

Nordiske Studier i Leksikografi

14



# Nordiske Studier i Leksikografi

14

Rapport fra den 14. Konference  
om Leksikografi i Norden

Reykjavík 30. maj–2. juni 2017

Redigeret af

Ásta Svavarsdóttir  
Halldóra Jónsdóttir  
Helga Hilmisdóttir  
Þórdís Úlfarsdóttir

Skrifter udgivet af Nordisk Forening for Leksikografi

Skrift nr. 15

I samarbejde med  
Árni Magnússon-instituttet for islandske studier

Reykjavík 2018

*Nordiske Studier i Leksikografi 14*

© Nordisk Forening for Leksikografi og forfatterne, 2018

Redaktion: Ásta Svavarsdóttir, Halldóra Jónsdóttir, Helga Hilmisdóttir  
og Þórdís Úlfarsdóttir

Sats: Bessi Aðalsteinsson

Bidragydere til sats: Redaktørerne

Omslag: Ásta Svavarsdóttir og Bessi Aðalsteinsson

Tryk: Leturprent ehf

ISBN: 978-9979-654-48-3

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

Udgivet med støtte fra:

Nordplus Sprog

Stofnun Árna Magnússonar í íslenskum fræðum, Reykjavík



## Forord

Den 14. Konference om leksikografi i Norden blev afholdt i Reykjavík i dagene 30. maj til 2. juni 2017, arrangeret af Nordisk forening for leksikografi i samarbejde med Árni Magnússon-instituttet for islandske studier.

Konferencens tema var *Ordbøger og sproglige resurser* og de tre indbudte plenarforelæsere beskæftigede sig med dette fra hver sin synsvinkel. Det samme gjaldt mange af forelæserne i de to parallelle sessioner, mens problematikken i andre indlæg var af almen leksikografisk karakter. Ialt blev der holdt 31 forelæsninger på konferencen og desuden præsenteredes 6 poster. Det kan trygt hævdes at tilsammen vidner indlæggene om bredden og mangfoldigheden i nordisk leksikografi. Det faglige program fandt sted i Askja, Islands universitets smukke naturvidenskabernes bygning.

I tillæg til det faglige program arrangeredes et socialt program som anses som ikke mindre vigtigt i NFL-sammenhæng fordi det giver medlemmerne lejlighed til at møde kolleger og stifte nye bekendtskaber under uformelle forhold. Denne gang bestod denne del af programmet bl.a. af en reception i Nordens hus, en ekskursion med historiske, kulturelle og geologiske indslag rundt omkring på Suðurnes og en festmiddag i Iðnó, et gammelt teater og forlystelsessted midt i byen.

Konferencerapporten indeholder 29 artikler udarbejdet på grundlag af indlæg på konferencen, inklusive alle plenarforelæsningserne og de fleste andre foredrag og poster. Langt de fleste artikler er skrevet på et af de nordiske sprog dansk, norsk bokmål, nynorsk eller svensk, mens nogle er på engelsk. Alle artikler har et engelsk resumé. Bagest i rapporten trykkes conferenceprogrammet med en liste over de udstillede poster. Den lokale konferencekomité – Ásta Svavarsdóttir, Halldóra Jónsdóttir, Helga Hilmisdóttir og Þórdís Úlfarsdóttir – har også haft ansvaret for redigeringen af rapporten og Árni Magnússon-instituttet for islandske studier står for udgivelsen i samarbejde med NFL.

NFL-konferencen 2017 har modtaget økonomisk støtte fra mange instanser, både til arrangementet og til udgivelsen af konferencerapporten: Nordplus Sprog, Clara Lachmanns Fond, Letterstedtska föreningen og de nordiske ambassader i Reykjavík, foruden Árni Magnússon-instituttet for islandske studier.

Redaktørene takker alle som har bidraget til udgivelsen af denne rapport og retter en varm tak til forfatterne for et godt samarbejde under redigeringsfasen.

Reykjavík, maj 2018

*Ásta Svavarsdóttir  
Halldóra Jónsdóttir  
Helga Hilmisdóttir  
Þórdís Úlfarsdóttir*



# Indholdsfortegnelse

*Forord*

## Plenarforelæsere

*Jón Hilmar Jónsson*

*Íslenskt orðanet*: Tekstbasert kartlegging og presentasjon  
av leksikalske relasjonar 1

*Bolette Sandford Pedersen*

Semantisk processering og leksikografi 18

*Matthew Whelpton*

The Variability of Semantic Categories:  
An Experiment in Extensional Semantics 29

## Sessioner

*Sturla Berg-Olsen, Knut E. Karlsen & Dagfinn Rødningen*

Norsk historisk normeringsordbok 47

*Lars Brink*

Adjektiv eller adverbium? 56

*Rósa Elín Davíðsdóttir*

LEXIA – en islandsk-fransk online ordbog: Udfordringer og løsninger 63

*Kristine Gunn Eide*

Syntaktiske opplysninger i ettspråklige ordbøker  
med en gruppe intransitive verb som eksempler 71

*Oddrun Grønvik & Rune Kyrkjebø*

Språksamlingane og leksikografisk arbeid ved Universitetet i Bergen 80

*Oddrun Grønvik & Christian-Emil Smith Ore*

Bokmål og nynorsk samindeksert – Metaordboka som verktøy  
for jamføring og utforsking av ordtilfang 87

*Asgerd Gudiksen & Henrik Hovmark*

Omålsordbogen som leksikografisk, sproglig og kulturhistorisk resurse 96

<i>Anna Helga Hannesdóttir &amp; Benjamin Lyngfelt</i> Kontrastiv konstruktionsgrammatik och tvåspråkig lexikografi. Konstruktioner och konstruktionsekvivalenter i <i>Islex.se</i>	104
<i>Carsten Hansen &amp; Jakob Steensig</i> Ord i samtaler – findes de?	113
<i>Helga Hilmisdóttir &amp; Marjakaisa Matthíasson</i> Natur och digitala, tvåspråkiga ordböcker: En studie av finska som sökspråk i ISLEX	120
<i>Henrik Hovmark</i> Udflytning og nedarvning som formidlingsstrategi i ordbøger	129
<i>Jonas Jensen, Henrik Lorentzen, Sanni Nimb, Mette-Marie Møller Svendsen &amp; Lars Trap-Jensen</i> Thaipiger, muskelhunde og fulde svenskere: nedsættende ord, stereotyper og ligestilling i <i>Den Danske Ordbog</i>	141
<i>Ellert Þór Jóhannsson &amp; Simonetta Battista</i> Middelaldertekster som sproglig ressource	152
<i>Ellert Þór Jóhannsson, Simonetta Battista, Þorbjörg Helgadóttir &amp; Aldís Sigurðardóttir</i> Redigering af adjektiver i ONP Online	162
<i>Lennart Larsson</i> Får flickor vara vildbasare? Om könstillhörighet och genusangivelser i SAOL och hos Dalin	169
<i>Petri Lauerma</i> Finska språkets lånordförråd i ljustet av Martti Rapolas ordsamling	178
<i>Patrick Leroyer &amp; Henrik Køhler Simonsen</i> Sprogressourcer kontra ordbøger: Hvad er bedst?	186
<i>Pär Nilsson</i> Bildliga betydelser i SAOB	195
<i>Carina Nilstun</i> Orddannelsestypen konversjon i leksikografien	205



<i>Bodil Rosqvist &amp; Bo-A. Wendt</i> Gammalt projekt och nya resurser – SAOB:s hantering av nytilkomna elektroniska textmängder	214
<i>Jørgen Schack</i> Perfektum participium og ordklassen adjektiver	222
<i>Jóhannes B. Sigtryggsson</i> Nye islandske retskrivningsregler og den islandske retskrivningsordbog	230
<i>Emma Sköldberg &amp; Lena Wenner</i> Amatörlexikografiska insatser på sajten <i>Folkmun.se</i>	237
<i>Steinþór Steingrímsson, Sigrún Helgadóttir &amp; Eiríkur Rögnvaldsson</i> An Icelandic Gigaword Corpus	246
<i>Steinþór Steingrímsson, Ari Páll Kristinsson &amp; Halldóra Jónsdóttir</i> <i>Málið.is</i> : A Web Portal for Information on the Icelandic Language	255
<i>Ágústa Þorbergsdóttir</i> Den islandske termbanks forskellige resurser	262
<i>Konferenceprogram</i>	273



Plenarforelæsere



# Íslenskt orðanet: Tekstbasert kartlegging og presentasjon av leksikalske relasjonar

*Jón Hilmar Jónsson*

The period 1995 to 2005 saw the publication of three closely related dictionaries of Icelandic collocations, constructions and multi-word expressions. The topic of the first one was usage and variation in individual words, but the focus in the other two was semantic cohesion and semantic classification. The data collected for these dictionaries forms the basis of a new online dictionary, published under the title Íslenskt orðanet (The Icelandic Word Web). In this work, single- and multiword lemmas are combined into a list of semantically distinct lexical units. A multi-functional lexicographic description, showing the relations between entries, is created by analysing usage in electronic texts, paying special attention to coordinate constructions. The hypothesis is that the semantic relations between entries can be extracted from the syntactic and paradigmatic relations. Data on semantic proximity is collected by the use of sets of synonyms and classification of concepts, with attention to the similarity in usage between lexical units. This multifaceted approach is a combination of different aspects of lexicography, and the final result is that Íslenskt orðanet does the work traditionally divided between different types of dictionaries.

## 1. Innledning

Vi som nå tilhører den eldre generasjon av nordiske leksikografer, er så heldige å ha fått oppleve den revolusjon leksikografien og leksikografisk arbeid har gjennomgått de siste to årtiene, og for mange av oss har det vært lærerikt å få oppdage verdien av den digitale ordboken i praktisk ordboksarbeid. I år er det nettopp 20 år fra utgivelsen av *Nordisk leksikografisk ordbok* (NLO) der det leksikografiske begrepsinventar og terminologi ble presentert, bl.a. på bakgrunn av et da nylig utkommet internasjonalt hovedverk om leksikografien i tre store bind, *Wörterbücher* (1989–1991). I NLO får den digitale ordboken for så vidt en kort omtale under termen **elektronisk ordbok**, men samtlige termer og begreper tar ellers utgangspunkt i den trykte ordboken. De to sentrale nordiske håndbøkene om leksikografi, *Handbok i lexicografi* av Bo Svensén fra 2004 og *Ord og ordbøker* av Ruth Fjeld og Lars Vikør fra 2008, inneholder begge to et særskilt kapittel om elektroniske ordbøker, men heller ikke her blir det leksikografiske begrepsinventaret satt i en digital sammenheng. Det er godt forståelig med hensyn til at vi her er midt i en utvikling som Trap-Jensen (2014) nevner *den digitale revolusjon*, og som følger etter en annen revolusjon på dette feltet, *korpusrevolusjonen*. Men etter hvert er det blitt klarere at sentrale sider av leksikografien, ikke minst typologi

og strukturer, ikke lenger kan defineres uten hensyn til de krav og muligheter som følger av den digitale revolusjon.

I denne spennende tiden har det vært en interessant utfordring å få anledning til å videreføre et ordboksprosjekt som opprinnelig resulterte i en trykt ordbok med den hensikt å levere en omarbeidet og utvidet ordbok i digital form. Det blir mitt utgangspunkt i denne artikkelen, for de siste årene har jeg vært opptatt av å utvikle en digital ordbok som nå er blitt tilgjengelig på Árni Magnússon instituttets nettsted under overskriften Íslenskt orðanet. Denne ordboken bygger i høy grad på de ordbøker jeg arbeidet med i lang tid tidligere, og deres materiale, men den har samtidig fått et nytt og utvidet datagrunnlag i form av digitale tekster og korpusmateriale.<sup>1</sup>

Motivet for å utvikle dette ordboksprosjektet med sikte på en digital ordbok, henger sammen med den trykte ordbokens store begrensninger. Disse begrensningene vil gi ulike utslag i de enkelte ordbokstyper, men når det, som i dette tilfellet, dreier seg om ordbøker med en bred og sammensatt funksjon der fokuseringen gjelder levende språkbruk, blir de ganske tydelige. Lemmaseleksjonen blir sterkt avhengig av kravet om redusering og kondensering, og det samme gjelder for de mikrostrukturelle enhetene der lemmaenes kontekstuelle egenskaper blir behandlet og belyst. Dette går ikke minst ut over de språklige data som kan innhentes fra databaser og tekstsamlinger, og skal bidra til å gi et variert og detaljert bilde av de enkelte lemmaer. Dette forholdet førte i sin tid til den konklusjon at det genuine formålet ville kunne styrkes og utvides, og at det leksikografiske materialet prosjektet bygger på, ville kunne spille en mer aktiv rolle innenfor rammen av en digital ordbok der ulike informasjonstyper kan støtte og utdype hverandre, og der brukeren kan velge ulike innfallsvinkler til materialet.

Det er naturligvis et prinsipielt spørsmål om, og i tilfelle i hvilken grad, en ny digital ordbok skal bygge på materiale som tidligere er blitt strukturert og fremstilt innenfor rammen av et trykt ordboksverk. Det har bl.a. å gjøre med det forholdet at en trykt ordbok i hovedsak vil være knyttet til en bestemt funksjon, mens en digital ordbok i høy grad kan dekke ulike funksjoner og brukerbehov. Her kan det naturligvis spille en rolle hvilken ordbokstype det dreier seg om, og om struktureringen i den trykte ordboken lett kan modifiseres og tilpasses en mer fleksibel brukertilgang. Det kan bety at de komponenter som står sterkest i den trykte ordboken, ikke minst en utførlig og detaljert betydningsbeskrivelse, ikke automatisk vil ha en dominerende posisjon i en digital sammenheng.

Konstrueringen av en bredt sammensatt ordbok i digital form reiser spørsmålet om hvor mange og ulike funksjoner og brukerbehov en slik ordbok kan oppfylle. Svaret kompliseres når man innser at en digital ordbok ikke har noen klare

<sup>1</sup> Gjennom dette arbeidet har jeg fått uunnværlig støtte fra mine medarbeidere. Jeg vil spesielt fremheve at det datatekniske arbeidet som her er blitt gjennomført, har hatt en avgjørende betydning for prosjektet, der Ragnar Hafstað og Þórdís Úlfarsdóttir la grunnen, og der Bjarki Karlsson kom aktivt inn i den siste fasen med det oppdrag å konstruere en ny webside, og bidro med mange gode råd og løsninger.

grenser, og at den kontinuerlig kan utvides og kompletteres. Hermed vil det være avgjørende om brukeren går inn på å godkjenne den digitale ordbokens overflod og grenseløshet og ser på disse egenskapene som klare fordeler overfor den trykte ordbokens kortfattethet og begrensninger.

Artikkelen beskriver hvordan denne digitale ordboken er blitt utpreget, i hvilken grad den bygger på de eldre ordbøkene og deres materiale, og hvordan den form- og funksjonsmessig adskiller seg fra sine trykte forgjengere. Etter en grov presentasjon av de eldre ordbøkene blir det drøftet hvordan den digitale presentasjonsformen forutsetter en ny lemmamodell og en utvidet begrepsklassifisering. Deretter beskrives det hvordan den faktiske språkbruken i tekstsammenheng blir den sentrale ledetråden i analysen av ordboksmaterialet, og hvordan det preger den leksikografiske beskrivelsen der de enkelte lemmaene blir belyst i forhold til de ulike relasjoner de har til andre lemmaer. Dette relasjonsbildet deles opp i adskilte informasjonstyper som hver for seg får sin egen beskrivelse. Artikkelen avsluttes med et sammendrag der utviklingen av prosjektet blir skildret og vurdert.

## 2. Tre beslektede ordbøker

Íslenskt orðanet har sine hovedrøtter i arbeidet med tre beslektede ordbøker som utkom i perioden 1994–2005 (se Jón Hilmar Jónsson 2014). Det første produktet i dette arbeidet, *Orðastaður* (1994, 2. utvidet utgave 2001), kan karakteriseres som en islandsk konstruksjons- og kollokasjonsordbok som også inneholder en morfologisk komponent med sammensetninger. I alt dekker denne ordboken knapt 50 000 ordforbindelser og bortimot 100 000 sammensetninger, der brukertilgangen går via en lemmaliste med ca. 11 000 enkeltord. *Orðastaður* ble etterfulgt av *Orðaheimur* (2002), en ordbok som gir en begrepsklassifisert oversikt over islandsk fraseologi med spesiell vekt på idiomer og faste ordforbindelser. Her består hovedlemmalisten av drøyt 800 alfabetisk sorterte begrepslemmaer som da dekker de enkelte begrepsgrupper, men for å sikre en problemfri tilgang til den aktuelle informasjon, inneholder ordboken en alfabetisk indeks der samtlige innholdsord i frasene opptrer som lemmaer for de fraser de forekommer i. Her dreier det seg om drøyt 30 000 ordforbindelser, med en viss overlapping med konstruksjonsordboken.

De to ordbøkene er naturligvis nært beslektet, men ut fra brukernes forutsetninger er det en viktig forskjell. I kollokasjons- og konstruksjonsordboken gjelder oppslaget primært bruken og forekomsten av selve oppslagsordet i ulike leksikalske sammenhenger. I den fraseologiske begrepsordboken derimot er brukeren ute etter informasjon om leksikalsk og uttryksmessig variasjon i forhold til et gitt begrep eller en begrepssammenheng, og oppslagsordet blir oppfattet som et hensiktsmessig middel til å nå frem til den. Det dreier seg altså om den klassiske forskjellen mellom en semasiologisk og en onomasiologisk ordbok.

Bruken av en indeks over frasene i begrepsordboken gjorde det også mulig og

aktuelt å indeksere ordforbindelsene i kollokasjonsordboken, og man innså at en slik kombinasjon av de to så nært beslektede ordbøker, også kunne være fordelaktig for brukerne. Det førte til en ny utgave under tittelen *Stóra orðabókin um íslenska málnotkun* (2005) der de to ordbøkene ble slått sammen med en kombinasjon av ord- og begrepslemmaer og en felles fraseologisk indeks.

Samtidig med disse ordbøkene ble det bygget opp et omfattende register over ordforbindelser av ulike slag (*Skrá um orðasambönd*) som forekommer i bruks-eksempler fra skriftspråksarkivet til Orðabók Háskólans, i alt drøyt 130 000 forbindelser der det naturlig nok også fins en viss overlapping med de fraseologiske ordbøkene.

Under dette arbeidet ble det i økende grad dratt nytte av maskinell innsamling og behandling av leksikografiske data. Samtidig ble man klar over at en leksikografisk beskrivelse av denne typen, der fraseologi i vid forstand og begrepsmessige sammenhenger står i fokus, og der den ideale brukeren er ute etter informasjon om mangfoldet og dynamikken i bruken av ordene, krever en videre og mer fleksibel ramme.

### 3. Overføring til digital form

Arbeidet med en digital ordbok som bygger på, forener og kompletterer dette leksikografiske materialet og gir det en ny presentasjonsform, har nå pågått i ca. 10 år med varierende kraft og i varierende tempo. Det har resultert i webordboken *Íslenskt orðanet* som ble lansert i oktober 2016, og som i og for seg utgjør et ferdig produkt uten at selve ordboksarbeidet kan sies å være ferdig. Det forholdet er ikke ukjent fra andre ordbøker av denne typen der ordboksteksten gjenspeiler framdriften av en leksikografisk analyse av det innhentede materialet. Det må både leksikografen og brukerne finne seg i og faktisk glede seg over, for en digital ordbok krever en kontinuerlig oppdatering og fornyelse.

Man innså fort at i en digital fremstilling må det være aktuelt og hensiktsmessig å inkludere betydningsbærende fraser og ordforbindelser i selve lemmalisten, og gi dem dermed en mer fremtredende posisjon. Dermed blir de likestilt med enkeltordene og kan inngå i ulike relasjoner med dem. Det endrer også lemmalistens karakter for flerordslemmaene som semantisk entydige leksikalske enheter, vil gi tonen og kreve en entydiggjøring av samtlige lemmaer, dvs. også av ettordslemmaene.

Denne operasjonen førte til en kolossal utvidelse av lemmalisten som dermed omfattet drøyt 200 000 lemmaer, der bl.a. alle sammensetninger i den første ordboken var inkludert, samt alle innholdsord i ordforbindelser foruten en stor mengde flerordslemmaer. Men det betyr ikke at de alle sammen automatisk blir aktive lemmaer i *Íslenskt orðanet*, for her må de oppfylle visse krav om „aktivitet“, dvs. at de forekommer i de grupperinger og/eller relasjonstyper ordnettet er sammensatt av. I databasen fins det altså en god del „passive“ lemmaer som vil bli aktivisert så snart de får en relevant tilknytning.



For brukeren betyr dette at et enkelt søkeord i sin oppslagsform vil fremkalle en liste med alle de lemnaer det forekommer i. Som det fremgår av Figur 1 forener søkeordet *nótt* 'natt' en stor mengde flerordslemnaer der ordet (i ulike bøyingsformer) er inkludert, i alt omtrent 150 lemnaer. Her fins det ulike strukturer der verbaler og adverbialer av ulike slag dominerer. Hvis oppslaget gjelder et bestemt flerordslemma, vil det lett kunne lokaliseres takket være en fast sorteringsregel, men brukeren kan naturligvis også komme direkte frem til det tilsiktede lemmaet ved å inntaste en sammensatt streng.

<b>nótt</b> kvk	HUGTAK	langt fram á <b>nótt</b>
að <b>nóttu</b> til		láta <b>nótt</b> sem nemur
að <b>nóttunni</b>		láta þar <b>nótt</b> sem nemur
að úthallandi <b>nóttu</b>		leggja <b>nótt</b> með degi
aðra <b>nótt</b>		leggja <b>nótt</b> við dag
alla liðlanga <b>nóttina</b>		leggja saman <b>nótt</b> með degi
alla <b>nóttina</b>		leggja saman <b>nótt</b> og dag
		lengi <b>nætur</b>

Figur 1: Søkeordet *nótt* 'natt' forener alle de lemnaer der ordet (i varierende former) inngår som en del av lemmastrengen. Her vises det to bruddstykker av denne listen som i alt omfatter bortimot 150 lemnaer.

Flerordslemnaenes dominans er spesielt karakteristisk for verbene med alle sine ulike konstruksjonsmønstre og faste idiomer. Figur 2 viser noen bruddstykker av den lemmalisten søkeordet *sitja* 'sitte' fremkaller, i alt drøyt 300 lemnaer.

<b>sitja</b> eftir	<námið> <b>situr</b> á hakanum
<b>sitja</b> eftir með sárt ennið	<skreytingin> <b>situr</b> <vel, rétt>
<b>sitja</b> eins og aða í leiru	<stjórnin> <b>situr</b> að völdum
<b>sitja</b> eins og assa	<stjórnin> <b>situr</b> við stýrið
<b>sitja</b> eins og á nálaroddum	<vinnan> <b>situr</b> í fyrirrúmi
<b>sitja</b> eins og brúða	

Figur 2: Søkeordet *sitja* 'sitte' forener alle de lemnaer der ordet inngår som en del av lemmastrengen. Her vises det bruddstykker av denne listen som i alt omfatter drøyt 300 lemnaer.

Lemmatismen av betydningsbærende fraser og konstruksjoner krevde en systematisk og standardisert fremstilling som for en stor del kunne ledsages av det arbeidet som ble gjennomført i de trykte ordbøkene. Men omstruktureringen gjelder ikke bare flerordslemnaene, for grunntanken i det konseptet som ligger bak ordnettet, er å få til et aktivt samspill mellom de ulike komponenter det er sammensatt av. Forutsetningen er at syntagmatiske og dels også morfologiske relasjoner gjenspeiler semantiske relasjoner mellom de leksikalske enhetene. Dette

relasjonsforholdet kommer for så vidt allerede til uttrykk i ulike sammenhenger i de trykte ordbøkene, der leksikalsk variasjon i konstruksjonsmønstre og sammen-setninger forener semantisk beslektede ord.

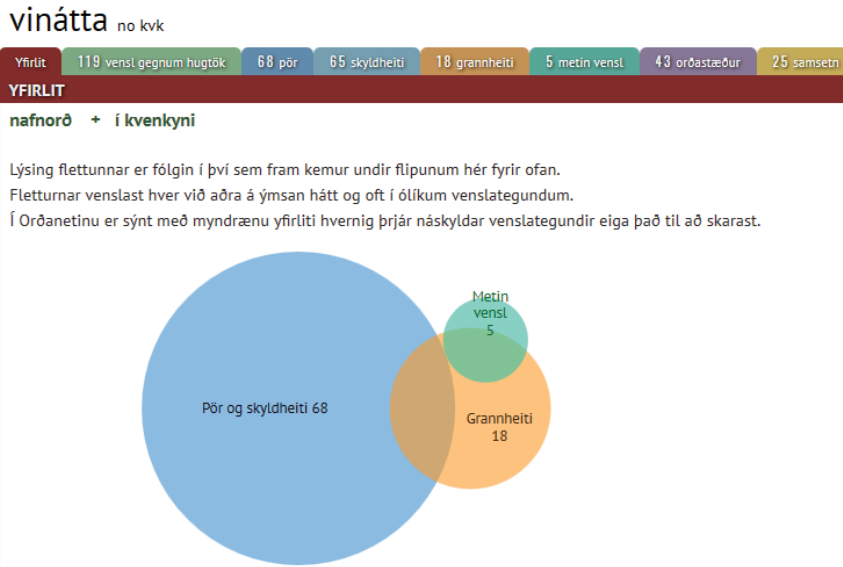
På bakgrunn av disse overveielserne ble det en sentral oppgave i den første fasen å utvide begrepsklassifiseringen, både i form av en bredere begrepsdekning og en bredere inkludering av lemmaene, også med hensyn til enkeltordene. Et viktig ledd i denne oppgaven besto i å forsyne flerordslemmaene med grammatiske tagger for dermed å komme øye på semantisk beslektede lemmaer av samme strukturmønster.

Ved den første gjennomgangen av materialet i ordnettet ble det foretatt en omfattende markering av synonymer, og da også blant flerordslemmaene der en felles argumentstruktur er et viktig kriterium. Synonymgruppering er en sentral del av den semantiske kartleggingen, men problemet er at den bygger for det meste på skjønn og intuisjon uten direkte støtte fra den faktiske språkbruken, og at det kan bli vanskelig å unngå at flertydighet blant lemmaene har en forstyrrende effekt. I tillegg kommer at en synonymgruppering bare vil dekke en del av lemmaene. Dermed var det et klart behov for å innlede en framgangsmåte som kunne forene synonyme og andre nært beslektede leksikalske enheter ut fra en konkret innsikt i deres opptreden i faktisk språkbruk.

#### 4. Lemmaenes relasjoner som adskilte informasjonstyper

Overføringen av ordboksmaterialet til digital form åpnet veien for gjennomgripende strukturendringer og viktige innovasjoner i selve ordboksarbeidet som samtidig har gitt brukerne en utvidet og mer systematisk tilgang til lemmaene. Hensikten er å belyse lemmaenes karakter ved å kartlegge deres forhold og relasjoner til andre lemmaer for dermed å skaffe seg et bilde av leksikalsk sammenhengende felt. Her spiller lemmatiseringen av flerordsenhetene en avgjørende rolle, men en utvidet begrepsklassifisering og en systematisk tagging av flerordslemmaene er også av stor betydning.

Som det fremgår av Figur 3 består denne nye inngangen til lemmaene av opptil åtte overskrifter som utgjør ulike innfallsvinkler til lemmaets plassering og relasjoner. Beskrivelsen av lemmaet *vinátta* 'vennskap' innledes med et diagram som viser størrelsesforhold og en eventuell overlappning mellom tre sentrale informasjonstyper. Ellers vil brukeren gå videre ved å velge en relevant informasjonstype.



Figur 3: Et oversiktsskilt med adskilte informasjonstyper utgjør inngangen til beskrivelsen av de ulike lemnaer i Íslenskt orðanet. Tallene ved de enkelte typer viser hvor mange lemnaer som er relatert til lemnaet *vinátt* 'vennskap'. Diagrammet viser størrelsesforhold og overlappning mellom tre sentrale informasjonstyper.

Blant disse overskriftene dreier det seg dels om relasjonar som allerede står sentralt i de trykte ordbøkene, de som kommer til uttrykk i kollokasjoner og sammensetninger. Når det gjelder de begrepsmessige relasjonene, får de økt tyngde og en mer nyansert kartlegging ved å bli sett i sammenheng med faktisk språkbruk, noe som gir en ny innsikt i lemnaenes plassering i forhold til beslektede lemnaer. I tillegg kommer grammatisk markering av flerordslemnaene i form av syntaktiske tagger.

#### 4.1. Tekstbaserte relasjonar

Man innså snart at den leksikografiske beskrivelsen ville styrkes ved å gi ordnes opptreden i tekstuell sammenheng en mer aktiv og synlig rolle for dermed å avdekke leksikalske relasjonar og vise hvordan de enkelte lemnaer er plassert i forhold til de lemnaer som opptre i en samme eller liknende semantisk rolle. For å oppnå dette har ordnettet trukket inn materiale fra nasjonalbibliotekets store digitale tekstsamling, *Tímarit.is*, med tekster fra islandske aviser og tidsskrifter, og det har også dratt nytte av en grammatisk markert korpus med tekster av blandet innhold fra en tiårsperiode etter 2000, *Mörkuð íslensk málheild*. En mer effektiv tilgang til tekstene og greiere søkerutiner ville være ønskelig for å få til en bedre

dekning av lemmaene. Men uansett vil det alltid være tidkrevende å innhente og analysere språklige data for å belyse lemmaenes semantiske karakter.

De tekstbaserte data fra disse digitale kildene har i økende grad preget ordnettets innhold og karakter. De får også en fremtredende plass i det sammensatte bildet av informasjonstyper brukeren står overfor ved de enkelte oppslag.

#### 4.1.1. Parataktiske forbindelser

Den informasjonstype som er sterkest forankret i faktisk språkbruk og som samtidig er mest karakteriserende for ordnettet, har den islandske overskriften *Pör* ((ord-)par). Her dreier det seg om en leksikalsk relasjon som gjennom oppbyggingen av ordnettet har fått økt oppmerksomhet og gjelder lemmaenes forekomst i parataktiske forbindelser, forbindelser som *svefn og hvíld* 'søvn og hvile', *hvíla sig og sofa* 'hvile og sove', *áhugasamur og bjartsýnn* 'interessert og optimistisk', *erfiður og óþægilegur* 'vanskelig og ubehagelig'. Utgangspunktet er at slike forbindelser, foruten i svært mange tilfeller å forbinde semantisk beslektede leksikalske enheter, i forhold til de enkelte lemmaer vil avdekke paradigmatisk relasjoner som i høy grad gjenspeiler semantiske likheter, der det konstante leddet virker som en magnet for partnere av samme semantiske karakter. De ulike partnere vil også gjerne uttrykke konnotative og modifierende betydningselementer i det ord eller den ordforbindelse de står sammen med, og på den måte ha en leksikografisk betydning.

Det innbyrdes forholdet mellom partnerne i en parataktisk forbindelse varierer. Det kan simpelthen dreie seg om en konstellasjon av ulike referenter: *borð og bekkir* 'bord og benker', *glös og bollar* 'glass og kopper'. I leksikografisk sammenheng er de forbindelser mer interessant som på en presiserende måte forener semantisk nær beslektede ord: *daufur og fálátur* 'sløv og stille', *nálægð og samvera* 'nærhet og samvær'. Her er den innbyrdes rekkefølgen nokså fri, mens den vil være mer eller mindre bunden når forbindelsen gjenspeiler en logisk prosess eller utvikling: *gamall og farlama* 'gammel og bevegelseshemmet', *atvinnuleysi og matarskortur* 'arbeidsløshet og matknapphet'.

Helhetsbildet av partnerne til et gitt lemma vil gjerne utdype dets semantiske karakter ved å trekke fram konnotative betydningselementer. Adjektivet *einmana* 'ensom' er f.eks. knyttet til partnere som *dreyminn* 'drømmende', *óframfærinn* 'tilbakeholden', *bældur* 'hemmet' og *aumkunarverður* 'ynkelig', som alle sammen fremhever aktuelle egenskaper til ordet *einmana*. Substantivet *mælska* 'veltalenhet' er på samme måte knyttet til ord som *ákafi* 'iver', *fyndni* 'humor' og *innblástur* 'inspirasjon', som i sin tur kan anses å være integrerte elementer i ordet *mælska*.

De parataktiske forbindelsene utgjør dermed en viktig ledetråd ved kartleggingen av lemmaenes innbyrdes relasjoner. Men det som er utslagsgivende for deres sentrale posisjon i ordnettet, er at de forekommer sammen med så vel enkeltord som flerordsforbindelser, og dermed får de en samlende og koordinerende funk-

sjon i en ordboksmodell der enkeltord og betydningsbærende ordforbindelser er likestilt som lemnaer.

#### 4.1.2. Leksikalsk samhörighet

Under overskriften *Pör* gis det en oversikt over lemnaets forekomst i parataktiske forbindelser i den grad det allerede er blitt registrert. Ved å kartlegge forekomsten av samtlige relaterte lemnaer i samme type forbindelser gis det informasjon om antall felles partnere med hvert enkelt lemma og dermed en viss pekepinn om lemnaenes nærhet. Her kommer det paradigmatisk aspektet altså inn i bildet, og resultatet av analysen blir synliggjort under overskriften *Skyldheiti* (Leksikalsk samhörighet) i form av en sortert liste med samtlige lemnaer vedkommende lemma knyter seg til i slike forbindelser.

Som det fremgår av Figur 4 kan listen sorteres på to ulike måter, etter antall felles partnere i denne type forbindelser, eller proporsjonelt etter hvor stor del av de partnere det tilknyttede lemnaet omfatter, er felles. Dette blir også synliggjort i form av et diagram der de felles partnere også blir listet opp som direkte lenker. Analysen viser at lemnaet *einmanaleiki* 'ensomhet' har de fleste felles partnere (95) med *einsemd* 'ensomhet' og deretter følger *vonleysi* 'håpløshet', *vanliðan* 'ubehag', *angist* 'sterk angst', *uppgjöf* 'resignasjon' og *depurð* 'nedstemthet'. En proporsjonal vurdering gir et annet bilde, der bl.a. ord som *tilgangsleysi* 'meningsløshet', *tómleiki* 'tomhet' og *söknuður* 'savn, lengsel' står relativt høyt.

#### Skyldheiti 156

RÖÐUN: -		Fjöldi sameiginlegra félaga	
Sameiginlegir félagar	Hlutfall	Myndrit	Skyldheiti
95 (af 173)	54,9%		<a href="#">einsemd</a>
81 (af 209)	38,8%		<a href="#">vonleysi</a>
81 (af 384)	21,1%		<a href="#">vanliðan</a>
58 (af 140)	41,4%		<a href="#">angist</a>
53 (af 215)	24,7%		<a href="#">uppgjöf</a>
51 (af 105)	48,6%		<a href="#">depurð</a>
49 (af 141)	34,8%		<a href="#">lífsleiði</a>
49 (af 86)	57,0%		<a href="#">dapurleiki</a>
49 (af 140)	35,0%		<a href="#">sektarkennd</a>
47 (af 195)	24,1%		<a href="#">firring</a>
44 (af 203)	21,7%		<a href="#">rótleysi</a>
42 (af 85)	49,4%		<a href="#">sorg</a>

Figur 4: Lemnaet *einmanaleiki* 'ensomhet' inngår i paradigmatisk relasjon til andre lemnaer gjennom felles partnere i parataktiske forbindelser. Her bestemmer antallet felles partnere (kolonne 1) rekkefølgen av de relaterte lemnaene. Kolonne 2 viser proporsjonen av de parataktiske forbindelser med vedkommende lemma ordnettet inneholder. Kolonne 3 gir adgang til en liste med de konkrete partnere.

Denne analysen gir den leksikografiske beskrivelsen et dynamisk preg og plasserer de enkelte lemmaer i forhold til et større semantisk felt de tilhører, eller i hvert fall knytter seg til. Innhentingen av de parataktiske forbindelsene er også et aktivt middel til en målrettet komplettering av lemmalisten der nye lemmaer, så vel enkeltord som flerordsenheter, umiddelbart får rotfeste og legitimering gjennom en partner som allerede tilhører lemmalisten, samtidig som eksisterende lemmaer får nye belegg og et klarere bilde.

