</i>	2009-	100.000
DSL	<i>Gammeldansk Ordbog (GO)</i>	2015-	30.000
<i>Grøn baggrund = Ordbøger med dansk opslagsord</i>			
<i>Blå baggrund = Ordbøger med latinsk opslagsord</i>			
<i>Grå baggrund = Ordbøger udenfor projektet, der fungerer som "rygrad" for linkningen.</i>			

Tabel 1: Oversigt over de forskellige ordbøger der indgår i linkningen.

2. Ordbogslinkning: metoder og værktøjer

Artiklerne i de projektrelevante ordbøger linkes på lemmaniveau i superordbogen. Udgangspunktet for, eller rygraden i, linkningen er Den Danske Ordbog (DDO) og Ordbog over det danske Sprog (ODS), samt i mindre omfang supplementet til Ordbog over det danske Sprog (ODS-S) og Gammeldansk Ordbog (GO). For vokabularerne med latinske opslagsord opbygger vi forsøgsvis en mindre, latinsk superordbog, som en parallel til den danske, med Jensen & Goldschmidt fra 1920 (latin-dansk, 1. udgave udkom i 1886) som udgangspunkt.

Etymologi er det primære kriterium, og artikler med fælles etymologi knyttes sammen, også selvom der er stor variation i opslagsformerne; således samles fx *blugsel*, *blussel*, *blyghsel* og *blysel* alle i superartiklen *blusel*. Tilsvarende placeres *kamp* ('kampesten') og *kamp* ('krig, tvekamp') i hver sin superartikel.

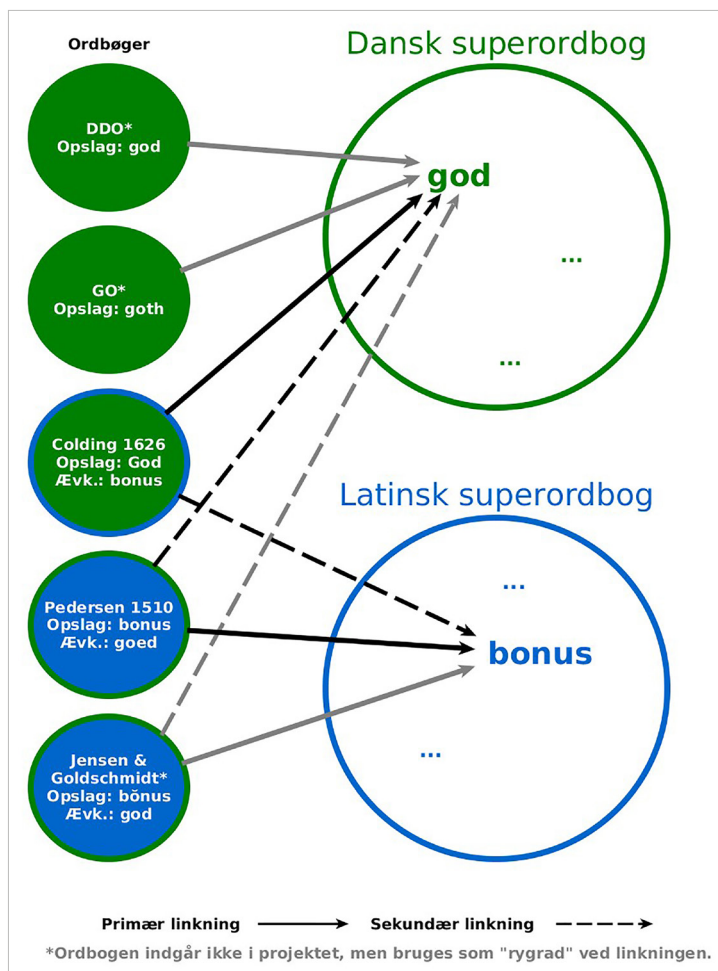
Hvis tiden tillader det, vil vi desuden forsøge at foretage en sekundær linkning, hvor artikler fra de bilingvale ordbøger via ækvivalenterne linkes til matchende artikler i målsproget.

Figur 1 viser et eksempel på linkning i form af lemmanet 'god' for hhv. dansk og latin. Her kan det ses hvordan DDO og GO fungerer som rygrad for linkningen, idet de udgør den knude som opslaget i *Dictionarium Herlovianum* (Col-ding 1626) linkes til.

Ved arbejdet med at etablere links til brug for superordbogen benyttes tre former for linkning af de leksikalske data: fuldautomatisk, halvautomatisk og manuel linkning. En redegørelse af de forskellige linkningsformer følger her.

2.1. Fuldautomatisk linkning

Artikler hvor der er match på form og ordklasse og hvor der ikke foreligger homografi, har vi valgt at linke fuldautomatisk med de potentielle fejl det kan medføre. Ord med datering i nyere tid fjernes automatisk i denne proces, så vi undgår at fx det gammeldanske substantiv *biker* 'bæger', linkes til DDO's *biker* 'motorcyklist'.



Figur 1: Figuren viser linkningsmuligheder for lemmaet 'god'.

2.2. Halvautomatisk linkning

For artikler hvor der ved inddragelse af flere kriterier end den rene opslagsform automatisk kan lokaliseres et eller flere mulige mål for linkningen, gennemføres en halvautomatisk linkning. Ved hjælp af et specialudviklet værktøj, navngivet *Konnektor*, kan redaktøren manuelt vælge og linke den eller de relevante målartikler. Kriterierne der bruges ved lokalisering af mulige målkandidater, omfatter:

- match på opslagsform og ordklasse (fx homografer)
- match på variantform der optræder i ordbogsartiklen

- match på genereret variantform, fx aa/å, x/ks, æ/e, ju/iu/y
- match på artikelindholdet generelt, baseret på strenglighed mellem tekstekstrakter fra kilde- og målartikel.

Artikler der skønnes ikke at have nogen matchende målartikel, gives et særligt mærke under linkningen i linkningsværktøjet, hvorefter de lægges ind som nye selvstændige lemmer, så vidt muligt normaliseret til moderne dansk for at lette fremtidig linkning. Tilsvarende sættes et særligt mærke ved problematiske artikler, hvilket typisk drejer sig om artikler hvor redaktøren ikke har et sikkert bud på linkningen. Endelig er der i værktøjet mulighed for manuelt at slå op blandt målartiklerne, og ad den vej linke manuelt til en målartikel der ikke er blandt de automatisk fundne kandidater.

2.3. Om Konnektor

Den halvautomatiske del af linkningen udføres ved hjælp af det specialudviklede værktøj Konnektor. Programmet blev oprindeligt udviklet til lemmatisering af ord i en række TEI-opmærkede middelaldertekster; til hvert ord blev vist en række mulige lemmakandidater, fundet ved at matche formen mod diverse ordlister vedr. perioden under prioriteret inddragelse af bøjningsparadigmer, lydregler, strenglighed m.m. Siden har værktøjet fået flere funktionaliteter, og det er blevet udvidet til at kunne arbejde på forskellige typer af input, så længe et simpelt XML-format overholdes. Således kan alle opgaver der består i at vælge 0, 1 eller flere mål til et givet udgangspunkt udføres i Konnektor, og programmet bruges i dag i flere DSL-projekter, både af sproglig, historisk og litterær art.

Tre eksempler på typiske situationer der opstår ved halvautomatisk linkning ved brug af Konnektor følger her.

• **Ét match fundet**

Kildeartiklen i Kalkars Ordbog er *uge*, sb.: ”samme som islandsk vika”. Konnektor finder match på form og ordklasse, og kun én mulig målartikel, nemlig *uge*, sb. Linkningen gennemføres ved at acceptere mulighed 1:

- 1) *uge*, sb. /DDO: 1. periode som strækker sig over syv på hinanden følgende [dage].

- **Flere matches fundet**

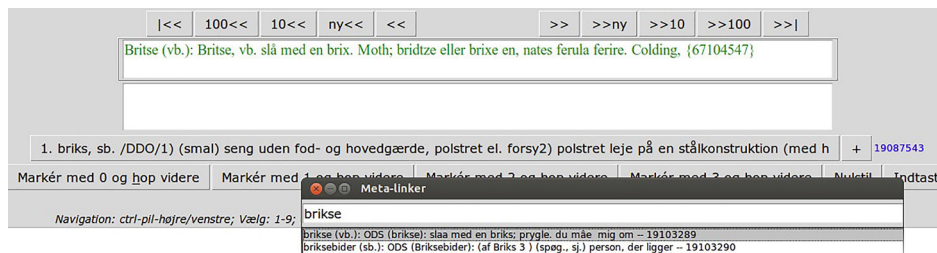
Kildeartiklen i Kalkars Ordbog er *skræppe*, sb. med betydningen 'taske'. Der er match på både form og ordklasse ved de første tre bud fra Konnektor, her findes dog flere mulige målar artikler:

- 1) *skræppe*, sb. /ODS: person, der skræpper op, taler meget og er støjende
- 2) *skræppe*, sb. /DDO: op til 1 m høj (flerårig) ukrudtsplante
- 3) *skræppe*, sb. /ODS: vadsæk, ransel olgn. ell (nu især) madkasse
- 4) *skræbe*, vb. om mennesker: græde (stærkt og vedholdende)
- 5) *skræppe*, vb. /DDO: give høje, skingre og gennemtrængende skrig

Mulighed 3 vælges, da kildeartiklens "taske" svarer til 'vadsæk, ransel' i målar-tiklen.

- **Ingen matches fundet**

Kildeartiklen i Kalkars Ordbog er *britse*, vb.: 'slå med en brix', se figur 2. Der foreslås ingen korrekt målarartikel i Konnektor, som kun foreslår *briks*, sb: '(smal) seng uden fod- og hovedgærde'. Det mistænkes at varianten *brikse* kan findes; *brikse* findes via et manuelt opslag i Konnektor, og der linkes til den korrekte artikel.



Figur 2: Halvautomatisk linkning af lemmaet *britse*, vb. i linkningsværktøjet Konnektor.