#### 4.1.3. Kollokasjoner

Kollokasjoner, med den islandske overskriften *Orðastæður*, utgjør en selvstendig informasjonstype i ordnettet. De er for en stor del hentet fra den første trykte ordboken, *Orðastaður*, og ordforbindelsesregisteret til *Orðabók Háskólans*, og har til funksjon å demonstrere typisk og karakteriserende bruk av vedkommende lemma. Men de bidrar også med materiale til en semantisk gruppering av det ordtilfanget lemmaene omfatter i form av kollokasjoner og spiller dermed en aktiv rolle i den semantiske analyse (jf. avsnitt 4.3). Kollokasjonene blir fremstilt som en ordnet liste der hver enkelt kollokasjon blir lenket til andre aktuelle lemmaer.

Selv om ordnettet ivaretar den funksjon å gi oversikt over kollokasjoner, inntar de her en forholdsvis beskjeden plass. Det beror bl.a. på at den informasjon de formidler, er dels inkorporert i flerordslemmaene og de tekstbaserte enhetene som opptre som konkrete eksempler på de parataktiske forbindelser.

#### 4.2. Skjønnsbaserte relasjoner

Kartleggingen av de paradigmatiske relasjonene henger nøye sammen med den informasjonstype som dekkes av overskriften *Metin vensl* (Skjønnsbaserte relasjoner), der brukeren har tilgang til den klassifisering og gruppering som er blitt foretatt under redigeringen av materialet og er basert på redaktørens innsikt og skjønn. Denne overskriften dekker i første rekke synonymymer (*Samheiti*) i tradisjonell forstand, men her finner vi også antonymer (*Andheiti*) og redaksjonelle henvisninger (*Stíklur*) til beslektede eller på en annen måte aktuelle lemmaer. Men synonymene står her sentralt, og der vil man naturlig nok støte på lemmaer som foruten å anses å være synonym til et gitt lemma, samtidig opptre sammen med det i en parataktisk forbindelse. En slik overlapping er ganske vanlig og i og for seg også naturlig i ordnettet med hensyn til ulike innfallsvinkler til grupperingen og analysen. Det er derimot viktig å understreke at der det dreier seg om absolutt synonymi, vil en parataktisk forbindelse neppe være inne i bildet. Ordet *bill* 'bil' vil neppe forekomme i forbindelsen *\*bifreiðar og bilar*, og på samme måte virker *deyja* 'dø' malplassert i forbindelsen *\*deyja og drepast*. Det samme gjelder for forbindelser med varierende støtteverb: *\*grípa í taumana og taka í taumana* 'gripe inn'. I ordnettet kommer de skjønnsbaserte synonymene sterkt inn når det gjelder

et forholdsvis sjeldent ordforråd, sammensetninger med felles ordledd (ord som *valdafíkn* 'maktbegjær' med synonymene *valdafýsn*, *valdagirnd*, *valdagræðgi* og *valdasýki*) og flerordsforbindelser med utbredt leksikalsk variasjon (*setjast að* <þar> 'bosette seg <der>' med synonymer som *búsetja sig* <þar>, *festa byggð* <þar>, *setja búðir sínar* <þar> og *taka sér bólfestu* <þar>). Her vil parataktiske forbindelser være mindre aktuelle som direkte bindeledd.

I ordnettet spiller de skjønnsbaserte synonymene en viktig rolle ved kartleggingen av de semantiske relasjoner. Det er verdt å påpeke at som entydige leksikalske enheter vil de kunne komme klarere til uttrykk enn i en tradisjonell synonymordbok der lemmaenes flertydighet kan ha en forstyrrende effekt.

### 4.3. Tekst- og skjønnsbaserte relasjoner

Ved å fremheve det tekstbaserte aspektet gir ordnettet et mer nyansert bilde av lemmaenes semantiske karakter enn det er realistisk å få til gjennom en grov begrepsklassifisering og synonymgruppering alene. Hensikten er at brukeren, ut fra sin innsikt og på bakgrunn av analysen, kan vurdere hvilke lemmaer som er nærmest beslektet med et gitt lemma.

Dette er motivet for den informasjonstype i ordnettet som har overskriften *Grannheiti* der det dreier seg om en skjønnsbasert vurdering og gruppering av de forbindelser som er innhentet som tekstbaserte enheter (parataktiske forbindelser og kollokasjoner) under de enkelte lemmaer. Her blir forbindelser med synonyme eller nært beslektede partnere gruppert sammen for å inngå i større grupperinger der vedkommende enheter forekommer.

Slike konstallasjoner blir gjennomgått maskinelt for å gi et samlet bilde av de lemmaer som på denne måten forekommer sammen med et gitt lemma. Her kommer synonymi ganske sterkt til uttrykk, og for en stor del dreier det seg om nært beslektede lemmaer som ikke kommer inn i det semantiske blikkfeltet på en annen måte. Lemmaene blir sortert etter antall samforekomster, uten at det har en avgjørende betydning, for én forekomst er også fullt legitim. Figur 5 viser en del av det ordtilfanget som på denne måten står i en synonym relasjon til lemmaet *eigingirni* 'egoisme'. I alt viser *eigingirni* 25 „grannheiti“ der *síngirni* og *sjálfselska* forekommer i de fleste grupperinger (hhv. 14 og 13). Til nærmere orientering gis det tilgang til de forbindelser som ligger bak vedkommende sammenkobling.

## Grannheiti 25

RÖÐUN: - Fjöldi merkingarvensta +

Tilvik	Sjá tilvik	Grannheiti
14		síngirni
13		sjálfselska
9		sérhyggja
7		sérdrægni
7		sérgæska
5		sérplægni
4		sérgæði
4		sjálfshyggja
3		eigingirnd
3		sjálfsdekur

Figur 5: Informasjonstypen „Grannheiti“ bygger på en gruppering av synonyme eller nært beslektede partnere til de enkelte lemmaer i parataktiske forbindelser eller kollokasjoner. Forekomsten av enkelte lemmaer som partnere blir gjennomgått maskinelt for å generere en frekvenssortert liste over alle de lemmaer et gitt lemma blir tilknyttet. Lemmaet *eigingirni* ’egoisme’ blir på denne måten knyttet til 25 „synonyme“ lemmaer, der *síngirni* og *sjálfselska* forekommer i de fleste grupperinger. Kolonnen i midten gir adgang til den konkrete tilknytningen.

Etter som den semantiske analysen har gått fremover, har denne informasjonstypen fått en mer fremtredende plass. En slik kombinasjon av en tekstbasert kartlegging og en skjønnsbasert behandling og vurdering av det materialet den bidrar med, gir faktisk et høyt interessant resultat i en ordboksmodell der lemmaene og deres posisjon kan betraktes og vurderes ut fra ulike synsvinkler. Her gjenspeiles ganske tydelig hvor viktig det er å få til en strukturering der de enkelte lemmaer og de forbindelser de omfatter, på en interaktiv måte bidrar med materiale som er med på å belyse forhold som kan gjelde fjerntliggende deler av lemmabestanden. Her inntar de parataktiske forbindelsene en sentral posisjon og bidrar til at brukeren i mange tilfeller vil kunne få et variert og sammensatt bilde av lemmaenes semantiske egenskaper og relasjoner ved å sammenlikne tre adskilte men beslektede informasjonstyper. Denne tilnæringsmåten understreker også hvor viktig det er å ta mer hensyn til i hvor høy grad betydningen av de leksikalske enheter karakteriseres av uklare grenser og semantisk vaghet (se Fjeld & Vikør 2008:80–84). Det kan den digitale ordboken demonstrere på en hensiktsmessig og klargjørende måte, mens den trykte ordboken tenderer til å vise klare semantiske grenser.

Det gjenstår mye arbeid før denne informasjonstypen har fått den tilskattede betydning i ordnettet. Men den vil helt sikkert få økt vekt og vise seg å være fordelaktig sammenliknet med en mer tradisjonell og „manuell“ synonymgrup-



pering. Det beror bl.a. på at her kommer tekstfrekvens inn i bildet som et ekstra og veiledende element.

#### 4.4. Morfologiske relasjoner i sammensetninger

Blant ettordslemmaene utgjør sammensetninger en selvstendig informasjonstype. Her gis det oversikt over lemmaets sammensetninger der både selve sammensetningene og de enkelte sammensetningsledd utgjør lenker til vedkommende lemma, så fremt de tilhører den aktive lemmalisten. Denne komponenten er opprinnelig hentet fra *Orðastaður* (1994) der sammensetningene faktisk får større oppmerksomhet ved delvis å bli gruppert etter semantisk slektskap.

#### 4.5. Syntagmatiske relasjoner i flerordslemmaer

Flerordslemmaer utgjør nesten halvparten av lemmabestanden i ordnettet, omkring 100 000 lemmaer. Deres karakter og mangfoldighet kommer klarest til uttrykk gjennom selve lemmalisten der et enkelt søk kan fremkalle en lang liste med lemmaer som inneholder søkeordet. En fast sorteringsregel er med på å fremheve strukturelle likheter innenfor en slik liste. For en stor del dreier det seg om verbale forbindelser, men adverbialer og nominaler inntar også en stor plass foruten adjektiviske ledd. Strukturelt får vi her en stor variasjon som er blitt kartlagt i form av grammatiske tagger, noe som gir et interessant klassifiseringsgrunnlag i ulike sammenhenger, også utenfor ordnettet. I ordnettet blir denne markeringen synliggjort som en særskilt informasjonstype under overskriften *Setningargerð* (Syntaktisk struktur) der det gis oversikt over og umiddelbar tilgang til samtlige lemmaer av vedkommende strukturtype.

#### 4.6. Begrepsmessige relasjoner og begrepsklassifisering

Jeg kommer da til den mest overgripende informasjonstypen som dekker størstparten av de aktive lemmaene, dvs. begrepsklassifiseringen, som i ordnettet har den islandske overskriften *Vensl gegnum hugtök* (Relasjoner gjennom begreper). I de ordbøker ordnettet bygger på, inngår begrepsoverskriftene i den alfabetiske lemmalisten sammen med ordlemmaene. Der kreves det en komplett indeksering av ordforbindelsene for å sikre en effektiv tilgang til det tilsiktede begrepet i lemmalisten ut fra hvilken som helst forbindelse begrepet dekker. I ordnettet gjelder i og for seg det samme prinsippet, men her foretar man direkte søk i lemmalisten for å komme på det riktige sporet. Begrepsoverskriftene fungerer samtidig som ordlemmaer, men de får et særskilt symbol som markerer at vedkommende lemma også har en begrepsfunksjon.

Det er ennå ikke blitt foretatt en helhetlig og systematisk begrepsklassifisering av lemmabestanden i ordnettet på den måte vi kjenner til fra de tradisjonsrike

begrepsordbøkene. Vi er ganske nylig, ved utgivelsen av *Den Danske Begrebsordbog* (2014), blitt minnet om hvor levedyktig denne tradisjonen er. En slik systematisering er naturligvis fordelaktig på mange måter, og hensikten er å gi den større oppmerksomhet i ordnettet. Grunntanken er likevel at det ikke er det samme behovet for et gjennomtenkt begrepssystem i en digital ordbok av denne type, der lemmaenes interaktive relasjoner spiller en så sentral rolle. I alt inneholder ordnettet knapt 5 000 begrepsoverskrifter som sammenlagt dekker bortimot 150 000 lemmaer, der de abstrakte begrepene generelt vil være bedre dekket enn de konkrete. Her gjenstår det altså en god del arbeid før begrepsklassifiseringen kan sies å ha fått sin endelige karakter.

**VINÁTTA (119)**

**Flettur undir hugtakinu**

- Óháð orðflokkum 119 +

afla sér vína  
aldavinátta  
aldavinkona  
aldavinur  
alúðarvinátta  
alúðarvinur  
bernskuvínur  
binda vínáttu við <hann, hana>  
bindast vínáttuböndum við <hann, hana>

**Venslaðar flettur í gegnum pör**

- Óháð orðflokkum 275 +

aóðáandi  
aóðáun  
aóstandandi1  
aóstandandi2  
afvopnun  
alúð  
andstæðingur  
áhangandi  
áhugamaður1

**Tengd HUGTÖK**

- Raðað eftir vægi 137 +

FÉLAGSTENGL 75  
ÆTTARTENGL 56  
SAMKENND 40  
STUÐNINGUR 29  
VÆNTUMPYKJA 27  
FJÖLSKYLDA/FJÖLSKYLDUHAGIR 25  
BÚSETUTENGL 23  
VINSEMD 23  
KUNNINGSSKAPUR 22

Figur 6: Lemmaenes begrepsklassifisering blir delt opp i tre spalter. Spalten til venstre inneholder lemmaer (i alt 119) som dekkes av begrepsoverskriften VINÁTTA (Vennskap). Spalten i midten utvider feltet ved å trekke inn parataktiske partnere til de lemmaene som tilhører begrepet. Spalten til høyre inneholder en verdisortert liste med lenker til beslektede begreper.

Det som særpreger ordnettet i denne forbindelse, er at begrepsklassifiseringen kan bli utvidet i et aktivt samspill med de parataktiske forbindelser lemmaene i den aktuelle begrepsklasse forekommer i. Som det fremgår av Figur 6 deles beskrivelsen opp i tre spalter. Den første spalten inneholder de lemmaer som direkte faller under begrepet VINÁTTA (Vennskap), der brukeren kan se hele ordtilfanget i alfabetisk rekkefølge eller velge en bestemt ordklasse. Spalten i midten utvider feltet ved å trekke inn parataktiske partnere til de lemmaene som tilhører begrepet. Det ordtilfanget er naturligvis av ulik karakter, dels står det klart utenfor begrepet, i andre tilfeller dukker det opp lemmaer der betydningen har klare berøringspunkter med vedkommende begrep. På denne måten får begrepsklassifiseringen en mer aktiv funksjon som en integret del av det leksikografiske konseptet.

Denne tanken om å utvide synsfeltet på en logisk og målbevisst måte, gjelder også for innholdet i den tredje spalten der vi har lenker til beslektede begreper. Som veiledning blir begrepene verdisortert, dvs. de begrepene som anses å ha mest felles med vedkommende begrep, står øverst. Ved sorteringen tas det hensyn til to faktorer, for det første at vedkommende begrep dekker en del av lemmaene

i den første spalten, og for det andre at de inneholder lemmaer som parataktisk er knyttet til disse lemmaene. En beregningsmodell som bygger på disse faktorene bestemmer da den innbyrdes rekkefølgen. I samsvar med det blir begrepene FÉLAGSTENGL (Sosial kontakt), ÆTTARTENGL (Slektskapsbånd), SAMKENND (Medfølelse) og STUÐNINGUR (Støtte) ansett å være nærmest beslektet med begrepet VINÁTTA. Denne begrepslisten vil i og for seg og til en viss grad kunne erstatte en systematisk oversikt over begrepene, for den gir brukeren alltid en viss orientering om det aktuelle og relevante begrepskomplekset.

## 5. Sammendrag

Arbeidet med de leksikografiske data Íslenskt orðanet bygger på, har foregått i mange og ulike faser med skiftende fokusering. I den første fasen gjaldt oppmerksomheten ordenes syntagmatiske relasjoner i konstruksjoner og kollokasjoner samt de morfologiske forbindelser i sammensetninger. Ordenes opptreden ble skildret ut fra et semasiologisk perspektiv og brukertilgangen var begrenset til en alfabetisk lemmaliste med ettordslemmaer. I neste fase kommer betydningsbærende ordforbindelser inn i bildet, og det semantiske aspektet får økt vekt. Lemmalisten blir utvidet med særskilte begrepslemmaer som gir adgang til en omfattende begrepsklassifisering av ordforbindelsene. Brukernes tilgang til ordboksteksten blir sikret gjennom en alfabetisk indeks med ettordslemmaer som i betydelig grad avdekker leksikalske og syntaktiske mønstre.

Det digitale skiftet markerer det avgjørende skillet i denne utviklingen og gjør det aktuelt å gi betydningsbærende ordforbindelser en mer fremtredende funksjon i form av flerordslemmaer. Deres presentasjonsformer blir standardisert, og de får en grammatisk markering i form av syntaktiske tagger slik at de kan sorteres på syntaktisk grunnlag. Samtidig får det semantiske aspektet økt oppmerksomhet, både gjennom en utvidet begrepsklassifisering av så vel ettords- som flerordslemmaer og en omfattende synonymgruppering.

I den siste fasen trekkes det inn nytt tekstmateriale i betydelig omfang for å underbygge en tekstbasert analyse av lemmaenes syntagmatiske, paradigmatiske og semantiske relasjoner og deres innbyrdes samspill. Målet er å få til en mer fin-kornet analyse av synonymi og semantisk nærhet ved å kartlegge lemmaenes opptreden i parataktiske forbindelser og kollokasjoner. For å få dette realisert kreves det en forbedret brukertilgang der de enkelte relasjonstyper blir tilgjengelige og kan vurderes i forhold til hverandre, og der søkeordet i betydelig grad kan forene leksikalsk sammenhengende lemmaer.

Oppmerksomheten i begynnelsen omkring ulike typer ordforbindelser har betydd at de fortsatt inntar en sentral plass som leksikalske enheter i ordnettet. Deres styrke ligger ikke minst i hvordan de på avgjørende punkter har preget utformingen av den leksikografiske beskrivelsen. For det første er de, som generelt semantisk entydige enheter, toneangivende ved entydiggjøringen av lemmaene, og

dermed blir det onomasiologiske perspektivet prioritert. For det andre motiverer de til å trekke inn og analysere de parataktiske forbindelsene, med hensyn til at de karakteriserer bruken av både enkeltord og betydningsbærende ordforbindelser. Det har i sin tur satt et sterkt preg på lemmalisten og innhentingen av nye lemmaer som i høy grad foregår i forbindelse med kartleggingen av de leksikalske relasjoner. En slik lemmaliste er åpen for kontinuerlig komplettering uavhengig av eldre krav om leksikaliseringsgrad, frekvens eller andre begrensende faktorer.

I Íslenskt orðanet blir den digitale ordbokens grenseløshet og kompletteringsbehov ansett som en positiv egenskap som er med på å gjenspeile åpenheten og dynamikken i den faktiske språkbruken. Den oppfatning innebærer en appell til brukernes innsikt og bedømmelsesevne, samtidig som den gir uttrykk for at en ferdig og avsluttet leksikografisk beskrivelse er mer et ideal enn en virkelighet.

## Litteratur

### Ordbøker og digitale ressurser

*Den Danske Begrebsordbog* (2014). Hovedredaktør: Sanni Nimb, Prosjektleder: Lars Trap-Jensen, Redaktører: Henrik Lorentzen, Liisa Theilgaard, Thomas Troelsgård. Det Danske Sprog- og Litteraturselskab.

*Íslenskt orðanet*. <<http://ordanet.arnastofnun.is>> (26.9.2017).

*Mörkuð íslensk málheild*. Stofnun Árna Magnússonar í íslenskum fræðum. <[www.arnastofnun.is/page/morkud\\_islensk\\_malheild](http://www.arnastofnun.is/page/morkud_islensk_malheild)> (26.9.2017).

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

Orðaheimur (2002) = Jón Hilmar Jónsson: *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandaskrá*. Reykjavík: JPV útgáfa.

Orðastaður (1994) = Jón Hilmar Jónsson: *Orðastaður. Orðabók um íslenska málnotkun*. Reykjavík: Mál og menning. [2. utgave 2001: JPV útgáfa.]

*Skrá um orðasambönd*. Stofnun Árna Magnússonar í íslenskum fræðum. <[www.arnastofnun.is/gagnasofn\\_ordasambond](http://www.arnastofnun.is/gagnasofn_ordasambond)> (26.9.2017).

Stóra orðabókin (2005) = Jón Hilmar Jónsson: *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.

*Tímarit.is*. Landsbókasafn Íslands – Háskólabókasafn. <<https://timarit.is/>> (23.3.2018).

## Annen litteratur

- Fjeld, Ruth Vatvedt & Lars S. Vikør (2008): *Ord og ordbøker*. Kristiansand: Høyskoleforlaget.
- Jón Hilmar Jónsson (2014): Oppbyggingen av en relasjonsbasert islandsk ordboksbase. I: *LexicoNordica* 21, 39–59.
- Svensén, Bo (2004): *Handbok i lexikografi*. Stockholm: Nordtedts Akademiska Förlag.
- Trap-Jensen, Lars (2014): Leksikografisk tradition og fornyelse: tre revolusjoner på 100 år? I: Ruth Vatvedt Fjeld & Marit Hovdenak (red.), *Nordiske studier i leksikografi 12*. Rapport fra Konferanse om leksikografi i Norden, Oslo13.–16. august 2013, 42–68.
- Wörterbücher (1989–1991) = *Wörterbücher/Dictionaries/Dictionnaires*. Edited by Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta. HSK 5. Berlin/New York: Walter de Gruyter.

Jón Hilmar Jónsson  
professor emeritus  
Árni Magnússon-instituttet for islandske studier/Islands universitet  
Laugavegur 13  
IS-101 Reykjavík  
jhj@hi.is

# Semantisk processering og leksikografi

*Bolette Sandford Pedersen*

The paper argues that lexicographical knowledge is crucial for the development of intelligent language technology, i.e. technology that deals with language in a nuanced and all-compassing way. In this context, an overview is given of some of the developments and acknowledgements that we have achieved through a number of cooperative projects at The University of Copenhagen (Center for Language Technology) and The Danish Society for Language and Literature. Main achievements encompass a number of interlinked Danish language technology resources that rely on the same sense inventory as *Den Danske Ordbog*, and which comply with international standards. Reference is also made to our work on developing principled methods for sense clustering as well as to our experiments with partial language transfer from English to Danish when developing the Danish FrameNet.

## 1. Intelligent teknologi kræver sprogviden – ikke mindst leksikalsk viden

Intelligent teknologi står foran det helt store gennembrud, og semantisk processering udgør en vigtig komponent i denne udvikling. Semantisk processering kan helt kort defineres som behandling af indholdsdimensionen i talt eller skrevet sprog. Det kan indbefatte automatisk entydiggørelse af flertydige ord i kontekst, emneklassifikation og beregning af hvem i teksten der gør hvad, hvor og hvornår. Øvelser der kan lyde banale og ligetil, men som kun de allerfærreste sprogteknologer og robotter i dag mestrer på bare et minimalt niveau og kun for de store sprog. Virtuelle assistenter, informationssøgerobotter og automatisk resumering er alle eksempler på nogle af de systemer der kræver sprogviden, og som allerede nu er ved at blive en del af dagligdagen.

I det inviterede foredrag ved NFL 2017 i Reykjavík argumenterede jeg for vigtigheden af at den viden vi repræsenterer og genererer i de leksikografiske miljøer, gøres egnet til og tilgængelig for brug i teknologien; herunder særligt i systemer der arbejder med semantisk viden, ofte omtalt som systemer med kunstig intelligens. Ordbøger er ikke bare systematiske samlinger af ord med information om ordforrådets bøjning og syntaks; de er i sig selv kulturelle vidnesbyrd om det samfund de er udviklet i. Den viden bør på længere sigt indgå med alle nuancer i den teknologi vi udvikler til brug i både privat- og arbejdsliv.

I denne artikel vil jeg i det store og hele følge foredragets tematik idet jeg vil argumentere for hvorfor det er vigtigt at vi i de relativt små sprogsamfund som de

nordiske er særligt bevidste om dette fokus på udnyttelse og tilgængeliggørelse – og samtidig forsøge at beskrive nogle af de muligheder og udfordringer der ligger i denne tilgang.

For mange af sprogmiljøerne i Norden udgjorde de såkaldte METANET-rapporter fra 2012<sup>1</sup> noget af en øjenåbner – eller i hvert fald et referencepunkt for hvordan vi på dét tidspunkt mente at vi var stillet rent sprogligt i forhold til den udbredte digitalisering og den nye teknologi i vores samfund. Vi vurderede at flere af de europæiske sprog var temmelig ringe stillet i forhold til de digitale udfordringer, og at der var et stort behov for opgradering. Dette gjaldt i høj grad de nordiske sprog der gennemgående lå i feltet med ”ringe eller fragmentarisk støtte”. Siden rapporterne er der sket en hel del udvikling også for de nordiske sprog, men gabet til engelsk er imidlertid stadig voksende.

Mit udgangspunkt for artiklen er baseret på mit eget arbejde i de senere år med udvikling af teknologiske sproressourcer for dansk. Arbejdet er udført i et mangeårigt samarbejde mellem Center for Sprogteknologi på Københavns Universitet (KU) og Det Danske Sprog- og Litteraturselskab (DSL). Afsnit 2 præsenterer i kort form de ressourcer der er tale om, og forklarer overordnet hvordan de forholder sig til *Den Danske Ordbog* (DDO) og *Den Danske Begrebsordbog* (Begrebsordbogen) som er udviklet ved DSL. I afsnit 3 kommer jeg ind på begreberne *standardisering* og *sprogransfer* – igen med udgangspunkt i de erfaringer vi har gjort os med vores danske data, særligt DanNet og Dansk FrameNet. Afsnit 4 omhandler specifikt udfordringerne omkring betydningsinventaret og beskriver vores arbejde med at udvikle en principiel metode til generering af betydningsklynger; en tilgang som letter den automatiske entydiggørelse betragteligt. Endelig kommer jeg i afsnit 5 ind på udviklingen i Europa og beskriver perspektiverne i et nyt infrastrukturprojekt, ELEXIS, som skal se på netop nogle af de problemstillinger som jeg har rejst i artiklen.

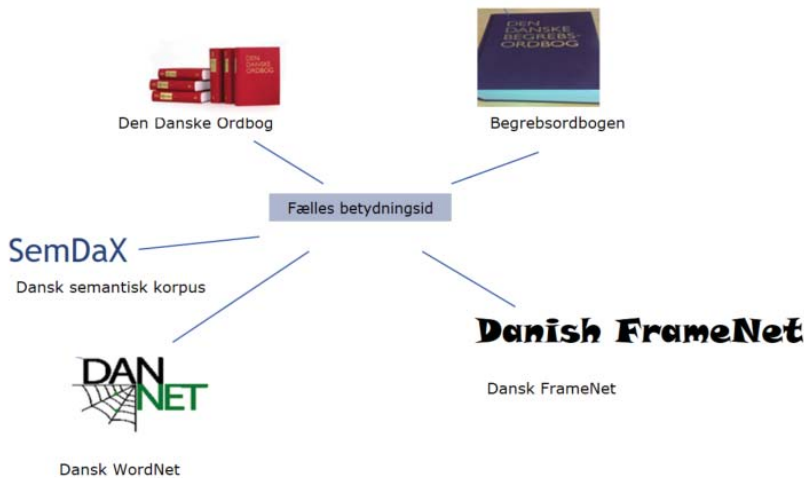
## 2. Flere sproressourcer – samme betydningsinventar

Med udgangspunkt i flere forskningsrådsprojekter<sup>2</sup> er KU og DSL i gang med at udarbejde en portefølje af sprogteknologiske ressourcer der i større eller mindre grad hænger sammen med og/eller trækker på DDO og Begrebsordbogen. Fælles for disse ressourcer er at de er udarbejdet med reference til de samme betydnings-id'er forstået på den måde at alle ressourcerne tager udgangspunkt i det betydningsinventar og den betydningsstruktur som er anlagt i DDO (se Figur 1).

<sup>1</sup> Metanetrapporterne kan ses på <<http://www.meta-net.eu/whitepapers/overview>>.

<sup>2</sup> Følgende fonde har støttet følgende involverede projekter: ”DanNet – et leksikalsk-semantic ordnet for dansk” 2005–2008: Det Danske Forskningsråd, ”DK-CLARIN” 2008–2011: Forsknings- og Innovationsstyrelsens nationale pulje for forskningsinfrastruktur, ”META-NORD” – EU-projekt 2011–2013 (ICT PSP Program), ”Semantic Processing across Domains” 2013–2017: Det Danske Forskningsråd og ”Fra Begrebsordbog til Framenet” 2016–2017: Carlsberg Fondet.

Det betyder ikke at der ikke i flere af ressourcerne arbejdes med grovere betydningsinventarer, i wordnettet DanNet<sup>3</sup> kan der fx godt være flere underbetydninger i samme *synset*,<sup>4</sup> og fx er det semantiske korpus, SemDaX,<sup>5</sup> opmærket med et graderet betydningsinventar spændende fra meget grove betydningsdistinktioner til meget finkornede. Men referencen til DDO's struktur er intakt, dvs. de grovere inventarer udgøres af *klynger* af DDO-betydninger sådan at linket til den samme basis ikke går tabt. Det fælles referencepunkt gør det langt nemmere at udvikle ressourcerne i samme takt, og det gør nemmere at bygge nye fordi man i højere grad kan genbruge den information som skal være tilgængelig i mere end én ressource.



Figur 1: Sammenhængende danske sprogressourcer med et fælles betydningsinventar/id.

Jeg vil ikke her komme nærmere ind på de enkelte sprogressourcers opbygning og struktur men henviser til følgende artikler der hver især beskriver dem i nærmere detaljer: *DanNet*: Pedersen et al. (2009), *SemDaX*: Pedersen et al. (2016), og *Dansk FrameNet*: Nimb & Pedersen (2016) og Nimb et al. (2017).<sup>6</sup>

<sup>3</sup> *DanNet*, det danske wordnet, kan downloades fra <<http://wordnet.dk>>. Det kan browses fra <<http://wordties.cst.dk/wordties-dannet/>>.

<sup>4</sup> Synsets er sæt af synonymer, som udgør den bærende enhed i wordnets.

<sup>5</sup> *SemDaX*, det semantisk opmærkede korpus, kan downloades fra github: <<https://github.com/coastalcph/sem dax>>.

<sup>6</sup> Man kan desuden læse om Begrebsordbogens udnyttelse af og sammenhæng med DanNet og *Den Danske Ordbog* i Nimb (2013, 2016).



### 3. Sprogtransfer og standardisering af sproglige data

I forbindelse med at resultaterne fra METANET-rapporterne igennem de forløbne år er blevet analyseret og fortolket i forskellige sproglige og politiske sammenhænge, er begreberne *sprogtransfer* og *standardisering af sproglige data* blevet endnu mere centrale end tidligere. Nok er vores ordbøger unikke, og nok indeholder de meget viden om det samfund de er udarbejdet i, som kan være relevant også i teknologien, men alligevel er der sproglige genveje man både kan og bør slå når et givent sprogsamfund skal rustes til den digitale tidsalder.

På KU og DSL har vi arbejdet bevidst med disse to begreber idet vi allerede i begyndelsen af vores samarbejde var meget opsatte på at overholde eksisterende standarder for ressourcer til sprogteknologi. Således er de danske WordNet og FrameNet direkte kompatibelt med de tilsvarende ressourcer der bygges for andre sprog.<sup>7</sup> Det giver en række fordele idet der er en hel del principielle detaljer man ikke behøver tage stilling til, og fordi man kan overtage og lære en del af de ressourcer der allerede er bygget. Endelig indgår man ved at anvende en international standard i et forsknings- og udviklingsfællesskab på internationalt niveau som hele tiden bevæger sig videre og afprøver nye tilgange på ens data, hvilket er en stor fordel fordi de hermed efterprøves og valideres.

Lad mig også give to eksempler på vores arbejde med *transfer* af sprogviden fra engelsk til dansk, men vel at mærke i en kontekst hvor grunddataene stadig er danske. I Nimb & Pedersen (2016) og Nimb et al. (2017) viser vi at semantiske rolleinventarer (som Communicator, Recipient mv.) identificeret på baggrund af engelske korpusanalyser fungerer fint på danske data.

Her har vi taget udgangspunkt i Begrebsordbogens xml-opmærkede grundmanuskript og udnyttet de synonymklynger man finder for verberne til relativt hurtigt og effektivt at generere et dansk FrameNet hvor grupper af synonyme danske verber tilskrives associerede semantiske rammer (frames) adapteret fra Berkeley Framenet (Ruppenhofer et al. 2016). I Pedersen et al. (2018b) går vi skridtet videre med sprogtransfer-tilgangen idet vi også eksperimenterer med at kombinere de danske FrameNet-data med rolleopmærkede engelske korpora.

Vi viser her at det er muligt til en vis grad at maskinlære fra engelske rolleopmærkede data til dansk, men at læringen af dansk forbedres betydeligt når den kombineres med træning på danske data (i dette tilfælde med opmærkede data fra SemDaX).

Dette kan lyde ukontroversielt for professionelle sprogfolk, men i dag hvor meget sprogteknologi i praksis *kun* anvender sprogtransfer (og fx benytter sprogmodeller trænet på engelsk til danske systemer), er det vigtigt at kunne påvise at ja, vi kan maskinlære på tværs af sprog, men kvaliteten højnes betragteligt når vi kombinerer læringen med sprogspecifikke data på målsproget – også selv om der som i vores tilfælde kun er temmelig begrænsede data til rådighed.

<sup>7</sup> Der forestår dog stadig standardiseringsarbejde, fx konvertering af DanNet til LMF (jf. <<http://www.lexicalmarkupframework.org/>>).

#### 4. Betydningsinventaret – graduering af detaljeringsgrad

Jeg har nævnt at vi forholder os til den samme betydnings-id i alle vores ressourcer (nemlig DDO's), og at vi betragter dette som en styrke ved vores samlede portefølje. Det betyder dog ikke at vi har løst problemet med hvordan vi definerer det rette niveau af betydningsnuancering til sprogteknologiske formål. En væsentlig hindring for en direkte og helt umiddelbar udnyttelse af almindelige ordbøger til sprogteknologi er lige præcis deres ofte meget finkornede og ikke altid lige systematiske betydningsinddelinger. Betydningerne er meget svære at skelne imellem automatisk. Faktisk udgør betydningsinventarets kompleksitet i konventionelle ordbøger efter min vurdering den største hindring overhovedet for automatisk udnyttelse af den semantiske viden der ligger gemt her.

Dette virker i nogen grad paradoksalt idet den menneskelige bruger som regel føler sig rigtig godt informeret når han/hun konsulterer en ordbog; hvorfor er det da så vanskeligt at udnytte i teknologiske sprog tjenester? Sagen er selvfølgelig at betydningsinventaret udgør et kontinuum af svært afgrænselige betydninger som selv i konkrete teksteksempler er svære at blive enige om og at definere entydigt. Dette faktum lever vi som regel helt fint med, og vi opfatter ikke umiddelbart ordbogens klart afgrænsede hoved- og underbetydninger som værende i uoverensstemmelse med sprogets praksis. Snarere tværtimod: ordbogen hjælper os som regel fint til at få overblik over spændvidden af et ords betydning. Men når vi arbejder med ordbøgernes betydningsinddelinger i direkte sammenhæng med løbende tekst eller tale, bliver det klart at der på ingen måde er tale om et én-til-én-forhold.

I de sprogteknologiske miljøer har man derfor i de sidste årtier eksperimenteret meget med hvordan betydningsinventaret kan gøres anvendeligt til semantisk processering (se fx Izquierdo et al. (2009), McCarthy et al. (2016) og Pedersen et al. (2018a)). Forskningen går ad flere veje: fra udtræk og sammenlægning af leksikografiske betydningsbeskrivelser fra forskellige kilder (også kaldet vidensbaseret "sense clustering") til en ren korpusbaseret tilgang hvor man forsøger at drage slutninger fra de enkelte betydningsforekomster og herudfra inducere ordets betydningsprofil (også kaldet "word sense induction", som i McCarthy et al. (2016)). Ved KU og DSL har vi arbejdet med den først nævnte tilgang, og vi har arbejdet dels med at håndopmærke korpuseksempler, dels med automatisk entydiggørelse ved hjælp af maskinlæring (Martinez et al. 2015). Øvelsen går ud på at finde det rette nuanceringsniveau, ikke mere finkornet end at vores annotører og algoritmer kan håndtere det, og ikke mere grovkornet end at det stadig giver betydningsmæssig mening. Det vil sige så der stadig skelnes mellem de betydninger som er væsentlige for at sprog tjenesterne kan fungere tilfredsstillende.

I vores arbejde har vi taget spørgsmålsvar-systemer (a la Siris personlige assistent på iPhone eller a la Watson, IBM's verdenskendte robot) som vores *prototypiske* case for hvilket detaljeringsniveau det er relevant at kunne håndtere. Spørgsmålsvar-systemer udmærker sig ved at kræve en mere nuanceret afdækning af hvad der foregår i en ytring, end fx almindelige søgjetjenester. Omvendt

er det ikke nødvendigt for et spørgsmålsvar-system at fange alle de sproglige nuancer som der fx kræves i oversættelse, hvor den fulde betydning ideelt set skal gengives på målsproget.<sup>8</sup>

De seneste resultater (Pedersen et al. 2018a) tyder på at det er muligt at finde et passende leje mellem DDO's fine betydningsdistinktioner og de ontologiske "primitiver" som er angivet i DanNet, og dermed skabe et noget grovere inventar som annotører kan blive enige om når de tekststopmærker, og som også algoritmerne kan håndtere.

	<b>Eksperiment 1</b> Antal DDO- betydninger og enighed		<b>Eksperiment 2</b> Antal betydningsklynger og enighed		<b>Eksperiment 3</b> Antal betydningsklynger og enighed	
<i>Selskab</i>	10	0,57	6	0,69	5	<b>0,91</b>
<i>Plads</i>	13	0,73	9	0,79	6	<b>0,74</b>
<i>Slag</i>	17	0,80	11	0,81	9	<b>0,79</b>
<i>Skud</i>	12	0,61	12	0,61	11	<b>0,69</b>
<i>Skade</i>	6	0,73	5	0,68	4	<b>0,80</b>
<i>Kort</i>	10	0,71	4	0,77	3	<b>0,78</b>
<i>Vold</i>	9	0,33	7	0,67	5	<b>0,77</b>
<i>Hul</i>	14	0,72	11	0,61	7	<b>0,78</b>
<i>Blik</i>	7	0,59	6	0,61	4	<b>0,61</b>
<i>Model</i>	8	0,74	7	0,84	6	<b>0,89</b>

Tabel 1: Tre opmærkningseksperimenter med polyseme substantiver med forskellig detaljeringsgrad og forskellig annotørenighed (IA).

I vores eksperimenter satte vi to annotører til at opmærke ti meget polyseme og samtidig højfrekvente substantiver (*selskab*, *plads*, *slag*, *skud*, *skade*, *kort*, *vold*, *hul*, *blik* og *model*) i tekster fra forskellige teksttyper og domæner. Vi udtrak 100 + 15n eksempler for hvert af de polyseme ord, hvor *n* repræsenterer antallet af betydninger i DDO. Det vil sige at jo flere betydninger der er beskrevet for et ord, desto flere eksempler opmærkes der.

I det første eksperiment indgår alle DDO's betydninger i tagsættet. I det andet eksperiment reducerer vi derimod antallet af betydninger ved at klynge underbetydninger inden for én hovedbetydning sammen hvis de er af samme ontologiske type, fx Human.<sup>9</sup> Det betyder at DDO-betydningerne *selskab\_2* og *selskab\_2a* der hhv. dækker betydningerne 'gruppe af personer som foretager sig noget i fællesskab' og 'gruppe af personer som er samlet til en fælles social aktivitet, fx en fest eller en middag', lægges sammen. I det sidste eksperiment slår vi også betydninger sammen som går på tværs af hovedbetydninger hvis de er af samme ontologi-

<sup>8</sup> Til maskinoversættelse bruger man desuden særlige teknikker til entydiggørelse, idet man her har en særlig "videnbase" at trække fra så at sige, nemlig tidligere oversatte tekster.

<sup>9</sup> DanNet anvender EuroWordNets topontologi (jf. Vossen et al. 1991).

ske type, fx `PhysicalLocation` ved *plads* i betydningen ‘plads til noget’ (`plads_1`) og ‘siddeplads’ (`plads_2`).

For alle tre eksperimenter beregnede vi annotørenigheden ved hjælp af Krippendorffs alpha (Krippendorff 2011). Et af de særlige træk ved denne beregningsmetode i forhold til fx almindelig procentberegning er at enighedskoefficienterne her korrigeres for sandsynlighed. Dvs. man modregner det faktum at det alt andet lige er nemmere at blive enige om få betydninger end mange. Det betyder at de tre annotørenighedskolonner i tabel 1 skulle være direkte sammenlignelige.

Værdierne varierer fra 0 til 1 hvor 0 er fuldstændig uenighed og 1 er fuldstændig enighed. Det ses tydeligt at enigheden øges for de grovere betydningsklynger (markeret med fed). Faktisk er det først i tredje eksperiment at annotørenigheden nærmer sig noget praktisk anvendeligt (0.78) idet man inden for semantisk processing antager at resultater med  $\alpha \geq .67$  er nogenlunde acceptable.

Vores grundpræmis i denne metodeudvikling er at maskinen ikke kan forventes at skelne bedre mellem betydninger end de menneskelige annotører kan. Og vi har med den menneskelige annotation påvist at der i sandhed er tale om en temmelig vanskelig opgave siden vi faktisk kun for ét af ordene (*selskab*) kommer over 0.90 i enighed selv med de grove betydningsklynger. Maskinlæringseksperimenterne understøtter vores overordnede antagelse; også her opnås de bedste resultater (også sammenlignet med en tilfældig sammenligning (*random baseline*)) med de reducerede betydningsklynger (Pedersen et al. 2018b:5).

Vores undersøgelse ser kun på et udvalg af meget polyseme substantiver, og selv om vores klyngemetode er principiel, dvs. kan udføres automatisk på hele ordforrådet, så ved vi endnu ikke om den er praktisk anvendelig for den største del af substantiverne (som gennemgående ikke er nær så polyseme som dem vi har undersøgt her). Vi ved heller ikke hvordan metoden virker på adjektiver og verber; måske skal man her identificere nogle andre træk at lave klynger ud fra end blot de ontologiske.

## 5. Europæiske tiltag

Hvordan ser det ud på europæisk plan mht. at genanvende viden i eksisterende ordbøger til teknologi? På trods af flere store europæiske projekter op gennem nullerne initieret af netop de datalingvistiske miljøer, så som EAGLES, PAROLE og SIMPLE (Lenci et al. (2000), Calzadori et al. (2002) m.fl.) går det ikke imponerende godt. Selv om disse projekter bragte initiativer som the Text Encoding Initiative<sup>10</sup> og Lexical Markup Framework<sup>11</sup> med sig, har forhindringer i form af intellektuelle rettigheder og meget individuelle tilgange i de forskellige leksikografiske miljøer endnu vanskeliggjort det helt store gennembrud.

<sup>10</sup> TEI: <<http://www.tei-c.org/>>.

<sup>11</sup> LMF: <<http://www.lexicalmarkupframework.org/>>.

I 2013 oprettedes imidlertid det europæiske netværk e-Lexicography (EnEL)<sup>12</sup> som en såkaldt EU COST Action. Dette initiativ blev igangsat for at forbedre adgangen til anbefalede ordbøger og gøre dem mere almindeligt kendte for et større publikum. Netværket lagde ud med 34 medlemmer fra 20 lande, men voksede i 2017 til 285 medlemmer fra 31 lande. I forbindelse med netværket opstod der et klart behov for en bredere og mere systematisk udveksling af ekspertise til etablering af fælles standarder og løsninger til udvikling og integration af leksikografiske ressourcer og for at udvide anvendelsesområdet til at indbefatte sprogteknologi og digital humaniora.

Som en udløber af ENEL-bestræbelserne blev der derfor etableret et konsortium, og i august 2017 blev forslaget til en infrastruktur under navnet ELEXIS udvalgt til finansiering som et projekt under Horizon 2020. De vigtigste mål for denne nye infrastruktur er i) at fremme samarbejde og vidensudveksling mellem forskellige leksikografiske forskningsmiljøer bl.a. for at mindske kløften mellem de mindre sprogmiljøer og dem med avanceret e-leksikografisk erfaring, ii) at arbejde med strategier, værktøjer og standarder for uddragning, strukturering og sammenkobling af leksikografiske ressourcer, iii) at lette adgangen til standarder, metoder, leksikografiske data og værktøjer, og sidst men ikke mindst iv) at tilskynde til en open access-kultur i den leksikografiske verden i overensstemmelse med de henstillinger der nu fremføres både fra EU-Kommissionens side og på nationalt plan mange steder, bl.a. i det danske forskningsministerium.

Det sprogteknologiske aspekt har et særligt fokus i projektet, og da både KU og DSL deltager som danske partnere i projektet, forventer vi at resultaterne fra det hidtidige danske samarbejde i høj grad vil kunne udnyttes og videreudvikles de næste 4 år. Vi håber desuden at kunne bidrage til udvikling af nye metoder og værktøjer til harmonisering af ordbogsformater og til automatisk segmentering og identifikation af ordbogsstruktur der muliggør udtræk og sammenkædning af ordbogsindhold – også på tværs af sprog. Endelig indgår der i projektet en arbejdsplan der skal se på netop metoder for at udvikle graduerede betydningsinventarer som er håndterbare til automatisk entydiggørelse.

Jeg vil gerne afslutningsvis nævne to andre europæiske initiativer under EU-programmet ”Connecting Europe Facility” som også Dansk Sprognævn deltager i sammen med KU. Det drejer sig om European Language Resource Coordination<sup>13</sup> og eTranslation TermBank; begge initiativer der skal fremme genbrug og open access til sproressourcer særligt til brug for EU’s maskinoversættelsessystem.

Til sammen vidner alle disse initiativer om en stigende interesse og forståelse for hvorfor det er vigtigt at gøre både leksikografiske, terminologiske og andre sprogdatabaser direkte tilgængelige for teknologi.

---

<sup>12</sup> Jf. <<http://www.elexicography.eu/>>.

<sup>13</sup> ELRC-SHARE: <<https://elrc-share.eu/>>.

## Litteratur

### Ordbøger og sprogteknologiske ressourcer

- DDO = *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab. <<http://ordnet.dk/ddo>> (1.11.2017).
- Begrebsordbogen = Nimb, S., H. Lorentzen, T. Troelsgård, L. Theilgaard & L. Trap-Jensen (2014): *Den Danske Begrebsordbog*, København: Det Danske Sprog- og Litteraturselskab.

### Anden litteratur

- Calzolari, Nicoletta, Antonio Zampolli & Alessandro Lenci (2002): Towards a Standard for a Multilingual Lexical Entry: The EAGLES/ISLE Initiative. I: *CICLing 2002: Computational Linguistics and Intelligent Text Processing*. Berlin/Heidelberg: Springer, 264–279.
- Izquierdo, Rubén, Armando Suárez & German Rigau (2009): An empirical study on class-based word sense disambiguation. I: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. The Association for Computational Linguistics, 389–397.
- Krippendorff, K. (2011): Agreement and Information in the Reliability of Coding. I: *Communication Methods and Measures* 5(2): 93–112.
- Lenci, A., N. Bel, F. Busa, N. Calzolari, E. Gola, M. Monachini, A. Ogonowski, I. Peters, W. Peters, N. Ruimy, M. Villegas & A. Zampolli (2000): SIMPLE: A general framework for the development of multilingual Lexicons. I: *International Journal of Lexicography* 13(4): 249–263.
- Martínez Alonso, Héctor, Anders Johannsen, Sussi Olsen, Sanni Nimb, Nicolai Hartvig Sørensen, Anna Braasch, Anders Søgaard & Bolette Sandford Pedersen (2015): Supersense tagging for Danish. I: *Proceedings of the 20th Nordic Conference of Computational Linguistics. NODALIDA 2015*. Linköping Electronic Conference Proceedings #109, Linköping: ACL Anthology/Linköping University Electronic Press, 21–29.
- McCarthy, D., M. Apidianaki & K. Erk (2016): Word Sense Clustering and Clusterability. I: *Computational Linguistics* 42(2):245–275.
- Nimb, Sanni (2013): Leksikalsk-semantisk information i en ny dansk begrebsordbog. I: Dorthe Duncker, Anne Mette Hansen & Karen Skovgaard-Petersen (red.), *Betydning og Forståelse. Festskrift til Hanne Ruus*. København: Selskab for Nordisk Filologi, 251–266.
- Nimb, Sanni (2016): Der er ikke langt fra tanke til handling. Om semantiske typer og systematisk polysemi i Den Danske begrebsordbog. I: *Danske Studier* 2016. Universitets-Jubilæets danske Samfund, 25–59.

- Nimb, Sanni & Bolette S. Pedersen (2016): Fra begrebsordbog til sprogteknologisk ressource: verber, semantiske roller og rammer – et pilotstudie. I: Asgerd Gudiksen & Henrik Hovmark (red.), *Nordiske Studier i Leksikografi* 13. Skrifter udgivet af Nordisk forening for Leksikografi, Skrift nr. 14, 405–415.
- Nimb, Sanni, Anna Braasch, Sussi Olsen, Bolette Sandford Pedersen & Anders Søgaard (2017): From Thesaurus to FrameNet. I: *Electronic Lexicography in the 21st century. Proceedings of eLex 2017 conference*, Leiden, 1–22.
- Pedersen, Bolette S., Sanni Nimb, Jørg Asmussen, Nicolai Sørensen, Lars Trap-Jensen & Henrik Lorentzen (2009): DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. I: *Language Resources and Evaluation* 43(3):269–299.
- Pedersen, Bolette S., Lars Borin, Markus Forsberg, Neeme Kahusk, Krister Lindén, Jyrki Niemi, Niklas Nisbeth, Lars Nygaard, Heili Orav, Eiríkur Rögnvaldsson, Mitchell Seaton, Kadri Vider & Kaarlo Voionmaa (2013): Nordic and Baltic Wordnets Aligned and Compared through ”WordTies”. I: *Proceedings from the 19th Nordic Conference on Computational Linguistics (NO-DALIDA 2013)*. Linköping Electronic Conference Proceedings 85. Linköping: Linköping University Electronic Press.
- Pedersen, Bolette S., Anna Braasch, Anders Johanssen, Héctor Martínez Alonso, Sanni Nimb, Susi Olsen, Anders Søgaard & Nicolai Hartvig Sørensen (2016): The SemDaX Corpus – sense annotations with scalable sense inventories. I: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož: European Language Resources Association (ELRA), 842–847.
- Pedersen, Bolette S., Manex Agirrezabal, Sanni Nimb, Sussi Olsen, Ida Rørmann (2018a): Towards a principled approach to sense clustering – a case study of wordnet and dictionary senses in Danish. In: *Proceedings of Global WordNet Conference 2018*, Singapore. <[http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018\\_paper\\_27.pdf](http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018_paper_27.pdf)> (1.4.2018).
- Pedersen, Bolette S., Sanni Nimb, Anders Søgaard, Mareike Hartmann & Sussi Olsen (2018b): A Danish FrameNet Lexicon and an Annotated Corpus Used for Training and Evaluating a Semantic Frame Classifier. Accepteret ved *LREC 2018*, Japan.
- Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. R. Johnson, C. F. Baker & J. Scheffczyk (2016). *FrameNet II: Extended Theory and Practice*. <[https://framenet.icsi.berkeley.edu/fndrupal/the\\_book](https://framenet.icsi.berkeley.edu/fndrupal/the_book)> (1.11.2017).
- Vossen, P. (red.) (1999): *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers.

Bolette Sandford Pedersen

Professor, ph.d.

Center for Sprogteknologi, Institut for Nordiske Studier og Sprogvidenskab

Københavns Universitet

Njalsgade 136, DK-2300 København S

[bspedersen@hum.ku.dk](mailto:bspedersen@hum.ku.dk)



# The Variability of Semantic Categories: An Experiment in Extensional Semantics<sup>1</sup>

*Matthew Whelpton*

The nature and variability of lexical semantic categories remains an enduring topic of controversy and debate. The Evolution of Semantic Systems (EoSS) project sought to address such issues by collecting naming data from 50 Indo-European languages, using a fixed stimulus array covering four domains: colours, body parts, containers, and spatial relations. This inspired subsequent work in Iceland on heritage language (North American Icelandic) and sign language (Icelandic Sign Language). This paper reports key findings relating to Icelandic. The least variability is found with body parts and colours, both of which are supported by specialised neuro-psychological architecture for visual and proprio-perception.

## 1. Introduction

To the naive observer, few aspects of language are more accessible than the meaning of words referring to concrete objects and their observable characteristics. A cup and bowl on a table or the red of berries against green leaves seem to present themselves naturally for classification and labelling. However, it is well known in the fields that have concerned themselves with word meaning—linguistics and lexicography, philosophy, psychology, and anthropology amongst them—that this transparent link between word meaning and the observable world is illusory. The term *red* in English is hemmed in by *orange* and *yellow* and *purple* and *pink*, whereas the Ejagham language of Nigeria and Cameroon has a three-term system: énjàgà ‘black/green/blue’, ébàré ‘white’, ébí ‘red/yellow’ (Kay et al. 2009:227); a *house* in English is not the same as a *hús* in Icelandic, the former being restricted to domestic dwellings and the latter not.

Such questions concerning the variability of semantic categories and how they

---

<sup>1</sup> This work stems from the Evolution of Semantic Systems project which received financial support from the Max Planck Gesellschaft. The EoSS data for Icelandic was collected by the author. I would like to thank Linnaea Stockall for the use of the British English data. Data for North American English and North American Icelandic was collected as part of the *Heritage Language, Linguistic Change and Cultural Identity* project (PIs: Höskuldur Práinsson and Birna Arnbjörnsdóttir) funded by the Icelandic Research Fund. Data for Icelandic Sign Language was collected as part of the *Colour in Context* project (PI: Matthew Whelpton) funded by the University of Iceland Research Fund. All views, errors and omissions are mine solely.

change is the focus of the Evolution of Semantic Systems (EoSS)<sup>2</sup> project (Majid, Jordan & Dunn 2011). How variable are semantic categories across languages? Do different kinds of categories differ in the degree to which they vary between languages? For instance, are categories for spatial relations between objects more variable crosslinguistically than categories for the concrete objects themselves? Are categories for attributes of objects more variable than categories for parts of those objects? How do such patterns of variation relate to the processes of historical change?

Unusually for a typological study, EoSS chooses to look at a single family of related languages (Indo-European) rather than a selection of typologically diverse languages. This allows for the use of statistical methods (similar to those used in biology for phylogenetic research) to quantify the degree of variation in the family group. Naming data was collected for four semantic categories—body parts, colours, containers, and spatial relations—in 50 Indo-European languages. Use of fixed stimulus arrays allows for comparisons across a uniform extensional space.

This paper reports results relating to Icelandic, including EoSS-inspired work on heritage Icelandic and Icelandic Sign Language. Section 2 sets the experimental work in the context of debates concerning the nature and variability of semantic categories. Section 3 sets out the methodology of the EoSS project. Section 4 reports the main EoSS findings for Germanic as well as related work on Icelandic. The findings support the view that semantic categories do indeed differ in the degree of variation that they exhibit. Colour and body-part naming are more constrained by neuro-psychological factors than containers, which are more susceptible to cultural influences. Spatial relations vary at least as much as containers but are arguably susceptible to different cognitive constraints.

## 2. Semantic Categories

The classical theory of (lexical) concepts holds that to identify the meaning of a word, you identify the set of things to which it can truly be applied (the extension); you identify the extension by asking questions concerning its defining characteristics (the intension), as in the parlour game *Animal? Vegetable? Mineral?* Each concept will have a set of necessary and sufficient conditions which define its range. This view goes back to Plato's *Euthyphro* and Aristotle's *Categories* and runs through Western philosophy (cf. Locke, Descartes, Hume, Frege and Russell).

This view has been famously challenged in philosophy, psychology and linguistics. The philosopher Wittgenstein (1953) challenges the feasibility of isolating a set of necessary and sufficient conditions for every lexical concept using an argument from family resemblances. The psychologist Rosch (1973) challenges the sharp categorial nature of lexical concepts with her work on prototypes. In a

<sup>2</sup> <<http://www.mpi.nl/departments/other-research/research-consortia/eoss/aims>>.

similar spirit, the linguist Labov (1973) illustrates the fuzzy boundaries between the common container categories, *bowl*, *cup* and *vase*, a study reproduced for Icelandic by Höskuldur Þráinsson (1979).

Underlying such controversies are assumptions concerning the sources of stability or variability in semantic categories. Universalists argue for strong shared constraints on semantic categories. The constraints might be imposed by the environment, for instance the claim that certain aspects of our experience “cry out to be named” (Berlin 1992:290); or they may reside in the cognitive system itself in the form of innately given primitive concepts (cf. Fodor 1975). Relativists on the other hand argue for the free construction of semantic categories, most famously by the linguistic system itself (Sapir 1912, 1929, Whorf 1956), a view going back to Saussurean notions of semiotic value (Saussure 1922).

Majid et al. (2015) point in particular to an ongoing controversy concerning the relative importance of external stimuli and internal cognitive constraints. The psychologist Gentner argues that directly perceptible individuated entities vary less than abstract relational entities (Gentner 1981, Gentner & Boroditsky 2001): external stimuli trump internal cognitive constraints. The linguist Talmy argues that closed class functional terms vary less than open class substantive terms (Talmy 1983, cf. Landau & Jackendoff 1993, Haspelmath 2003): internal factors strongly constrain abstract functional categories which are not readily subject to conditioning by external experience. To address such questions, an experimental extensional approach is taken, in which the lexical categorisation in different languages of a fixed stimulus array is compared statistically.

### 3. Methodology

#### 3.1. Languages and participants

Icelandic was one of the 50 European languages for which data was elicited as part of the EoSS project. Section 4 reviews some studies relating to Icelandic, resulting from this original research, including comparisons with Dutch, Frisian, Swiss German, Standard High German and Norwegian Bokmål. This work inspired further research in two projects in Iceland.

The *Heritage Language, Linguistic Change and Cultural Identity* project (2013–2015; PIs: Höskuldur Þráinsson and Birna Arnbjörnsdóttir), funded by the Icelandic Research Fund, was concerned with the current status of the heritage language, North American Icelandic (NAI), spoken by descendants of emigrants from Iceland, primarily in 19th century, who settled in Canada and the United States (Arnbjörnsdóttir 2006). As part of a wide ranging set of research assessing the status of NAI, from grammatical properties to cultural context, the project collected naming data for both NAI and the English spoken in the same communities (North American English, NAE). Colour naming data for NAI and NAE was

collected in 2014 in Manitoba, Saskatchewan and North Dakota. Naming data in the other three EoSS categories was collected in 2013 in Manitoba. This data was supplemented with the EoSS data for British English, kindly supplied by Linnaea Stockall.

The *Colour in Context* project (2014-2016; PI: Matthew Whelpton), funded by the University of Iceland Research Fund, collected colour naming data from the Icelandic Sign Language community. The aim was to allow researchers to compare Icelandic colour naming practices with those of a community in Iceland speaking a typologically unrelated language, as well as with a typologically related variant of Icelandic which had developed in a foreign country (North American Icelandic).

Table 1 shows information on the participants for each language reported here.

Language (Project)	Participants (Female)	Mean Age	Researchers
Icelandic (EoSS)	21 (10)	29	Matthew Whelpton; Þórhalla Guðmundsdóttir Beck
Frisian (EoSS)	23 (13)	20	Pieter Duijff; Arjen Versloot
Swiss German (EoSS)	20 (10)	26	Raphael Berthele; Martina Zimmerman
Standard High German (EoSS)	20 (10)	21	Cornelia van Scherpenberg; Michael Dunn; Fiona Jordan
Norwegian (Bokmål) (EoSS)	20 (10)	28	Áshild Næss
Dutch (NL) (EoSS)	21 (16)	22	Wendy van Ginkel; Fiona Jordan; Michael Dunn
British English (EoSS)	20 (9)	22	Linnaea Stockall; Euphemia Snell
North American English (Heritage)	51 (36)	67	Matthew Whelpton; Þórhalla Guðmundsdóttir Beck
North American Icelandic (Heritage)	30 (19)	78	Kristín Jóhannsdóttir; Íris Edda Nowenstein; Matthew Whelpton; Þórhalla Guðmundsdóttir Beck
Icelandic Sign Language (Colour in Context)	21 (14)	48	Kristín Lena Þorvaldsdóttir; Matthew Whelpton; Þórhalla Guðmundsdóttir Beck

Table 1: Overview

Two points require immediate comment. Participants in the EoSS project were primarily undergraduates and the mean age of participants was in the 20s. NAI

on the other hand is a heritage language which is dying. The participants were therefore considerably older (mean age 78). The English participants were selected from the same communities as the NAI speakers and were often their carers, who brought them to the experimental site. Though younger than the NAI speakers, the NAE speakers are therefore also considerably older than the EoSS participants: mean age 67. Icelandic Sign Language (ISL) is spoken by a very small community in Iceland and it was not possible to find 20 volunteers if participation was restricted by age; the mean age of participants was 48.

This means that age is a potentially confounding factor in interpreting differences between the languages. However, this only serves to underline the striking similarities in colour naming reported in section 4.4.

### 3.2. Experimental stimuli

Data was collected in each of the four semantic categories, following a standardised elicitation task, as specified by the EoSS project protocol (Majid et al. 2011). Each participant was presented with a series of stimuli and asked to name them. The stimuli were presented in a fixed random order. It was emphasised that they should use the first word that came to mind and that we were interested in simple everyday language. In addition, participants completed a focal colour task and a colour blindness task. Responses were audio-recorded (video-recorded in the case of sign language) and transcribed and coded by the researcher afterwards.

#### 3.2.1. Body parts

Stimuli in the body part task comprised 90 line drawings: 70 of the human body viewed front and back; 20 of the head and face viewed from the front. A red dot on each line drawing identified the area to be named. The stimuli were developed for the EoSS project by the principal investigators (Jordan, Dunn & Majid 2009). Participants are asked “The man has a dot on his...?”.

#### 3.2.2. Containers

Stimuli in the containers task comprised 67 colour photographs of household containers, primarily kitchen containers, which were developed by and used in Ameel et al. (2005). Items were photographed from a fixed distance against a neutral background with a ruler in the foreground. Participants are asked “What is this called?”

### 3.2.3. Spatial relations

Stimuli in the spatial relation task comprised 71 line drawings from the Topological Relations Picture Series (Bowerman & Pederson 1992). Each drawing shows a Figure in orange set against a Ground in black. Participants are asked “Where is the X?”.

### 3.2.4. Colour naming

Stimuli in the colour naming task comprised 84 Munsell colour chips. Four colours were achromatic (black-grey-white). The remaining 80 vary in hue, brightness and saturation. There are 20 equally spaced hues at four degrees of brightness. Saturation varies so that colours are generally at the maximal possible chroma for that point in the colour space. Only Munsell certified colour sheets were used. As light conditions critically affect perception, colours were presented under a daylight bulb simulating the full spectral range of daylight.

### 3.2.5 Focal colours

After colour names had been elicited, the participant was presented with all 84 colours in a 4x21 array, ordered by hue and brightness. Participants were then asked to point to the best example of a series of basic colour terms (e.g. “Please point to the best example of red”). A standard list of basic colour terms was used for both varieties of English; the list for Icelandic Sign Language was taken from Rannveig Sverrisdóttir and Kristín Lena Thorvaldsdóttir’s (2016) detailed study. No standard list was available for Icelandic so a rapid listing test was run to determine a salient set of colour terms (Corbett & Davies 1997, cf. Berlin & Kay 1999). Given constraints on data collection with the heritage community of North American Icelandic speakers, it was not possible to run such an elicitation test; the list for Icelandic was used. The consequences of this decision are discussed in detail in Guðmundsdóttir Beck & Whelpton (To appear). All five languages had the eleven basic colour terms standardly listed for English (Berlin & Kay 1999, Kay et al. 2009).

### 3.2.6. Colour blindness test

Finally, all participants took a colour blindness test. The British English and Icelandic participants took the Waggoner (2002) colour blindness test. The North American and Sign Language participants took the 10-plate Ishihara Test for Colour Blindness (The Isshinkai Foundation 2005). Results from participants who were found to be colour blind were then excluded from the final analysis.

### 3.3 Coding

The complete response given by each participant to each stimulus was transcribed into a standardised spreadsheet as the full response. Sign language responses were transcribed using the standard transcription for ISL, including information on mouthing and facial expression, as well as the fingered spelling of items. That full response was then coded as a main response, which was the transparent morphological and semantic head of the word used to name the stimulus. For instance, if a participant responded “it’s a black coffee cup” (full response), then *cup* was coded as the main response because *cup* is the morphological head of the compound *coffee cup* and the semantic relation between the compound and the morphological head is transparent: a coffee cup is a kind of cup (hyponym relation). If the participant did not respond, gave an unintelligible response, or responded in a way unrelated to the stimulus, an error was coded. From the main responses, a dominant term was also identified for each stimulus: a dominant term is a term that is used more often than any other term to name a stimulus.

In some cases, the choice of a coded main response was problematic, usually because the relationship between the target phrase in the full response and the morphological head was not (entirely) semantically transparent. For instance, the compound noun *ashtray* is made up morphologically of two nouns *ash+tray*. However, most speakers do not accept that an ashtray is a kind of tray and ashtrays can take a form which is quite different from the form of trays. In this case, the whole nominal term *ashtray* was coded as the main response.

A number of coding decisions deserve special comment here. The standard Icelandic translation for *orange* is *appelsínugulur*, literally *orange+yellow*. The standard Icelandic translation for *purple* is *ffólublár*, literally *violet+blue*. These terms were coded as independent lexemes. This decision was taken for two reasons. First, most Icelandic speakers are not comfortable saying that *appelsínugulur* is a kind of *gulur* or that *ffólublár* is a kind of *blár*. Second, both *appelsínugulur* and *ffólublár* appeared prominently in the rapid listing task discussed in section 3.2.5 and therefore appear to be salient to speakers as independent lexical items.

Another significant coding decision relates to the EoSS protocol as a whole. In the responses to the spatial relation stimuli, it was decided to code only prepositions as main responses. In some cases, this has a potentially significant distorting effect on the analysis. So for instance one stimulus shows a fence as the Figure surrounding a house as the Ground. The participant is asked “Where is the fence?”. A common Icelandic response exemplifies the problem that arises.

- |                                 |                |           |
|---------------------------------|----------------|-----------|
| 1) girðingin                    | umkringir      | húsið     |
| fence=the                       | around+circles | house=the |
| “the fence surrounds the house” |                |           |

The Icelandic verb for *surround* includes an incorporated prepositional particle

*um* ‘around’. However, there is no prepositional phrase *um húsið* ‘around the house’. This was therefore coded as an error. The study reported in section 4.2 seeks to address this deficit in the standard protocol.

The coding of heritage language responses posed a special challenge. Given the fragile state of heritage NAI, there was enormous variation in the phonological form of responses and often also in the morphological class attributed to nouns. For instance, NAE *eyebrow* and Icelandic *augabrún* might appear in NAI responses as *augabrún*, *augnbrún*, *eygabrúr*, *augabrýr*, or *eygabrár*. It was decided to code such continua of variation by a single standard form as the main response. Similarly, NAE *arm* and Icelandic *handleggur* (masc) might appear as *armur* (masc), *armið* (neut) or *armini* (fem). These were coded by a single standardised main response. It was however possible to set the processing script to recognise all such variants as separate main responses.

Conversely, there were cases where the variant forms potentially signified diverging semantic categories, especially the use of English versus Icelandic terms (e.g. both *bowl* and *skál*). It was decided to code these as separate main responses but with an option in the processing script of treating these as a single main response.

## 4. Icelandic and the EoSS results

### 4.1. Statistical analysis for the EoSS Germanic languages

A special volume of *Language Sciences* appeared in 2015 dedicated to results from the Germanic group of languages in the EoSS project. Majid et al. (2015) offer an overview of the results, including a statistical analysis of naming patterns in the Germanic languages, which address the issue of relative variability of semantic categories.

Following Malt et al. (1999), the authors conducted simple Pearson correlations on all possible pairwise similarity matrices in each of the semantic categories. To take body parts as an example: there are 90 line drawing stimuli in the body part task. For each individual participant, the response for a pair of stimuli was compared, e.g. the response to stimuli 5 and the response to stimuli 62. If the same main response was given to name both stimuli the value of 1 was assigned; otherwise a value of 0 was assigned. For each individual participant, there was therefore a 90x90 matrix representing which stimuli were grouped under the same name and which were not. The matrix for participants in a language were then averaged to give a value between 0 and 1 representing the proportion of participants in that language who assigned each pair of stimuli the same name. Simple Pearson correlations were then run on each pair of languages within a particular semantic category to see how similarly two languages divided up the extensional space lexically. All languages were significantly positively correlated in all categories



(at  $p = .01$ , one-tailed); however, there was an interesting difference in the degree of variability of categories. Colour and body parts correlated on average over 0.9, containers over 0.8, and spatial relations over 0.7 (Majid et al. 2015:7).

The same data was then subjected to a Principal Component Analysis (again following Malt et al. 1999). This technique identifies a small number of components and quantifies how much of the variation in the data can be attributed to each component. For all four domains, a single factor accounted for most of the variance: body parts (94.5%), colour (93.8%), containers (83.9%), spatial relations (75.8%). For all but the spatial relations the eigenvalue scores were higher than 1 only for the first component, suggesting that a single component was enough to account for variation in the data; 2 were sufficient for spatial relations (Majid et al. 2015:8).

The authors observe that this is consistent with the Gentner perspective rather than the Talmy perspective, in so far as abstract spatial relations are more variable than containers, though the high correlation for colours and body parts is surprising from that perspective (Majid et al. 2015:9). However, a number of factors undermine support for the Gentner perspective. If the crucial factor is individuation of concrete objects as opposed to abstract relations, then one would expect containers to show the least variation, perhaps followed by body parts, which are less easily individuated than cups and bowls but still have clear potential boundaries for segmentation at the joints (Majid 2010:60). Colours are an abstract quality representing a continuum of physical properties; spatial relations are also abstract, representing a continuum of relative positions in physical space.

The grouping of body parts and colours is not however surprising from another perspective: both involve categories for which we have specialised neuropsychological architecture (visual and proprio-perception). Containers on the other hand are a cultural artefact, the general constraints on which are largely functional (convenient ways of eating, drinking, storing, cooking and distributing consumables). The similarities reflect both the broad functional constraints on the artefacts and the shared cultural history of many of the Indo-European languages. But there is still ample room for cultural variation, even for instance within a sub-group such as a Western Europe. Dutch for instance set itself apart from the other Germanic languages by having a container, *kom*, which is a container for consuming soup. It cuts across the typical division into cups and bowls that emerged strongly in the Germanic data (see Whelpton, Guðmundsdóttir Beck & Jordan 2015 for more detailed discussion). This suggests support for the Talmy perspective, in which external physical objects are more susceptible to cultural and functional influences on categorisation. If factors such as proprioception and visual perception play a strong role in constraining categories in the body parts and colour categories then this can be viewed as supporting a certain kind of internal constraint on categorisation.

The main problem for the Talmy approach is of course the high variability of the spatial relations relative to the other categories. However, the second compo-

ment in the Principal Component Analysis (Majid et al. 2015:8, Figure 1) appears to split the Germanic languages into two tight clusters: on the one hand, Netherlands Dutch, Belgium Dutch, Frisian, German, Swiss German; on the other hand, Danish, English, Faroese, Icelandic, Luxembourgish, Norwegian, Swedish. This looks like a typological split of the kind that one might expect in a model such as Talmy's where languages differ in the way that information is coded but only in constrained ways; for instance, the difference between post-Latin Romance languages which favour Path-conflation in motion events as opposed to other Indo-European languages which favour Coevent-conflation (e.g. Manner-conflation) (Talmy 2000:27). In section 4.2, we see in more detail how a subgroup of the EoSS Germanic languages differ precisely in the way that they code spatial information in the verb and preposition.

#### 4.2. Spatial relations in five Germanic languages

As mentioned in section 3.3, the standard coding protocol in EoSS was problematic for cases where spatial information was coded in a prepositional particle incorporated into the verb rather than in an independent prepositional head. This was particularly marked for languages like German which make extensive use of prepositional particles. The focus on independent prepositional heads also potentially obscures other ways of encoding spatial information within the predicate.