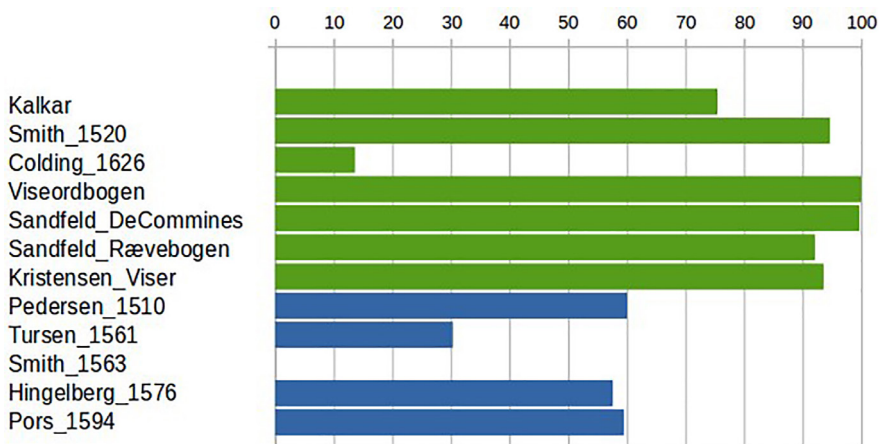
2.4. Manuel linkning

Artikler der ikke kan matches ved automatisk eller halvautomatisk linkning, samt artikler for hvilke der ikke er fundet nogle forslag til match, udskydes til manuel linkning. Denne linkning foretages som en senere "anden runde" af en kollega med speciale i 1500-tallets danske sprog. I praksis foregår linkningen som en fornyet kørsel i Konnektor, med den forskel at de fleste af disse problematiske

artikler først analyseres, fx ved kontrolopslag i andre værker, og derefter enten linkes manuelt til en eksisterende målartikel, eller markeres som nyt lemma. Et eksempel kunne være *tog* fra Kalkars Ordbog, i den artikel der begynder med: ”Tog, no.; i (på) t. (svf på tok) o: i blinde, løst hen”. Her bliver redaktøren præsenteret for ti forslag til match, hvoraf ingen giver mening, og frem for at bruge tid på analyse og kontrolopslag, sendes artiklen videre til en senere manuel linkningsrunde.

3. Status på projektet

Linkningen er i gang for de fleste ordbøger, og figur 3 viser status for linkningen i oktober 2019. I en del af de ældre ordbøger er lemmatisering og artikelstruktur inkonsekvent, og en større del af artiklerne har derfor ikke kunnet linkes under de hel- og halvautomatiske linkningsrunder.



Figur 3: Oversigt over status i procent for linkningen på projektet i oktober 2019.

4. Perspektiver

I og med at flere større ordbogsressurser – herunder DDO og ODS – indgår i den overordnede danske metaordbog, er grunden lagt for en bedre udnyttelse af ordbogsdata i DSL’s onlineresurser. Hvor opslag på tværs af ressourcer hidtil primært har været helt strengbaseret, kan nu også de linkede artikler (med afvigende opslagsformer) inddrages. Linkningen af ordbogsressurserne giver desuden mulighed for at levere nye typer af dataudtræk til brug i sprogteknologiske projekter,

hvor data fra forskellige resurser kombineres. Dette er allerede aktuelt, idet DSL deltager i det fælleseuropæiske projekt ELEXIS, bl.a. med leverance af data til forsøg med automatisk matching af betydningsbeskrivelser på tværs af sprog og perioder.

For tiden kan der kun søges enkeltvis i DSL's ordbogsressurser eller via ordbogsportaler som fx sproget.dk. En oplagt mulighed for videre udnyttelse af linkningen i fremtiden vil være opbygningen af en ny, samlet og integreret ordbogsressource for alle DSL's ordbøger. Ligesom den nye superordbog kan hjælpe brugere med interesse i reformationstidens sprog, vil lignende ordbøger med fokus på andre perioder kunne være til gavn for alle der arbejder med det ældre danske sprog.

Litteratur

Ordbøger m.m.

Colding, Poul J. (1626): *Dictionarium Herlovianum*. København: Salomon Sartor.

DDO = *Den Danske Ordbog*. Onlineversion: <www.ordnet.dk/ddo> (oktober 2019).

Duncker, Dorthe (1999): Visernes vej: Sammenhæng mellem visebøger? I: Flemming Lundgreen-Nielsen & Hanne Ruus (red.): *Svøbt i mår: dansk folkevisekultur 1550-1700*. København: C.A. Reitzel.

ELEXIS: European lexicographic infrastructure <www.elex.is>.

GO = *Gammeldansk Ordbog*. Onlineversion: <www.gammeldanskordbog.dk> (oktober 2019).

Hingelberg, Poul Nielsen (1576): *Vocabulorum variorum expositio*. København: Mads Vingaard.

Jensen, J. Th. & M. J. Goldschmidt (1920): *Latinsk-Dansk Ordbog, 2. udgave*. Onlineversion: <www.latinskordbog.dk> (oktober 2019).

Kalkar, Otto (1976): *Ordbog til det ældre danske sprog (1300-1700), 2. udgave, bind I-VI*. København: Universitets-Jubilæets danske Samfund, Akademisk Forlag. Onlineversion: <www.kalkarsordbog.dk> (oktober 2019).

Kristensen, Marius (1931/1978-79): *Ordbog til Danske Viser fra Adelsvisebøger og Flyveblade 1530-1630*. København: DSL, Gyldendal/C.A. Reitzel.

ODS = *Ordbog over det danske Sprog*. Onlineversion: <www.ordnet.dk/ods> (oktober 2019).

ODS-S = *Supplement til ODS*. Onlineversion: <www.ordnet.dk/ods> (oktober 2019).

Pedersen, Christiern (1510): *Vocabularium ad usum dacorum*. Paris: anon.

- Pors, Mads (1594): *De nomenclaturis Romanis*. Frankfurt a. M.: Joh. Feyerabend.
- Sandfeld, Kristian (1919): *Ordbog til Hans Mogensens Oversættelse af Philippe de Commynes Memoirer*. København: DSL, Gyldendal.
- Sandfeld, Kristian (1923): *Ordbog til En Ræffue Bog, Herman Weigeres Oversættelse af Reinke de Vos*. København: DSL, Gyldendal.
- Smith, Henrik (1520): *Hortulus synonymorum*. København: Melchior Blumme.
- Smith, Henrik (1563): *Libellus Vocum Latinarum*. København: Hans Zimmermann.
- Tursen, Jon (1561): *Vocabularius rerum*. København: Lorenz Benedicht.

Anden litteratur

- Grønvik, Oddrun & Christian-Emil Smith Ore (2018): Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforskning av ordtilfang. I: *Nordiske studier i leksikografi* 14. Reykjavik: Nordisk forening for leksikografi, s. 87-95.
- Hudeček, Lana & Milica Mihaljević (2019): Croatian Web Dictionary – Mrežnik – Linking with Other Language Resources. I: Kosem, I., T. Zingano Kuhn, M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubiček, S. Krek & C. Tiberius (red.): *Electronic lexicography in the 21st century*. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., s. 72-98.
- Ore, Christian-Emil (2001): Metaordboken – et rammeværk for Norsk Ordbok. I: *Nordiske studier i leksikografi* 5. Gøteborg: Nordisk forening for leksikografi, s. 250-270.

Mette-Marie Møller Svendsen
assisterende redaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
mettemariesvendsen@gmail.com

Nicolai Hartvig Sørensen
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
nhs@dsl.dk

Thomas Troelsgård
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
tt@dsl.dk

Tilblivelsen af fagordbøger og termsamlinger i Island

Ágústa Þorbergsdóttir

The aim of this article is to present the circumstances behind the compilation of dictionaries and databases for specialist and technical terms in Iceland, and to explain what parties have been involved in initiating this work. Also discussed is whether these resources were created for specific user groups, and if so, how these resources might fulfil the particular needs of these given user groups. The corpus of published specialist dictionaries and databases was surveyed and the motivation behind the creation of each resource was identified. The conclusions of this research show that there are various types of databases containing specialist terms in Icelandic and that the reasons for their existence can be diverse. Most prominently, it is seen that specific user groups were targeted when technical terms were compiled in conjunction with teaching. The conclusions also show that it is common for motivations behind the development of resources of this type to overlap and therefore that it can be difficult to determine what the primary or secondary motivation might be in each case.

1. Indledning

Denne artikel omhandler tilblivelsen af fagordbøger og termsamlinger i Island. Dette emne er ikke blevet undersøgt tilstrækkeligt, og derfor er formålet med artiklen at opnå en bedre forståelse og nye erkendelser på det islandske fagsprogsområde.

Mit hovedspørgsmål er: Hvorfor er fagordbøger og termsamlinger blevet publiceret i Island? For at kunne besvare dette spørgsmål opstilles tre underspørgsmål:

- Hvem tager initiativ til at udarbejde fagordbøger og termsamlinger?
- Hvilke motiver ligger bag tilblivelsen af disse ressourcer?
- Rettes fagordbøger og termsamlinger mod en speciel brugergruppe?

For at besvare spørgsmålene har jeg anvendt den metode at gennemgå publicerede islandske fagordbøger, læse indledningskapitlerne i disse og eventuelle artikler om dem. Endvidere har jeg undersøgt om man kan finde motiv ud fra selve fagordsressourcerne. Jeg benyttede mig af et register der omfatter 50 fag-

ordssamlinger fra artiklen *Íslenskar íðorðaskrár* (Baldur Jónsson 1987:21-24) som dækker 60 år (1927-1987). Angående fagsprogsmateriale publiceret efter 1987 undersøgte jeg især trykte fagordbøger, f.eks. en publikationsserie fra Islandske sprogævn,¹ og jeg har også gennemgået de godt 60 termsamlinger som findes i den islandske termbank (idord.arnastofnun.is).

2. Hvem tager initiativ til at udarbejde fagordbøger og termsamlinger?

I Island er det almindeligt at såkaldte termudvalg tager sig af termarbejde, og mange sådanne udvalg er blevet oprettet indenfor de sidste årtier. Ligeledes er det ret almindeligt at enkeltpersoner har taget initiativ til at udarbejde termsamlinger, hvor i mod termsamlinger på myndighedernes initiativ er forholdsvis få.

2.1. Termudvalg og individer

Allerede i 1890'erne foreslog digteren Einar Benediktsson at man oprettede et udvalg som havde den opgave at danne nye ord af islandske ordstammer. Tanken var at hvis mange personer hver for sig danner forskellige ord over det samme, er det usikkert om nogen af disse ord etablerer sig i sproget. (Einar Benediktsson 1891:56-57). Det har længe været en udbredt holdning i Island, at det er hensigtsmæssigt at en gruppe mennesker arbejder sammen om terminologien inden for de forskellige fag.

I henhold til det blev det første af mange udvalg oprettet af Ingeniørforeningen 1919. Denne ordning eksisterer stadig, og det kan nævnes at det nyeste af den slags udvalg, som er inden for fagområdet historie, blev oprettet i 2019.

Initiativ til at udarbejde fagordbøger og termsamlinger tages først og fremmest af disse termudvalg, og i Island har over 50 termudvalg eksisteret i længere eller kortere tid, og de kan være af forskellig slags. Det mest almindelige er at de personer der vælges til termudvalgene er nogle som man regner med at kunne have indflydelse på sprogbrug og termforrådets udvikling indenfor det pågældende område.

Det er heller ikke ualmindeligt at specialister, universitetslærere eller oversættere udarbejder termsamlinger på eget initiativ, og ofte er det i forbindelse med deres arbejde (jfr. Sigrún Helgadóttir & Kristján Árnason 1991:8-10). Her kan som eksempel nævnes en termsamling i maskinteknik som lærere inden for området har udarbejdet primært for deres studerende (SJÓ-1998:3).

¹ www.islenskan.is/um-islenska-malnefnd/utgafur.

2.2. Myndighederne

Da de første udvalg blev oprettet, fandtes der ingen institutioner i Island, der havde kapacitet til at tage sig af termarbejde. Tiltaget måtte derfor komme fra individer, og man kan derfor sige, at termudvalg og individer har sat kursen, hvad termarbejdet angår (jfr. Sigurður Jónsson 2007:158, 162).

Myndighedernes involvering af et såkaldt neologiarbejde² begyndte i 1950'erne, da Altinget (parlamentet) uddelte midler til dette arbejde. Som resultat af arbejdet udkom fire neologismehæfter (Nýyrði I-IV). De indeholdt først og fremmest fagord knyttet til søfart, landbrug og luftfart. Undervisningsministeriet publicerede disse hæfter og derudover en teknisk ordbog (Tækniordasafn), der udkom i 1959.

De islandske myndigheder oprettede i 1964 det islandske sprognævn der bl.a. havde til opgave at registrere nye ord (termer). Sprognævnet varetager ikke længere termarbejde, men udgav tidligere nogle fagordbøger i samarbejde med vedkommende termudvalg.

Árni Magnússon-instituttet for islandske studier er center for termarbejde i Island og ifølge loven om instituttet (nr. 40/2006), skal det have samarbejde med termudvalg og andre som arbejder med fagspecifik terminologi. ÁM-instituttet administrerer en termbank (Íðorðabankinn) hvor de fleste islandske termsamlinger er publiceret, deriblandt termsamlinger fra termudvalg.

Undtagelsesvis har myndighederne taget initiativ til at udarbejde termsamlinger. Ét eksempel på et sådant initiativ fra myndighederne er en fagordbog om luftfart (*Flugorðasafn*) som blev udgivet i kølvandet på en diskussion der startede i 1985 om at det engelske sprog prægede i stort omfang diskussionen om luftfart.

3. Motivation

Jeg har gennemgået publicerede islandske fagordbøger og termsamlinger og forsøgt at analysere hvilken motivation ligger bag deres tilblivelse. Der findes ikke altid et indledningskapitel eller præcise informationer i disse værker, hvor man f.eks. kan få tydelige oplysninger om redaktørernes hensigt. Hensigten kan dog analyseres ud fra andre faktorer, f.eks. hvordan emnet fremstilles. Det undersøges også om motivationen har ændret sig gennem tiden, og om de i nogle tilfælde muligvis drejer sig om flere end en motivation der knyttes sammen i en ordbog eller om motivationen er afgrænset inden for hver enkelt publikation.

² I Island blev der først efter 1980 skelnet mellem neologismer og termer. Indtil da blev det islandske ord *nýyrði* (nyord) brugt for såvel neologismer som termer.

3.1. Sprogrensning

Det ældste publicerede fagsprogsmateriale bærer præg af sprogrensning, men da disse blev udarbejdet, bekæmpede man aktivt dansk indflydelse på det islandske sprog. Der er tale om tosprogede ordlister uden definitioner hvor nye islandske ord er blevet foreslået i stedet for de danske ord som var udbredte i islandsk. Langt op i forrige århundrede førte man en sprogpolitik der gik ud på at friholde islandsk for påvirkning fra andre sprog, især dansk, men senere mest fra engelsk. Forestillingen om et rent sprog hvor neologismer, der er dannede af islandske ordstammer blev foretrukket frem for låneord, var almindelig. Det ovennævnte udvalg, som ingeniørforeningen oprettede, blev til at begynde med kaldt *málhreinsunarnefnd* (sprogrensningsudvalg), hvilket henviser til at det bekæmpede aktivt dansk indflydelse på islandsk og foreslog nye islandske ord i stedet for de danske ord som var udbredte i islandsk sprogbrug. Dette udvalg valgte senere selv at anvende betegnelsen *orðanefnd* (ordudvalg) (jfr. Guðmundur Finnbogason 1919:53-54). Sprogrensning er ikke længere drivkraft bag islandsk fagsprogsarbejde.

3.2. Udvidelse af ordforrådet

Nye begreber tilføjes konstant med samfundsudviklingen og nye disciplin indenfor videnskaben, og den hensigt er tydelig at ordforrådet ønskes udvidet med islandske ord (termer) frem for låneord.

Dette kan tydeligt ses i arbejdet med udvikling af ordforråd hos mange termudvalg som lægger stor vægt på at finde islandske ord for innovationer, og at forhindre brugen af låneord.

Sprogrensning (*málhreinsun*) er tæt beslægtet med sprogrøgt (*málrækt*), men der kan dog skelnes mellem de to ideologier. Sprogrensning er purisme, og i denne sammenhæng har den til formål at bevare islandsk som et rent sprog, og finde nye islandske ord i stedet for udenlandske ord.

Med sprogrøgt lægges der større vægt på at islandsk bør kunne anvendes på alle områder, men ikke kun at rense dansk eller engelsk fra det islandske sprog. Det er i overensstemmelse med islandsk sprogpolitik, hvor der lægges vægt på at bidrage til at der altid findes et ordforråd for at imødekomme nye forhold, og sikre så vidt som muligt at islandsk bliver brugt på så mange områder som muligt (Mennta- og menningarmálaráðuneytið 2009).

Som eksempler på termsamlinger, der har som hensigt at udvikle ordforrådet for at der kan skrives og diskuteres på fuldgyldigt islandsk kan her nævnes en IT-termsamling (TO-1983) og en miljøteknologisk termsamling (UÍO-2007). I indledningskapitlerne til disse termsamlinger (TO-1983:10; UÍO-2007:7) fremgår at målet er at brugerne kan anvende islandsk i disse fagområder.

3.3. Dokumentation af sprogbrug og lettilgængelige termer

Dokumentation af termer som anvendes på pågældende områder er undertiden formålet for termarbejdet. Specialiserede termer findes som regel ikke i almindelige ordbøger, og i indledningskapitler i mange fagordbøger og termsamlinger fremhæves betydningen af at islandske fagord findes registrerede. Det er f.eks. af stor betydning for studerende at kunne finde frem til islandske fagord og deres engelske ækvivalenter. I mange tilfælde har de islandske termer ingen lighed med de internationale ord, og derfor er det vigtigt at kunne slå dem op. Et eksempel på dette er en matematisk termsamling (EÍOS-1997). Det fremgår bl.a. af indledningen til denne samling at hvis termer ikke er tilgængelige vil følgen være upræcis brug af islandske termer, og derfor vil det være mere sikkert at bruge det udenlandske ord (EÍOS-1997:v).

3.4. Koordinering og effektivt sprogbrug

I indledningen til fagordbøger og termsamlinger nævnes konsekvent sprogbrug ofte som et formål. Her bestræber man at skabe sproglig klarhed på fagområder, og dermed at effektivisere kommunikationen. Holdningen til koordination er dog meget forskellig. Nogle redaktører er bevidste om at deres arbejde skal være normativt, og de lægger vægt på en terminologi hvor begrebsindhold er udtrykt ved definitioner, og at der er enighed om de foretrukne termer. Et eksempel på dette er den førnævnte miljøtekniske termsamling (UÍO-2007:7-10).

Disse to sidste motivationer, dokumentation af sprogbrug og koordinering, knyttes ofte sammen. Publikationen i sig selv gør termerne tilgængelige, på den måde fører den til koordinering, hvor i mod at hvis termerne ikke findes registreret nogen steder kunne konsekvensen være upræcise islandske termer eller at man anvendte udenlandske ord.

3.5. Undervisning og indlæring

Betingelsen for at alle fag kan undervises på islandsk er at der findes islandske termer. Det er grunden til tilblivelsen af nogle fagordbøger og termsamlinger. De skal være hensigtsmæssige hjælpemidler for dem der læser engelske fagbøger, men samtidig skal tilegne sig det islandske ordforråd. Her kan f.eks. nævnes en termsamling inden for maskinteknik der primært er beregnet til studerende, f.eks. med hensyn til valg af termer (SJÓ-1998).

I tæt forbindelse med undervisning som baggrund for termarbejde er indlæring. Herunder hører fagordbøger eller termsamlinger som er rettet mod specifikke læringsmiljøer. Der er tale om en slags definitionsordbøger, som kan give

de studerende og andre interesserede mulighed for at fordybe sig i det pågældende emne. Eksempler på islandske termsamlinger, som kan være en støtte for indlæring er termsamlinger i skriftserien *Íðorðarit Stofnunar Árna Magnússonar í íslenskum fræðum* (Ágústa Þorbergsdóttir 2018:262-270).

3.6. Forbindelse med andre projekter

Nogle termsamlinger er blevet udarbejdet i forbindelse med diverse nordiske eller internationale termprojekter, dvs. islandske termer bliver tilføjet til dem. I disse tilfælde drejer det sig om et organiseret termarbejde og en systematisk termdannelse inden for specifikke emneområder. Der dannes islandske ord som modsvarer fremmede ord på bestemte termlister, dvs. forudsætningen for denne udvikling er ikke at der er opstået et stærkt behov i samfundet for nøjagtig disse ord. Et eksempel på dette er den elektrotekniske termsamling *Raftækniord í Íðorðabankinn*.