Berthele et al. (2015) therefore decided to recode the EoSS data in five Germanic languages to recognise four kinds of spatial coding: complex prepositional forms (e.g. *up against*), three primary posture verbs (*sit, stand, lie*), other coevent verbs (e.g. *hang*), and perfective resultatives (e.g. *be wound around*). The data was recoded for Frisian, Icelandic, Norwegian (Bokmål), Swiss German and Standard German. For our purposes, the results concerning complex prepositions and posture verbs will illustrate the important point.

All five languages made available the option of using complex prepositional constructions. A line drawing depicting a ladder leaning against a wall elicited complex prepositional constructions in all five languages (Berthele et al. 2015:86–87). However, Icelandic differed sharply from all four other languages, including Norwegian, in how frequently it used such complex prepositional combinations (Berthele et al. 2015:87, Figure 1). Even in simple static locational descriptions, complex prepositional elements are often used (Berthele et al. 2015:88, ex. 13).

- 2) Kanína=n er inni í búri=nu  
 rabbit=the is inside in hutch=the  
 “The rabbit is in the hutch.”

With respect to the primary posture verbs (*sit, stand, lie*), Icelandic was at the opposite end of the scale and it was Frisian which differed sharply from the other four languages, being the only language in which use of a posture verb was more

common overall than lack of a posture verb (Berthele et al. 2015:91, Figure 3). The high proportion of posture coding in Frisian was driven largely by use of the verb *sitte* 'sit', whereas in the other four languages the three posture verbs were used in roughly equal proportions. For Icelandic, Standard German and Swiss German, the use of *sit* cognates was largely limited to animate figures in a prototypical posture (forming a low block profile) with firm contact along the bottom to the ground (e.g. cat on a mat). In both Frisian and Norwegian, the specifically postural elements have been bleached, leaving the sense that the Figure is closely contained by or in close contact with (often adhered to) the Ground (e.g. the gum sits under the table). In Frisian, this bleaching has gone far enough to allow use of the verb with cases of containment-plus-suspension (e.g. the fish sits in the fish bowl). In Frisian (but not Norwegian), this has led to a strong preference to use *sitte* rather than the copular in appropriate contexts (Berthele et al. 2015:92–93).

Close analysis of the coding of spatial information therefore suggests that variation in this domain can be sharp but is not completely free. The variation is structured along an number of dominant lines: a preference for coding information in the prepositional versus the verbal domain; a preference for coding vector orientation or posture/attachment. A statistical analysis of the variation will show marked differences between languages but this will potentially obscure the fact that the variation is generated by a small number of coding choice points. This is exactly the kind of contrast that Talmy predicts. This kind of structured variation is very different from the kind of variation we find in the domain of container naming.

### 4.3. Containers in Dutch and Icelandic

The EoSS work on containers is built on previous research by Malt and Ameel (Malt et al. 1999, 2003, Ameel et al. 2005). Malt et al. (1999) have shown that speakers of English, Spanish and Chinese who name a set of containers in rather different ways, nevertheless sort them into groups in very similar ways: lexical categories are not the same as, and do not condition, conceptual categories. Malt et al. (2003) suggest that container naming shows complex interactions of linguistic naming conventions and cultural history with more general factors. They find neither the free cross-cutting of categories that one would expect from a structuralist-relativist perspective nor the kind of general uniformity with nesting for more specific descriptions, that one would expect from a universalist perspective. Rather, the complexity of the data requires what they call a mix-and-match approach.

Ameel et al. (2005) consider the naming patterns of Belgian French-Dutch bilinguals in comparison to their monolingual peers. They find that the naming patterns of bilinguals are distinct from those of their monolingual peers and tend towards convergence, though bilinguals have not developed a single uniform naming system. This kind of variation is very different from the structured variation we found with spatial information coding.

One claim in Malt et al. (2003:35) caught the particular attention of the EoSS group. Not only did they find significantly different container naming categories for the three languages but also a significant difference in the number of terms: for 60 stimuli, 5 Chinese terms, 7 English terms, and 15 Spanish terms. They suggested that this was influenced by productive morphology in Spanish which lead to elaboration of container names. Spanish has productive instrumental suffixes (-ero/-era/-or) for describing objects used for performing a particular action, e.g. *mamadera* naming an object for sucking on. In the EoSS data, the least related languages for container naming were Icelandic and Dutch. We noticed that Dutch made productive use of diminutive morphology in this naming domain but Icelandic did not. We therefore decided to test Malt et al.'s hypothesis with respect to Icelandic and Dutch (Whelpton et al. 2015).

Dutch has been noted for its “fondness for the diminutive” (Brachin 1985:63). The diminutive -(t)je serves a range of functions, including smallness (*kopje* ‘espresso cup’), individuation (*broodje* ‘bread roll’), portioning (*een biertje* ‘a glass of beer’), and the expression of speaker attitudes such as endearment, modesty, or contempt. Our Dutch data includes 12 diminutives. Icelandic has diminutives but they are not especially productive. The diminutive -lingur expresses smallness or youth (*gríslingur* ‘piglet’); the diminutive -s(l)a expresses endearment and is common in child-directed speech (*tásla/tása* ‘toesie’). There were no diminutives in our Icelandic container naming data. The question is then whether the division of the extensional space reflects this difference.

Icelandic has a few broad inclusive terms in the domain (e.g. *skál* ‘bowl’ and *bolli* ‘cup’) including at least small and medium sized items. There are then a number of more specialized terms, particularly at the larger end (e.g. *fat* ‘platter’ and *kanna* ‘mug/tankard’). Dutch tends to separate off the smaller items with the diminutive. In some cases, the diminutive separates off the small items (*schaaltje* ‘(small) bowl’) and the non-diminutive covers the medium-to-large items (*schaal* ‘bowl/dish’); in some cases there is simply a more even spread across the space: *kopje* ‘espresso cup’, *mok* ‘mug’, *pul* ‘tankard’.

This might lead one to expect strong prototypical centres for the diminutive terms, especially terms like *kopje* ‘espresso cup’ whose root *kop* ‘cup’ does not occur at all as a dominant term. However, Icelandic and Dutch differ strikingly in the items that receive the highest inter-speaker agreement in naming. The Icelandic items that receive highest inter-speaker agreement would typically be named with a diminutive in Dutch (*kopje*, *schaaltje*). However, it is the non-diminutive terms describing larger items that have the highest inter-speaker agreement in Dutch (e.g. *schaal* ‘bowl/dish’ and *bord* ‘plate’).

Productive diminutive morphology can influence boundary placement in the extensional space but does not produce a proliferation of terms and/or fine-grained nesting within the domain. Despite its salience, the diminutive is not associated with its own clear prototypical exemplar.

#### 4.4. Colours in Icelandic, English and Icelandic Sign Language

Guðmundsdóttir Beck & Whelpton (To appear) report on colour-naming data from five languages: Icelandic (ICE), North American Icelandic (NAI), North American English (NAE), British English (BRE), and Icelandic Sign Language (ISL).

Following Majid et al. (2015) and Malt et al. (1999), they conduct simple Pearson correlations<sup>3</sup> of the similarity matrices. All the languages are extremely highly correlated (mean 0.92, at  $p = .01$ ). This is particularly striking given the wide variation in age of participants for the five languages and in the typological diversity introduced by the inclusion of Icelandic Sign Language. To the extent that the small variation between language pairs suggests anything, it is interesting to note that the top end of the ranking includes languages sharing a cultural context: Icelandic and Icelandic Sign Language (0.942), typologically diverse but sharing a common cultural environment; and North American Icelandic and North American English (0.939), different languages spoken by members of the same communities; British and North American English also rank highly (0.94). Similarly, the Principal Component Analysis showed that 93.6% of the variance was accounted for by a single component and only the first component received an eigenvalue score greater than 1.

It should be noted that the authors discuss the kinds of variability that emerge beyond this broad statistical analysis. Even in dominant naming patterns some interesting differences emerge. All five languages share 10 dominant terms that may be treated as rough equivalents: green, blue, purple, pink, brown, yellow, red, orange, grey, white. However, even in dominant naming we begin to see diversity. Speakers of both North American and British English used the term *peach* twice as a dominant term (for the same stimuli). For one of these, Icelandic speakers used *húðlitur* ‘skin-color’ (a term that occurred recurrently in the EoSS data; see Zimmermann et al. (2015) for discussion). In addition, the two Englishes each introduced one additional term of its own but in a different area of the colour palette. North American English introduced *turquoise* as the dominant term for one color in the blue-green area, while British English introduced *maroon* for one color in the dark brown-red area. In their in-depth discussion, Guðmundsdóttir Beck & Whelpton (To appear) explore these kinds of diversity further.

---

<sup>3</sup> A preliminary statistical analysis was conducted on a snapshot of the data from May 2016 by the Social Sciences Institute of the University of Iceland (Tryggvadóttir & Jónsdóttir 2016); I would like to thank the authors for their invaluable analytical work and discussions. The analysis presented here is based on a snapshot of the data from August 2017. The analysis is conducted in R version 3.3.2 (2016-10-31). I would like to thank Michael Dunn, Joe Jalbert and Helgi Guðmundsson for their help with the R analysis. All errors and misunderstandings remain mine solely.

## 5. Conclusion

The picture that emerges is a complex one. However, it is clear that different semantic categories vary to different extents, even in closely related languages. The least variable categories are colours and body parts. These are domains for which humans have specialised neuropsychological architecture (visual perception and proprioception). The same does not go for cultural artefacts such as containers or particular configurations of spatial relations. Although culture is an important factor in all these domains, it cannot be the sole reason for the high similarity of colour naming and body part naming, as containers at least should be just as susceptible to cultural factors in a West European context. Further, the variability in the domain of containers and spatial relations is of a potentially different type. While container naming is affected by a complex interaction of cultural, linguistic and formal factors, the variability in the domain of spatial relations may well be structured by a smaller number of coding choices which may reflect other kinds of internal constraint.

Although more detailed investigations and a broader range of techniques are necessary to establish a fuller picture, it is clear that experimental extensional semantics provides a useful tool in probing some of the deeper questions of lexical semantics.

## References

- Ameel, Eef, Gert Storms, Barbara C. Malt & Steven A. Sloman (2005): How bilinguals solve the naming problem. In: *Journal of Memory and Language* 53, 60–80.
- Arnbjörnsdóttir, Birna (2006): *North American Icelandic – the Life of a Language*. Winnipeg: University of Manitoba Press.
- Berlin, Brent (1992): *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies*. Princeton, N.J.: Princeton University Press.
- Berlin, Brent & Paul Kay (1999): *Basic Colour Terms: Their Universality and Evolution*. Stanford, California: CSLI Publications.
- Berthele, Raphael, Matthew Whelpton, Åshild Næss, & Pieter Duijff (2015): Static spatial descriptions in five Germanic languages. In: *Language Sciences* 49, 82–101.
- Bowerman, Melissa & Eric Pederson (1992): Topological relations picture series. In: Stephen C Levinson (Eds.): *Space stimuli kit 1.2: November 1992*. Nijmegen: Max Planck Institute for Psycholinguistics, 51.
- Brachin, Pierre (1985): *The Dutch Language: A Survey*. Leiden: Brill.
- Corbett, Greville G. & Ian R. L. Davies (1997): Establishing basic color terms: measures and techniques. In: CL Hardin & Luisa Maffi (Eds.): *Color Categories in Thought and Language*. Cambridge University Press, 197–223.

- Fodor, Jerry A. (1975): *The Language of Thought*. New York: Crowell.
- Gentner, Dedre (1981): Some interesting differences between verbs and nouns. In: *Cognition and brain theory* 4, 161–178.
- Gentner, Dedre & Lera Boroditsky (2001): Individuation, relativity, and early word learning. In: Melissa Bowerman & Stephen C Levinson (Eds.): *Language acquisition and conceptual development*. Cambridge: Cambridge University Press, 215–256.
- Guðmundsdóttir Beck, Þórhalla & Matthew Whelpton (To appear): Universals and Variability of Color-Naming in Icelandic, Icelandic Sign Language and North American Icelandic. In: Ida Raffaelli Daniela Katunar & Barbara Kerovec (Eds.): *Lexicalization patterns in color naming: a cross-linguistic perspective*. John Benjamins.
- Haspelmath, Martin (2003): The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In: Michael Tomasello (Eds.): *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Lawrence Erlbaum Associates, 211–42.
- Jordan, Fiona M., Michael Dunn & Asifa Majid (2009): Body Part naming booklet. Developed for the EoSS project.
- Kay, Paul, Brent Berlin, Luisa Maffi, William R. Merrifield & Richard Cook (2009): *The World Color Survey*. Stanford, California: Center for the Study of Language and Information.
- Labov, William (1973): The boundaries of words and their meanings. In: CJN Bailey & RW Shuy (Eds.): *New Ways of Analyzing Variation in English*. Washington: Georgetown University Press, 340–373.
- Landau, Barbara & Ray Jackendoff (1993): Whence and whither in spatial language and spatial cognition? In: *Behavioral and Brain Sciences* 16, 255–265.
- Majid, Asifa (2010): Words for parts of the body. *Words and the Mind: How Words Capture Human Experience*. Oxford: Oxford University Press, 58–71.
- Majid, Asifa, Fiona Jordan & Michael Dunn (2011): *Evolution of Semantic Systems Procedures Manual*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Majid, Asifa, Fiona Jordan & Michael Dunn (2015): Semantic systems in closely related languages. In: *Language Sciences* 49, 1–18.
- Malt, Barbara C., Steven A. Sloman & Silvia P. Gennari (2003): Universality and language specificity in object naming. In: *Journal of Memory and Language* 49, 20–42.
- Malt, Barbara C., Steven A. Sloman, Silvia Gennari, Meiyi Shi & Yuan Wang (1999): Knowing versus Naming: Similarity and the Linguistic Categorization of Artifacts. In: *Journal of Memory and Language* 40, 230–262.
- Rosch, Eleanor H. (1973): Natural categories. In: *Cognitive Psychology* 4, 328–350.
- Sapir, Edward (1912): Language and Environment. In: *American Anthropologist* 14, 226–242.

- Sapir, Edward (1929): The Status of Linguistics as a Science. In: *Language* 5, 207–214.
- Saussure, Ferdinand de (1922): *Cours de linguistique générale* (C Bally, A Sechehaye, and A Riedlinger, Eds.). Paris: Payot.
- Sverrisdóttir, Rannveig & Kristín Lena Þorvaldsdóttir (2016): Why is the SKY BLUE? On colour signs in Icelandic Sign Language. In: Ulrike Zeshan & Keiko Sagara (Eds.): *Semantic Fields in Sign Languages Colour, Kinship and Quantification*. Berlin, Boston: De Gruyter, 209–250.
- Talmy, Leonard (1983): How language structures space. In: Herbert L Pick Jr. & Linda P Acredolo (Eds.): *Spatial Orientation: Theory, Research and Application*. New York: Plenum Press, 225–282.
- Talmy, Leonard (2000): *Towards a Cognitive Semantics 2: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.
- The Isshinkai Foundation (2005): *Ishihara's Design Charts for Colour Deficiency of Unlettered Persons*. Tokyo, Japan: Kanehara Trading.
- Tryggvadóttir, Guðný Bergþóra & Guðbjörg Andrea Jónsdóttir (2016): *Colour naming*. Félagsvísindastofnun Háskóla Íslands.
- Waggoner, T.L. (2002): Quick six colour vision test pseudoisochromatic plates. *Colour vision testing made easy*. Good-Lite company.
- Whelpton, Matthew, Þórhalla Guðmundsdóttir Beck & Fiona M. Jordan (2015): The semantics and morphology of household container names in Icelandic and Dutch. In: *Language Sciences* 49, 67–81.
- Whorf, Benjamin Lee (1956): Science and linguistics. In: Stuart Chase & John B Carroll (Eds.): *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. New York: MIT Press, 207–219.
- Wittgenstein, Ludwig (1953): *Philosophical investigations* (GEM Anscombe, Tran.). New York: Macmillan.
- Zimmermann, Martina, Carsten Levisen, Þórhalla Guðmundsdóttir Beck, & Cornelia van Scherpenberg (2015): Please pass me the skin coloured crayon! Semantics, socialisation, and folk models of race in contemporary Europe. In: *Language Sciences* 49, 35–50.
- Þráinsson, Höskuldur (1979): Hvað merkir orðið bolli? [What does the word cup mean?]. In: *Íslenskt mál og almenn málfræði* 1, 96–106.

Dr. Matthew Whelpton  
 Professor of English Linguistics  
 University of Iceland  
 Brynjólfsgata 1  
 107 Reykjavík  
 whelpton@hi.is



Sessioner



# Norsk historisk normeringsordbok

*Sturla Berg-Olsen, Knut E. Karlsen & Dagfinn Rødningen*

In this paper we present ideas for a historical dictionary of the two Norwegian written standards, Bokmål and Nynorsk. Such a dictionary resource would compile the necessary information from disparate sources in one place and make it available and easily searchable. The dictionary would contribute to the documentation of the recent history of written Norwegian. Furthermore, it would meet the needs of the interested public and serve as an important reference tool for the Language Council of Norway when considering adjustments of the norms in the future. The project is complicated by the many orthography reforms throughout the 20<sup>th</sup> century and the multitude of changes in orthography and inflection for both Norwegian standards.

## 1. Innleiing

I denne artikkelen vil vi diskutere ein idé vi har om å opprette ein database med informasjon om rettskrivingshistoria til moderne norsk, med arbeidstittellen *Norsk historisk normeringsordbok* (NHN). Vi er alle tre tilsette i Språkrådet i Noreg, som er statens fagorgan for språkspørsmål. Som forgjengarane Norsk språknemnd (1952–1972) og Norsk språkråd (1972–2005) har Språkrådet mellom dei viktigaste oppgåvene sine å stå for normeringa av dei offisielle målformene bokmål og nynorsk og å informere folk om språkspørsmål, inkludert normtilhøve.

Nordmenn flest er svært interesserte i språkspørsmål, og etterspørselen er stor etter nettressursar der folk kan finne svar på spørsmåla sine. Vi meiner ein database som gjer det mogleg å følge stave- og bøyingsmåten av enkeltord gjennom normeringshistoria vår, kan fylle eit behov både for den språkinteresserte nordmannen og for Språkrådet.

## 2. Kvifor treng vi ei norsk normeringsordbok?

Svartenesta i Språkrådet får ofte spørsmål som handlar om rettskriving. Ikkje sjeldan har spørsmåla ein historisk dimensjon ved seg, til dømes ved at spørjaren lurar på når ein viss skrivemåte vart innført, eller når det ikkje lenger vart lov å skrive slik eller slik. Også grammatiske spørsmål kan ha den same historiske dimensjonen og handle om tidspunkt for innføring eller avskaffing av bøyings- eller avleiingsendingar i skrift.

Det er ikkje så rart om folk stiller oss desse spørsmåla. Noreg har ei moderne språkhistorie der språkreformer og rettskrivingsvedtak har kome tett. Etter at Stor-

tinget i 1885 vedtok at Noreg skal ha to likestilte norske skriftspråk, har vi til saman hatt tolv ulike skriftnormalar for norsk, seks for bokmål og seks for nynorsk. Frå 1917 og meir enn eit halvt hundreår framover var skriftnormalane dessutan prega av mange valfrie alternativ både i rettstaving og i bøyning og avleiing av ord. Det var særleg målet om å føre bokmål og nynorsk nærmare kvarandre, den såkalla samnorskpolitikken, som førte til dei mange valfrie formene, og tanken var at om ein gjorde det valfritt å skrive orda likt eller likare i begge målformene, ville folk nytte seg av det. Ein annan måte å føre språka saman på, var å forby skrivemåtane som var mest ulike i målformene, og også det vart gjort.

Etter at samnorskpolitikken vart nedtona mot slutten av nittenhundretalet og etter kvart avvikla, har det vore eit mål å redusere talet på valfrie alternativ i begge målformene, men det er framleis relativt stor valfridom i både bokmål og nynorsk. Valfridomen er i dag i hovudsak begrunna med pedagogiske argument og identitetsargument. Det skal eit stykke på veg vere mogleg for alle å skrive om lag slik dei snakkar.

Dagens normer kan ein slå opp i ordbøkene for å finne, men det kan vere vanskeleg for folk å få eit godt oversyn over rettskrivingshistoria til eit ord. Skrivemåten kan ha endra seg, eventuelt i fleire steg, og ein eldre skrivemåte kan igjen ha vorte tillaten etter å ha vore i strid med offisiell rettskriving ein periode. I dag har vi ikkje nokon stad der denne informasjonen er samla og søkbar.

Også for oss i Språkrådet vil ein base med slik søkbar informasjon bli ei god hjelp. Dels vil vi kunne få eit nyttig verktøy å bruke når vi svarar på spørsmål frå folk, men ei normeringsordbok kan i tillegg gi viktig informasjon for Språkrådet i arbeidet med den vidare normeringa av bokmål og nynorsk. Kunnskap om tidlegare praksis kan vere ein sentral del av grunnlaget for nye vedtak. Ei historisk normeringsordbok vil derfor utgjere ein samla ressurs det er stort behov for, men som vi ikkje har tilgang til i dag.

### 3. Førebilete i Danmark og Sverige

Både i Danmark og Sverige er det oppretta databasar der ein kan søke etter noko av informasjonen vi er ute etter, for dansk og svensk. Danske RO<sub>hist</sub> inneheld om lag 64 000 opplagsord, medan svenske SAOL<sub>hist</sub> har om lag 200 000.

Det er Dansk Sprognævn som står bak RO<sub>hist</sub>. I basen kan ein søke i dei fem utgåvene av *Rettskrivningsordbogen* frå 1955 til 2012. Desse er søkbare frå basens søkefelt. I tillegg er fem eldre rettskrivningsordbøker frå 1872 til 1946 skanna slik at ein kan slå opp og bla i dei, men dei er ikkje søkbare frå søkefeltet. Søkeresultatet blir presentert i ei tidslinje der berre det reine oppslaget kjem opp, utan fleire opplysningar. Fulle opplysningar får ein på ein faksimileversjon av ordboksida som ein kjem til ved å klikke på ordet i den aktuelle utgåva. Det er berre stavemåten av ordet som blir vist i tidslinja. Det er ikkje mogleg å få opp noko samla oversyn over orda med bøyings-, uttale- eller tydingsopplysningar.

SAOL<sub>hist</sub> er ein søkbar database over alle oppslagsord som har funnest i minst éi utgåve av *Svenska Akademiens ordlista* frå 1874 til 2015. Basen inkluderer også *Ordbok öfver svenska språket* av A. F. Dalin (1850–1853). Det ein først og fremst kan finne i basen, er opplysningar om korleis ord vart stava i dei ulike utgåvene, når dei kom inn i SAOL første gongen og eventuelt når dei forsvann ut. Visinga er om lag den same som i RO<sub>hist</sub>, og det er berre stavemåten som er vist i tidslinja. Heller ikkje her får ein noko samla oversyn over bøyning eller tyding.

Oppsettet med vising i tidslinje er eit format det kan gå an å bygge på i NHN. Det vi særleg kan tenke oss å utvikle vidare, er moglegheitene for å få opp fleire opplysningar i tidslinja, slik at ein ikkje treng å gjere fleire operasjonar for å skaffe seg eit bilete over rettskrivings- og bøyingshistorikken til eit ord.

## 4. Kjelder til normering av norsk

Eit hovudmål med å lage ei norsk historisk normeringsordbok er å samle informasjon som er spreidd over mange kjelder, på éin stad. Normeringa av dei to norske målformene har opp gjennom åra skjedd på fleire måtar, og ein kan òg snakke om fleire nivå av normering. I dokumentet *Retningslinjer for normering* frå 2015 trekkjer Språkrådet opp eit skilje mellom eksplisitt normering og implisitt normering. Det eksplisitt normerte ordtilfanget finst ifølgje dokumentet i desse kjeldene (Språkrådet 2015:6):

- ordbanken for bokmål og ordbanken for nynorsk
- tidlegare rettskrivingsvedtak og avgjerder som ikkje strir mot den seinaste normeringa
- ordlister og ordbøker til skulebruk som Språkrådet har godkjent
- normative publikasjonar frå Språkrådet

Med implisitt normering er det tenkt på språket i dei læremidla (først og fremst lærebøker) som vart gjennomgatte og godkjende av Norsk språknemnd og Norsk språkråd mellom 1952 og 2000. Når ei lærebok kom på trykk i denne perioden, var språket i læreboka per definisjon godkjent som normert.

Nedanfor ser vi litt nærare på dei ulike typane normeringskjelder som finst og som er aktuelle som grunnlagsmateriale for NHN.

### 4.1. Innstillingar og vedtak i høve rettskrivingsreformer

I samband med alle dei større rettskrivingsreformene på nittenhundretalet vart det gjort forarbeid av ei eller fleire nemnder. Nemndene publiserte rapportar og innstillingar der dei gjorde greie for endringane dei ønskte, og – i varierende grad

– for argumentasjonen bak endringsframlegga. I den grad desse innstillingane faktisk vart vedtekne, vil dei sjølvsagt danne eit svært sentralt grunnlagsmateriale for normeringsordboka. Samstundes går innstillingane ikkje i djupna slik ei rettskrivingsordbok eller -ordliste kan gjere. På mange måtar kan ein seie at innstillingane i samband med reformene staka ut dei store linjene, medan det vart ei oppgåve for ordlisteredaktørane å meisle ut dei ulike reformene i praksis.

## 4.2. Godkjende ordlister og ordbøker

Som nemnt i retningslinjene for normering (Språkrådet 2015) er det Norsk ordbank som i dag er den sentrale kjelda for gjeldande normering av norsk. Ordbanken er samansett av to databasar, ein for bokmål og ein for nynorsk, som begge inneheld ei stor mengd lemma med informasjon om normert skrivemåte og bøyning. I skrivande stund er det registrert drygt 150 000 lemma i ordbanken for bokmål og ca. 250 000 lemma i ordbanken for nynorsk. Til samanlikning inneheld *Bokmålsordboka* berre ca. 60 000 lemma og *Nynorskordboka* ca. 90 000 lemma. Begge desse ordbøkene hentar i dag informasjon om staving og bøyning frå ordbanken. I Norsk ordbank er det òg lagt til rette for registrering av historisk normering. Førebels er dette systemet utnytta i liten grad, men det er mogleg å spore nyare endringar, til dømes endringar i samband med den siste reforma for nynorsk i 2012.

For historiske normeringsdata er ordlister og ordbøker som har vore godkjende av Norsk språknemnd, Norsk språkråd og Språkrådet til skulebruk, heilt sentrale kjelder. Etter kvar rettskrivingsreform har det vore behov for nye ordlister, og som nemnt ovanfor er det i stor grad nettopp skuleordlistene som har tolka normene og meisla dei ut i detalj lemma for lemma. Ordlistar til skulebruk er stort sett nokså tynne, men vi har òg meir omfattande ordlister og ordbøker, som *Norsk rettskrivningsordbok* av Sverdrup og Sandvei, først gitt ut i 1940 og seinare døypt om til *Tanums store rettskrivningsordbok*, og *Bokmålsordboka* og *Nynorskordboka*, som begge kom ut første gong i 1986.

For at ein skal kunne dra nytte av ordlister og ordbøker i ei historisk normeringsordbok, er det ein føresetnad at dei er digitaliserte. Ved Universitetet i Oslo er det allereie gjort mykje for å digitalisere og indeksere ordlister frå 1900-talet, jf. Grønvik & Ore (2018; denne utgivinga). I del 5 nemner vi døme på ordlister og ordbøker som bør tene som grunnlag for normeringsordboka.

## 4.3. Einskildvedtak

I tillegg til dei store rettskrivingsreformene har det frå 1950-åra og frametter vorte gjort ei rekkje normeringsvedtak som omfattar anten einskilte ord eller grupper av ord. I nokre tilfelle er det òg snakk om systemendringar som omfattar svært mange ord, til dømes alle hokjønnsstaktiv i bokmål eller alle sterke hokjønnsstaktiv i nynorsk.

Einskildvedtaka er kunngjorde i årsmeldingane frå Norsk språknemnd, Norsk språkråd og Språkrådet, som vart publiserte på papir frå 1952 til 2006. Årsmeldingane er digitaliserte og ligg per i dag tilgjengelege i pdf-format på internett, men dei er ikkje søkbare. Språkrådet arbeider for å gjere årsmeldingane og vedtaka meir tilgjengelege og søkbare i ein såkalla vedtaksdatabase. Planen var opphavleg å lenke normeringsvedtaka til Norsk ordbank, men i ein del tilfelle viser det seg vanskeleg å knyte eit vedtak til eitt konkret lemma. Nokre av vedtaka gjeld svært mange lemma, og nokre av dei er utforma som opne reglar som kan omfatte ei ukjend mengd lemma, slik som dette frå årsmeldinga for 2000, side 13:

Verb som er laga av substantiv med valfri enkeltkonsonant/dobbelkonsonant i utlyd, skal ha obligatorisk dobbeltkonsonant, eksempelvis *handikap(p) > handikappe*

Informasjon om rettskrivingsvedtak frå 2014 og framover er i dag å finne på Språkrådets nettsider, men desse må òg innarbeidast i vedtaksbasen og sjølvsagt takast med i NHN.

#### 4.4. Andre kjelder – operativ normering

Som nemnt ovanfor, nemner retningslinjene for normering også «normative publikasjonar frå Språkrådet» som ein type kjelde for normering av norsk. Eksempel på slike publikasjonar er Språkrådets termwiki, Språkrådets svarbase og listene med geografiske og historiske namn som er publiserte på Språkrådets nettsider.

I tillegg til dette må ein rekne med det ein kan kalle for *operativ normering*. Dette er utsegner om tilrådde skrivemåtar som Språkrådet gir, først og fremst i svar på språkspørsmål på e-post eller i telefonen. Det kan vere tilfelle der ordbøkene og dei eksisterande normeringsvedtaka ikkje gir klare svar, men der sekretariatet i Språkrådet vurderer det slik at det ikkje er naudsynt å ta vegen om formell sakshandsaming i fagrådet for normering og språkoobservasjon og Språkrådets styre. Eit konkret døme på operativ normering er det når Språkrådet på spørsmål tilrådde skrivemåten *RIB* eller *ribb* for den båttypen som på engelsk heiter *rigid-hulled inflatable boat*. Desse skrivemåtane er komne inn i Norsk ordbank og har òg fått artiklar i *Bokmålsordboka* og *Nynorskordboka*, og dei må såleis seiast å vere normerte. Òg utsegner som berre står i eit e-postsvar, må vurderast å ha ei viss normerande kraft.

### 5. Utval av kjelder til ordboka

Nasjonalbiblioteket i Noreg har gjort ca. 250 000 bøker digitalt tilgjengelege i eit nettbibliotek. Bøkene blir publiserte som faksimile (biletfil) og er søkbare (Nasjonalbiblioteket). Bokutvalet omfattar også norske ordbøker og ordlister, som med

få unntak er digitalt tilgjengelege for alle med norsk IP-adresse. Denne ressursen kan spele ei viktig rolle i eit normeringsordbokprosjekt. I tillegg gir denne tenesta frå Nasjonalbiblioteket på ein enkel måte oversikt over det som er gitt ut av norske einspråklege ordbøker og ordlister, og det er dei som er interessante i denne samanhengen.

Som vi peikte på i del 2, har vi fram til i dag hatt til saman tolv forskjellige skriftnormalar for bokmål og nynorsk. Etter ei reform vil det naturleg nok vere behov for å informere om endringar i nye utgåver av ordlister og ordbøker, og såleis har normeringshistoria medverka til at denne sjangeren er heller talrik. Det er derfor ei viktig oppgåve å peike ut dei verka som skal danne grunnlaget for normeringsordboka i den forstand at dei uttrykkjer den offisielle skriftspråknorma for si tid.

Ser ein på utvalet av ordbøker og -lister, så varierer dei sterkt i omfang, frå korte ordlister til tjukke ordbøker. Mange har knappe bøyingsopplysningar, kanskje er berre kjønnet på substantivet gitt opp, mens andre gir opp heile bøyingsparadigme. Dei fleste hevdar autoriteten sin gjennom ei offisiell godkjenning (av departementet eller eit organ som Språkrådet).

Det er dermed fleire faktorar å ta omsyn til når ein skal peike ut referanseverka, men omfanget, både av oppslagsord og bøyingsinformasjon, vil vere eit viktig kriterium. I tillegg må informasjonen sjølvstøtt vere pålitande.

Det krev god kjennskap til norsk normeringshistorie for å vurdere kva for ordlister og ordbøker for nynorsk og for bokmål som kan utgjere grunnlaget for normeringsordboka. Det må gjerast eit utval for kvar reform, og for nokre av normalane kan det vere aktuelt å ta med fleire referanseverk som kan utfylle kvarandre. For dei fleste aktuelle verka er det mogleg å sjå faksimile av originalane i nettbiblioteket til Nasjonalbiblioteket. Når det gjeld dei tidlegaste reformene, er det ikkje lenger knytt opphavsrett til dei aktuelle referanseverka, og for dei siste reformene (1981 og 2005 for bokmål, og 1959 og 2012 for nynorsk) ser vi at *Nynorskordboka* og *Bokmålsordboka* er dei sentrale referanseverka. Dette er verk som Språkrådet er deleigar i, og dermed er spørsmålet om opphavsrett enklare. Når det gjeld bokmål, så må det nemnast at det vi i dag kjenner som *Tanums store rettskrivningsordbok* (og som har komme i talrike utgåver frå 1940), ville vore eit naturleg referanseverk. Utgåvene av denne ordboka finst ikkje fritt tilgjengelege i nettbiblioteket, men målet må vere å få til ein avtale om tilgang. Det er mogleg for forlag å reservere seg mot å bli innlemma i nettbiblioteket, og det er truleg tilfellet her. *Bokmålsordboka* kom første gong i 1986 og spegla då 1981-normalen, mens Tanum kunne vore ei nyttig kjelde for bokmålsnormalane frå 1938 og 1959.

## 6. Korleis ordboka kan utformast

Vi skal gi nokre døme på korleis informasjon frå referanseverka kan stillast saman til ein artikkel i normeringsordboka. Her er det all mogleg grunn til å leggje vekt



på *kan*, for det er mange ulike løysingar, alt etter kor mykje ein er villig eller har høve til å investere av ressursar.

Figur 1 viser normeringshistoria til substantivet *sykkel*, som er døme på eit ord med ei enkel normeringshistorie, presentert i tabellform. Dette minner om måten informasjon er presentert på i SAOL<sub>hist</sub> og RO<sub>hist</sub>.

1907	1910	1917	1938	1941	1959	1981	2005	2012
<i>cykkel</i>		<i>sykkel</i> (bm) <i>sykkel</i> (nn)					<i>sykkel</i> <i>sykkelen</i> <i>sykler</i> <i>syklene</i>	<i>sykkel</i> <i>sykkelen</i> <i>syklar</i> <i>syklane</i>

Figur 1: Døme på korleis eit ord med enkel normeringshistorie kan presenterast.

Her er informasjonen ekspandert med fulle bøyingsparadigme i gjeldande normal, men dette kan sjølvsagt forenklast slik at det berre er oppslagsforma som er synleg og klikkbar. Oppslaget viser at den normerte skrivemåten var *cykkel* i bokmålsrettskrivinga frå 1907, og at han vart norvagensert til *sykkel* i begge målformene i 1917. Når det ikkje er gitt opp noko informasjon om skrivemåten i nynorsk før 1917, heng det saman med at ordet ikkje er å finne i dei valde referanseverka før den tid. Det er elles lagt inn hyperkoplingar i framstillinga slik at klikk på oppslagsordet vil føre brukaren til det valde referanseverket. Med denne løysinga vil ein berre få fram dei opplysningane som kjem fram i ordliste-/ordboksoppslaget. Annan implisitt informasjon, som til dømes tillaten infinitivsending, må ein då søke opp i forord eller andre framstillingar av reforma. Det er ei løysing som krev mykje av brukaren, men som er mindre ressurskrevjande å lage.

Det andre dømet vi skal sjå nærare på, er verbet *stjele/stele*, eit ord med ei temmeleg omfattande normeringshistorie. Her er det så mykje informasjon som skal formidlast, at tidslinje i tabellform ikkje eignar seg, jf. figur 2. Det har skjedd talrike endringar frå dei første normalane fram til i dag i begge målformene.

2012 (nn)	stele/stela (+ kløyvd inf.) – stel – stal – stole
2005 (bm)	stjele – stjeler – stjal – stjålet
1981 (bm)	stjele [stjela] – stjeler – stjal – stjålet stele [stela] – steler – stal – stålet
1959 (nn)	stele/stela [+ kløyvd infinitiv] – stel [steler] – stal – stole/stoli
1941 (nn)	stela – stel – stal – stole
1941 (bm)	stjele – stjeler – stjal – stjålet
1938 (nn)	stela (+ kløyvd inf.) [stele] – stel [steler] – stal – stoli

1938 (bm)	stjele [stjela] – stjeler – stjal – stjålet stele [stela] – steler – stal – stålet
1917 (bm)	stjele – stjeler – stjal – stjålet
1917 (nn)	stela/stele [kløyvd inf. for elevar] – stel/steler – stal – stole
1907 (bm)	stjæle – stjæler – stjal – stjaalet
1901 (nn)	stela – stel – stal – stole

Figur 2: Døme på korleis ord med kompleks normeringshistorie kan presenterast.

Ein viktig grunn til dette er den skiftande statusen til infinitivsendinga, særleg i nynorsk, men også i bokmål. Infinitivsendinga i nynorsk kan som kjent anten vere *-a* eller *-e*. Som skrivar kan ein velje, og så held ein på same endinga gjennom heile teksten. I tillegg kjem systemet vi kjenner som kløyvd infinitiv, der infinitivsendinga skifter mellom *-a* og *-e* etter gamle kvantitetsforhold. Dette systemet er levande i enkelte moderne dialektar, og det er framleis mogleg å reflektere systemet i nynorsk skriftmål i dag. I bokmål fall kløyvd infinitiv ut i 2005. Også i denne framstillinga er det laga hyperkoplingar til dei aktuelle referanseverka, men igjen kan ein merke seg at ein del av normeringsinformasjonen ikkje er å finne på oppslagsplass, men er generelt formulert, ofte i føreordet. Det gjeld typisk informasjon som omfattar ei heil ordgruppe, som til dømes at det er høve til å bruke både *e*-infinitiv og *a*-infinitiv i nynorsk.

## 7. Konklusjon

Det norske språksamfunnet manglar ei einskapleg framstilling av normeringshistoria til enskildord. Både som del av grunnforskinga for å dokumentere den nyare språkhistoria og som praktisk ressurs for å kunne svare når Språkrådet får spørsmål om historiske endringar i norma, er det viktig å gjere den fragmenterte informasjonen som finst på feltet, tilgjengeleg. Den norske normeringshistoria med hyppige rettskrivingsendringar talar i seg sjølv for ei samla framstilling. SAOL<sub>hist</sub> og RO<sub>hist</sub> kan fungere som gode føredøme for ei norsk normeringsordbok.

Det må peikast ut referanseordbøker og -ordlister som uttrykkjer den operative norma for kvar reform, og i tillegg må årsmeldingane til Språkrådet utgjere ein del av grunnlaget for normeringsordboka. Det er viktig å streke under at mykje informasjon om skrivemåte og bøyning blir oppfatta som implisitt informasjon. Det inneber at mykje (og viktig) formalinformasjon ikkje kjem fram i ordlisteoppslaget/ordboksartikkelen, men må i beste fall leitast opp i rettleiinga/forordet til ordlista/-boka. Dette kan vere problematisk dersom skrive- og bøyingsmåte skal søkast fram maskinelt.

Etter at språksamlingane og ordbøkene vart overførte frå UiO til UiB, ligg det

no til rette for at det kjem eit leksikografmiljø ved UiB. Eit normeringsordbok-prosjekt bør sjåast i samanheng med oppbygginga av eit slikt miljø.

## Litteratur

### Ordbøker

*Bokmålsordboka*. <<http://ordbok.uib.no>> (september 2017).

*Nynorskordboka*. <<http://ordbok.uib.no>> (september 2017).

RO<sub>hist</sub>. <<http://rohist.dsn.dk/>> (september 2017).

SAOL<sub>hist</sub>. <<http://spraakdata.gu.se/saolhist/>> (september 2017).

*Tanums store rettskrivningsordbok*. (2015). 10. utgave. Oslo: Kunnskapsforlaget.

### Annan litteratur

Grønvik, Oddrun & Christian-Emil Ore (2018): Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforskning av ordtilfang. I: *Nordiske Studier i Leksikografi* 14 (denne utgivinga). Reykjavík: NFL & Stofnun Árna Magnússonar í íslenskum fræðum, 87–95.

Nasjonalbiblioteket. <<https://beta.nb.no/search>> (september 2017).

Norsk ordbank for bokmål. <<http://inger.uib.no/perl/search/search.cgi?appid=72&tabid=1106>> (september 2017).

Norsk ordbank for nynorsk. <<http://inger.uib.no/perl/search/search.cgi?appid=73&tabid=1116>> (september 2017).

Språkrådet (2015): *Retningslinjer for normering. Nynorskversjon*. <<http://www.sprakradet.no/globalassets/spraka-vare/norsk/retningslinjer-for-normering-nynorskversjon.pdf>> (september 2017).

Språkrådets termwiki. <<http://www.termwiki.sprakradet.no/wiki/Hovedside>> (september 2017).

Årsmeldingar frå Norsk språknemnd, Norsk språkråd og Språkrådet. <<http://inger.uib.no/perl/search/search.cgi?appid=241&tabid=3173>> (september 2017).

Sturla Berg-Olsen  
seniorrådgivar, dr.art.  
sturla.berg.olsen@sprakradet.no

Knut E. Karlsen  
seniorrådgivar, ph.d.  
knut.e.karlsen@sprakradet.no

Dagfinn Rødningen  
seniorrådgivar, cand.philol.  
dagfinn.rodningen@sprakradet.no

Språkrådet  
Postboks 8107 Dep  
NO-0032 Oslo

# Adjektiv eller adverbium?

Lars Brink

Ordnung muß sein in our word classes. I intend to solve the old issue whether *smukt* (beautifully), *pænt* (nicely) etc. in fx *De danser smukt – Han spiller pænt* (They dance beautifully – He plays nicely) are adverbs (derived with a *-t* from adjectives), or adjectives in the neuter used as adverbs. When I was in secondary school, I learnt they were adverbs, whereas my university teacher, Erik Hansen, pontificated that they be best analyzed as neuter adjectives. Galberg Jacobsen consider them to be adverbs. The issue has lexicographic consequences. If the above *smukt* is an adverb, it is a different word from *smuk* and must have its own entry, whereby all instances of the adverb must be removed from the entry of *smuk*. If not, everything is united under *smuk* adj. The problem is solved by scrutinizing the difference between inflection morphemes and derivation morphemes. The root morpheme (*smuk-*) has the very same meaning in the two functions so it has to be a meaning analysis of the morpheme *-t* that settles the matter.

## 1. Indledning

Er *smukt* (svensk *vackert*) i *De danser smukt* et adjektiv (i neutrum, i adverbial funktion) eller et adverbium (afledt af adjektivet *smuk* med suffikset *-t*)? Dét er et gammelt grammatisk stridsspørgsmål. De fleste universitetsgrammatikere (men ikke fx Galberg Jacobsen 2013:38) stemmer for adjektiv. Det må vi have afklaret. Og hvad er mere naturligt end at fokusere på forskellen mellem bøjning og afledning?

## 2. Bøjning vs. afledning generelt

### 2.1. Sikre afledninger

Lad os starte med de sikre, dvs. de helt ukontroversielle tilfælde af afledning: *sund – sundhed – sundhedsmæssig*, *Shakespeare – Shakespearesk*, *høj – højde* og meget andet. Disse ændringer kalder alle for afledning. Og grunden er:

Slutmorfemet har dén eneste egenskab, der ifølge alle grammatikere skiller afledning fra bøjning: Det kan få ordet til at skifte ordklassefunktion. Som fx *-agtig*, *-dom*, *-eri*, *-skab*, nul<sup>1</sup> (*spring*), *-me* (*rødme*). Alle andre forskelle mellem, hvad

<sup>1</sup> Eller som *et skrig*, *et køb*, *en kvind* (humoristisk), *en leg*, *et spil*, *et „sgu“* (et eksemplar af bandeordet *sgu*).

der normalt kaldes fleksiv og affiks, er kun tendentielle. Fx: meget åbne ordgrupper vs. mindre åbne; svagtryk vs. bitryk/hovedtryk; ren syntaktisk betydning vs. ydre verdensbetydning (fx *-en* bestemt fleksiv vs. *ur-* (fx i *urkraft*)). Slutmorfebet *-eri*, *-skab* osv. viser altså positivt funktionsskiftet, fx viser *bageri* alene i kraft af sin endelse, at det er et substantiv modsat den verbale rod.

Kriteriet kan dog kun udpege afledning, ikke bøjning, thi mange morfemer, der traditionelt i enighed kaldes affikser, kan ikke ændre ordklassefunktionen: *mikro-*, *mav-se* (svensk *magelilla*), tysk *-lein*. Men hvor et morfem som sagt kan ændre ordklassefunktionen, er det ifølge alle et affiks.

Nogen vil nu sige, at vort spørgsmål er løst, for *-t* i *smuk-t* fremtryller jo adverbial funktion. Men dét tæller ikke, for hvis dette *-t* i virkeligheden er et neutrumsbøjnings-*t*, så er det slet ikke *-t*'et, der skaber ændringen. (I modsætning til *-t* i *smukt* kan de sikre suffikser (som de ovennævnte) ikke forveksles med et fleksiv).

De sikre suffikser kan tage nye fleksiver til sig: *landskab-er*, *højde-n*, *sydligere*. Disse fleksiver hører til den nye ordklasse, alt falder altså på plads. En rod kan endog tage nye fleksiver – fleksiver, der ellers ikke hører til roden i dens grundfunktion – til sig uden et manifest suffiks: *nu'et*, *valsede* (af *vals*), svensk *lillan*; i så fald regner man gerne med »nul-afledning« (nul-suffiks + fleksiv). Vores *smukt* kan ikke tilføje nogen bøjningsendelse, for det kan adverbialer nu engang ikke (bortset fra komparativ); problemet er stadig uløst.

## 2.2. Sikre ikke-afledninger

Lad os så se på de »sikre« tilfælde af ordklassefunktionsændring, der normalt ikke regnes som afledning til nyt ord.

*Farum* i *Farum Rådhus* regnes aldrig som et andet ord end *Farum*. Det forbliver et proprialt substantiv. *Iron* i engelsk *iron bar* er stadig et substantiv, ikke et adjektiv. *Juleaften* i *Jeg mødte hende juleaften* er det kære gamle substantiv, ikke et adverbium. *Av* i *Det gør av!* er stadig en interjektion. Og *før* er ikke tre forskellige ord, men ét ord, der optræder som konjunktion hhv. præposition hhv. adverbium. Ligesom adjektivet *hårrejsende* i *hårrejsende billig* ikke er forvandlet til et adverbium. (Med hensyn til *før* er der dog nogen uenighed). Og hvis man endelig gik ind på dette, at ren syntaksforskel – uden formændring – kan ændre ordklassen, hvor skulle grænsen så sættes? Der er syntaktiske småforskelle overalt. Adjektiver har tre forskellige normalsyntakser – tre forskellige ordklasser? Verber kan stå finit og infinit – to ordklasser? Dette er ikke særlig kontroversielt.

Det er altså klart, at et ord, der er helt uforandret (ingen nye affikser/fleksiver) ikke siges (og ikke bør siges) at skifte ordklasse, selvom dets forhold til sætningens andre ord (dets syntaks) skifter til en stilling typisk for andre ordklasser. Selv ikke, hvor det (måske lidt søgt) kan siges at skifte betydning: *Farum* er en by. *Farum* i *Farum Rådhus* betyder strengt taget 'som hører til Farum by'. *Før* i *Vi har før oplevet det* betyder 'før noget'. I eksotiske sprog uden bøjning og ret megen afledning springer ordene ofte voldsomt mellem funktioner, der hos os varetages

af forskellige ordklasser. En rod kan uforandret betyde 'stor mand' og 'stor' og 'kraftigt' (adverbielt). Det ville være sygt at operere med tre ord («squinting grammar»); der er derimod tale om ét ord i tre syntaktiske stillinger og tre betydninger.

Selv hvor et ord uden formændring både har en særlig syntaks og en hertil hørende gedigen sær betydning, er der ifølge alle ordbøger tale om ét og samme ord uden ordklasseskift: *Gud* forbliver et proprium i *Gud, hvor er jeg sulten!*, hvor det står interjektionelt. Talrige verber har kraftigt betydningsskift knyttet til forskellige syntakser, uden at man taler om afledning: *Jeg vælter bilen – Bilen vælter. Stikke* betyder noget helt for sig i *stikke af*, osv. Intet ved selve ordet (*vælde, stikke*) viser funktions- og betydningsskiftet.

### 2.3. Tre særtilfælde

*Ung* i *den unge* 'den unge person' og *blind* i *en blind* regnes traditionelt i ordbøger som adjektiver, dvs. ikke som afledte substantiver. Og de bøjes da også med de sædvanlige adjektivfleksiver: *den unge, de blinde*, aldrig *\*ungen, \*blindene*. (Vi taler her ikke om den ægte elliptiske brug: *et gammelt ansigt er ofte smukkere end et ungt. – Både seende killinger og blinde...* (Brink & Schack 2014)).

Ifølge Jørgen Schack og mig er dette galt. De er substantiver, for de får i virkeligheden substantivbøjning. I *en ung* regnes *en* jo for en perifrastisk artikel, og her syfter den netop på *ung*, hvorimod den i *en ung pige* slet ikke syfter på adjektivet, men på hele nominalet. Tilsvarende *de blinde*: *de* er et substantivfleksiv i bestemt pluralis, der ligesom i *de blinde vinkler* syfter på hele nominalet, ikke på adjektivet. Der er også en tredje bøjningsform, den nøgne: *Der var unge i klubben* i går parallelt med *Der var børn i klubben*, idet vi jo i pluralis ikke behøver ubestemt artikel (*nogle*). Artikler og nøgenformen er klart nok substantivfleksiver, hvorved de pågældende ord (*ung, blind* m.fl.) i den nævnte brug altså har et konkret mærke, der signalerer den substantiviske brug.

Participier inddeles med Jørgen Schack (2018) i følgende to grupper, hvortil jeg føjer en første, som Schack helst ikke vil kalde participier:

- 1) »Participier«, der ikke modsvares af verber, *fodsløbende, forfløjen, tilstundende, håndholdt, bevinget* osv., er naturligvis fødte adjektiver, ikke verber. Det er tvivlsomt, om de bør kaldes participier, når de aldrig står verbalt. Men de opfattes måske verbalt: *håndholdt* 'som er blevet holdt i hånden' osv.
- 2) Participier med modsvarende verbum, men i en ikke-processuel (egen-skabs-)betydning forskellig fra verbets procesbetydning, fx *begejstret, interesseret, strålende*. De er adjektiver: For det første, fordi de ikke længere rummer de verbale procesfleksiver i aktiv (*-ende*) hhv. passiv (*-et*) – der er jo ingen aktiv hhv. passiv betydning til stede. Morfembestanden har ændret sig ved bortfaldet af morfemerne *-et, -ende*: Et

morfem er jo kun til stede, når dets betydning er til stede – og tilbage er blot et par lyde, der tilfældigvis også optræder som morfemer andetsteds, ganske som *-t* i *stolt* tilfældigvis minder om et neutrums-*t*. For det andet bøjes disse participier – attributivt – med adjektivfleksiv: *det begejstrede barn – de interesserede damer* (-ende er dog ubøjeligt), ligesom de prædikativisk kan få adjektivfleksivet *-e* (der hos mange, fx undertegnede, er obligatorisk): *De var ikke interesserede*, hvorimod den rene verbale procesbetydning på moderne dansk er uden bøjning af participiet prædikativt (modsat svensk og nynorsk): *Bukserne er lige (blevet) lappet* (proces) – *nu er de lappede som moden kræver det* (egenskab – Erik Hansens fremragende eksempler<sup>2</sup>). Som det ses, er et givet verbum ikke bundet til en given betydning (proces/ikke-proces), en stor del kan frit afledes til ikke-processuelle adjektiver, men også forblive procesverber. De uafledte, præteritum participier (*de er lige lappet*) bør med Schack (2018) kaldes supiner.

- 3) Endelig er der participier, hvis betydning må opfattes som en verbal procesbetydning: *flyene er aflyst – gaden er fejlet – slavinderne var solgt*. De kan i moderne dansk – ligesom de ovenstående i procesbetydning – ikke bøjes i pluralis i prædikativ, og det passer jo fint med betydningen: De er (stadig) verber. Men attributivt bøjes de stadig i dansk. (Med tendens til ubøjet form på slagterskilte o.l.: *småhakked kyllinger*; der er meget lille fonetisk forskel på *hakked* og *hakkede*): *aflyste fly, den fejede gade, de solgte brude, det nedtagne skilt* (med »stærk« part.; yngre ofte: *nedtagede*). Disse attributive participier må kaldes afledte adjektiver (fleksivet viser adjektivfunktionen), men det er forståeligt, at traditionen samler dem med de nys nævnte prædikative (*flyet er aflyst*) og regner dem som verber. Der er i ordbøger intet lemma a la »aflyst adj. -e«.

Endelig særtilfældet *kæmpe* sb. (*en kæmpe – kæmpen – kæmper* osv.), svensk *jätte*, versus *en kæmpe is*. Alle ordbøger regner det som to ord, substantiv hhv. adjektiv. Tilsyneladende er det uforandret, men reelt ikke: Formen *kæmpe* står som substantiv i singularis ubestemt, »klart« markeret af sit nul-fleksiv. I *kæmpe is* er der intet nul-fleksiv, for betydningen 'singularis ubestemt' foreligger slet ikke; ordet er ubøjeligt. Substantivet har mistet sine fleksiver, har skiftet betydning og er blevet til et adjektiv.

Heller ikke de sikre ikke-afledninger hjælper os. Nemlig igen på grund af *-t*'s tvetydighed i adverbialt *god-t, smuk-t* osv. Hvis dette *t* er et neutrums-*t*, er der intet suffiks, der kunne vise adverbialfunktionen, hvorved ordet ville vedblive at være et adjektiv; og hvis *t* er et adverbialsuffiks, er ordet en sikker selvstændig afledning, altså et adverbium. Men hvad er det?

<sup>2</sup> I personlig kommunikation.

Løsningen er den følgende: *-t* kan ikke være neutrumsfleksivet, thi dets neutrumsbetydning mangler. I sætninger som *De danser smukt* signalerer *-t* ikke, at der findes et neutrum-singularis-substantiv i sætningen, som *smukt* syfter på. Der er faktisk slet ikke nogen substantiver i nævnte sætning. Et bøjningsmorfem skal naturligvis som ethvert andet morfem have sin bestemte betydning for at kunne identificeres som netop dette morfem. Er problemet så løst? Ikke helt! Adjektivi-sterne har en smutvej:

Et bøjningsmorfem kan naturligvis som alle andre morfemer have betydningsvarianter. *Glas* betyder dels 'materialet glas', dels 'drikkeglas'. Præteritums morfemet *-ede* betyder snart 'i fortiden', snart 'non-realis' (*Var jeg en Rype i Dag, løfted jeg Vingen til Slag...*), og *-en* i fx *hesten* betyder snart 'den pågældende hest nævnt i foregående sætning, eller udpeget af mig, eller den mest nærliggende i vort samtaleunivers', snart 'enhver' fx i *Hesten er et sky dyr*.

## 2.4. Polysemi eller homonymi

Kriteriet for 'betydningsvarianter', *polysemi*, er, at betydningerne føles beslægtede, i modsat fald foreligger *homonymi*.

Jeg har moret mig med at se nærmere på dette føleri – og sammenlignet mine følelser med Nudansk Ordbogs (NDO).

Jeg er enig med NDO i, at følgende betydninger er homonyme, dvs. tilhører hvert sit ord: *bor* ('redskab' – 'grundstof'), *tak* ('taksigelse' – 'lille spids'), *bog* ('tekstmasse' – 'bøgefrø'; samme oprindelse), *bryde* ('brække' – 'bryde sig om'), *stær* ('fugl' – 'øjensygdom'). Og jeg er enig med NDO i, at følgende hører til det samme, polyseme ord: *kor* ('sanggruppe' – 'kirkedel'), *rejse* vb. ('foretage lang tur' – 'oprejse'), *moster* ('mors søster' – 'ikke-ung dame'), *krone* ('kongekrone' – 'mønt' – 'trækroner'), *klare* vb. ('gennemføre' – 'overvinde' – 'gøre klarere'), *marsvin* ('delfin' – 'gnaver'), *pande* ('stegepande' – 'ansigtsdel'), *loft* (svensk 'innertak' – svensk 'vind'), *svamp* ('sporeplante' – 'havdyr' – 'vaskeredskab'), *stamme* ('træstamme' – 'menneskegruppe' – 'orddel'), *et led* ('del af kæde' – 'knogleforbindelse' – 'låge'; forskellig oprindelse).

Men vi divergerer mht. følgende, som jeg føler er samme ord: *stand* ('tilstand' – 'samfundslag' – 'salgsbod'; i NDO tre ord), og mht. følgende, som jeg føler som ubeslægtede homonymer: *linse* 1 ('lysbryder' – 'pastil' – 'kage') vs. *linse* 2 ('frø'). De har alle samme oprindelse og NDO henfører dem alle til ét og samme ord.

Det fremgår af mine eksempler, hvilke betydninger der har fælles oprindelse, og hvilke der har forskellig oprindelse, men lad mig slå fast, at det historiske ingen betydning bør have for den synkrone, leksikografiske analyse, hvad den heller ikke har, hverken for NDO eller mig.

Det må efter disse eksempler stå klart, at neutrumsbetydningen og adverbialbetydningen dårligt kan være mere forskellige:

Neutrumsbetydningen er: 'Denne stamme syfter på et bestemt neutrum-singu-



laris-ord i sætningen', fx i *Et smukt barn legede på stranden. – Smukt var barnet jo. – Barnet var, smukt og ugenert, blevet alles yndling. – Aftenskåret gjorde det [barnet] smukt.* Neutrumsbetydningen er en ren syntaktisk betydning; den udsiger kun noget om ordenes betydningsforhold til hinanden. Neutrumsbetydningen signalerer intet om, hvorvidt *t*-ordet udgør et primært sætningsled.

Adverbialbetydningen er: 'Denne rod udsiger måden, forholdet foregår på'. Fx i *De danser smukt.* Den udsiger både noget syntaktisk: at *t*-ordet er et primært sætningsled, og noget leksikalsk: at *t*-ordet udtrykker den måde, som forholdet foregår på (uden at sige noget om de optrædendes skønhed).

Hermed er problemet løst. *-t* i *De danser smuk-t* kan ikke være det samme morfem som neutrums-*t* i *Barnet er smuk-t.* For dels er neutrumsbetydningen ikke til stede i *De danser smuk-t*, dels er de to betydninger, adverbialbetydningen og neutrumsbetydningen, så totalt forskellige, at der ikke kan foreligge nogen *t*-polysemi.

Dér, hvor der slet ingen forskel er på adjektivisk og adverbial funktion, fx i *De danser strålende*, skifter ordet ikke ordklasse (jf. ovenfor: Noget i ordet skal vise den nye syntaks), ganske som *handelsvejen* forbliver et substantiv i *Han gik handelsvejen* og *Farum* forbliver et substantiv i *Farum Rådhus.*

Men mht. adjektiver, der mangler *t*-form, men som stadig bøjes i numerus/bestemthed: *stolt(e)*, *bondsk(e)* (svensk også *bondskt*), må der være tale om afledning fra adjektiv til adverbium: *De danser bondsk*, thi her har *bondsk* mistet sit nul-fleksiv, der jo i adjektivet betyder 'ental, ubestemt', idet betydningen 'ental, ubestemt' overhovedet ikke foreligger i den adverbial brug. Altså: Syntaksændring signaleret af bøjningsændring. Uanset hvad man mener om nul-morfemer, må alle erkende, at adjektiviske bøjningsformer som *bondsk* signalerer 'ental, ubestemt', men at adverbialt *bondsk* ikke er den samme bøjningsform, eftersom hin betydning totalt mangler.

Ord med valgfrit *-t*, fx i *vældig(t) god*, er naturligvis (grads)adverbier, når de har *-t*. Men også uden *-t* må de efter min opfattelse være adverbier: De afviger bøjningsmæssigt fra adjektivet (hvis nul-form betyder 'fælleskøn, singularis, ubestemt') som beskrevet ovenfor.

Naturligvis ligner neutrumsfleksivet *-t* og adverbialsuffikset *-t* hinanden så meget, at eftertænksomme sprogbrugere vil ane en historisk forbindelse (som er sand). Sådanne ligheder mellem forskellige morfemer er almindelige i sprogene. I Brink (2000) kalder jeg dem simulante. *Pap* ligner *papir*, *mand* ligner *menneske*, *han* ligner *hun* uden, at nogen regner dem for indeholdende samme morfem, skønt de faktisk går tilbage til ét og samme morfem respektivt.

### 3. Konklusion

Vi kan altså roligt i vore ordbøger udskille adverbierne fra adjektivartiklerne, hvad der altid er fordelagtigt for både forfatter og læser ved længere artikler. Altså

udskille adverbier som *godt, langt, smukt, tappert, stort* og mange flere. Vi udskiller fx *langt* adv., der dels har *t*-morfemets adverbielle grundbetydning: *langt ude i skoven*, dels har en helt anden, gradsbetydning: *langt bedre*. Ligesom vi udskiller *længst* adv. fra *længst* adj., der er en del af *lang*. Som adverbium er *længst* blevet ubøjeligt – *disse længst forsvundne byer* – som adjektiv har det to bøjningsformer, *længst* og *længste*.

Men naturligvis udskiller vi kun etablerede adverbier. Det er en dødssynd at optage uetablerede lemmaer, hvad også kun de færreste gør. Adverbier som *selvhjulpent, kompensatorisk* (som adverbium), *liechtensteinsk* (som adverbium), *slyngelagtigt, mafiøst, sterilt* hører ingen steder hjemme, da de som adverbier ikke er etablerede (lagrede og huskede) for nogen. *t*-afledning er overordentlig produktiv, ligesom fx afledning med *-agtig*. Man behøver naturligvis heller ikke, selv i en stor ordbog, at tage alle sådanne med, som kun er etablerede hos de færreste.

## Litteratur

### Ordbøger

NDO = Nudansk ordbog = Christian Becker-Christensen et al. (1998): *Politikens Store Nye Nudansk Ordbog*, 1. udg. København: Politiken.

### Anden litteratur

Brink, Lars (2000): Kongen af Danmark s bolsjer. I: *Nydanske studier* 26–27. København: Dansklærerforeningen, 35–42.

Brink, Lars & Jørgen Schack (2014): Tale er sølv. Om begrebet syntaktisk ellipse. I: *Danske Studier* 2014, 45–75.

Galberg Jacobsen, Henrik & Peter Stray Jørgensen (2013): *Håndbog i Nudansk*. 6. udg. København: Politikens Forlag.

Schack, Jørgen (2018): Perfektum participium og ordklassen adjektiver. I: *Nordiske studier i Leksikografi* 14 (denne udgiving). Reykjavik: NFL & Stofnun Árna Magnússonar í íslenskum fræðum, 204–211.

Lars Brink  
professor i dansk sprog & litteratur  
Elmevænget 30  
2880 Bagsværd  
labri@webspeed.dk

# LEXIA – en islandsk-fransk online ordbog: Udfordringer og løsninger

*Rósa Elín Davíðsdóttir*

This article describes a new Icelandic-French online dictionary in the making, LEXIA. The dictionary is based on an already established dictionary database for the Icelandic language that was compiled parallel to the making of the ISLEX dictionary with Icelandic as a source language and Danish, Norwegian, Swedish, Faroese and Finnish as target languages. LEXIA contains 50.000 headwords and numerous examples, collocations and idioms. The dictionary's target group consists of both Icelanders and French speaking users. The paper discusses some of the challenges that the elaboration of the dictionary represents. Furthermore, the paper explores the influence of a new target language, French, on the database and changes that the target language entails on both the lemma list and on elements in the microstructure of the dictionary such as examples.

## 1. Indledning

LEXIA er en ny islandsk-fransk online ordbog under udarbejdelse. Ordbogen er baseret på en den islandske ordbogsbase som blev udarbejdet af Árni Magnússon-instituttet for islandske studier i forbindelse med redigeringen af ISLEX-ordbogen. ISLEX er en nordisk netordbog med islandsk som kildesprog og dansk, svensk, nynorsk og bokmål som målsprog. Nu sidst er færøsk tilføjet, og arbejdet med finsk som målsprog er godt i gang. Brugere har taget godt imod ISLEX-ordbogen, hvilket blandt andet fremkommer af en brugerundersøgelse der blev forelagt brugere (Jónsdóttir & Úlfarsdóttir 2016:101).

LEXIA er et samarbejde mellem Árni Magnússon-instituttet hvor det redaktionelle arbejde for kildesprogets database finder sted (foruden al programmering og teknisk vedligeholdelse) og Vigdís Finnbogadóttir-instituttet for fremmedsprog som har ansvaret for det redaktionelle arbejde med målsproget, fransk<sup>1</sup>.

Arbejdet med fransk som målsprog begyndte i slutningen af 2015, og ifølge

---

<sup>1</sup> LEXIA er også et europæisk samarbejdsprojekt mellem Islands Universitet, Sorbonne Universitet (Université Paris-Sorbonne) hvor Karl Gadellii, professor i skandinavisk lingvistik, er den ansvarlige person og Göteborgs Universitet hvor Anna H. Hannesdóttir, professor i nordiske sprog, er ansvarlig som rådgivende specialist i leksikografi. Foruden arbejdet med ordbogen, drejer samarbejdsprojektet sig om kurser i leksikografi og lingvistik ved Islands Universitet og ved Sorbonne Universitet, arrangeret og udført af Karl Gadellii sammen med undertegnede.

planen skal arbejdet være færdigt i begyndelsen af 2019. Ligesom ISLEX-ordbogen, skal LEXIA være gratis tilgængelig online, men LEXIA får sin egen selvstændige hjemmeside.

Ordbogens database vil her kort blive beskrevet (se nærmere om ISLEX-ordbogen: Sigurðardóttir et al. (2008), Jónsdóttir & Úlfarsdóttir (2012) og Úlfarsdóttir (2014)).

Databasen er en beskrivelse af moderne islandsk og indeholder ca. 50.000 opslagsord samt et stort antal sprogbrugseksempler (ca. 30.000) og ca. 10.000 ordforbindelser, både kollokationer og faste udtryk.

Desuden indeholder databasen 3.000 billeder, samtlige islandske opslagsords udtale som lydfiler, og der er links til en hjemmeside med de islandske ords bøjningsparadigmer (BIN). Ordbogsbasen er under stadig udvikling og revision på Árni Magnússon-instituttet. Den eksisterer desuden nu som en etsprogsordbog med definitioner og en foreløbig version af denne ordbog er blevet offentliggjort på *málið.is*.

Heri artiklen vil der blive gjort rede for nogle af de udfordringer som den franske målsprogsredaktion har mødt i arbejdet med denne nye islandsk-franske ordbog, og hvilke ændringer der må laves når dette sprogpar, islandsk og fransk, stilles over for hinanden i en elektronisk ordbog.

## 2. LEXIA's brugergrupper

Behovet for en ny islandsk-fransk ordbog er stort eftersom den eneste ordbog på markedet med islandsk som kildesprog og fransk som målsprog er fra året 1950 (Boots 1950). En fransk-islandsk ordbog udkom derimod for "kun" 22 år siden (*Frönsk-íslensk orðabók* 1995).

Der undervises i fransk som tredje fremmedsprog i islandske gymnasieskoler (efter engelsk og dansk) og i nogle folkeskoler som valgfag. Desuden undervises der i fransk på Islands Universitet – fra begynderniveau til ph.d.-niveau. Målgruppen består derfor hovedsagelig af islændinge som studerer fransk (helt fra begyndere til viderekomne), og islændinge som bruger fransk i fritiden eller professionelt. Ordbogen er derfor en traditionel produktionsordbog. Det vil sige at ordbogen skal hjælpe den islandske bruger med at løse problemer i forbindelse med tekstproduktion på fransk.

Ordbogen kan endvidere bruges af fransktalende brugere til at forstå islandsk som en traditionel receptionsordbog. Her kan nævnes at der undervises i islandsk ved Sorbonne Universitet og ved universitetet i Caen.

Det er endvidere vores håb at LEXIA vil blive brugt af oversættere der oversætter fra islandsk til fransk. I dag udkommer i stigende grad islandsk litteratur oversat til fransk i Frankrig.

LEXIA sigter således mod to brugergrupper på samme måde som ISLEX (Sigurðardóttir et al. 2008:780).

### 3. Udfordringer og løsninger

Arbejdet med målsproget fransk stiller store krav til ordbogsredaktørerne der på den ene side skal finde de rigtige ækvivalenter og på den anden side tænke på de forskellige brugergrupperes behov (se nærmere Jónsdóttir & Úlfarsdóttir 2012, Rauset et al. 2012, Hannesdóttir 2012, Hilmisdóttir & Martola 2015).


I denne artikel vil udfordringerne med at finde passende ækvivalenter ikke blive beskrevet nærmere, men problemerne skyldes især forskelle mellem islandsk og fransk kultur og følgelig forskellig brug af begreber i de to sprog. Det kan dog nævnes her at en af de største udfordringer har været at finde et tilsvarende udtryk på fransk til de islandske ordforbindelser. Det gælder især faste udtryk med overført betydning men også kollokationer. Der er tit en større forskel mellem islandsk og fransk end mellem islandsk og de øvrige nordiske målsprog hvis vi bortser fra finsk. Det er for eksempel umuligt at oversætte kollokationen *takk fyrir síðast* til fransk, men den er almindelig både i islandsk og i svensk, dansk og norsk. Som figur 1 viser, blev vi nødt til at give en meget lang forklaring på denne islandske kollokation.

takk interj.


 **prononciation**

 **merci**


*má bjóða þér köku? - já takk*

 **tu veux du gâteau ? - oui, volontiers**


*viltu kaffi? - nei takk*

 **tu veux du café ? - non merci**

**takk fyrir síðast**

 **formule de politesse, utilisée pour remercier son interlocuteur du moment passé ensemble la dernière fois qu'on l'a vu**

**takk fyrir <gjöfina>**

 **merci pour <le cadeau>**

Figur 1: Opslagsordet *takk* 'tak' i LEXIA med ækvivalenter og oversættelser til fransk. Her er den franske brugergrænseflade blevet valgt med fransk som metasprog.

#### 3.1. Nogle eksempler på ændringer af lemmalisten

Anna Hannesdóttir (2014:196–197) har påpeget at skellet mellem kildesprog og målsprog har flyttet sig med de elektroniske ordbøger og med de nye søgemuligheder. Det vil sige at nu kan man slå ord og ordforbindelser op i målsproget og således få adgang til kildesprogets ordforråd. Dette er uden tvivl en fordel og et supplement til tosprogsordbøgernes rolle og nytte, det vil sige i forhold til papirudgaverne. På den måde kan fransktalende brugere anvende LEXIA til at finde frem til hvordan man siger noget på islandsk, og dermed kan LEXIA fungere ikke blot som receptionsordbog for fransktalende brugere, men også som produktionsord-

bog. Ikke desto mindre går ordbogsbeskrivelsen ud fra kildesproget, islandsk, og ord og ordforbindelser i målsproget, fransk, afhænger af det som enten forekommer som lemma i ordbogen eller som sprogbrugseksempel eller ordforbindelse.

Den franske redaktion er sig bevidst om de udvidede brugsmuligheder der findes i elektroniske tosprogsordbøger, og vi ønskede for eksempel at det islandske sammensatte ord *smjördeigshorn* 'croissant' blev tilføjet lemmalisten således at det franske ord *croissant* kunne komme med i ordbogen. Nogle gange drejer det sig om ord som vi tilfældigt opdager at de mangler i lemmalisten. For eksempel manglede *sléttujárn* 'glattejern' i lemmalisten, men derimod var *krullujárn* 'krøllejern' med. Der er også ord som målsproget kalder på, for eksempel *kampavínglas* 'champagneglas'. Her drejer det sig om et enkelt sammensat ord på islandsk, men den franske ækvivalent er ikke så transparent for den islandske bruger eftersom den ikke indeholder ækvivalenten for ordet *glas* 'glas', *verre*, men derimod *flûte* eller *coupe à champagne*. *Flûte* bruges om et højt og slankt glas, men *coupe* derimod når glasset er bredt.