3.7. Tilfældige termsamlinger

I termbanken findes der eksempler på termsamlinger som karakteriseres af en lidt tilfældig samling termer og ordforbindelser. Der er især tale om biprodukt af andre projekter, en slags ordlister fra oversættere, men ikke systematisk termarbejde. Selv om sådanne ordlister er ustrukturerede kan de være til gavn for brugere.

I andre tilfælde er der tale om en arbejdsopgave som er blevet udført sideløbende med et andet projekt. Her kan fremfor alt nævnes forberedelsen til udgivelsen af en engelsk-islandsk ordbog omkring 1980 og den islandske encyklopædi omkring 1990 som blev ophav til nogle termsamlinger. For eksempel er grundlaget for termsamling i meteorologi i *Íðorðabankinn*, en termsamling på knap 500 ord, som blev bearbejdet til udgivelsen engelsk-islandsk ordbog omkring 1980. Publikationen af disse to førnævnte bøger har uden tvivl haft stor indflydelse på ordforråd i videnskab såvel som på andre områder med koordinering af islandske fagord hvor der før havde været stor inkonsekvens i brug af fagord mellem offentligheden og specialister inden for de forskellige fagområder (jf. Dóra Hafsteinsdóttir & Sigríður Harðardóttir 2001:19).

4. Rettes fagordbøger og termsamlinger mod en speciel brugergruppe?

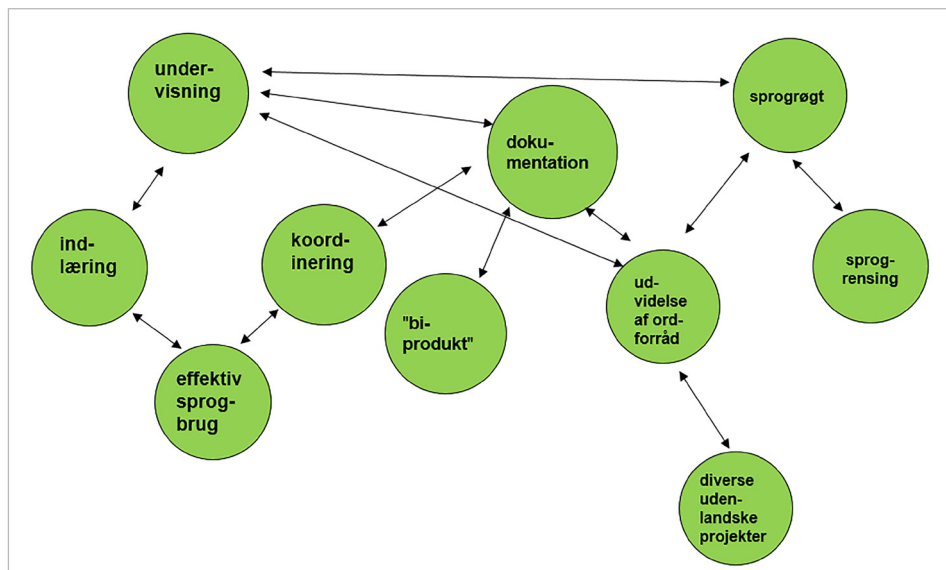
Jeg har undersøgt om man kan registrere at fagordbøger eller termsamlinger er rettet mod en speciel brugergruppe og hvorvidt de svarer til dens individuelle behov.

Efter min mening identificeres brugergruppen tydeligst i termsamlinger som har undervisning og indlæring som motiv. Det er meget tydeligt at de er rettede mod studerende. De færreste fagordbøger og termsamlinger har en præcist afgrænset brugergruppe. Termsamling i statistik rettes f.eks. ikke kun mod studerende, men også dem der skriver om statistik i alle former for rapporter og artikler. Når der er tale om sundhedsområdet er specialiseret personale og klienter også en del af brugergruppen. I det hele taget er der behov for islandske fagord og at de findes registrerede og let tilgængelige. En brugerundersøgelse fra 2017 i forbindelse med Ídorðabankinn viste at foruden studerende er oversættere, forskere og specialister en rigtig stor brugergruppe (jfr. Ágústa Þorbergsdóttir 2017:101-103).

5. Konklusion

Jeg har gennemgået tilblivelsen af fagordbøger og termsamlinger i Island. Som fremgår i nærværende artikel har de såkaldte termudvalg og deres arbejde været karakteristisk for islandsk termarbejde. Der er stadig intet der indikerer, at termudvalgene vil blive nedlagt inden for de næste mange år.

Jeg har forsøgt at analysere hvilke motiver ligger bag islandske fagsprogsressurser, og ifølge min undersøgelse er der mange motiver. Man kan registrere, at motiverne har ændret sig i tidens løb. Karakteristisk for de ældste termsamlinger var, at de skulle bidrage til afskaffelse af danske ord i islandsk, for eksempel den ældste terminologisamling, udgivet i 1928 af ingeniørforeningens ordudvalg (*Ídorðasafn*). Nu lægger man større vægt på, at fagord er registrerede og dermed let tilgængelige. Oprettelsen af en islandsk termbank i 1997 var begyndelsen til en ny epoke, og gjorde publicering af termsamlinger lettere og billigere. Med termbanken blev det realistisk at publicere termsamlinger i det lille islandske sprogsamfund. Det var en stor ændring, og mange termsamlinger blev publiceret efterfølgende.



Figur 1: Motivationerne er forbundet med hinanden.

Der lægges stor vægt på at man kan anvende islandsk på alle områder, og ordforrådet skal beriges med islandske ord i stedet for fremmedord. Der skal kunne tales og skrives om alt på det islandske sprog. I Island har man en lang tradition for at danne islandske ord frem for at bruge de internationale termer.

Som vist i figur 1, er mange af motivationerne, som ligger bag islandske fagsprogsressurser forbundet med hinanden. Som eksempel herpå kan nævnes at koordination af islandske termer, som er en almindelig motivation for fagsprogsressurserne, medvirker til effektiv sprogbrug.

I fagsprog er en præcis terminologi vigtig for en effektiv kommunikation uden misforståelser. Også de tilfældige termsamlinger som er opstået som bi-produkter har en kobling til dokumentation på den måde at de fører til at termer findes tilgængelige.

De forskellige motivationer kan være som to sider af samme mønt, og det kan gøre det vanskeligt at fastslå, hvad der er en primær eller sekundær motivation.

Litteratur

Ordbøger og termresurser

- EÍSO-1997 = *Ensk-íslensk stærðfræðiorðaskrá ásamt íslensk-enskum orðalykli*. (1997): Reynir Axelsson (red.) Reykjavík: Íslenska stærðfræðafélagið. Háskólaútgáfan.
- Flugorðasafn*. Íslenskt-enskt, enskt-íslenskt. (1993): Jónína M. Guðnadóttir (red.). (Rit Íslenskrar málnefndar 7.) Reykjavík: Íslensk málnefnd.
- Íðorðabankinn*. Stofnun Árna Magnússonar í íslenskum fræðum. <www.idord.arnastofnun.is> (oktober 2019).
- Íðorðasafn* (1928): Orðanefnd Verkfræðingafjelagsins. Sjerprentun úr Tímariti V.F.Í. Reykjavík.
- Nýyrði I-IV* (1953, 1954, 1955, 1956): Sveinn Bergsveinsson (red. Nýyrði I), Halldór Halldórsson (red. Nýyrði II-IV). Reykjavík: Menntamálaráðuneytið.
- TO-1983 = *Tölvuorðasafn*. Íslenskt-enskt, enskt-íslenskt (1983): Orðanefnd Skýrslutæknifélags Íslands. (Rit Íslenskrar málnefndar 1.) Reykjavík: Hið íslenska bókmenntafélag.
- Tækniorðasafn* (1959): Sigurður Guðmundsson. Reykjavík: Menntamálaráðuneyti.
- UÍO-2007 = *Umhverfistækni – Íðorðabók* (2007): Einar B. Pálsson (red.) Reykjavík: Háskólaútgáfan.
- SJÓ-1998 = *Sjómennsku- og vélfræðiorðasafn*: enskt-íslenskt, íslenskt-enskt. Eysteinn Sigurðsson, Franz Gíslason, Tryggvi Gunnarsson. Reykjavík: Iðnú.

Anden litteratur

- Ágústa Þorbergsdóttir (2017): Brugeranalyse af ordbanken. I: *Nordterm* 20, 98-106.
- Ágústa Þorbergsdóttir (2018): Íðorðarit Stofnunar Árna Magnússonar í íslenskum fræðum. I: *Orð og tunga* 20, 159-163.
- Ágústa Þorbergsdóttir (2019): Nye perspektiver og udfordringer for terminologi i en islandsk kontekst. I: *Nordterm* 21, 47-51.
- Dóra Hafsteinsdóttir & Sigríður Harðardóttir (2001): Íslenska alfræðiorðabókin. Lýsing á viðamiklu orðabókarverkefni. I: *Orð og tunga* 5, 13-22.
- Baldur Jónsson (1987): Íslenskar íðorðaskrár. I: *Málfreginir* 1, 21-24.
- Einar Benediktsson (1891): Íslenzk orðmyndan. *Sunnanfari*, 56-57. Reykjavík.
- Guðmundur Finnbogason (1919): Orðanefnd Verkfræðingafjelagsins. I: *Tímarit Verkfræðingafjelags Íslands* 4(6), 53-55.

Mennta- og menningarmálaráðuneytið (2009): *Íslenska til alls. Tillögur Íslenskrar málnefndar að íslenskri málstefnu samþykktar á Alþingi 12. mars 2009*. <hdl.handle.net/10802/4405> (maj 2019).

Sigrún Helgadóttir & Kristján Árnason (1991): *Áætlun um tilhögun iðorðastarfssemi*. Tillögur. Reykjavík: Íslensk málnefnd, Menntamálaráðuneytið.

Sigurður Jónsson (2007): *Det vilda tänkandet och det kultiverade. Isländsk fackspråklig språkvård med tyngdpunkt på första hälften av 1900-talet*. Acta Wasaensia, Nr. 184. Vaasa: Universitas Wasaensis.

Ágústa Þorbergsdóttir

M.A.