Indtil nu har vores forslag til nye lemmaer været tilfældige og sker kun hvis vi falder over noget, men i slutfasen af ordbogsarbejdet ville det være interessant og nødvendigt at undersøge dette systematisk, for eksempel ved at sammenligne LEXIA med lemmalisten i en mellemstor fransk etsprogsordbog. Dermed kan vi undersøge hvilke almindelige franske ord der helt mangler i LEXIA og tage stilling til om det er passende eller muligt at råde bod på dette, for eksempel ved at tilføje sprogbrugseksempler (se Hannesdóttir 2014:207). Der vil dog altid være lakuner i målsproget og nemmere at give en præcis beskrivelse af kildesproget i en tosprogsordbog.

### 3.2. Nogle problemer knyttet til mikrostrukturen

Sprogbrugseksemplerne udvælges fra kildesproget med det formål at vise en typisk brug af de islandske opslagsord (se Jónsdóttir & Úlfarsdóttir 2012:359). Dette er dog ikke altid tilfældet i en tosprogsordbog. Ifølge Atkins og Rundell (2008:506–507) udvælges en stor del af sprogbrugseksemplerne i en tosprogsordbog der skal benyttes til produktion på et fremmedsprog, med målsproget i tankerne.

Men i tilfældet med ISLEX var det nødvendigt at udvikle ordbogsbasen alene ud fra kildesproget, islandsk, eftersom ordbogsbasen var tiltænkt mange målsprog fra starten. Her kan nævnes at de nordiske målsprogsredaktioner kunne foreslå ændringer til sprogbrugseksemplerne (se Hannesdóttir 2014:200, Rauset et al. 2012:515–516).

Projektlederen på den finske ISLEX, Helga Hilmisdóttir, har sammen med Nina Martola (2016) påpeget vanskelighederne ved at komme ind i projektet senere end de øvrige målsprog eftersom det redaktionelle arbejde på de andre målsprog allerede var overstået. Af den grund var det ikke muligt at foretage ændringer i den islandske base.

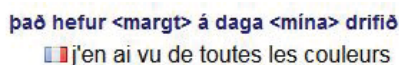
Vi i den franske redaktion har ikke det problem fordi LEXIA vil få sin egen hjemmeside, og vi nyder godt af at kunne foreslå ændringer i samarbejde med ordbogsbasens hovedredaktør, Þórdís Úlfarsdóttir.

Som Rauset et al. (2012:515–516) påpeger, kan der være andre ord end opslagsordet i et sprogbrugseksempel som giver problemer når det drejer sig om at oversætte eksemplet til målsproget. Denne problematik reflekteres fx i det islandske substantiv *systkini* ‘søskende’ der ikke har nogen ækvivalent på fransk. En måde at løse det på er at undgå ordet hvis det forekommer i sprogbrugseksempler under andre opslagsord end *systkini*.

### 3.3. Ordforbindelser og kollokationer i LEXIA

Ordforbindelser og kollokationer præsenteres på den måde at sætningsdele som kan skiftes ud sættes i skarp parentes (dvs. parenteser i formen < >). Subjekt og objekt i ordforbindelser præsenteres med pronomen i tredje person (det vil sige *hann* eller *hún* i singularis og *þeir*, *þær* og *þau* i pluralis). Denne opsætning er hentet fra Jón Hilmar Jónssons (2005) etsprogsordbog, *Stóra orðabókin um íslenska málnotkun*, og metoden stammer fra ham.

Der kan være vanskeligheder forbundet med at bruge denne opsætning når det drejer sig om at oversætte til et andet sprog. Figur 2 viser et eksempel hvor det ikke er muligt at have en skarp parentes på fransk. Her drejer det sig om den islandske ordforbindelse *það hefur < margt > á daga < mína > drifið* (oversættelsen til dansk i ISLEX lyder ‘<jeg> har oplevet <en hel del> i <mit> liv’). På fransk findes der et tilsvarende fast ordforbindelse, *en voir de toutes les couleurs*, men der er ikke noget som svarer til *margt* ‘en hel del’ og som kan skiftes ud. Desuden er apostrof i fransk heller ikke velegnet til parenteserne: <j> ‘*en ai vu de toutes les couleurs*’.



það hefur < margt > á daga < mína > drifið  
 j'en ai vu de toutes les couleurs

Figur 2: en del af artiklen *dagur* ‘dag’ i LEXIA som viser ordforbindelsen *það hefur < margt > á daga < mína > drifið*.

Der er dog ofte sammenfald mellem de to sprog, og figur 3 viser nogle eksempler på faste ordforbindelser med ord eller sætningsled inden for skarpe parenteser som kan skiftes ud på samme måde i islandsk og fransk (her er der tale subjekter eller objekter inden for skarpe parenteser).

<b>&lt;fréttin&gt; er vatn á myllu &lt;andstæðinganna&gt;</b>
▣▣ <la nouvelle> apporte de l'eau au moulin <des opposants>
<b>&lt;henni&gt; rennur kalt vatn milli skinns og hörunds</b>
▣▣ <elle> a des sueurs froides
<b>setja &lt;fangann&gt; á vatn og brauð</b>
▣▣ mettre <le prisonnier> au pain sec et à l'eau
<b>&lt;andi breytinga&gt; svifur yfir vötnunum</b>
▣▣ il y a <du changement> dans l'air

Figur 3: En del av artiklen *vatn* 'vand' i LEXIA som viser fire faste ordforbindelser på islandsk med franske oversættelser. Disse ordforbindelser er blevet oversat til dansk i ISLEX som: 1) <nyhederne> er vand på <modstandernes> mølle, 2) det løber <hende> koldt ned ad ryggen, 3) dømme <fangen> til vand og brød (eller sætte <fangen> på vand og brød), og 4) der er <forandring> i luften.

Vi på den franske redaktion har drøftet om vi skulle udelade brugen af skarpe parenteser i fransk. Vi har valgt at beholde disse parenteser (for ord eller sætningsled som kan skiftes ud), men undlade dem i de tilfælde hvor de ikke passer i forhold til fransk. Vi har ønsket at beholde disse oplysninger om kildesproget da især den fransktalende bruger hermed får oplysninger om hvordan den islandske kollokation eller ordforbindelse er opbygget.

Et andet problem i forbindelse med præsentationen af ordforbindelserne og oversættelsen af dem til fransk, opstår for eksempel når et dativobjekt eller indirekte objekt<sup>2</sup> skiftes ud med et personligt pronomen. I disse tilfælde flyttes pronomenet op foran verbet, for eksempel <lui> écrire 'skrive til <hende>' i stedet for écrire à <sa femme> 'skrive til <sin kone >'. Vi spekulerer på om denne præsentation fremstår klart nok for brugeren eftersom han ikke kan se at man skal anvende præpositionen à med verbet écrire 'skrive'. Hvordan vi har løst det, kan man se i figur 4 nedenfor.

<b>skrifa &lt;henni&gt;</b>
FALLSTJÓRN: ÞÁGUFALL
▣▣ <lui> écrire
▣▣ écrire à <quelqu'un>
<b>skrifa &lt;henni&gt; bréf</b>
FALLSTJÓRN: ÞÁGUFALL + ÞOLFALL
▣▣ <lui> écrire une lettre
▣▣ écrire une lettre à <quelqu'un>

Figur 4: En del af artiklen *skrifa* 'skrive' i LEXIA.

<sup>2</sup> I fransk grammatik bruges termen 'indirekte objekt' (fr. *complément d'objet indirect*) om en sætningsled der betegner den eller det som verbalhandlingen angår. Objektet er forbundet med verbet ved hjælp af en præposition som udvælges af verbet.



Vi prøver os frem med at have begge dele (se figur 4), det vil sige en oversættelse på fransk med pronomen i tredje person og så verbet i infinitiv med det ubestemte pronomen *quelqu'un* 'nogen' som objekt (i overensstemmelse med fransk ordbogstradition) for at det tydeligere skal fremtræde at præpositionen *à* skal anvendes med verbet.

#### 4. Afsluttende bemærkninger

Det er en stor fordel at kunne starte arbejdet med en færdigredigeret islandsk ordbogsbase som alligevel er fleksibel. Vi kan i nogle tilfælde søge i ISLEX for at se redaktørernes løsninger (undtagen de finske oversættelser) selv om problemerne ikke altid er de samme.

Den franske redaktion nyder godt af at LEXIA får sin egen selvstændige hjemmeside, hvor der endnu ikke er noget andet målsprog. Som nævnt her i artiklen er det til en vis grad muligt at tilpasse opslagsordene til målsproget i samarbejde med ordbogsbasens hovedredaktør, og foreslå at sprogbrugseksemplerne ændres eller fremstille ønsker om at tilføje sprogbrugseksempler med det formål at bedre at kunne vise forskellen på islandsk og fransk. Med ordbogens produktionsfunktion som formål så ville det have været en fordel hvis databasen kunne have endnu flere sprogbrugseksempler valgt ud fra en kontrastiv synsvinkel. Men det er spørgsmål om tid og finansiering og til hvilken grad man kan tilføje oplysninger i ordbogen.

I fremtiden er det så muligt at LEXIA får flere målsprog, som for eksempel engelsk, tysk eller spansk, og så bliver redaktørerne for disse sprog mere bundet end den franske redaktion i forhold til sprogbrugseksemplerne og de faste ordforbindelsers og kollokationers opsætning i ordbogsbasen.

Redaktionens erfaring er, at det er en ideel metode at arbejde på målsproget i samarbejde med kildesprogredaktionen og at arbejde ud fra en kontrastiv synsvinkel.

#### Litteratur

##### Ordbøger

BÍN = *Beygingarlýsing íslensks nútímamáls*. Reykjavík: Árni Magnússon-instituttet for islandske studier. <<http://bin.arnastofnun.is>> (august 2017).

Boots, Gerard (1950): *Íslensk-frönsk orðabók*. Reykjavík: Ísafoldarprentsmiðja. *Frönsk-íslensk orðabók* (1995). Þór Stefánsson (red.). Reykjavík: Örn og Örlygur; Paris: Dictionnaires Le Robert.

ISLEX = *ISLEX-ordbogen*. Þórdís Úlfarsdóttir (hovedredaktør). Reykjavík: Árni Magnússon-instituttet for islandske studier. <[www.islex.is](http://www.islex.is)> (august 2017).

Jónsson, Jón Hilmar (2005). *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV.  
*málið.is*. Reykjavík: Árni Magnússon-instituttet for islandske studier. <<http://malid.is>> (august 2017).

## Anden litteratur

- Atkins, B.T. Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Hannesdóttir, Anna Helga (2012): Ekvivalensrelationer i tvåspråkig lexicografi. I: *LexicoNordica* 19, 39–58.
- Hannesdóttir, Anna Helga (2014): Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi. I: *Nordiske studier i lexicografi* 12. Oslo, 193–211.
- Hannesdóttir, Anna Helga (2015): What is a Target Language in an Electronic Dictionary? I: *Proceedings of the eLex 2015 conference, 11–13 August 2015, Herstmonceux Castle, United Kingdom*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., 236–249.
- Hilmisdóttir, Helga & Nina Martola (2015): Flerspråkig lexicografi och obeslaktade språk erfarenheter från finska Islex. I: *Nordiske studier i lexicografi* 13. København, 143–153.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2012): ISLEX – en flersproget nordisk ordbog. I: *Nordiske studier i lexicografi* 11. Lund, 353–366.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2016): Hvem bruger ISLEX og hvordan? I: *LexicoNordica* 23, 89–104.
- Rauset, Margunn, Anna Helga Hannesdóttir & Aldís Sigurðardóttir (2012): Ein-, to- eller fleirspråkleg ordbok? I: *Nordiske studier i lexicografi* 11. Lund, 512–523.
- Sigurðardóttir, Aldís, Anna Helga Hannesdóttir, Halldóra Jónsdóttir, Håkan Jansson, Lars Trap-Jensen & Þórdís Úlfarsdóttir. (2008): ISLEX – An Icelandic-Scandinavian Multilingual Online Dictionary. I: Elisenda Bernal & Janet DeCesaris (red.): *Proceedings of the XIII Euralex International Congress. Barcelona 2007*, 779–789.
- Úlfarsdóttir, Þórdís (2014): ISLEX – A Multilingual Web Dictionary. I: *LREC 2014 Conference Proceedings*, 2820–2825.

Rósa Elin Davíðsdóttir  
 redaktør og projektleder, ph.d.  
 Vigdís Finnbogadóttir-instituttet for fremmedsprog  
 Brynjólfsögðu 1  
 107 Reykjavík  
 red1@hi.is

# Syntaktiske opplysninger i ettspråklige ordbøker med en gruppe intransitive verb som eksempler

*Kristine Gunn Eide*

In Norwegian, more and more traditionally transitive verbs are being used in intransitive constructions. This development has been going on for the past 100 years, and one natural consequence of this syntactic change is that sentences that seem perfectly correct to younger speakers, are regarded by older speakers as ungrammatical. In this presentation, I argue that some grammatical phenomena that depend on the transitive/intransitive distinction, such as presentational sentences, cannot be captured by grammatical rules alone, and that as such, they belong in a dictionary. I present possible ways to deal with different levels of grammaticality in the Dictionary of the Norwegian Academy (NAOB) and argue that newer grammatical developments represented by younger speakers belong in this dictionary, even though they are considered ungrammatical by older speakers of Norwegian.

## 1. Innledning

På norsk er det en dreining mot at stadig flere verb som tidligere var transitive kan brukes intransitivt og i presenteringssetninger. Dette er en pågående endring, slik at setninger som kan være helt normale for en yngre person, kan virke ugrammatikalske for en som er eldre. Et illustrerende eksempel på denne endringen, og hvilken irritasjon den kan avstedkomme, er et leserbrev til Aftenpostens språkspalte hvor en irritert leser hadde kommet over setninger som «Tross advarsler satset Gardermoen på glass som er farlig når det knuser.» Og «Dette er glass som fortsatt sitter på plass etter at det har knust.»

Til dette svarte avisens faste språkspaltist Per Egil Hegge følgende:

Brevskriveren forlanger bruk av rødblyanten her, og de fleste i min aldersgruppe (65 og over) ville nok imøtekomme kravet. For oss er verbet «knuse» transitivt, det krever et objekt, som i «han knuste bilruten». Men for dem som er yngre, og som tidligere nevnt blir det stadig flere av dem, lyder det ikke galt å si at «bilruten knuste» når vi eldre ville lure på hva den da knuste, og vi ville «oversette» det med at bilruten ble knust. Forfatteren Lars Saabye Christensen, som har passert 50, skriver i romanen «Modellen» at «det følte som øyet knuste». Det er litt problematisk å påstå at Saabye Christensen ikke behersker norsk. Så her har vi nok et knusende eksempel på både at og hvordan språket forandrer seg - ved at noe som først blir oppfattet som feil, blir det etablerte og dermed det riktige. (Hegge 2008)

Nå er det ikke slik at denne intransitive bruken av *knuse* uten videre kan overføres til andre, liknende verb. For eksempel vil ikke en person som godtar at «vinduet knuste» automatisk godta at «vinduet smadret», selv om man vanskelig kan finne noen objektiv grunn til at semantisk og ellers syntaktisk like verb som *smadre* og *knuse* skulle ha forskjellig syntaks akkurat på dette punktet<sup>1</sup>. I de følgende avsnittene viser jeg eksempler på noen verb som i varierende grad brukes intransitivt og i presenteringssetninger, og diskuterer hvordan de kan presenteres i en dels normgivende, dels dokumenterende ordbok som *Det Norske Akademis ordbok* (NAOB).

## 2. Intransitivitet i endring

Jeg laget et lite spørreskjema som jeg delte ut til mine 20 leksikografkollegaer ved NAOB. Det er ingen vitenskapelig undersøkelse, men resultatene gir en pekepinn på hva som kan godtas som kurante konstruksjoner, hva som definitivt ikke kan godtas, og i hvilke tilfeller det er snakk om en generasjonskløft. Tabell 1 viser resultatene fra NAOBs samlede redaksjon, og viser stor uenighet rundt hva som godtas eller ikke på norsk. I tabell 2 og 3 har jeg delt kollegiet i to: den yngste halvparten (de under 50) og den eldste (de over 50). Det kommer tydelig frem at det bare er personer i den eldre generasjonen som ikke umiddelbart godtar *knuse* brukt intransitivt, og som svarer at setningen er litt rar. Likeledes ser vi at den intransitive setningen med verbet *smadre*, ikke godtas av noen, men at yngre mennesker ikke er totalt negative, siden mange har svart «kanskje» på akkurat denne setningen. *Spjære* er stort sett godtatt av alle, mens *avslutte* er mest akseptert blandt de under 50.

	Ja	Kanskje	Nei
Vinduet knuste under flyttingen av drivhuset.	X	x	
Bilen smadret inn gjennom ruten.		x	X
Hun bøyde seg frem og buksen spjæret	X	x	
Filmen avslutter med et roterende kameraskudd.	x	X	x

Tabell 1: Intransitiv bruk av tradisjonelt transitive verb i moderne norsk. Informantene (redaksjonen til NAOB) ble bedt om å krysse av for «ja» hvis de syntes setningen var god, «kanskje» hvis de syntes den var litt rar og «nei» hvis de syntes den hørt helt feil ut. Stor X betyr at overvekten av svarene befinner seg i en kolonne, liten x betyr at flere enn en har satt kryss i kolonnen, og ingen x betyr null eller én person.

<sup>1</sup> I et såpass nær beslektet språk som dansk, er det (ifølge DDO) omvendt: *knuse* er bare transitivt, mens *smadre* også er intransitivt.

	Ja	Kanskje	Nei
Vinduet knuste under flyttingen av drivhuset.	X	X	
Bilen smadret inn gjennom ruten.			X
Hun bøyde seg frem og buksen spjæret <sup>2</sup>	X	x	
Filmen avslutter med et roterende kameraskudd.	x	X	x

Tabell 2: Svar fra de i redaksjonen som er over 50.

	Ja	Kanskje	Nei
Vinduet knuste under flyttingen av drivhuset.	X		
Bilen smadret inn gjennom ruten.		X	X
Hun bøyde seg frem og buksen spjæret <sup>2</sup>	X		
Filmen avslutter med et roterende kameraskudd.	X	X	

Tabell 3: Svar fra de i redaksjonen som er under 50.

Hvis vi sammenlikner verbene *knuse*, *spjære* og *smadre*, kan man følge Per Egil Heggens aldersanvisninger fra sitatet i innledningen og litt grovt si at hvis man er under 70 kan man bruke verbet *knuse* intransitivt, og fint si «vinduet knuste», mens hvis man er over 70 må man ty til en passivkonstruksjon for å uttrykke det samme, slik som i eksempel (1). Når det gjelder *spjære*, er det svært få som i dag ville insistere på passivkonstruksjon, og for mange vil det være direkte feil å si «buksen ble spjæret» med mindre en person faktisk rev i buksen slik at den revnet (2). Faktisk virker det som om man må være nesten 80 for å si at den intransitive bruken er direkte feil. Verbet *smadre* kan derimot ikke brukes intransitivt (3), selv om resultatet av Petters handlinger i (1) og (3) i begge tilfeller er et vindu som er slått i stykker. (I avsnitt 4 skal vi likevel se at det finnes litterære belegg for at dette verbet også brukes intransitivt)

- 1) Petter knuste vinduet
  1. Under 70 – intransitivt: Vinduet knuste
  2. Over 70 – passiv: Vinduet ble knust
- 2) Petter spjæret buksen
  1. Under 80? – intransitivt: Buksen spjæret
  2. Over 80? – passiv: Buksen ble spjæret
- 3) Petter smadret vinduet
  1. \*vinduet smadret

<sup>2</sup> Bare én person svarte at dette er en feil setning. Personen fyller 80 i disse dager. Etersom det er den eneste personen i NAOBs redaksjon i denne aldersgruppen, ble jeg nødt til å spørre andre 80-åringere «utenfor huset», og de kunne bekrefte at for dem var setningen hun bøyde seg fram og buksen spjæret ugrammatisk.

### 3. Intransitivitet og presenteringssetninger

Presenteringssetninger brukes for å introdusere nye referenter i en samtale eller tekst, og kan bare brukes hvis verbet er intransitivt. Likevel gjelder ikke dette alle intransitive verb. For mens (4), (5) og (6) er allment aksepterte presenteringssetninger, er noen verb avhengige av bundet adverbial når de brukes i slike konstruksjoner. Et verb som *arbeide* (7) fungerer ikke i en presenteringssetning uten at setningen inneholder informasjon om hvor arbeidet foregår, som i (8).

- 4) Det kom en mann i går
- 5) Det ringte en mann i går
- 6) Det knuste et vindu i går
- 7) \*Det arbeidet en mann i går
- 8) Det arbeidet en mann på åkeren

Går vi til grammatikkene, finner vi regler som f.eks. «For at eit verb skal kunne brukast i presenteringssetningar i norsk, må det kunne brukast utan objekt. Det vil seie at det anten må vere eit intransitivt verb, eller eit transitivt verb av den typen som også kan stå utan objekt» (Faarlund et al. 1997:830). Derfor er presenteringssetning i (9) mulig, mens i (10) forventer vi at det er feil.

- 9) Vinduet knuste → det knuste et vindu
- 10) \*vinduet smadret → \*det smadret et vindu

Det er likevel ikke slik at alle intransitive verb kan brukes i presenteringssetninger, og det er ingen automatikk i at en person som godtar *vinduet knuste* også godtar *det knuste et vindu*. Tabell 4, 5 og 6 viser i hvilken grad redaksjonsmedlemmene i NAOB godtar presenteringssetninger med verbene i tabell 1.

	Ja	Kanskje	Nei
Det knuste et vindu under flyttingen av drivhuset.	X	X	x
Det smadret en istapp rett foran meg.	x	x	X
Kan jeg ringe deg om fem minutter? Det avslutter en film akkurat nå som jeg gjerne vil se ferdig.			X

Tabell 4: Aksept for presenteringssetninger i redaksjonen totalt.

	Ja	Kanskje	Nei
Det knuste et vindu under flyttingen av drivhuset.		X	X
Det smadret en istapp rett foran meg.	x		X
Kan jeg ringe deg om fem minutter? Det avslutter en film akkurat nå som jeg gjerne vil se ferdig.			X

Tabell 5: Aksept for presenteringssetninger hos redaksjonsmedlemmer over 50.

	Ja	Kanskje	Nei
Det knuste et vindu under flyttingen av drivhuset.	X	x	
Det smadret en istapp rett foran meg.	x	x	X
Kan jeg ringe deg om fem minutter? Det avslutter en film akkurat nå som jeg gjerne vil se ferdig.			X

Tabell 6: Aksept for presenteringssetninger hos redaksjonsmedlemmer under 50.

Det samme mønsteret ser vi i presenteringssetningene med de samme verbene: setningen med *knuse* godtas ikke av de over 50 (selv om heller ikke de er totalt negative), mens den er akseptert blant de yngre. Setningen med *smadre* (som er hentet fra en litterær tekst) er i liten grad akseptert, selv om det også her er enkelte som synes den er ok. Presenteringssetningen med *avslutte* virker fullstendig uakseptabel, selv om den er godtatt brukt intransitivt, særlig blant de yngre.

#### 4. Syntaktiske opplysninger i ettspråklige ordbøker

Som vi ser av sammenhengen mellom intransitivitet og presenteringssetninger, er det slik at noen grammatikkregler på norsk forutsetter at man kan norsk. For å vite om man kan lege en presenteringssetning med et verb, må man vite om verbet kan brukes intransitivt. Men, som vi ser av f.eks. *avslutte*, er det ingen éntil-én-korrespondanse mellom intransitivitet og presenteringssetninger. En annen opplysning som er vanskelig å få ut av grammatikker, er hvilke verb som krever steds- og tidsadverb eller annen utfylling, og evt. hva slags utfylling, slik som for eksempel *arbeide*.

Informasjon om transitivitet, presenteringssetninger og evt. obligatoriske adverbialledd hører hjemme i en ordbok fordi ingen norske grammatikkregler fanger opp skillet mellom f.eks. *smadre* og *knuse*. Det er heller ingen semantiske regler som tilsier en forskjell mellom *smadre* og *knuse*.

Dersom det ikke finnes generelle grammatiske regler for hvilke syntaktiske konstruksjoner et verb kan opptre i, er det et godt argument for at slik informasjon hører hjemme på et leksikalsk nivå, på lik linje med transitivitet og intransitivitet som vel er den syntaktiske informasjonen som oftest oppgis.

I det følgende vil jeg presentere de fire måtene *Det Norske Akademis Ordbok* oppgir slike syntaktiske opplysninger på (definisjoner, formaliserte konstruksjoner, redaksjonelle eksempler og sitater), og diskutere hvilke måter som er mest hensiktsmessige.

##### 4.1. Definisjoner

Syntaktiske opplysninger i definisjonen oppgis ofte nesten ubevisst (fra leksikografens side) og implisitt for leseren ved at definisjonen er entydig transitiv, f.eks. *slå i stykker* eller entydig intransitiv *gå i stykker*, som i (11)

## 11) knuse

1. slå i stykker
2. gå i stykker

Den entydig intransitive definisjonen gir likevel ikke informasjon om verbet kan inngå i en presenteringskonstruksjon.

## 4.2. Formelle konstruksjonopplysninger og syntaktiske betegnelser

Dette er den tradisjonelle, eksplisitte måten å gi grammatikkopplysninger på. Den oppgis enten i form av syntaktiske betegnelser (transitiv, intransitiv, refleksiv osv) eller i form av formaliserte setninger som i (12) og (13). NAOB bruker regelmessig syntaktiske betegnelser, men formaliserte setninger som oppgir hva slags type konstruksjoner et ord kan opptrå i, vurderes brukt i en senere fase av prosjektet. Den er brukt gjennomført i *Den Danske Ordbog* (DDO).

## 12) knuse

1. slå i stykker  
**noen** knuser **noe**
2. gå i stykker  
**noe** knuser

## 13) smadre

1. ødelegge (voldsomt); knuse  
**noen** smadrer **noe**

Meg bekjent, finnes det ingen studier om ordboksbrukere oppfatter syntaktiske betegnelser forskjellig fra formaliserte setninger, men det er ikke usannsynlig at grammatikkopplysninger i form av formaliserte setninger oppfattes som mer normgivende enn både implisitt bruk av definisjoner og eksplisitte grammatiske termer, som i større grad kan oppfattes som beskrivende. En ordbok som NAOB, som skal fungere som dokumentasjonsordbok samtidig som den skal oppgi normal, korrekt språkbruk, trenger en måte å skille det rent beskrivende fra det normgivende. Verbet «å smadre» kan faktisk brukes intransitivt på norsk, som sitatet i (14) viser. Likevel er ikke dette nødvendigvis en allment godtatt bruk, slik at en grammatikkanvisning som blir oppfattet som normgivende (f.eks *\*noe smadrer*), ikke er hensiktsmessig under (14), betydning 2.

## 14) smadre

1. ødelegge (voldsomt); knuse  
**noen** smadrer **noe**
2. gå i knas  
*Jeg har sett istapper smadre rett foran meg* (Aasvik 2012)



### 4.3. Redaksjonelle eksempler

Redaksjonelle eksempler er enkle setninger eller ordkonstellasjoner som illustrerer allmenn bruk, og som kan gi samme type informasjon som formaliserte setninger, uten å oppfattes som grammatikkregler i samme grad. Ved å erstatte de formaliserte ubestemte pronomenene med vanlige ord, sier man at dette er en vanlig bruk av ordet, men ikke at det nødvendigvis *må* brukes på denne måten. I tilfeller hvor det for eksempel er uenighet om hvorvidt et intransitivt verb kan brukes i presenteringssetninger, kan et redaksjonelt eksempel virke mindre absolutt enn en formalisert konstruksjon. I (15) er det dessuten satt opp en presenteringssetning som redaksjonelt eksempel, for å vise at en slik konstruksjon er mulig. Alternativet ville vært en formalisert setning, f.eks. ‘**det** [foreløpig subjekt] knuser **noe** [ubestemt, ny]’, hvor *det* er definert som ‘foreløpig subjekt’ og subjektet må være ubestemt og inneholde ny informasjon.

#### 15) knuse

1. slå i stykker  
**noen** knuser **noe**  
*Petter knuste vinduet*
2. gå i stykker  
**noe** knuser  
*vinduet knuste*  
*det knuste et vindu da vi flyttet drivhuset*

#### 16) smadre

1. ødelegge (voldsomt); knuse  
**noen** smadrer **noe**  
*Carina smadret bilen*
2. gå i knas  
*Jeg har sett istapper smadre rett foran meg (Stig Aasvik (2012) Indre anliggender)*

I (17) har jeg brukt et redaksjonelt eksempel for å vise at presenterings-setningen tar et stedsadverbial (et redaksjonelt eksempel som for så vidt også kunne vært formalisert ‘**det** [foreløpig subjekt] arbeider **noen** [ubestemt, ny] + [sted]’

#### 17) arbeide

1. utføre arbeid, jobbe  
**noen** arbeider  
*det arbeider en mann på åkeren*

#### 4.4. Sitater

Litterære sitater kan absolutt brukes for å dokumentere en konstruksjon i bruk, men er ikke like tilgjengelige for en bruker. De forutsetter at brukeren kjenner forfatterens stil og alder, siden de illustrerer all mulig bruk, også foreldet bruk for noen (yngre) aldersgrupper, dersom man tar med eldre forfattere.

### 5. Når skal syntaktiske endringer opptas i en ordbok?

Typisk for språkendringer, og transitivitet er under endring i norsk, er at, mens de pågår, vil alltid begge varianter være i omløp. På hvilket tidspunkt tar vi inn endringen i ordboken? Nyord tas inn i den grad man forventer at de er mer enn døgnfluer, og blir en del av allmennspråket. Det samme gjelder nye betydninger av allerede eksisterende ord. Hvis man argumenterer slik jeg gjør her, for at noen syntaktiske konstruksjoner hører hjemme på lemmanivå, vil en naturlig konsekvens være å bruke de samme kriteriene for å ta inn ny syntaks. To hovedproblemer oppstår ved å bruke en slik tilnærming: For det første, er det langt mer komplisert å finne frekvens på syntaktiske konstruksjoner enn på enkeltord og kollokasjoner, som er lett søkbare i alle tekstdatabaser. For det andre, er syntaktiske endringer langt mer generasjonsbestemte enn nyord (dette gjelder ikke bare verbalsyntaks, det gjelder en lang rekke andre syntaktiske nyvinninger). Den syntaktiske endringen det er snakk om her, har pågått i over 100 år, slik f.eks. eksemplet med «buksen spjæret» illustrerte. Den pågår fremdeles. Spørsmålet om når en syntaktisk nyvinning skal tas inn i en ordbok, blir dermed mer komplisert enn for (i hvert fall en del) nyord. Den kompliseres ytterligere ved at det er folks språkfølelse som utfordres. Hvis noen oppfatter f.eks. en setning som *vinduet knuste* som direkte feil, og at bruken av en slik setning egentlig skriker på rødblyanten, ligger det mye smerte i å ta den inn i en ordbok, selv når halvparten av språkbrukerne anser den som det eneste riktige. Samtidig, hvis vi ikke tar med en slik konstruksjon, gir vi ikke en korrekt beskrivelse av gjeldende språkbruk, og ordboken er utdatert allerede fra utgivelsesdatoen.

For å dokumentere syntaktiske konstruksjoner er det i mange tilfeller nødvendig med et syntaktisk tagget korpus (eller et program som analyserer det tekstkorpuset man allerede har, slik man finner hos f.eks. INESS). For å finne intransitive verb er det helt nødvendig i norsk. Presenteringssetninger er enklere å finne. Likevel inneholder søkeresultatene mye støy, slik at å regne ut frekvens av syntaktiske konstruksjoner er svært tidkrevende, hvis ikke umulig, og forutsetter bedre taggede korpus enn det vi har tilgjengelig i dag. Derfor står leksikografens språkfølelse sentralt når man skal bestemme hvilke nyvinninger som skal med.

## 6. Oppsummering

Syntaktiske konstruksjoner som ikke kan utledes fra grammatiske regler regnes for å høre hjemme på lemmanivå, og dermed i en ordbok.

Syntaktisk endring bør fanges opp i en ordbok så snart som råd er ved korpus-søk, på lik linje med nyord, hvis ikke er den rett og slett ikke oppdatert.

Syntaktisk endring som godtas av flertallet blant de under en viss alder, f.eks. de under 50, skal inn enten som redaksjonelle eksempler eller i en formalisert form.

## Kilder

### Ordbøker

DDO = *Den Danske Ordbog*. <ordnet.dk/ddo> (30. mai 2017).

NAOB = *Det Norske Akademis ordbok*. <naob.no> (30. mai 2017).

### Annen litteratur

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

Hegge, Per Egil (2008): "Språket Vårt". I: *Aftenposten* 17.01.2008.

INESS = *Infrastructure for the Exploration of Syntax and Semantics*. <clarino.uib.no/iness> (30.mai 2017).

Aasvik, Stig (2012): *Indre Anliggender*. Oslo: Cappelen Damm.

Kristine Gunn Eide, ph.d.  
Kunnskapsforlaget, NAOB  
Gullhaug Torg 1, Nydalen, 0484 Oslo  
kristine.gunn.eide@gmail.com

# Språksamlingane og leksikografisk arbeid ved Universitetet i Bergen

*Oddrun Grønvik & Rune Kyrkjebø*

The Norwegian Language Collections comprise 130 years of scholarly documentation of the Norwegian language, speech and literature, names and general language, from Old Norse to today's dialects and standard languages. In November 2015 it was decided to move the Language Collections from the University of Oslo to the University of Bergen. The two year period 2016–2017 was set as the transit project period. This article discusses the transfer process, including applications for funding revision projects for the lexicographical databases *Bokmålsordboka* and *Nynorskordboka* (<ordbok.uib.no>).

## 1. Innleiing

Hausten 2015 vart det avgjort at Universitetet i Bergen (UiB) skulle ta over ansvaret for dei norske språksamlingane som var bygd opp ved Universitetet i Oslo (UiO) gjennom om lag 130 år, frå professor Johan Storm gav ut *Kortere Ordliste* til fullføringa av *Norsk Ordbok*. Denne materialmengda blir heretter kalla Språksamlingane. Overføringa av ansvaret vart markert ved at Kunnskapsdepartementet flytta 6 millionar av grunnløyvinga frå UiO til UiB frå og med budsjettåret 2016.

Grunngjevinga er politisk like mykje som fagleg: Språksamlingane utgjer så store verdiar for Noreg og for norsk språk og kultur at heilskapen må haldast ved like.

Språksamlingane ved UiO dekkjer norsk talemål, norrøn skrift, skriftmåla bokmål og nynorsk, og norske namnesamlingar, særleg stadnamn. Språksamlingane femner om fleire referansebibliotek, delsamlingar og samlingsprodukt på papir og i digital form. Det arkivtekniske vedlikehaldet av Språksamlingane ved UiO har vore mangelfullt i lang tid, og fullstendig oversyn over papirsamlingane mangla. Dei elektroniske samlingane var i jamføring i god stand. Overføringa til elektronisk format har vara sidan tidleg 90-tal, etter prioritering frå forskingsmiljøa og i samarbeid med informatikarar med vitenskapleg kompetanse (Aukrust & Hodne 1998). Ved flyttinga frå UiO til UiB var det mest elektronisk tilfang for nynorsk og målføre. Eit oversyn over dei viktigaste komponentane finst i Ebeling et al. (2014).

I denne artikkelen står leksikografien i fokus, sidan reetablering av ordbøkene har vore ei hovudsak ved UiB i åra 2016–2017. Det tyder ikkje at samlingane for norrønt, målføre eller namn er mindre viktige, og i framtida vil dei få den plassen dei fortener i publikasjonar frå Språksamlingane ved UiB.