Stofnun Árna Magnússonar í íslenskum fræðum

Laugavegi 13

IS-101 Reykjavík

agusta.thorbergsdottir@arnastofnun.is

Normering af ord uden normering – i teori og praksis

Lars Trap-Jensen

By law, the Danish Language Council is authorized to decide the spelling of Danish vocabulary and publish their decisions in the official Danish spelling dictionary (RO), which contains a lemma list of c. 64,000 entries. As *The Danish Dictionary* (DDO) covers more than 100,000 lemmas but not all the entries of RO, it follows that more than 40 per cent of the DDO entries cannot be looked up in the spelling dictionary. In this article, I discuss how the editorial staff navigate in situations where no norm is available. This involves theoretical considerations of what aspects of language belong to the domain of orthography and how far the normative mandate extends, as well as more practical decisions in order to determine how far the use of analogy can be applied and what counts, in the eyes of the users, as a spelling error.

1. Indledning

Udgangspunktet for denne artikel har været et praktisk problem for redaktionen af *Den Danske Ordbog* (herefter DDO). Problemet er formentlig kendt af enhver leksikograf der arbejder med sprog hvis ordforråd normeres af en central instans: det at man skal tage stilling til opslagsformer af ord der enten ikke i forvejen har en normeret form, eller hvor der kan opstå tvivl om normeringens grundlag og rækkevidde.

Et eksempel kan illustrere problemet: ordet *svedtrøje* (en undertrøje af svedtransporterende materiale, brugt til sport) hvor en konsultation i DDO's korpus viser at lemmaet realiseres i to former: *svedtrøje* (24) og *svedetrøje* (6)¹. Ordet optræder ikke i *Retskrivningsordbogen* (herefter RO) udgivet af Dansk Sprog­nævn, som er den autoriserede kilde til stavning af danske ord, og leksikografen stilles derfor over for flere valg: Skal kun den ene eller begge former tages med i ordbogen? Skal de betragtes som forskellige, men synonyme ord? Eller bør de snarest regnes for to varianter af samme ord og derfor behandles under samme opslagsord, og i givet fald: Er de da sidestillede varianter, eller skal den ene form have forrang for den anden?

¹ En ordform efterfulgt af et tal i parentes angiver her og i det følgende antallet af forekomster i DDO's interne korpus *Bakspejlet*. Dette korpus indeholder lidt over 1 milliard løbende ord (oktober 2019).

Selvom ordet ikke optræder i RO, kan man godt udlede noget af den. Det fremgår eksempelvis af opslaget *sved* at det danner sammensætninger med 0 i fugen: ”**sved** *sb.*, -en, *i sms.* *sved-*, *fx* *svedlugt*, *svedtransporterende*”. Samtidig kan man se af ordbogens alfabetiske liste at den også har opslagsordene *svedekur*, *svederem* og *svedetur*. Det må forstås sådan at disse ord af Sprognævnet vurderes at være dannet til verbet *svede*. Eftersom både former med *sved-* og *svede-* findes og bruges, har DDO’s redaktion derfor vurderet at begge former er acceptable, men på baggrund af frekvensforholdene gjort *svedtrøje* til hovedform og *svedetur* til variantform. Resultatet er derfor blevet som beskrevet i figur 1.

Generelt er problemet nok større i Danmark end eksempelvis i Norge og Sverige fordi den danske retskrivningsordbog er mindre omfattende end den tilsvarende norske (*Tanum*) og svenske retskrivningsordbog (SAOL). RO rummer ca. 60.000 opslagsord, mens DDO har lidt over 100.000, og derfor sker det ikke sjældent at DDO’s opslagsord savner modsvarighed i RO og dermed er uden officiel normering. Man kan i den situation spørge: Hvilken status har disse ord da når det gælder normering, og hvilken leksikografisk praksis er mest hensigtsmæssig når nye ord optages i ordbogen uden at have et normeringsmæssigt forlæg?²

svedtrøje substantiv, fælleskøn

også i formen: **svedetur**

BØJNING -n, -r, -rne

UDTALE  

OPRINDELSE kendt fra 1978

Betydninger

tynd trøje af et svedtransporterende materiale til at have tættest på kroppen (under en anden trøje) når man dyrker sport

SE OGSÅ [sweatshirt](#)

ORD I NÆRHEDEDEN [sauna](#) | [svededug](#) | [svederem...vis mere](#)

De unge mennesker kan ikke nøjes med en trøje, som vi kunne. Nu er der både shorts, strømper, svedtrøjer og trøjer med i tasken [JyVestk2004](#)

Figur 1: *svedtrøje* i DDO.

² For andre normeringsmæssige aspekter af DDO’s praksis se også Trap-Jensen (2002).

2. Hvad siger loven?

Først kan man se på det formelle aspekt. I Danmark er retskrivningen reguleret af lov om dansk retskrivning, der i al sin enkelhed lyder:

§ 1. Dansk retskrivning fastlægges af Dansk Sprognævn og offentliggøres i Dansk Sprognævns retskrivningsordbog.

Stk. 2. Ved fastlæggelsen af dansk retskrivning følger Dansk Sprognævn de regler, der er fastsat i eller i medfør af lov om Dansk Sprognævn. (Retskrivningsloven, Lov nr. 332 af 14-5-1997)

Derudover findes også en bekendtgørelse (Bekendtgørelse af lov om Dansk Sprognævn, Lovbekendtgørelse nr. 217 af 13-2-2015) der regulerer Dansk Sprognævns opgaver, og som bl.a. præciserer at kun ikkeprincipielle ændringer kan foretages af nævnet på egen hånd.

Lovteksten er prisværdig kortfattet og umiddelbart klart formuleret. Én ting der dog kan give anledning til tvivl, er hvor langt bemyndigelsen strækker sig. Dansk Sprognævn er bemyndiget til at fastlægge dansk retskrivning, men det fremgår ikke af selve lovteksten hvad der menes med retskrivning. Det kan derfor være nyttigt at se på hvordan begrebet forklares i ordbøgerne. Her er nogle definitioner fra udvalgte ordbøger for dansk og dets nabosprog:

- *retskrivning*: sæt af regler og principper for den rigtige stavning af ord og den korrekte tegnsætning i et sprog (*Den Danske Netordbog*)
- *retskrivning*: (regler for) den rigtige stavning af ord og den rigtige tegnsætning i et sprog (DDO)
- *retskrivning*: den skriftlige gengivelse af ord som er i overensstemmelse med en vedtaget måde at gøre det på (*Nudansk Ordbog*)
- *rättskrivning*: (övning i) att stava rätt · särsk. som moment i skolundervisningen (*Svensk ordbok*)
- *rettskriving*: tradisjonell eller offisielt fastsatt skrivemåte for ordene i et språk (*Bokmålsordboka*)
- *rettskrivning*: tradisjonell eller (offisielt) fastsatt skrivemåte for ordene i et språk; norm for skrivemåte og bøyning (NAOB)
- *Rechtschreibung*: nach bestimmten Regeln festgelegte Schreibung der Wörter, Orthographie (DWDS)
- *orthography*: correct or proper spelling; spelling according to accepted usage or convention (OED)

Hvad der går igen i de her citerede ordbøger, er at retskrivning henregnes til det skrevne sprog alene: hvordan ordene staves. De danske ordbøger nævner også tegnsætning, mens NAOB som den eneste angiver at også ordenes bøjning er en del af retskrivningen.

Jeg synes det giver god mening at retskrivning alene vedrører det skrevne sprog. Skriftsproget er tillært i modsætning til det talte sprog, som ethvert barn lærer hvis det får det rigtige input fra omgivelserne. Derfor skal man være varsom med at gribe ind over for menneskers naturlige sprog, hvorimod det giver god mening at det tillærte sprog, skriftsproget, standardiseres så kommunikation kan foregå mest effektivt.

Forskellen på normering af skrift og tale kan illustreres med eksemplerne *album* og *job*. Når man skal afgøre om bestemt form ental af *album* skrives *albumet* eller *albummet*, er der alene tale om en normering af skriftsproget, og det falder derfor helt klart under retskrivningen. Indtil 2001 var begge skrivemåder tilladt på dansk, men formen med enkelt *m* blev afskaffet i 2001 så det nu kun kan skrives *albummet*.

Det forholder sig anderledes med *job* eller *jobs* som flertalsform af *job*. Her er der ikke tale om normering af skriftsproget alene, for den der skriver ”jobs”, siger formentlig også ”jobs”. Og dermed har vi bevæget os ud over det rent skriftsproglige. Dette er Sprognævnets medarbejdere selvfølgelig opmærksomme på og forsøger at navigere mellem systematik og usus efter bedste evne.

På samme måde må DDO's redaktion forsøge at lægge sig fast på hvor langt den mener at nævnets normeringsmandat kan udstrækkes, og hvilken normeringsmæssig status der kan tillægges ord der ikke står i RO. Kan redaktionen fx tillade sig at analogisere eller forsøge at gætte sig til hvad Sprognævnet ville beslutte hvis de skulle normere et bestemt ord?

Disse spørgsmål lader sig ikke besvare helt entydigt eftersom forskellige typer af situation kræver forskellige svar. De enkelte typer vil derfor blive gennemgået i de følgende afsnit.

3. Ord der ikke står i *Retskrivningsordbogen*

Mange unormerede ord kan henføres til bestemte typer som systematisk er underlagt den samme ortografiske regel i RO. Her vælger DDO som regel også at lade reglen gælde for ord der ikke står i RO. De vigtigste tilfælde er:

- 1) Initialforkortelser, der efter RO systematisk er normeret med både store og små bogstaver: *DVD* eller *dvd*, *BMI* eller *bmi*. DDO ekstrapolerer derfor og angiver også andre forkortelser af denne type som dobbeltformer selvom de ikke optræder i RO: *FAQ* eller *faq*, *HDMI* eller *hdmi* osv.

- 2) Gruppesammensætninger. Her anfører RO på alfabetisk plads kun muligheden med bindestreg mellem de to hovedled: *væg til væg-tæppe*, *jord til luft-missil* osv. Derudover henviser opslagene til § 57.7 i retskrivningsreglerne, hvoraf det fremgår at det også er tilladt at sætte bindestreger mellem alle ordene i en gruppesammensætning, dvs. *væg-til-væg-tæppe*, *jord-til-luft-missil* osv. Da begge stavemåder dermed er godkendte, har DDO's redaktion valgt at oplyse om dette på alfabetisk plads ved at gøre dem til dobbeltformer: *væg til væg-tæppe* eller *væg-til-væg-tæppe*. Det samme gør vi for ord af samme type som ikke står i RO: *a la carte-ret* eller *a-la-carte-ret* og *arbejd efter reglerne-aktion* eller *arbejd-efter-reglerne-aktion*.
- 3) Ord skrevet med tal, som i RO ikke anføres på alfabetisk plads. Det fremgår dog af retskrivningsreglerne at man kan bruge taltegn i stedet for bogstaver (§ 57.4 om bindestreg i sammensætninger med taltegn). Tallene er imidlertid ikke egentlige sproglige tegn på linje med bogstaver, og sammensætninger med tal behandles derfor ikke som dobbeltformer. DDO's redaktion angiver derfor en udbredt skrivemåde med tal som en variantform med kommentaren "også i formen": *hundredårig*, også i formen: *100-årig* og *femdørs*, også i formen *5-dørs* osv.
- 4) Sammensætninger med *proprier*, som efter retskrivningsreglernes § 12.10.b valgfrit kan skrives med stort eller lille begyndelsesbogstav. De angives i DDO derfor som dobbeltformer selvom RO også her viser det indirekte ved kun at have sammensætninger med lille begyndelsesbogstav under opslagsordet, suppleret med en henvisning til retskrivningsparagraffen. DDO har altså *balkanland* eller *Balkanland*, *beatle* eller *Beatle* og *hollywoodfilm* eller *Hollywoodfilm*.

Udover at sammenligne med retskrivningsreglerne ser vi også på ord der minder om det aktuelle ord. Hvis der er tale om rækkesammensætninger, føler vi os på nogenlunde sikker grund når et ord føjer sig til en række af ord som i øvrigt er konsekvent behandlet. Et eksempel er sammensætninger med *linje*. Skrivemåden *linie* blev afskaffet som officiel form i 2001 således at kun formen med *j* i dag er godkendt, og det gælder også alle sammensætninger med *linje*. Derfor ekstrapolerer vi og anfører alle opslagsord i DDO med *-linje* på tilsvarende vis, også dem der ikke står i RO: *linjeafstand*, *linjeforløb*, *linjefunktion* osv., mens udbredte former med *linie-* anføres som uofficielle.

I tilfælde hvor det gælder valg af fugeelement, fx *rødfiskefilet* overfor *rødfiskefilet*, er det lidt vanskeligere. Her kan man ikke nødvendigvis slutte fra forekomster som *fiskefilet* og *laksefilet* (som står i RO) til at der så også skal være *e-* i fugen ved en flerleddet sammensætning. Som nævnt ovenfor (i afsnit 2) er der desuden noget andet end ren stavning på spil. I dette tilfælde fordeler usus sig nogenlunde ligeligt mellem formerne, og derfor er begge ord taget med i DDO, ikke som dobbeltform, men som variantformer, dvs. med løsningen "også i formen: *rødfiskefiler*", hvorved det signaleres at begge former anses for acceptable, men at de er ikke officielle dobbeltformer i RO.

4. Ord der står i *Retskrivningsordbogen*

Når det gælder ord som står i RO, kommer tolkningen af Sprognævnets bemyndigelse til at få visse konsekvenser.

Fugeelement i sammensætninger er som nævnt ikke noget rent skriftsprogligt, og hvis sprogbrugen udviser tilstrækkelig mange eksempler på et alternativt fugeelement, kan de være svære at afvise selvom formen ikke er normeret i RO. Eksempler fra DDO er *risvinseddike* (som står i RO) overfor *risvineddike* (ikke i RO) eller *fedtfinger* (i RO) overfor *fedtefinger* (ikke i RO). I disse tilfælde har DDO medtaget begge former, idet den form der ikke er i RO, er gjort til variantform.

En tilsvarende argumentation kan man gøre gældende når det drejer sig om morfologiske varianter. Heller ikke de er rent ortografisk motiveret, og hvis korpus viser at sprogbrugernes fornemmelse foretrækker *gambianer* frem for *gambier* eller *alabaster* for *alabast*, er det svært at afvise disse former. Det er dog ikke noget der er behandlet fuldstændig konsekvent i DDO, og man kan derfor godt støde på eksempler hvor to redaktører har truffet hvert sit valg. I de nævnte eksempler er *gambianer* således anført som uofficiel form, mens *alabaster* er gjort til variantform.

Beslægtet med morfologiske varianter er tilfælde hvor der opstår tvivl om hvorvidt en sammensætning er dannet til et substantiv, et verbum eller en afledning, og også dette fører undertiden til en redaktionelt lidt vaklende kurs. I RO er opslagsordene *fiskefartøj/fiskerfartøj* og *fiskekutter/fiskerkutter* dobbeltformer, og dette kunne være et argument for at bruge disse ord som skabelon for behandlingen af *fiskebåd/fiskerbåd*, men her har DDO alene medtaget *fiskerbåd*, hvorimod *fiskebåd* hverken optræder som dobbeltform eller variantform til trods for et ellers ganske solidt empirisk belægsmateriale: Formen *fiskebåd* har 229 forekomster i DDO's korpus. Det er formentligt gjort fordi RO alene har medtaget *fiskerbåd*. Derimod har DDO valgt at tage *murerske* med som en variantform af *murske* selvom RO kun registrerer *murske*. Det virker ikke konsekvent.

En tilsvarende usikkerhed og variation er almindelig ved låneord. RO har medtaget ordene *squaredance* og *breakdancer*, men ordene optræder i DDO's korpus ganske hyppigt med en fordansket form som andetled: *squaredans* og *breakdanser*. Også i dette tilfælde mener vi at den fordanskede form må accepteres, og DDO bringer derfor både den engelske form og den fordanskede som variantformer.

I andre tilfælde går tvivlen på om to former repræsenterer varianter af samme lemma eller snarere bør regnes som to ord med samme betydning. Sprogbrugen viser således både danske og engelske former af *kickbokser* (61), *kickboxer* (40) og *kickboksning* (163), *kickboxing* (236). I tilfældet *kickbokser/kickboxer* er det en rent skriftsproglig forskel, og da RO kun anerkender den danske form, anfører

DDO den engelske som en uofficiel variantform. Navnet på sporten anføres til gengæld af RO som dobbeltformer, og det gør DDO derfor også.

I tilfælde hvor noget andet end det rent skriftsproglige er på spil, er der forskellige muligheder: I et eksempel som *enårig/etårig* har RO to opslagsord med forskellige eksempler og gensidig henvisning. Det dækker over at fagfolk bruger *enårig* om planter, mens det i almensproget kan hedde både *enårig* og *etårig*. I DDO er det løst sådan at *enårig* har variantformen *etårig*, mens der under *etårig* er anført en betydning som er en henvisning til *enårig* i den botaniske brug (*etårig* har derudover andre betydninger hvor *enårig* ikke kan bruges).

Et andet eksempel er *hormonel/hormonal*, hvor RO har to opslagsord med samme betydningsangivelse ('som vedrører hormoner'). DDO bruger andre kriterier for homografi end RO, primært etymologi, og har derfor kun én artikel. Da ordene ikke er dobbeltformer i RO, er den ene form, *hormonal*, gjort til hovedopslagsform i DDO, mens den anden, *hormonel*, bringes som variantform. Et lignende tilfælde er *interpunktere/interpungere*, som i RO også er separate opslagsord med (næsten) samme betydningsangivelse, mens de i DDO behandles i én artikel (samme etymon trods lidt forskellige låneveje), med *interpunktere* som hovedform og *interpungere* som variantform.

I andre tilfælde er der mere oplagt tale om forskellige ord, fx *slåfejl* vs. *slagfejl* eller *nykøbingenser/nykøbinger*. Ikke desto mindre har RO sidstnævnte som dobbeltform, og i dette tilfælde har DDO af pragmatiske grunde valgt at gøre det samme, ligesom det er gjort for en række andre indbyggerbetegnelser med tilsvarende problemstilling: *burmeser/burmaner*, *hobroer/hobrogenser*, *kartaginienser/kartager*, *napolitaner/neapolitaner* m.fl.

Endelig er der en række tvivlstilfælde hvor man som redaktør må prøve at tolke RO. Denne har eksempelvis to opslagsord *visker* (med angivelsen 'vinduesvisker') og *whiskers* (med angivelsen 'kindsæk'), men hvilken opslagsform vælger man for ordet i betydningen 'trommebørste'? Usus er ikke i tvivl om at det skal være *whisker*, men RO er ikke til stor hjælp her. DDO har derfor oprettet tre artikler: udover RO's to artikler en tredje artikel *whisker* med trommestik-betydningen. Et tilsvarende eksempel er *boxer* eller *bokser* for betydningen 'kineser fra bokseropstanden' RO har kun *bokser* (med angivelsen 'person som dyrker boksesport') og *boxer* ('en hund'). DDO har derfor også her valgt at oprette en selvstændig artikel *bokser* med variantformen *boxer* til at beskrive betydningen 'kineser fra bokseropstanden'.

5. Diskussion og konklusion

Gennemgangen har vist hvordan DDO's redaktion håndterer dels det unormerede ordstof, dels normeret ordstof som kræver normeringsmæssig fortolkning. I de fleste tilfælde er det problemfrit fordi en stor del af DDO's ekstramateriale udgøres af sammensætninger og afledninger der består af normeringsmæssigt kendte elementer. Men indimellem kommer DDO i situationer hvor redaktionen er nødt til at træffe beslutninger som indebærer en form for normering, omend indirekte fordi ordbogen ikke har det officielle mandat til at gøre det. Det er især tilfældet i to principielle situationer.

For det første når vi kan konstatere variation i usus for ikkenormerede ord og derfor udøver normering ved at foretrække én eller flere former frem for andre, egentlig uanset hvor gode argumenter der er for at gøre det. Derfor er redaktionen tilbageholdende med at beskrive ikkenormerede ord på en stærkt normerende måde, fx ved at gøre dem til dobbeltformer eller beskrive nogle former som "uofficielle". I stedet foretrækker vi, som korpusbaseret ordbog, en deskriptiv fremgangsmåde hvor den mest udbredte form gøres til opslagsform, mens alternative og udbredte skrivemåder beskrives som variantformer med kommentaren "også i formen". Det kan undertiden give mange sideformer. Et eksempel er *pak-choi*, der i DDO har fem variantformer udover opslagsformen: *bok choy*, *pak choi*, *bok-choi*, *bok choi* og *pak choy*. Omvendt er vi også nødt til at vælge pragmatiske løsninger og ekstrapolere fra det normerede stof til det unormerede: Selvom *linjeafstand*, *linjelyn* og *linjenummer* ikke står i RO, beskriver DDO alligevel variantformerne *linieafstand*, *linielyn* og *linienummer* som uofficielle fordi *linje* og alle sammensætninger med *linje-* siden 2001 kun har været normeret med *j*. Og på samme måde tillader vi os at gøre de unormerede *chilisovs/chilisauc*e, *citronsovs/citronsauce* osv. til dobbeltformer fordi alle andre ord sammensat med *-sovs/-sauce* er dobbeltformer.