## 2. Flyttinga

### 2.1. Dei fysiske arkiva

På flyttetidspunktet var dei fysiske samlingane spreidde på fleire lokale ved UiO, ingen av dei innreidde som forskingsarkiv. Plassbehovet var 700 kvadratmeter, vekt viste seg å ligga mellom 70 og 80 tonn.

UiB gjennomførte flyttinga mai–august 2016. Arkivmaterialet fekk mellom-bels lagring i påvente av ny flytting våren 2017 (Myking 2017:42). Tida vart brukt til å skaffa oversyn og vurdere kva som kunne gå til fjernlager. Det vart lagt sær-leg vekt på å identifisera delsamlingar og sørgja for at dei vart slik registrerte og haldne saman. Såleis vart forskararkiva etter einskildforskarar som t.d. Amund B. Larsen og Hallfrid Christiansen haldne samla, likeins dei ulike referansebiblio-teka. Det synte seg elles at ompakking og ny oppstilling av Språksamlingane er eit kostbart arbeid.

Ordninga av dei fysiske samlingane er ikkje ferdig, men vil i framtida bli fag-messig utført under leiing av handskriftavdelinga ved Universitetsbiblioteket i Bergen. Per i dag er der meir enn to arkivstillingar knytte til arbeidet.

### 2.2. Elektroniske arkiv og verktøy

Dei elektroniske komponentane i Språksamlingane er store, komplekse, og er bygde opp bit for bit sidan Dokumentasjonsprosjektet (1991–1997). Premissen for overføringa av vitskaplege samlingar til elektronisk form er at den elektro-niske versjonen av samlingane skal ta vare på og vidareutvikla den vitskaplege analysen som ligg til grunn for ordninga av de ulike samlingane. Elektronisk til-rettelegging skal gjera samlingane meir tilgjengelege, både til forskning og andre føremål, ved å synleggjera dei faglege prinsippa attom oppbygging og endring. Ore (1998) gjer greie for korleis dette er gjort.

Eit anna prinsipp er samarbeidsmodellen, med vitskapleg kompetanse både hos filologar og IKT-utviklarar, og eit likeverdig ansvar for drift og ny utvikling. Frå oktober 2016 har IKT-avdelinga ved UiB drive nettportalen med NOB og BOB utan nemneverdig nedetid. Det har vore tilgang for redigeringsarbeidet, og tett samarbeid mellom redaktørar og programmerarar. Hausten 2017 kom det ein mobilapp for ordbøkene, i samarbeid mellom Språkrådet, UiB og eit utviklarfirma.

UiB har fått ansvar for eit stort og komplekst system av resursar, dels som elek-troniske dokument og del organisert i databaseform. Generelt gjeld det at svært mange av delane heng saman, og motivasjonen bak oppsett og system er fagleg. For UiB er det derfor vesentleg å finna løysingar som tek vare på vitskaplegheit og heilskap, i parallel med nødvendige oppgraderingar og forbetra brukartilgjenge. Det gjeld ikkje minst namnesamlingane, der det står mykje att å gjera.

Ein viktig føresetnad for UiB sitt vidare arbeid er at Språksamlingane skal ha

fagleg forankring i Det humanistiske fakultetet ved UiB via ei fagleg referansegruppe med medlemmer frå fleire institusjonar. Med fagleg støtte vil Universitetsbiblioteket og IKT-avdelinga stå for den daglege drifta av samlingane.

### 2.3. Folk og rekruttering

UiB tok mot Språksamlingane føresett at innsatsen skulle vera finansiert. Ramma var i 2015 sett til 6 millionar. Det er alt for lite til å byggja opp ein stab tilsvarande den Språksamlingane hadde ved UiO. I 1990 var det knytt meir enn tjue forskarårsverk til Språksamlingane, pluss administrasjon og vitskapelege assistentar. I 2015 var det framleis meir enn seks stillingar på professornivå knytt til Språksamlingane. I tillegg kom to årsverk ved Eining for digital dokumentasjon.

UiB har lagt fram eit oversyn over ønskt minimumsbemanning (Myking 2016). Oversynet omfattar elleve stillingar, seks av desse forskarstillingar, pluss fagleg-administrativ leing og administrasjon. Dette gjev budsjettet ikkje rom for, men det gjev eit mål å arbeida fram mot.

Det er tilsett ein redaktør for ordboksportalen (*Bokmålsordboka* (BOB) og *Nynorskordboka* (NOB) på nett), ein fagleg leiar for Språksamlingane og to IKT-ingeniørar. Denne gruppa tilsette blir komplett seint i 2017, når Peder Gammeltoft tiltrer som fagleg leiar. Peder Gammeltoft er namnegranskar, og tilfører etterspurd fagkunnskap til Språksamlingane.

## 3. Prosjekt og søknader

I dag er det vanleg å skilja mellom vanleg drift og større utviklingsprosjekt, også i akademia. I samsvar med dette har UiB inngått ein samarbeidsavtale om drift med Språkrådet om dei samlingskomponentane der Språkrådet er medeigar: BOB, NOB og Ordbanken. Driftsansvaret for heilskapen ligg til UiBs eigen samlingsstab (sjå pkt 2.3).

Ønsket om å få i gang to definerte, større leksikografiske prosjekt følgde med på lasset då Språksamlingane kom til UiB: (1) revisjon av første del av *Norsk Ordbok* (NO) bokstavbolken a–h og (2) innhaldsrevisjon av BOB og NOB.

Både prosjekt inneber vidare integrering mellom ordboksdatasasane og dei elektroniske materialresursane i Språksamlingane, og både prosjekt understrekar behovet for å sjå dei elektroniske løysingane for Språksamlingane som ein heilskap.

### 3.1. Norsk Ordbok a–h

UiB og Høgskulen i Volda søkte saman Kulturdepartementet om midlar til revisjon av NO a–h tidlegare i 2017. Søknadssummen var 160 millionar kroner over 10 år, 2018–2027. Denne søknaden har førebels ikkje ført fram.

Føremålet er å fullføre NO på nett ved å revidere innhald og form på alfabetstrekket a–h, og slik skapa ei komplett elektronisk NO. Delmål er å (1) skapa ein fritt tilgjengeleg, dynamisk ordboksdatabase a–å over alt norsk talemål og nynorsk skriftmål, med laupande oppdatering, (2) leggja til rette samverke mellom NO og dei andre ordboksbasane, (3) gjere NO søkbar med moderne rettskriving, (4) lenkje NO til Ordbanken, så kvart oppslagsord får bøyingsinformasjon.

Revisjonen av NO a–h er rekna som ei så tung oppgåve, endå verket ligg føre i førstegongsrevidert form. Revisjonen dekkjer omlag førti prosent av den totale tekstmengda, ei svær mengd i seg sjølv. Dei første banda av NO vart redigerte ut frå eit mykje mindre tilfang, og føresetnadene skifta under marsjen: Innleiingsvis rekna ein med at ordboka skulle bli på 4 til 6 band. Etter kvart skifte fokus til full utnytting av det tilfanget ein hadde, utan sterkt omsyn til plasskrav. Artikkelformatet vart tilpassa under marsjen, og vart ulikt brukt av ulike redaktørar. Det er sterkt påkravd å gje innhaldet i a–h-bolken formell og innhaldsmessig konsistens. Både lingvistikken og leksikografien har utvikla seg mykje, i stor mon av di data-teknologien har omforma faget. Dette skal ein ta omsyn til under revisjonen, slik at NO held på statusen sin som ei vitskapleg ordbok i tungvektsklassen. Ei meir fylllestgjerrande framstilling av saga om NO er å finna i Vikør & Wetås (2016) og Worren (2017).

Eit anna omsyn er at redaksjonen bør byggjast opp att snarast mogleg. Den måten som arbeidet med NO vart avslutta på, førte til at ein høgt skulert stab vart spreidd for alle vindar. Ein må derfor ha von om at revisjonsprosjektet NO a–h kan gjennomførast i nær framtid, medan kunnskapen er tilgjengeleg.

### 3.2. Bokmålsordboka og Nynorskordboka

BOB og NOB kom første gong i 1986, etter ei redigeringstid på 12 år. Arbeidet vart utført ved Norsk Leksikografisk Institutt i samarbeid med Norsk språkråd, og kosta om lag førti redaktørårsverk – mot opphavleg planlagt seks. Redigeringshistoria og sambandet med NO er gjort greie for i brev frå UiO til Språkrådet 12.12.2005. Ordbøkene vart redigerte i parallell, og Språksamlingane kom bae til gode (Hovdenak 2014:232).

Etter 1986 er det kome fleire prenta utgåver med mindre oppdateringar av innhaldet. Ordbøkene vart søkbare på nett frå 1994. I 2009 vart dei to ordbøkene lagde i portal med sams søkjerute. Frå 2012–2013 har bae ordbøkene lege i det same databaseformatet som NO. Ordartiklane hentar ordklasse og bøyning frå Norsk Ordbank. Ordbøkene er populære, og den forbetra funksjonaliteten på nett har sendt bruken til himmels.

Bae ordbøkene treng no full innhaldsrevisjon, tretti år etter førstegongsredigeringa. Eit pilotprosjekt i 2013 synte omfanget av det naudsynte redigeringsarbeidet. UiB søkte derfor Kulturdepartementet 28.2.2017 om midlar til revisjon. Denne søknaden er stetta med 6 millionar i tilskot i 2018 (Stortingsmelding 1 kap. 326, post 75).

Revisjonssøknaden tek utgangspunkt i at dei to ordbøkene er den viktigaste formidlingskanalen for kunnskap om normert norsk, i og utanfor Noreg. Det heiter at

Fokuset må ligge på at både bokmål og nynorsk er slitesterke bruksspråk i samfunnet, og dette hensynet må veie tungt i revisjonsarbeidet.

Søknaden set opp desse hovudpunkta: (1) revisjon og ajourføring av innhald, (2) standardheving på ordbøkene som normeringsverktøy, ettersom BOB og NOB er dei einaste ordbøkene for norsk som gjer greie for heile rettskrivinga, (3) samkøyring av lemmautval og parallelt inntak av nye ordartiklar, (4) utjamning i omfang, slik at BOB blir like stor som NOB (NOB har om lag 30 000 ordartiklar meir enn BOB).

Eit viktig føremål er oppbygging av eit fast leksikografmiljø ved UiB – og det høyrer med at UiB ser for seg ein eigendel på over tre millionar i året gjennom prosjektperioden.

I søknaden er det lagt vekt på at innhaldsrevisjonen av standardordbøkene må vere materialbasert. Dette føreset tettare integrering i materialdelen av Språksamlingane, og er naudsynt både for å dokumentere bruk og for å halda oppe den vitenskaplege standarden som BOB og NOB no har:

Det europeiske nettverket for elektronisk leksikografi (ENeL) vurderer det slik at BOB og NOB i dag tilfredsstillar de kvalitetskravene som gjelder vitenskapelige ordbøker. Dette kvalitetsstempelet forplikter. Det databaseformateret standardordbøkene ligger i, er designet med tanke på at alle artiklar skal være lenket til dokumenterte belegg på ordet. Dette betyr at alt ligger til rette for at BOB og NOB også i framtida vil tilfredsstillare dagens krav til vitenskapelig etterrettelighet uten at det vil kreve mye ekstra arbeidsinnsats. De teknologiske løysningene som ligger til grunn, gjør det også lett å bygge språksamlingene vidare ut med tanke på nye prosjekter og sikrer at samlingene blir tilgjengelige for allmennheten. (UiB (2017))

## 4. Konklusjon

Det leksikografiske arbeidet ved Universitetet i Bergen er både ei vidareføring og ein ny start. Målet er godt tilgjenge til samlingsinnhald av høg kvalitet, og med gode tekniske løysingar for redigering og datadeling. I 2018 kjem det eit leksikografisk fagmiljø ved UiB. Det vert ei fagleg styrking som er naudsynt om ein skal nå målet. Naudsynt er det òg å styrkje vidare den IKT-faglege utviklinga som samlingane – ikkje minst ordbøkene – treng. Likeins må den bibliotek- og arkivfaglege drifta satsast på, for å gjere tilgjengeleg innhaldet av samlingane. Dette



må alt skje innom ei større ramme av fagleg samarbeid – innanlands og vonleg i Norden.

## Litteratur

### Ordbøker

- BOB (1986–2016) = Landrø, M., & B. Wangensteen (1986–, 3. utg. 2005). *Bokmålsordboka*. Bergen: Universitetsforlaget. Redaktør nettutgåve 2013–2016 B. Kjelsvik. Redaktør nettutgåve 2016– T. Svardal. <<http://ordbok.uib.no/>> (15.10.2017).
- NO (1950–2016) = *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. I–XII. Oslo: Samlaget.
- NOB (1986–2016) = M. Hovdenak, L. Killingbergtrø, A. Lauvhjell, S. Nordlie, M. Rommetveit & D. Worren (2006): *Nynorskordboka* (1. utg. 1986. 2. utg. 1993. 3. utg. 2001. 4. utg. 2006). Oslo: Samlaget. Redaktør nettutgåve 2012–2016 Oddrun Grønvik. Redaktør 2016– Terje Svardal. <<http://ordbok.uib.no/>> (15.10.2017).
- Storm, Johan (1884): *Kortere Ordliste med Forklaring af Lydskriften*. Christiania.

### Annan litteratur

- Aukrust, Knut & Bjarne Hodne (red.) (1998): *Fra skuff til skjerm*. Om universitetenes databaser for språk og kultur. Oslo: Universitetsforlaget.
- Ebeling, S., K. Kristoffersen, Ø. Vangsnes & Å. Wetås (2014): Den forskings- og undervisningsmessige verdien av eit utval samlingsressursar ved ILN. Rapport og vedlegg 1–6. <[http://www.uib.no/sites/w3.uib.no/files/attachments/d-sak-1-2-samlet\\_samlingsrapport\\_0.pdf](http://www.uib.no/sites/w3.uib.no/files/attachments/d-sak-1-2-samlet_samlingsrapport_0.pdf)> (14.10.2017).
- Fjeld, Ruth Vatvedt og Marit Hovdenak (2014): *Nordiske studier i leksikografi* 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.–16. august 2013. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 13. Oslo.
- Hovdenak, Marit (2014): *Bokmålsordboka* og *Nynorskordboka* gjennom ein generasjon. I: Fjeld, R. V. & M. Hovdenak (2014), 229–246.
- Karlsen, H. U., L. Vikør & Å. Wetås (red.) (2016): *Livet er æve, og evig er ordet*. Norsk Ordbok 1930–2016. Oslo: Det Norske Samlaget.
- Kjelsvik, B., C.-E. S. Ore, L. Vikør, Å. Wetås & D. Worren (2017): *Målreising og morsmålsdokumentasjon*. Heidersskrift til Oddrun Grønvik ved 70-årsleitet. Oslo: Novus Forlag.

- Myking, J. (2016): *Språksamlingane og UiB – organisatoriske moment*. Nasjonalt arbeidsseminar, Bergen, 14.4.2016. <[http://www.uib.no/sites/w3.uib.no/files/attachments/jm\\_innlegg\\_arbeidsseminar\\_spraeksamingane\\_0.pdf](http://www.uib.no/sites/w3.uib.no/files/attachments/jm_innlegg_arbeidsseminar_spraeksamingane_0.pdf)> (15.10.2017).
- Myking, J. (2017): Språksamlingane. Årsrapport 2016. I: ILLE, UiB. *Årsrapporter 2016. Forskergruppene Samlingene Forskerskolene*, 42–44.
- Ore, Christian-Emil Smith (1998): Hvordan lage databaser for språk- og kultur-fag. I: Aukrust & Hodne (1998), 29–57.
- Stortingsmelding 1 (2017–2018). *Nasjonalbudsjettet 2018*. Kap. 326 post 75 Tilskudd til ordboksarbeid. <<https://www.regjeringen.no/no/dokumenter/prop.-1-kud>> (12.10.2017).
- Universitetet i Bergen og Høgskolen i Volda (2017): *Norsk Ordbok. Søknad om revisjon og oppdatering*. Brev til Kulturdepartementet 28.2.2017.
- Universitetet i Bergen (2017): *Prosjektsøknad – revisjon av Bokmålsordboka og Nynorskordboka*. Brev til Kulturdepartementet 28.2.2017.
- Universitetet i Oslo (2005): *Om revisjon av Nynorskordboka*. Brev til Språkrådet frå Insittutt for lingvistisk og nordiske studium 12.12.2005.
- Worren, Dagfinn (2017): Fleirgenerasjonsverket Norsk Ordbok – frå nasjonsbygger til grunnlovsjubileumsprosjekt. I: Kjelsvik et al., 37–84.
- Vikør, Lars S. & Åse Wetås (2016): Norsk Ordbok: om folket – av folket – for folket. Historia om ordboksverket fram til i dag. I: Karlsen et al., 15–39.

Oddrun Grønvik  
Forskar emeritus  
Universitetet i Oslo  
Institutt for lingvistiske og nordiske studium  
Boks 1102, Blindern  
N-0317 Oslo  
oddrun.gronvik@iln.uio.no

Rune Kyrkjebø  
Førstebibliotekar  
Universitetsbiblioteket  
Universitetet i Bergen  
Postboks 7808  
5007 Bergen  
Rune.Kyrkjebo@uib.no

# Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforsking av ordtilfang

*Oddrun Grønvik & Christian-Emil Smith Ore*

Two written standards of different origins are in use in Norway, Bokmål and Nynorsk. During the 20<sup>th</sup> century Bokmål and Nynorsk came closer to each other through a series of orthographic reforms. Results of this policy can only be studied in detail if the standards are linked through the lexical item. A suitable tool is the Meta dictionary, an electronic registry for the Norwegian lexicon which allows linking lexicographical evidence to several headword forms through one node, the Meta Dictionary entry. This paper discusses principles and methods for expanding coverage of the linked standards through adding new, machine-analysed materials.

## 1. Innledning

Bokmål og nynorsk gir som skriftmål hver sin innfallspurt til norsk språkbruk og skriftkultur. Enten en ser bokmål og nynorsk som to ulike skriftmål eller som to ulike skriftlige uttrykk for norsk talemål, så er de nærstående skriftspråkstandarder, og utgjør til sammen den skriftspråklige rammen som folk i Norge lever innenfor, med et historisk spenn fra dansk (før 1907) til Aasen-normalen (uttrykt i oppslagsordformen i Aasens *Norsk Ordbog* fra 1873).

Formålet med arbeidet som er beskrevet i denne artikkelen, er å legge til rette for en systematisk undersøkelse av skriftmålsutviklingen i norsk på 1900-tallet. Dersom en skal kunne uttale seg med en rimelig grad av sikkerhet om resultatene av norsk språkpolitikk etter 1900, må det foreligge språksamlinger som samordner både primærkilder og sekundærkilder for begge målformer på ordnivå.

## 2. Premisser

Inngangspunktet til vitenskapelig analyse av språk er alltid ord i kontekst, enten tilfanget er registrert som lyd eller tekst. Dette gjelder hele spekteret av lingvistiske deldisipliner, fra fonologi til stilistikk. Det registrerte tilfanget som skal undersøkes må være lett tilgjengelig, ordnet etter faste kriterier og ha adekvate metadata, med kilde og annen bakgrunnsinformasjon på plass.

Formålet er å legge til rette for synkron og diakron jamføring av rettskrivningen for bokmål og nynorsk. Det eneste funksjonelle er en samordning der den leksikalske gjenstanden («lexical item», Atkins & Rundell 2008:163 ff.) er fikspunkt,

representert ved grunnformen, dvs. den eller de ordformene som brukes som oppslagsform for artikler i ordbøker for morsmålsbrukere (jf Bergenholz et al. 1997, s. 132). Forutsetningen for en slik samordning er at egenskapene til den leksikalske gjenstanden er kjent, og at det er enighet om rangering av egenskapene.

Nedenfor er det antatt at en leksikalsk gjenstand har et tilhørende begrep eller et begrepssett som kan defineres med støtte i dokumentert språkbruk, og at bruken kan vises med brukseksempler. Det at det eksisterer dokumentert enighet om betydning og bruk, er nøkkelegenskapen til en leksikalsk gjenstand. En ordform som bare har formopplysninger, men som ikke har eller kan knyttes til en definisjon, kan ikke regnes som en leksikalsk gjenstand, fordi ordets betydning og bruk ikke er kjent.

I tillegg er det forutsatt at ordets formelle egenskaper er enten kjent eller mulige å avlede fra det en vet om språket for øvrig. De formelle egenskapene omfatter ordtype, ordklasse, bøyning, uttale og historie. Det skriftlige uttrykket for en og samme leksikalske gjenstand kan derimot ha ulik form etter kontekst. Rettskrivning i egentlig forstand – ortografi – kan bare eksistere innenfor rammen av et standardspråk. Men spennet av formuttrykk for leksikalske gjenstander omfatter alle former av ordet eller uttrykket som kan finnes i tolkbar tekst eller tale.

De mest systematiske ordbeskrivelsene finnes i større ordbøker der ordboksartiklene for leksikalske gjenstander omfatter alle egenskapene som er nevnt ovenfor.

### 3. Verktøy

Jamføring av rettskrivningene krever et felles register med maksimum ett sammenkoplingspunkt per leksikalsk gjenstand. I dette registeret må hver leksikalske gjenstand kunne ha både flere normerte grunnformer som oppslagsord, og mange ulike typer brukseksempler med hver sine metadata.

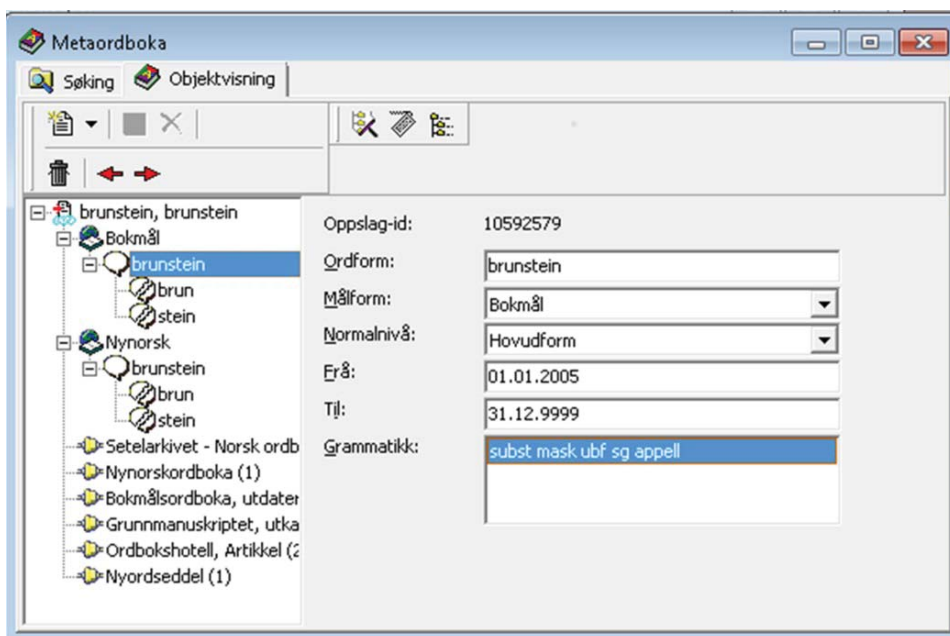
Tidligere brukte man kortarkiv der ordningsprinsippet ville vært den abstrakte størrelsen «den leksikalske gjenstanden med definisjon og brukseksempler», uttrykt ved en eneste foretrukket oppslagsform. Papirutgaven av Setelarkivet til Norsk Ordbok er et slikt arkiv. Oppslagsformen er satt til «Nynorsk, 1938-rettskrivningen, tradisjonelle former». Denne rettskrivningen var eneste mulige inngang til materialet i papirform dersom en ikke lagde ekstra kortarkiv som register til hovedarkivet. Men i den digitale verdenen er det mulig å lage mange innganger til et materiale.

For norsk språk eksisterer det i dag to verktøy som kan brukes som sammenkoplingsregister, Metaordboka og Norsk ordbank. Norsk ordbank er et fullform-system for bokmål og nynorsk. Det finnes to ordbanker, en for hver skriftstandard. De to ordbankene er ikke koplet sammen, men fungerer parallelt. Det andre verktøyet er Metaordboka, som har den funksjonaliteten en trenger for å lage et jamføringsregister for rettskrivningene, og som er brukt i arbeidet som er beskrevet i denne artikkelen (se også Ore 2000). Begge verktøyene er del av språksamlingene ved Universitetet i Bergen.

## 4. Metaordboka

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. Det er dermed mulig å bygge inn et historisk formregister per artikkel ved å kople til leksikalske ressurser for begge målformer, og indeksere dem. Hver artikkel kan ha ett eller flere oppslagsord i grunnform. Ordformer som er eller har vært normerte former i rettskrivningen, er brukt som oppslagsformer, mens kildene speiler den variasjonen som foreligger i skrift og tale over tid. Opplysningsskjemaet for oppslagsord i Metaordboka gjør at den ligger tett opp til fullformsystemet Norsk ordbank, men Metaordboka har ikke bøyingsinformasjon og regler for å generere fulle bøyningsskjema.

Artikkelkroppen består av referanser eller «hyperlenker» til ulike kildedatabaser. De fleste viser brukseksempler eller gir opplysninger om betydning, men det finnes også belegg som bare gir opplysning om forekomst, uttale, ordklasse, bøyning, ordlagingspotensiale eller liknende. I sin enkleste versjon er Metaordboka en sortering eller gruppering av tradisjonelle ordbokssedler. En metaordboksartikkel svarer da til en konvolutt eller eske med sedler som ansees å behandle samme oppslagsord.



Figur 1: Eksempel på metaordboksartikkelen for «brunstein».

Metaordboka samordner normert og unormert tilfang, tekst og transkribert tale. Materialet som er lenket sammen gjennom Metaordboka kan grovt kategoriseres som primærkilder og sekundærkilder.

Primærkildene omfatter (1) ordsedler, med metadataene oppslagsord, ordklasse, stedfesting eller litterær kilde og medarbeidersignatur. Belegg kan være tekst, lyd, bilde, video, (2) korpuskonkordanser.

Sekundærkildene omfatter utvalgte ordbøker, for eksempel grunn-manuskriptet til *Norsk Ordbok*, *Nynorsk etymologisk ordbok* (Torp 1919), og tidligere utgave av *Nynorskordboka* (NOB) og *Bokmålsordboka* (BOB). I tillegg kommer innholdet i Ordbokshotellet, som er beskrevet nedenfor.

## 5. Ordbokshotellet

«Ordbokshotellet» er både en søke- og framvisningsportal og et bibliotek for digital(isert)e ordbøker (Tvedt et al. 2007). Det inneholder nå 60 store og små dialektordsamlinger, ordbøkene Ivar Aasens ordbok (Aasen 1873) og *Norsk Landbruksordbok* (NLO) og fire rettskrivingsordlister for bokmål. Det utvides fortløpende og er Språksamlingenes hovedlager for digital(isert)e ordbøker. Hver ordbok/liste er en XML-kodet tekst i henhold til TEI P5 Dictionary. Siden graden av standardisering i foreleggene varierer, har hver artikkel ett eller flere normerte oppslagsord som brukes i koplingen til Metaordboka. Normeringspråket for dialektmaterialet er nynorsk.

For hver ordbok/liste er det angitt om den er normert til bokmål eller nynorsk. For koplingen til Metaordboka betyr det at en ordbok markert med nynorsk koples til en metaordboksartikkel som har oppslagsordet med språkform markert som nynorsk, og tilsvarende for bokmål.

## 6. Tidligere og nåværende bruk av Metaordboka

Metaordboka var i perioden 2002–2015 indeksert bare for nynorsk, etter 1938-normalen med tradisjonelle former. Formålet var å tilrettelegge resursene for bruk i Norsk Ordbok. Inndelingskriteriet er grunnform og ordklasse. En artikkel i Metaordboka ble ikke homografseparert selv om redigering i *Norsk Ordbok* førte til flere ordboksartikler.

I 2012 fikk *Norsk Ordbok* nettutgave. Dermed kom ønsket om å kunne søke i både *Norsk Ordbok* og Metaordboka med moderne rettskriving, helst fra både bokmål og nynorsk. Tiden var inne for å se på Metaordbokas utvidingspotensiale, ettersom den kunne ha flere oppslagsformer per artikkel, med markering av skriftform og gyldighetsperiode. Hver artikkel i *Norsk Ordbok* er dessuten lenket til en metaordboksartikkel som omfatter materialet NO-artikkelen ble redigert ut fra. I dag er *Norsk Ordbok* direkte søkbar bare i den målformen som ble brukt som

redaksjonsspråk. Men om metaordboksmaterialet får inngang både fra bokmål og nynorsk og i gjeldende rettskriving, vil det være langt enklere å tilrettelegge for at *Norsk Ordbok* også blir søkbar med dagens rettskriving (se også Ore 2012).

Første skritt var derfor å kople opp BOB mot Metaordboka på nytt. I forlengelse av dette ble også *Nyordsarkivet* for bokmål koplet opp mot Metaordboka. Oppkoplingen ble gjort halvautomatisk. De samordningsprinsippene som er beskrevet i neste avsnitt, gav grunnlag for en maskinell samkjøring. Bokmålsordene som manglet blant oppslagsordene i Metaordboka fikk nye rene bokmålsartikler i Metaordboka.

Dette førte til at Metaordboka fikk noe over 250 000 artikler med oppslagsform i normert bokmål. Av disse har om lag 70 000 oppslag både på bokmål og nynorsk. Ettersyn viser at denne forsiktige holdningen har gitt en rekke bokmålsartikler som bør slåes sammen med en tilsvarende nynorskartikkel. Men det krever manuell kontroll å sikre at dette blir riktig gjort.

## 7. Samordningsprinsipper for bokmål og nynorsk

Bokmål og nynorsk er nærstående skriftmålsstandarder, men langt fra identiske. Om en holder seg til samordning av grunnformer på leksikalsk nivå, forekommer det likevel at (1) en leksikalsk enhet er belagt i den ene målformen, men ikke den andre, (2) homografinnelingen er ulik, noen ganger med utgangspunkt i ortografi, andre ganger betydning eller bruksmåter. I den første sammenkoplingen av bokmål og nynorsk mot Metaordboka er det derfor valgt en restriktiv linje: bokmåls- og nynorsk materialet koples til samme MO-artikkel bare dersom det eksisterer en grunnform som er identisk for bokmål og nynorsk, og som er eller har vært gjeldende rettskriving. I kravet om identitet kommer også leddeling inn. Dersom den ene grunnformen er leddelt og den andre ikke, blir det ingen maskinell oppkopling. Tabell 1 viser hvordan systemet er satt opp.

I tabellen omfatter «full identitet» også at (1) bokmålssubstantiv med genus f og m svarer til nynorsk f, og at (2) nynorsk infinitivsending *-e* og *-a* svarer til bokmål infinitivsending *-e*.

I sammenkoplingsprosessen er det ikke tatt hensyn til etymologi – det er antatt at alle grunnformer i rettskrivningen har sine motsvar i nåværende norsk tale. Dermed forekommer det sammenkoplinger som dette:

Nynorsk: *botnlaus*, adj

Bokmål: *botnlaus* el. *botnløs* el. *bunnlaus* el. *bunnløs*, adj

Type	Bokmål	Nynorsk
full identitet	<i>kaste</i> , v	<i>kaste</i> v
identitet i en form	<i>legge</i> , v	<i>legge</i> el. <i>leggje</i> v
identitet i en form	<i>bo</i> el. <i>bu</i> , v	<i>bu</i> v

Type	Bokmål	Nynorsk
identitet i en form	<i>høgraud</i> el <i>høgrød</i> el <i>høyraud</i> el <i>høyrød</i> , adj	<i>høgraud</i> , adj
identitet i en tidligere form	<i>vanskelig</i> , adj	<i>vanskeleg</i> , adj (til 2012 også [ <i>vanskelig</i> adj])
identitet i en tidligere form	<i>høst</i> m (til 1981 også [ <i>haust</i> m])	<i>haust</i> m

Tabell 1: Kriterier for sammenkopling mellom bokmål og nynorsk

På noen punkter skiller Metaordboka strengere enn i de tilkoplta ordbøkene og ordlistene. I *NOB* og *BOB* forekommer det at ulike avledninger er ført som oppslagsord til samme artikkel. Det ser ut til særlig å gjelde substantiv som er avledet fra verb:

- bokmålsavledninger med *-else* m og *-ing* m
- nynorskavledninger med *-sel*, m el. *-sel*, f og *-sle*, f.

En rimelig tolking er at en i disse ordbøkene mener at det bare er snakk om rettskrivningsvarianter. At Metaordboka har et strengere skille, gjør det mulig å kople sammen identiske oppslagsord på *-ing* eller på *-sel* på en konsistent måte. For eksempel vil vi kunne kople bokmålsordene *ending* og *tilførsel* sammen med nynorskordene *ending* og *tilførsel* uten å plassere koplingene i sammenheng med bokmål *endelse* eller nynorsk *tilførsle*.

Metaordboka var i utgangspunktet en enkel systematisering av språklig grunnlagsmateriale uten homografseparering. Sammenkopling av bokmål og nynorsk tvinger fram bruken av leksikalsk gjenstand som koplingspunkt.

### III **bu** verb

1 ha fast tilhald (i kortare eller lengre tid); ha heimen sin på (ein stad); leve i lag

2 finnast, vere (IV)

3 lage til; gjere ferdig

Figur 2: Artikkel for verbet å *bu* i *NOB*

I *BOB* dekker verbet *bo* el. *bu*, v, betydning 1 og 2. Betydning 3 er skilt ut i et eget verb, *bu*. For å få tatt hensyn til denne oppdelingen må metaordboksmaterialet fordeles på to artikler, en for bokmål *bu* el. *bo*, v, nynorsk *bu*, v og en hvor både bokmål og nynorsk har oppslagsform *bu*, v.



## 8. Oppkopling av skoleordlister for bokmål

Metaordboka kan brukes til å lage en historisk oversikt over ordinventaret i norsk. Som en test på hvordan det kan gjøres, ønsket vi å se på dokumentasjonen av det sentrale ordforrådet i bokmål fra 1938 og utover. Testmaterialet var fire skoleordlister for bokmål. I 1938 og 1959 var det store og omfattende rettskrivningsreformer for både bokmål og nynorsk som omfattet både ortografi og ordforråd. To av ordlistene er fra 1939, en fra 1959 og en fra 1973. Ordlistene ble maskinelt behandlet slik at oppslagsordene med ordklasse ble hentet ut der det var mulig. Deretter ble listene med oppslagsord kjørt mot oppslagsordene i først bokmålsoppslagene i Metaordboka, og dersom det ikke var treff, ble nynorskoppslagene brukt. Omtrent ti prosent av oppslagsordene i ordlistene manglet ordklasse, eller gav flere tilslag i Metaordboka. De ble sjekket manuelt. Deretter ble de fire ordlistene gitt en enkel TEI-koding, lagt inn i Ordbokshotellet som bokmålskilder og koplet opp mot eksisterende artikler, eller lagt i nye artikler. Samordningsprinsippene for bokmål (og nynorsk) er beskrevet over.

De fire ordlistene gav tilsammen ca. 70 000 nye koplinger til Metaordboka, fordelt på ca. 30 000 metaordbokartikler. Drøyt 7 600 oppslagsformer finnes i alle fire ordlister. Noe over 12 600 oppslagsord finnes i bare en av ordlistene. Alle ordlistene har mange ord som ikke forekommer i de andre.

Ordliste	Utgitt	Antall oppslagsord	Ordform uten ordklasse	Antall tilslag i MO-artikkel ved maskinell analyse	MO-artikkel bokmål ikke funnet
Lange	1939	12 422	454	10 279	1 761
Eitrem	1939	16 626	2 561	12 054	2 721
Krogsrud & Seip	1959	16 657	559	14 178	1 903
Torvik	1973	23 535	837	20 153	2 957

Tabell 2: Oversikt over tilkoblede ordlister, automatisk tilslag og artikler som krevde manuell kontroll.