For det andet udøver vi normering når vi tolker retskrivningsreglerne og grænserne for Sprognævnets normeringsmandat. I begge tilfælde kan det føre til at DDO's beskrivelse afviger fra RO's, også for ord der står i RO. DDO har medtaget *europaparlamentariker/Europaparlamentariker* og *væg til væg-tæppe/væg-til-væg-tæppe* som dobbeltformer selvom RO kun har *europaparlamentariker* og *væg til væg-tæppe* på alfabetisk plads, mens den anden form kun fremgår af retskrivningsreglerne. Og når DDO har medtaget *fedtefinger* og *risvineddike* som acceptable varianter af *fedtfinger* og *risvinseddike*, er det primært fordi vi anser et hørbart fugeelement for at ligge udover det rent ortografiske. Det samme argument kan anvendes for bøjningsformer og genusangivelser: Stavemåden *albummet* som bestemt form ental frem for *albumet* (som var tilladt indtil 2001) falder klart under normeringsmandatet, men om *spoiler* i flertal skal

være *spoilere* eller *spoilers*, er ikke alene et ortografisk spørgsmål, og DDO anfører derfor formen med *-s* som en acceptabel mulighed selvom den ikke fremgår af RO (der dog også har en anden betydningsangivelse).

Dog gælder det også her at man må være pragmatisk og skønne hvordan det bliver opfattet af brugerne. DDO følger derfor RO ganske langt og beskriver fx den udbredte intetkønsform af *selfie* som ”uofficiel” skønt dette egentlig ikke vedrører ortografi. Der er et vist mål af skøn involveret når redaktørerne skal beslutte sig, og gennemgangen har da også vist at redaktionen ikke er helt konsistent når reglerne skal tolkes i de enkelte tilfælde.

Oftest konsulterer redaktionen i tvivlstilfælde kollegerne på Sprognævnet, navnlig når der er normeringsmæssigt interessante eller udfordrende forhold på spil. Det sker ikke sjældent at ordene efterfølgende optages i RO, og på den måde er udvekslingen til fordel for begge parter.

Litteratur

Ordbøger

Bokmålsordboka. Universitetet i Bergen og Språkrådet. <ordbok.uib.no> (oktober 2019).

DDO = *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab. <www.ordnet.dk/ddo> (oktober 2019).

Den Danske Netordbog. Henning Bergenholtz m.fl. Odense: Ordbogen A/S. <www.ordbogen.com> (oktober 2019).

DWDS = *Digitales Wörterbuch der deutschen Sprache*. Berlin Brandenburgische Akademie der Wissenschaften. <www.dwds.de> (oktober 2019).

NAOB = *Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur og Kunnskapsforlaget. <www.naob.no> (oktober 2019).

Nudansk Ordbog. Odense: Ordbogen A/S. <www.ordbogen.com> (oktober 2019).

OED = *Oxford English Dictionary*. Oxford University Press. <www.oed.com> (oktober 2019).

RO = *Retskrivningsordbogen*. Bogense: Dansk Sprognævn. <www.dsn.dk/ro> (oktober 2019).

SAOL = *Svenska Akademiens ordlista*. Göteborg: Svenska Akademien. <www.svenska.se/saol> (oktober 2019).

Svensk ordbok. Göteborg: Svenska Akademien. <www.svenska.se/so> (oktober 2019).

Tanum = Boye Wangenstein: *Tanums store rettskrivningsordbok*, 10. utgave. Oslo: Kunnskapsforlaget.

Anden litteratur

Bekendtgørelse af lov om Dansk Sprognævn (Lovbekendtgørelse nr. 217 af 13-2-2015). Kulturministeriet. <www.retsinformation.dk/Forms/R0710.aspx?id=167989> (oktober 2019).

Lov om dansk retskrivning (Lov nr. 332 af 14-5-1997). Undervisningsministeriet. <www.retsinformation.dk/forms/r0710.aspx?id=84769> (oktober 2019).

Trap-Jensen, Lars (2002): Normering og deskription i Den Danske Ordbog – mere eller mindre? I: Henning Bergenholtz & Sven-Göran Malmgren (red.): Oslo: *LexicoNordica* 9, s. 63-79.

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

Konferenzprogramm

15 konferensen om lexikografi i Norden

4–7 juni 2019

Ordböcker – av vem, för vem och varför?

Helsingfors universitet, Forsthuset, Unionsgatan 40

DAG 1 Tisdag 4.6.2019

- 18.00 • Mottagning på Institutet för de inhemska språken
Hagnäskajen 6
slutar 20.00

DAG 2 Onsdag 5.6.2019

- 08.15 • Registrering
Forsthuset, Unionsgatan 40, utanför sal 1
- 09.00 • Öppning
Sal 1 – Hanna Lehti-Eklund och Ulla-Maija Forsberg
- 09.15 • Plenarföreläsning
Sal 1 – ordförande: Caroline Sandström
Anna Helga Hannesdóttir (Göteborg):
Den tvåspråkiga lexikografins utmaningar förr och nu

10.15 • Kaffe

- 10.30 •
- | | | |
|--|--|--|
| <p>SEKTION 1 <i>sal 6</i></p> <p>Dagfinn Rødningen & Knut E. Karlsen:
Varemerkenavn i ordböcker</p> <p><i>ordförande: Emma Sköldberg</i></p> | <p>SEKTION 2 <i>sal 12</i></p> <p>Jonas Jensen, Henrik Lorentzen & Mette-Marie Møller Svendsen:
Lad brugeren komme til orde. Hvordan får ordbogen mest muligt ud af brugerhenvendelser?</p> <p><i>ordförande: Eva Thelin</i></p> | <p>TEMASEKTION <i>sal 8</i></p> <p>Santeri Junttila: En etymologisk nätordbok över det äldsta finska ordförrådet</p> |
|--|--|--|

	SEKTION 1 <i>sal 6</i>	SEKTION 2 <i>sal 12</i>	TEMASEKTION <i>sal 8</i>
11.00	<p>Margunn Rauset: Har bokmålsbrukere <i>bein</i> eller <i>ben i nesa</i> – eller er det <i>nesen</i>? Om idiomatisk variasjon i norske ordbøker og faktisk språkbruk</p> <p><i>ordförande: Emma Sköldberg</i></p>	<p>Karin Cavallin: Ngram i SAOB</p> <p><i>ordförande: Eva Thelin</i></p>	<p>Mikko Bentlin: Osmo Nikkiläs finsk-svenska etymologier på nätet</p> <p><i>ordförande: Santeri Junttila</i></p>
11.30	<p>Lars Trap-Jensen: Normering af ord uden normering – i teori og praksis</p> <p><i>ordförande: Emma Sköldberg</i></p>	<p>Mikkel Ekeland Paulsen: Sammensatte ord – et sammensatt problem</p> <p><i>ordförande: Eva Thelin</i></p>	<p>Hanne Lauvstad: NAOBs etymologier</p> <p><i>ordförande: Santeri Junttila</i></p>
12.00	Lunch		
13.00	<p>Oddrun Grønvik, Stig Jarle Helset & Johan Myking: Nettbrukaren i fokus: digitalisering og revisjon av Norsk Ordbok</p> <p><i>ordförande: Sturla Berg-Olsen</i></p>	<p>Halldóra Jónsdóttir & Þórdís Úlfarsdóttir: Omdannelsen af en flersproget til en monolingval ordbog</p> <p><i>ordförande: Elina Heikkilä</i></p>	<p>Miro Metsämuuronen: Färoisk lånordsforskning i historisk överblick</p> <p><i>ordförande: Santeri Junttila</i></p>
13.30	<p>Kristian Blensenius & Lena Rogström: Att hantera grammatisk förändring i en deskriptiv ordbok</p> <p><i>ordförande: Sturla Berg-Olsen</i></p>	<p>Petri Lauerma & Elina Palola: Rasmus Rask som katalysator för finska vetenskapliga ordböcker och finsk grammatik 1818–1819</p> <p><i>ordförande: Elina Heikkilä</i></p>	<p>Vesa Jarva: Suomen murteiden sanaston asteittaista etymologiaa</p> <p><i>ordförande: Santeri Junttila</i></p>
14.00	<p>Caroline Sandström & Harriet Eriksson: Suffixavledda verb av finskt ursprung i de svenska dialekterna i Finland</p> <p><i>ordförande: Sturla Berg-Olsen</i></p>	<p>Henrik Køhler Simonsen & Patrick Leroyer: Multimediale ordbøger: hvordan og hvorfor?</p> <p><i>ordförande: Elina Heikkilä</i></p>	<p>Johan Schalin: Helsingfors universitet Fenniska *litna 'borg, köping', än en gång</p> <p><i>ordförande: Santeri Junttila</i></p>
14.30	<p>Asgerd Gudiksen: Forbindelser og folketro i Ømålsordbogen</p> <p><i>ordförande: Sturla Berg-Olsen</i></p>	<p>Øivin Andersen: Enspråklig norsk ordbok for fremmedspråklige</p> <p><i>ordförande: Elina Heikkilä</i></p>	<p>Minerva Piha: Religiösa kontakter? Mytologiska lånord i sydsamiska från de skandinaviska språken</p> <p><i>ordförande: Santeri Junttila</i></p>

15.00	Kaffe		
15.30	SEKTION 1 <i>sal 6</i>	SEKTION 2 <i>sal 12</i>	TEMASEKTION <i>sal 7</i>
	Ellert Þór Jóhannsson & Simonetta Battista: Brugere, anvendelse og udvikling <i>ordförande: Ásta Svavarsdóttir</i>		Johan Schalin & Juha Kuokkala: Fenniska *šike-/*šiko- 'svett(as)' [poster] <i>Slutdiskussion för temasektionen</i>
16.00	Henrik Hovmark: Ømålsordbogens brugere – nu og i fremtiden <i>ordförande: Ásta Svavarsdóttir</i>	Poul Hansen: Hvordan har efterspørgslen efter tematiske dansk-svenske og svensk-danske netordlister forandret sig efter at majoriteten af brugerne nu søger ord via smarttelefoner? <i>ordförande: Lars Trap-Jensen</i>	Ágústa Þorbergsdóttir: Tilblivelsen af fagordbøger og termsamlinger i Island <i>ordförande: Kristin Magnussen</i>
16.30	Ulla-Maija Forsberg: En etymologisk slangordbok – för vem och varför <i>ordförande: Ásta Svavarsdóttir</i>	Carina Nilstun: Den elektroniske ordbokens innvirkning på den praktisk leksikografi <i>ordförande: Lars Trap-Jensen</i>	Irina Kudasheva: Students' terminology work: what about, whom for, and why? Compiling terminological glossaries as a part of students' translator competence acquisition <i>ordförande: Kristin Magnussen</i>
17.00	Erik Bäckerud, Pär Nilsson & Emma Sköldberg: Så används Svenska Akademiens ordböcker på nätet – till frågan om <i>barometerglas</i> <i>ordförande: Ásta Svavarsdóttir</i>	Annika Aalto, Leena Savolainen & Pia Taalas: Webaserad crowdsourcing inom SignWiki-ordboksarbete <i>ordförande: Lars Trap-Jensen</i>	Bente Selback: «Å nei, det ordet er ikkje lov på nynorsk! Eller ...?» Om parallell redigering av to norske ordbøker <i>ordförande: Kristin Magnussen</i>
18.00	Mottagning på Helsingfors universitet <i>Huvudbyggnaden, Fabiansgatan 33</i>		
	<i>slutar 20.00</i>		

DAG 3 Torsdag 6.6.2019

09.00 ● Plenarföreläsning
Sal 1 - ordförande: Klaas Ruppel

Iztok Kosem (Ljubljana):
 Responsive dictionaries: bringing dictionaries (and lexicographers) and users closer together

10.00 ● Kaffe och posterpresentationer

Oddrun Grønvik, Stig Jarle Helset & Johan Myking:
 Prosjekt NO-AH: fulldigitalisering av Norsk Ordbok

Halldóra Jónsdóttir, A.P. Kristinsson & Steinþór Steingrímsson:
 For whom? End users and lexicographical data on web portals

Monica von Martens:
 Redigeringsmiljön för Stora finsk-svenska ordboken – återbruk, samarbete och vidareutveckling

Einar Freyr Sigurðsson & Hinrik Hafsteinsson:
 A word frequency dictionary of Icelandic child-directed speech

Mette-Marie Møller Svendsen, Thomas Troelsgård & Nicolai Hartvig Sørensen: Salmesang og superordbog: ordbogslinkning i praksis

Bo-A. Wendt:
 SAOB – för vem och hur?

11.00 ● SEKTION 1 *sal 6* SEKTION 2 *sal 12* SEKTION 3 *sal 8* WORKSHOP *sal 7*

Oddrun Grønvik:
 Å dela ord i bitar –
 lingvistisk information
 uttryckt ved leddeling i
 ordbøker

ordförande: Pär Nilsson

Gunn Inger Lyse:
 Ut med *adamsslekt* og
 inn med *arveprinsesse*?
 Leksikografiske
 metoder i
 Revisjonsprosjektet
 for Bokmålsordboka
 og Nynorskordboka

*ordförande:
 Henrik Hotmark*

**Kristín Bjarnadóttir
 & Kristín Ingibjörg
 Hlynsdóttir:** Online
 Data on Icelandic
 Inflection: Descriptive
 or prescriptive: Why,
 for whom, by whom,
 and how?

*ordförande:
 Halldóra Jónsdóttir*

**Arbetsmöte om
 utvecklade av
 etymologiska
 nätresurser:
 Perehdyts
 etymologisessa
 verkkosanakirjassa
 julkaiseville ja
 avustaville.
 Arbetspråk: finska**

*ansvarig ledare:
 Juho Pystynen*

	SEKTION 1 <i>sal 6</i>	SEKTION 2 <i>sal 12</i>	SEKTION 3 <i>sal 8</i>	WORKSHOP <i>sal 7</i>
11.30	<p>Bo-A. Wendt: Hänvisning eller upprepning i samspelet mellan olika ordboksartiklar</p> <p><i>ordförande: Pär Nilsson</i></p>	<p>Jacqueline Levin & Malene Ullerlund Madsen: Ordbogen A/S 15 år med Ordbogen.com – fra banebrydende nyt koncept til konventionel ordbogsportal?</p> <p><i>ordförande: Henrik Hoemark</i></p>	<p>Tarja Heinonen: Is alien Finnish?</p> <p><i>ordförande: Halldóra Jónsdóttir</i></p>	<p>Arbetsmöte om utvecklande av etymologiska nätresurser: Perehdytys etymologisessa verkkosanakirjassa julkaiseville ja avustaville. Arbetspråk: finska</p> <p><i>ansvarig ledare: Juho Pystynen</i></p>
12.00	<p>Inger Schoonderbeek Hansen: Jysk Ordbog – af hvem, til hvem og hvorfor?</p> <p><i>ordförande: Pär Nilsson</i></p>	<p>Ásta Svavarsdóttir: Ordbøger og sprogstandardisering. Islandske ordbøger i 1800-tallet og deres rolle i standardiseringsprocessen</p> <p><i>ordförande: Henrik Hoemark</i></p>	<p>Henrik Köhler Simonsen: Augmented writing: nye muligheder og nye teorier</p> <p><i>ordförande: Halldóra Jónsdóttir</i></p>	
12.30	<p>Sturla Berg-Olsen & Daniel Gusfre Ims: Eit dilemma om lemma</p> <p><i>ordförande: Pär Nilsson</i></p>	<p>Lennart Larsson: Bruklighetsangivelserna i SAOB</p> <p><i>ordförande: Henrik Hoemark</i></p>		
13.00	Lunch			
14.15	Exkursjon och middag <i>Villa Hvitträsk, Kyrkslätt och konferensmiddag på Restaurang Haltia, Esbo</i>			
	<i>slutar 21.30</i>			

DAG 4 Fredag 7.6.2019

09.00	● NFL:s generalförsamling <i>Sal 6</i>		
10.00	● Kaffe		
10.30	● Plenarföreläsning <i>Sal 6 - ordförande: Ulla-Majja Forsberg</i> Nina Martola (Helsingfors): Den tvåspråkiga ordboken – ett absurt åtagande <i>slutar 11.30</i>		
	SEKTION 1 <i>sal 6</i> SEKTION 2 <i>sal 12</i>		
11.40	<table border="0" style="width: 100%;"> <tr> <td style="width: 50%; padding-right: 10px; vertical-align: top;"> <p>Tarja Korhonen & Anna Ryödi: Erfarenheter av nätprojektet för den finska dialektordboken</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p> </td> <td style="width: 50%; padding-left: 10px; vertical-align: top;"> <p>Riina Heikkilä & Tarja Larsson: Om sverigefinskt terminologi- och ordlistearbete</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p> </td> </tr> </table>	<p>Tarja Korhonen & Anna Ryödi: Erfarenheter av nätprojektet för den finska dialektordboken</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p>	<p>Riina Heikkilä & Tarja Larsson: Om sverigefinskt terminologi- och ordlistearbete</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p>
<p>Tarja Korhonen & Anna Ryödi: Erfarenheter av nätprojektet för den finska dialektordboken</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p>	<p>Riina Heikkilä & Tarja Larsson: Om sverigefinskt terminologi- och ordlistearbete</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p>		
12.10	<table border="0" style="width: 100%;"> <tr> <td style="width: 50%; padding-right: 10px; vertical-align: top;"> <p>Christian-Emil Ore: Å ta Hans Ross på ordet – Norsk Ordbog (1895) og Aasens Norsk Ordbog (1873) som en helhet</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p> </td> <td style="width: 50%; padding-left: 10px; vertical-align: top;"> <p>Sofia Björklöf & Saarni Laitinen: Användning av en korpus och en enspråkig ordbok som bas för den nya estnisk-finska ordboken</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p> </td> </tr> </table>	<p>Christian-Emil Ore: Å ta Hans Ross på ordet – Norsk Ordbog (1895) og Aasens Norsk Ordbog (1873) som en helhet</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p>	<p>Sofia Björklöf & Saarni Laitinen: Användning av en korpus och en enspråkig ordbok som bas för den nya estnisk-finska ordboken</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p>
<p>Christian-Emil Ore: Å ta Hans Ross på ordet – Norsk Ordbog (1895) og Aasens Norsk Ordbog (1873) som en helhet</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p>	<p>Sofia Björklöf & Saarni Laitinen: Användning av en korpus och en enspråkig ordbok som bas för den nya estnisk-finska ordboken</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p>		
12.40	● Lunch		
13.45	<table border="0" style="width: 100%;"> <tr> <td style="width: 50%; padding-right: 10px; vertical-align: top;"> <p>Elina Heikkilä & Jarkko Kauppinen: Nedsättande ord i finländska nätordböcker – vilka, hur och varför?</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p> </td> <td style="width: 50%; padding-left: 10px; vertical-align: top;"> <p>Louise Holmer: Klickande, skrivande och onödigt vattendrickande. Verbalsubstantiv på -ande i ordböcker och korpusar</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p> </td> </tr> </table>	<p>Elina Heikkilä & Jarkko Kauppinen: Nedsättande ord i finländska nätordböcker – vilka, hur och varför?</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p>	<p>Louise Holmer: Klickande, skrivande och onödigt vattendrickande. Verbalsubstantiv på -ande i ordböcker och korpusar</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p>
<p>Elina Heikkilä & Jarkko Kauppinen: Nedsättande ord i finländska nätordböcker – vilka, hur och varför?</p> <p style="text-align: right;"><i>ordförande: Oddrun Gronvik</i></p>	<p>Louise Holmer: Klickande, skrivande och onödigt vattendrickande. Verbalsubstantiv på -ande i ordböcker och korpusar</p> <p style="text-align: right;"><i>ordförande: Charlotta af Hällström-Reijonen</i></p>		
14.30	● Avslutning <i>Sal 6</i> <i>slutar 15.00</i>		