Den minste har om lag 11 prosent «særegne» ord. De to jamstore fra 1939 og 1959 har henholdsvis 18 og 8 prosent. Den største ordlisten, fra 1973, har flest både relativt og absolutt, med 29 prosent ord som ikke forekommer i de andre, men deler like fullt 16 500 ord med en eller flere av de 3 andre, spesielt med ordlisten fra 1959 (Krogsrud & Seip 1959). 85 prosent av oppslagsordene i denne er også med i 1973-ordlisten. For den jamstore ordlisten fra 1939 (Eitrem 1939) er 75 prosent av ordene med i 1973-ordlisten. En rask gjennomgang viser at svært mange av de 25 prosentene fra denne 1939-ordlisten som ikke er med i 1973-ordlisten, faktisk er rettskrivingsnøytrale ord som *lovmessig* og *metafysikk*.

Foreløpige resultater viser at de fleste oppslagsformene som ble introdusert i 1938, finnes i dagens bokmålsnorm. Det kreves mer arbeid for å få et fullt bilde, og det er mange spørsmål i tilknytning til dette arbeidet som fortjener videre utforskning. Eksperimentet med de fire ordlistene viser at det er overkommelig å lenke inn ordlister fra ulike perioder og på den måten gi et historisk overblikk over de leksikalske gjenstandene og rettskrivingsnormene.

Et tilsvarende overblikk over bruken krever tilgang til et historisk korpus. Metaordboka er utstyrt for å lenke til korpus og til fulle konkordanser. Norsk bokmål skilte lag med dansk skriftspråk for alvor i 1907. Et bokmålskorpus med tekster fra 1907 til i dag står høyt på ønskelisten.

## 9. Konklusjon

Det er her vist at det er både mulig og rimelig å legge til rette for en systematisk undersøkelse av skriftmålsutviklingen i norsk på 1900-tallet, ved å samordne primærkilder og sekundærkilder for begge målformer med den leksikalske gjenstanden som samordningspunkt. Dette forsøket viser ett resultat av slik samordning. Et annet er at arbeidet i de siste tiårene med språksamlingene for nynorsk også er til nytte for dokumentasjon og utforskning av bokmål, og dermed hele den norske skriftmålssituasjonen.

## Litteratur

### Ordbøker

- Aasen, Ivar (1873) *Norsk Ordbog med dansk Forklaring. Fjerde uforandrede udgave, (1918)*. Kristiania: Vestmannalaget og Cammermeyers forlag.
- Bergenholtz, Henning et al (1997): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- BOB = Wangensteen, Boye (2006): *Bokmålsordboka. Definisjons- og rettskrivingsordbok*. (3. utg.). Oslo: Kunnskapsforlaget.
- Eitrem, Hans (1939): *Rettskrivning 1938. Regler for skoler og privat bruk*. Oslo: Fabritius.
- Krogsrud, Torgeir & Didrik Arup Seip (1959): *Rettskrivningsregler*. Rettskrivningslære og ordliste etter „Ny læreboknormal 1959“. Oslo: Cappelen.
- Lange, August (1939): *Norsk rettskrivningsordliste*. Oslo: Tiden Norsk Forlag.
- NLO = Rommetveit, Magne (1978): *Norsk landbruksordbok I-II*, Oslo: Samlaget.
- NO = Norsk ordbok: ordbok over det norske folkemålet og det nynorske skriftmålet. 1966-2016. Bd. 1-12. Oslo: Det Norske Samlaget.

NOB = Marit Hovdenak, Laurits Killingbergtrø, Arne Lauvhjell, Sigurd Nordlie, Magne Rommetveit & Dagfinn Worren (2006): *Nynorskordboka* (4. utg.). Oslo: Samlaget.

*Nyordsarkivet*. Bokmål. Språksamlingane, Universitetet i Bergen.

Torp, Alf (1919): *Nynorsk etymologisk ordbok*. Kristiania: Aschehoug.

Torvik, Ingvald (1973): *Ordliste for alle*. Bokmål. Oslo: Universitetsforlaget.

## Annen litteratur

Atkins, Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Ore, Christian-Emil, (2000): Metaordboken – et rammeverk for Norsk Ordbok. I: *Nordiska studier i lexicografi* 5. Göteborg: Nordisk forening for lexicografi, s. 250—270.

Ore, Christian-Emil Smith (2012): Nettordbøker og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok. I: *Nordiska studier i lexicografi* 11. Rapport från Konferensen om lexicografi i Norden, Lund 24–27 maj 2011. Lund: Nordisk forening for lexicografi, s.486–499.

Tvedt, Lars Jørgen, Elisabeth Lien & Øyvind Eide (2007): Ordbokshotellet – varig lagring og formidling av norske ordsamlinger. I: Torben Arboe (red.), *Nordisk dialektologi og sociolingvistik*, Foredrag på 8. Nordiske Dialektologkonferens, Århus 2006. Peter Skautrup Centret for Jysk Dialektforskning, Aarhus Universitet, s. 379–388.

Oddrun Grønvik  
Forsker emeritus  
Universitetet i Oslo  
Institutt for lingvistiske og nordiske studier  
Boks 1102, Blindern  
N-0317 Oslo  
oddrun.gronvik@iln.uio.no

Christian-Emil Smith Ore  
Førsteamanuensis, dr.scient.  
Universitetet i Oslo  
Institutt for lingvistiske og nordiske studier  
Enhet for digital dokumentasjon  
Boks 1102, Blindern  
N-0317 Oslo  
c.e.s.ore@iln.uio.no

# Ømålsordbogen som leksikografisk, sproglig og kulturhistorisk resurse

*Asgerd Gudiksen & Henrik Hovmark*

As part of an extensive process of digitalization the previously edited volumes of *Ømålsordbogen* (The Dictionary of Danish Insular Dialects) (1–11, *a-lindorm*) are now available as one searchable xml file. We show how this new digital tool, even in its primitive state in terms of tagging, constitutes a valuable resource, which not only enables to explore the rich linguistic and cultural data in the dictionary in new ways but also saves time and brings more rigour to the editing process.

## 1. Ømålsordbogen anno 2017: omstilling og digitalisering

*Ømålsordbogen* (ØMO) er en af de store nationale videnskabelige dokumentationsordbøger i Norden (jf. fx *Norsk Ordbok*, *Ordbok över Finlands svenska folkmål*, mfl.). ØMO beskriver de traditionelle danske dialekter, dvs. talesproget blandt bønder og fiskere, på Sjælland, Lolland-Falster, Fyn og omliggende øer i perioden 1750–1950, med hovedvægten på tiden 1850–1920. Ordbogen medtager hele ordforrådet hos landbefolkningen og giver detaljerede beskrivelser af ordenes udtale, bøjning, syntaks, betydning og brug. Et særkende ved ordbogen er at den også giver indgående etnologisk-encyklopædiske beskrivelser af de redskaber, arbejdsprocesser og traditioner der hørte til ældre tiders bonde-, håndværker- og fiskerkultur, med henvisning til den folkelige terminologi der er knyttet til dem (jf. Gudiksen & Hovmark 2009). Under artiklen *bage* er der fx et afsnit på tre spalter om bagning i den gammeldags murede bageovn, og under *fastelavn* en skildring af dyster, optog, lege og gilder. Ordbogens oplysninger er sammen med de tilgrundliggende samlinger en vigtig del af dansk kulturarv.

Første bind af ordbogen udkom i 1992 efter en slags førstegenerationsdigitalisering i slutningen af 1980'erne, hvor man tog et særligt tekstbehandlingsprogram og et korpusværktøj i brug. Herefter er ordbogen udkommet i regelmæssigt tempo med et bind hvert andet år, senest bind 11 (*kurv-lindorm*) i 2013. Halvdelen af materialet er således redigeret. I 2013 var en teknisk omstilling imidlertid påkrævet, og man gik derfor i gang med en ny digitaliseringsproces finansieret af Carlsbergfondet. Processen har omfattet flere forskellige dele: udarbejdelse af en dania-font i unicode, et nyt korpusværktøj (i samarbejde med Det Danske Sprog- og Litteraturselskab), digitalisering af de trykte bind 1–11 (*a-lindorm*) til xml-format, indkøb og opsætning af et nyt databasebaseret redigeringsværktøj (iLEX) inkl. udarbejdelse af xml-skema samt publiceringsløsning. Både database

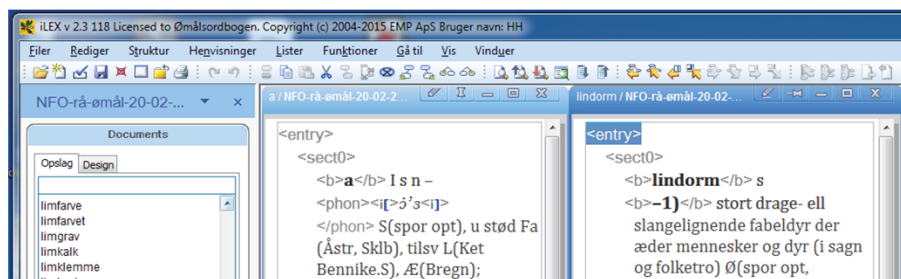
og publiceringsløsning forsøges så vidt muligt tilrettelagt så ordbogen fremover ikke er bundet til at udkomme på tryk, men også er forberedt til publicering på andre platforme. Hertil kommer at man er i gang med en digitalisering af ordbogens seddelsamling. Målet er en fulddigital udgivelsesproces. Bind 12 forventes at udkomme i 2018. Mange af de nye resurser foreligger indtil videre dog kun i forholdsvis rå former. Det gælder fx de digitaliserede trykte bind. De udgør dog allerede en værdifuld resurse, og i denne artikel vil vi give nogle eksempler på de muligheder som denne nye resurse giver.

## 2. Ømålsordbogen bind 1–11 som samlet, søgbar base

Da ØMO begyndte at udkomme i 1992, var det takket være udnyttelsen af det DOS-baserede tekstbehandlingsprogram T3, der rummede mulighed for at designe egne fonttegn. Det gjorde det muligt selv at udarbejde fonttegn i lydskriften *ania* og fremstille et manuskript med lydskrift, noget de daværende programmer ellers ikke gav mulighed for

T3 viste sig imidlertid også at have en bagside. Programmet kunne i praksis kun håndtere en lille datamængde ad gangen, typisk hvad der svarer til fx 10–15 sider i den trykte bog. Søgning i det samlede, udgivne materiale har derfor aldrig været muligt. Et lige så stort, og med tiden langt mere alvorligt problem var at filformatet havde sin egen, partikulære standard og kodning. De redigerede data blev hurtigt umulige at læse i og af andre programmer og kunne ikke umiddelbart konverteres. De redigerede data var dermed reelt ikke tilgængelige, og var heller ikke sikret i digital form.

Dette har heldigvis ændret sig. I forbindelse med omstillingsprocessen lykkedes det en medarbejder hos konsulentfirmaet EMP at knække koden til T3 ved hjælp af printfilerne, og data fra alle de trykte bind (*a–lindorm*), er blevet konverteret til unicode-tegn og foreligger nu som en let opmærket xml-fil som p.t. kan tilgås fra det nye redigeringsprogram, iLEX (ØMO-D). På Figur 1 ses således opslag på den første og sidste artikel i ØMO-D, *a* og *lindorm*, samt til venstre et udsnit af den alfabetiske ord- eller artikelliste.



Figur 1: Opslag i ØMO-D på artiklerne *a* og *lindorm* (første og sidste artikel i de hidtil trykte bind af ØMO).

En simpel søgning i basen kan foregå via ordhjulet til venstre i iLEX, men iLEX tilbyder langt flere søgemuligheder via et søgepanel i en funktionsmenu hvor både formelle søgeudtryk og ikke mindst xml-opmærkningen kan udnyttes til både søgning og statistik. Da basen med ØMO-D kun er minimalt opmærket efter skriftsnit (<b> (fed), <i> (kursiv), <phon> (lydskrift – dania), <spatieret> (i forb s p a t i e r e t s k r i f t) samt i tre meget grove sektionsinddelinger (<sect0>, <sect1> og <sect2>) (jf. Figur 1), er søgning i fritekst ofte den mest oplagte søgestrategi, men undertiden kan opmærkningen dog også være til hjælp til afgrænsede eller kombinerede søgninger (jf. afsnit 3).

### 3. ØMO-D som leksikografisk resurse

Flerordsforbindelser består som navnet siger af mere end ét ord, og en bruger kan komme i tvivl om hvor en sådan forbindelse behandles i en ordbog. Løsningen er normalt at en forbindelse kun behandles ét sted, dvs. i den traditionelle ordbog ordnet.dk efter enkeltord: i én artikel, men ofte med henvisning fra de andre artikler hvor forbindelsen potentielt kunne have været behandlet (idet funktionsord o.l. normalt ikke tæller med som potentielle opslagsord, fx Svensén 2004:247ff.). Dette princip følges også i ØMO, hvor princippet har en ekstra begrundelse, nemlig at der skal gives en samlet udbredelse for en flerordsforbindelse dér hvor den behandles, dvs. fx:

**lus** s ... i forb k n æ k k e l u s Ø(spor opt)  
**knække** v ... i forb k n æ k k e →l u s

For at kunne overholde princippet om at beskrivelse og udbredelsesangivelse kun gives ét sted, skal man have alle relevante belæg til rådighed når man går i gang, også hvis den artikel man står i, skal resultere i en henvisning og ikke en fuld beskrivelse (fx hvis man er i gang med at redigere artiklen *knække*). Behandlingen af flerordsforbindelser i ØMO stiller imidlertid redaktørerne over for nogle særlige udfordringer på grund af den måde som datamaterialet er organiseret på. Datamaterialet er indtil videre organiseret i en række forskellige delsamlinger, og en forbindelse kan være behandlet i flere samlinger. Tidligere redaktører var dog naturligvis opmærksomme på denne problemstilling, og derfor er mange belæg allerede forsøgt samlet under ét opslagsord – fx er ordsprog og talemåder, som ofte kan have en kompleks struktur med mange potentielle opslagsord og mange varianter, samlet i én særlig delsamling. Alligevel kan der gemme sig materiale flere steder. Der er også forskel på i hvor høj grad den enkelte redaktør har opfattet en flerordsforbindelse som fast og dermed behandlet den eksplicit som en forbindelse, eller fx ”blot” har medtaget forbindelsen som et citat. Og ulykkeligvis har man heller ikke haft knivskarpe principper for under hvilket ord (fx ”første substantiv”, ”første substantiv eller verbum” e.l.) en forbindelse skulle anbringes

i hverken samlinger eller ordbog: Man fulgte til en vis grad ODS, men ODS's organisering af forbindelser følger netop ikke noget entydigt princip af den netop nævnte type. Det kan altså være svært at vide om en flerordsforbindelse evt. allerede er blevet behandlet i de tidligere trykte bind.

Omstillingsprocessen vil kunne råde bod på denne vanskelighed, som i den nuværende produktionsproces ofte koster mange kræfter og tid for redaktørerne i form af gentagne kontrolopslag i seddelsamlingen og de trykte bind. For det første er der håb om at man fremover vil kunne søge alfabetisk på tværs af alle delsamlinger når digitaliseringen af seddelsamlingerne er afsluttet. For det andet giver ØMO-D allerede nu mulighed for at lave samlede undersøgelser af hvor og hvordan ord og ordforbindelser evt. er blevet behandlet i de tidligere trykte bind.

Ordet *lus* er forholdsvis velbelagt i ØMOs samlinger, og det seddelmateriale som havde *lus* som opslagsord, og som derfor dannede udgangspunkt for redaktørens arbejde, viste også at ordet indgår i en del ordsprog og talemåder. ØMO-D blev derfor taget i brug for at tjekke evt. yderligere relevant materiale i de trykte bind. Undersøgelsen viste hurtigt at der gemte sig materiale, ikke kun som udhævede forbindelser, men også i fx citatstof. I materialet under *lus* lå fx et belæg på talemåden *være rig på lus (og lopper) og lange negle*. Denne flerordsforbindelse rummer flere ord som potentielt kunne fungere som opslagsord, men kun ét, nemlig *lang*, lå tidligere i alfabetet. Et simpelt opslag på ”lange negle” viste imidlertid at talemåden ikke var behandlet under *lang*, men til gengæld at en variant var medtaget under *fattig*, som der nu bliver henvist til under *lus*. Søgningen fandt i øvrigt også en helt anden talemåde i artiklen *hip* (II), hvor *lus* var kombineret med *lange negle* – som ikke lå i materialet til *lus*.

Problemet ligger dog ikke kun i at der er vanskeligheder forbundet med at finde alt relevant stof og koordinere det konsekvent. Selve stoffet kan undertiden også være præget af meget stor variation. Fx var der i materialet til *lus* belæg på ordsproget *der er liv i kludene, sagde manden, han pillede lus af sin frakke*. Ordsproget havde varianten *hun skal nok få sat liv i kludene, som kællingen sagde, hun pillede lus af uldklokken*. Et opslag i ØMO-D viste at ordsproget – eller ordspørgene? – allerede var behandlet under *klud*, dog med mangelfuld udbredelse. Under *klud* fandtes imidlertid en tredje variant: *der er liv i kludene, sagde pjaltekræmmeren, lusene krøb omkring i pjalterne*. I denne slags tilfælde kan det være svært at pege på én grundlæggende form af et ordsprog eller en talemåde (som ville kunne styre udvælgelsen af et hovedopslagsord efter mere objektive kriterier).

Den meget specifikke søgning på ”lange negle” gav to meget konkrete resultater. En søgning på formen ”lus” og/eller på bøjningsformer heraf gav betydeligt flere resultater (over 200). Den sparsomme opmærkning efter skriftsnit i ØMO-D giver dog mulighed for at begrænse omfanget af irrelevant støj. En søgning inden for tagget <i> vil således begrænse søgningen til tekst i kursiv, hvilket fx vil omfatte citater og henvisninger i ØMO, og en søgning der inddrager tagget <b> (’fed’) vil give opslagsord og orddannelser som søgeresultat. Selvom den rene tekstsøgning på formen ”lus” med evt. bøjningsformer gav forholdsvis mange

resultater, viste det sig alligevel at være umagen værd at gennemgå resultaterne som altså kunne forekomme alle steder i en artikel. Ordet *lus* optrådte fx i en af definitionerne til ordet *bid* (I.8). Dette ord forekom ikke i materialet til *lus*, men brugtes metonymisk og uformelt om *lus*, og det vil være relevant at henvise til ordet og dermed koble materialet sammen i basen.

Samlet set bidrager ØMO-D med værdifulde oplysninger som sætter redaktørerne i stand til at gøre de fremtidige artikler mere konsekvente og udtømmende i forhold til det komplekse materiale. Muligheden for at søge elektronisk i ØMO-D er tidsbesparende og giver et mere fuldstændigt datagrundlag at arbejde ud fra.

#### 4. ØMO-D som sproglig resurse

ØMO har som sagt i sin egenskab af en stor dokumentationsordbog detaljerede beskrivelser af udtale, bøjning, syntaks, betydning og brug. Kildegrundlaget afviger på mange måder fra den typiske dokumentationsordbog, en ordbog over standardsproget som fx *Ordbog over det danske Sprog* (ODS). Ordbøger over standardsproget som ODS beskriver overvejende skriftsproget og bygger på excerpter af trykte kilder og – i nyere tid – på elektroniske korpora. Dialektordbøger som ØMO inddrager også trykte kilder, fx bondedagbøger, dialektlitteratur, etnologiske og folkloristiske skildringer. Men de vigtigste kilder til ordbogens seddelsamling er ordbogens eget indsamlede materiale. Materialet hidrører fra direkte optegnelser med dialekttalende meddelere foretaget af dialektforskere, meddelernes besvarelser af spørgelister og endelig ordsamlinger og ordbøger udarbejdet af lokale dialekttalende. Der er tale om et meget heterogent materiale der må vurderes kildekritisk i forhold til den situation det er blevet til i. Optegneren kan fx misforstå noget, eller meddelelserne kan huske forkert (ofte drejer det sig om forhold der ligger mange år tilbage i forhold til optegnelsestidspunktet). Det kræver altså en omfattende baggrundsviden at forstå oplysningerne fuldt ud. Mange sprogforskere vil derfor være mere tjent med de generaliseringer i ordbogen som redaktørerne efter kritisk vurdering er nået frem til end det måske lidt fortegnede billede de selv ville være nået frem til ved at gå direkte til kilderne.

Det har fra begyndelsen da også været tanken at ordbogen skulle kunne danne grundlag for (mindre) sproglige undersøgelser (jf. Pedersen 1994). Indtil for nylig måtte man gennembladre ordbogen for at samle materiale til en sproglig undersøgelse, som fx en undersøgelse af variation i verbalbøjningsmønstre (Gudiksen 2011) – en ret arbejdskrævende opgave, men dog overkommelig i modsætning til at skulle have etableret det samme overblik ved hjælp af grundmaterialet. Men vores nye resurse giver nye muligheder for sproglige undersøgelser. Ved hjælp af basen kan man fx lave visse syntaktiske undersøgelser. Ganske vist kan man ikke søge systematisk på syntaktiske konstruktioner, men man kan søge på grammatiske markører som ”upers” (dvs. upersonlig), ”i sb anv” (dvs. i substantivisk anvendelse), ”i adj anv” (dvs. i adjektivisk anvendelse) eller ”sideordnet”. En



hurtig søgning på ”upers” bringer bl.a. en anvendelse af *kaste* som er ukendt i standardsproget, nemlig ’blæse med kraftige vindstød’, der bl.a. er afhjemlet på Tåsinge med citatet *det er slemt til at kaste fra nordvest* (*kaste* III.6.1). Man kan på tilsvarende måde søge på betegnelser for lydlige fænomener som ”kortvokalstød”, ”v-stød”, ”enklitisk”. En søgning på ”enklitisk” fører bl.a. hen til konstruktionen *til den* ’afsted’, fx *vi skal te ’en* ’vi skal afsted’ (under *den* 2). Et andet søgeresultat er citatet: *Nu har jeg knaldfyret en hel klokke time og endnu er hæ’ d varmt*, en sammentrækning af ordet *her* med nægtelsen *itte* (en sideform til *inte*). Desuden fremkommer også den generelle beskrivelse af sådanne enklitiske forbindelser med nægtelsen under udtaleangivelserne ved *itte* (II), fx [*sgadə*] ’skal itte’.

## 5. ØMO-D som kulturhistorisk resurse

Omkring 1940 blev ØMOs seddelsamling opdelt i de tidligere omtalte delsamlinger, der – foruden delsamlinger som ordsprog og talemåder – omfatter emnesamlinger som fx *høst*, *pløjning*, *bagning* (og en restsamling for den del af ordforrådet der ikke kunne indpasses i emnerne). Formålet var at kunne foretage en samlet analyse af ordforrådet inden for de enkelte begrebsområder til brug for de etnologiske beskrivelser, der som tidligere nævnt er en central del af ordbogen, og udarbejde synonymkredse, herunder udvælge et hovedsynonym hvor alle synonymerne er samlet, og hvortil der henvises. *Lokum* er et eksempel på et hovedsynonym inden for emnet ”gården og dens indretning”. Artiklen afsluttes med synonymkredsen *das*, *hus*, *kloset*, *retirade*, *skidehus*, *vandhus*, *wc*. Der er altså også synonymer længere fremme i alfabetet end redaktionen er nået til, takket være forredaktionen af emnerne, og en søgning på *lokum* har ikke givet yderligere eksempler inden for det redigerede. Men artiklen *lokum* har inspireret os til en anden søgning der har bragt viden frem der lå og gemte sig i artiklerne.

Betydningsbeskrivelsen i artiklen *lokum* slutter med følgende måske lidt overraskende oplysning: ”indført fra omkr 1860, i ældre tid brugte man normalt grebningen i stalden”, og det inspirerede til at søge på ”indført”. Resultatet var 26 artikler hvor ”indført” indgår et eller andet sted i artiklen. Nogle af eksemplerne havde ikke relevans i denne sammenhæng, fx hvor ”indført” indgår i et citat. Men tilbage blev der 12 artikler med redaktionel markering. I tre tilfælde er der angivet et præcist årstal, nemlig hvor noget er indført ved lov: *femogtyveøre* (1873), *jagt-tegn* (1922) og *konfirmation* (1736.) I de øvrige tilfælde viser angivelserne at der er tale om gradvise ændringer. Det er karakteristisk for datidens bondesamfund at nye ting vinder frem i langsom takt. Det angives fx at *højstade* (en type bihus af træ) er indført ”ca. 1860”, *kødhakkemaskine* ”omkr 1880–1900” og *kalvesteg* ”i takt med indførslen af *komfuret*”.

Søgning på ”først alm”, ”først i brug”, ”først brugt” gav yderligere eksempler på forandring af bondekulturen, bl.a. *henkoge*, *høsegård*, *hårbørste*, *hårnål*, *jordbær*, *juletræ*. I mange tilfælde drejer det sig dog om ting som inden for andre

sociale lag har været kendt tidligere, men altså først bliver en del af bondekulturen i slutningen af 1800-tallet. ”Nyere” var topscoreren med 625 artikler og ”yngre” med 593 artikler. ”Nyere” bruges undertiden om udtalemæssige forhold, og ”yngre” ofte.

Under ”nyere” er der ikke kun ord for ”nye” sagforhold som *lokum*, *henkoge* og *jordbær*, men også nye ord for gamle ”sager”, fx: *kostald*, der afløser fx *nøds*, *kohus*, *kolade* (idet *stald* oprindeligt kun blev brugt om hestenes afdeling), og *idiot* brugt som skældsord (oprindeligt blev *idiot* kun brugt som betegnelse for en svært mentalt retarderet person). Arbejdet med at nå frem til tidsfæstelse af ord og fænomener er noget vi bruger en del tid på, og det er derfor glædeligt at man med de nye søgemuligheder hurtigt kan samle sådanne oplysninger sammen og nå frem til nye indsigter som vel at mærke ikke fremgår af de tidligere emneredaktioner. Søgning på ”nyere” og ”yngre” vil givetvis også kunne give muligheder for at etablere et mere nuanceret billede af den rigsmålspåvirkning af udtalen der sætter ind inden for ordbogens periode.

Citaterne i ØMO er ofte ikke kun sprogligt, men også kulturhistorisk interessante. Undertiden giver de også et indblik i holdninger og mentalitet i det gamle bondesamfund, noget vi naturligvis også lægger vægt på ved udvælgelsen af citater. I materialet til opslagsordet *københavnsk* var der to citater som blev udvalgt af den grund: *hun er bleven så kevihavnsk at hun er færdig at brække tungen* og *hun er blevet så københavnsk i de par måneder så hun næsten ikke er til at forstå*. Men kan billedet suppleres ved søgning på *København*? Ja, under *kloges* finder man fx citatet: *je ku jonok kloves ae* (kloges af, dvs. ’forstå’), *a ded vaa saan in rejti Københavner Sjaafelhed* (Holbæk, Nordvestsjælland).

Søgningen gav imidlertid især eksempler på en anden form for kontakt, at København er et sted man tager *ind* til med varer, fisk eller landbrugsprodukter for at sælge dem der, som fx i et citat under *dreng*: *Nogle måneder herefter sejlede drengen til København med en ladning sild* (”blandt fiskere kaldes de drenge til de bliver gift”) (Humblebæk, Nordsjælland). Citaterne om at bringe varer ind til København forekommer ret uforudsigeligt under følgende opslagsord: *dingster*, *distrikt*, *dreng*, *gå*, *have* (V.43.1), *herrefisk*, *klage*, *klods*, *kvase*, *københavnsko*, *københavnslæs*, *last*, *lave*, *levende*, og det ville næppe have været muligt at finde frem til citaterne uden elektronisk søgemulighed.

## 6. Afslutning

Der er ingen tvivl om at digitaliseringen af bind 1–11 har givet redaktørerne et godt hjælpemiddel ved redaktionen. Men mange andre vil også kunne have glæde af de nye muligheder for at søge på tværs i hele ordbogsmanuskriptet at dømme ud fra de småundersøgelser vi har beskrevet her, og det er noget vi vil tage med os i overvejelserne over hvordan en net-version af ordbogen skal komme til at se ud.

## Litteratur

### Ordbøger

- Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet* 1–12. 1966–2016. Oslo: Det norske Samlaget.
- ODS = *Ordbog over det danske Sprog* 1–28. Trykt 1919–1956, elektronisk <ordnet.dk/ods>. København: Det Danske Sprog- og Litteraturselskab.
- Ordbok över Finlands svenska folkmål*. Institutet för de inhemska språkens webpublicationer 33. <kaino.kotus.fi/fo>. 2013–. Helsingfors: Svenska litteratursällskapet i Finland.
- ØMO = *Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer* 1– (1992–). København: Center for Dialektforskning, Københavns Universitet & Universitets-Jubilæets danske Samfund.

### Anden litteratur

- Gudiksen, Asgerd (2011): Variation i morfologi: 2. svage verbalbøjning i de danske ømål. I: L.-E. Edlund et al. (red.): *Studier i dialektologi och socio-lingvistik: Föredrag vid Nionde nordiska dialektologikonferensen i Uppsala 18–20 augusti 2010*. Uppsala: Gustav Adolfs Akademi, 123–130.
- 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.
- Pedersen, Karen Margrethe (1994): Ordbog og encyklopædi. I: Anna Garde & Pia Jarvad (red.): *Nordiske Studier i Leksikografi 2*. LEDA (Foreningen af Leksikografer i Danmark), 248–256.
- Svensén, Bo (2004): *Handbok i leksikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.

Asgerd Gudiksen  
lektor, ph.d.  
gudik@hum.ku.dk  
Ømålsordbogen

Henrik Hovmark  
lektor, ph.d.  
hovmark@hum.ku.dk  
Ømålsordbogen

Center for Dialektforskning  
Institut for Nordiske Studier og Sprogvidenskab  
Emil Holms Kanal 2  
DK-2300 København S

# Kontrastiv konstruktionsgrammatik och tvåspråkig lexikografi. Konstruktioner och konstruktionsekvivalenter i *Islex.se*

*Anna Helga Hannesdóttir & Benjamin Lyngfelt*

In this paper we relate bilingual lexicography and contrastive construction grammar via bilingual constructicography, addressing the treatment of constructional and phraseological information in bilingual lexicography. After a discussion of the different notions of equivalence in lexicography and constructicography, we compare two online resources: the Swedish constructicon and the Icelandic-Swedish part of *Islex.se*. It is shown that although *Islex.se* contains a wealth of constructional information, this does not readily carry over across languages. A creative user can find correlations between constructions in Icelandic and Swedish, but these connections are indirect and somewhat *ad hoc*. Explicit links between the two languages mainly concern lemmas. In an online resource, it is clearly possible to also include systematic comparisons of phraseological information, and we argue that making use of constructionist resources such as constructicons would be helpful in this regard.

## 1. Inledning

Inom modern lexikografi, inte minst e-lexikografi, märks både ökade möjligheter och ökad ambition att inkorporera olika former av konstruktionell information, såsom valens, kollokationer, idiom och språkbruksexempel. Även inom grammatikforskningen ökar intresset för konstruktioner, och konstruktionsgrammatik växer fram som en allt starkare lingvistisk tradition. Särskilt relevant från ett lexikografiskt perspektiv är utvecklingen av s.k. konstruktikon, en typ av språkresurs som kanske enklast kan beskrivas som: lexikon, fast för grammatiska konstruktioner i stället för ord (se t.ex. Fillmore, Lee-Goldman & Rhomieux 2012, Lyngfelt, Borin m.fl. u.u.).

Till skillnad från lexikografiska resurser är konstruktikon inte bundna till lexikala ingångar och kan därmed täcka delvis annan information; exempelvis kan de lättare hantera språklig produktivitet (t.ex. Lyngfelt & Sköldberg 2013). Samtidigt innehåller de mycket information som lika gärna kan stå i en ordbok, och bör därför kunna utgöra en värdefull informationskälla för konstruktionellt och/eller fraseologiskt orienterad lexikografi.

Kopplingen mellan konstruktionsgrammatik och lexikografi har hittills i första hand prövats i den enspråkiga lexikografien (Lyngfelt & Sköldberg 2013, Blensienius 2017). I detta bidrag inriktar vi oss på hur konstruktionsgrammatik kan

bidra även till tvåspråkig lexikografi. Detta gör vi med utgångspunkt från två web-baserade resurser: *Svenskt konstruktikon* (Lyngfelt m.fl. 2014) och den isländsk-svenska delen i *Islex*.

Artikeln är disponerad enligt följande: i avsnitt 2 introduceras konstruktioner och konstruktikon. I avsnitt 3 jämförs vilken roll ekvivalens spelar inom lexikografi, konstruktionsgrammatik och konstruktikonutveckling, illustrerat i avsnitt 4 av en mer konkret jämförelse mellan några svenska och isländska konstruktioner. I avsnitt 5 granskas konstruktionsekvivalens såsom den kan utläsas i *Islex* och i avsnitt 6 hävdar vi nyttan av en resurs som konstruktikon inom den tvåspråkiga lexikografin.

## 2. Konstruktioner, konstruktikon och deras relevans för lexikografi

Inom konstruktionsgrammatik tillämpar man inte en traditionell uppdelning av språket i en grammatik och ett lexikon, utan betraktar hela lexikogrammatiken som ett nätverk av konstruktioner – ett konstruktikon. Konstruktioner definieras då typiskt som ”conventional, learned form-function pairings at varying levels of abstraction and complexity” (Goldberg 2013:17), vilket inkluderar generella grammatiska mönster, specifika lexikala uttryck och allt däremellan.<sup>1</sup> I likhet med ’grammatik’ och ’lexikon’ har det teoretiska konceptet ’konstruktikon’ också en praktisk motsvarighet i form av samlade beskrivningar av konstruktionsmönster, t.ex. *Svenskt konstruktikon* som är en fritt tillgänglig webbdatabas med svenska konstruktionsbeskrivningar (Lyngfelt m.fl. 2014). Utvecklingen av sådana konstruktikon bygger på en kombination av konstruktionsgrammatik och lexikografi, och kallas därför *konstruktikografi* (Lyngfelt, Borin m.fl. u.u.).

En grupp konstruktioner som har ägnats särskild uppmärksamhet i *Svenskt konstruktikon* är sådana som är partiellt schematiska, dvs. har både fasta och variabla led, t.ex. *FÖR* ANTAL TID *SEDAN* (ex. *för två veckor sedan, för många år sedan*) och *X FÖR X* (ex. *bit för bit, dag för dag*).<sup>2</sup> Sådana mönster, med både lexikala och grammatiska drag, är perifera – och i vissa fall svårhanterade – för både grammatik och ordböcker, medan de i ett konstruktikon är lika centrala som andra strukturer. En annan prioriterad, delvis överlappande grupp är mönster som visat sig problematiska för andraspråksinlärning.

Huruvida grammatiska konstruktioner är lexikografiskt relevanta beror på deras koppling till lexikala enheter, s.k. *kollostruktionella* relationer (t.ex. Stefanowitsch 2013). Typiskt relevanta är bland annat:

<sup>1</sup> Detta konstruktionsbegrepp är alltså betydligt vidare än vad som brukar tillämpas inom lexikografi. I exempelvis Oxford English Dictionary (*construction*, 5b) likställs konstruktioner snarast med verbs valensmönster.

<sup>2</sup> Konstruktionsbenämningarna är hämtade ur *Svenskt konstruktikon* och markeras här med kapitäl.

- konstruktioner som innehåller fasta innehållsord, t.ex. *FÖR X SKULL* (*för din skull, för spänningens skull*) och den ovannämnda *FÖR AN-TAL TID SEDAN*; valensmönster och idiom hör till denna grupp
- konstruktioner med lexikala led som kännetecknas just av att ingå i konstruktionen, t.ex. *BÅDE XP1 OCH XP2* och *DELS X DELS Y*
- konstruktioner som innehåller variabla led med tydliga restriktioner, t.ex. *EN TIDS AKTIVITET* (*tio minuters paus, tre veckors semester*) eller *REFLEXIV RESULTATIV* (*äta sig mätt, supa sig full, skrika sig hes*), detta därför att de aktuella variablerna ofta realiseras av vissa centrala/frekventa medlemmar av respektive kategori.

Den tredje typen visar att även fullt schematiska konstruktioner (som inte har några lexikalt fasta led) kan vara lexikografiskt relevanta. Å andra sidan finns det konstruktioner med fasta led som *inte* är det, nämligen när orden är så vanliga i förhållande till konstruktionen att den inte hör till ordens mer framträdande användningsområden, t.ex. *EN NP PÅ ENHET* (*en gädda på två kilo, en promenad på fem kilometer*). Det väsentliga i sammanhanget är alltså att orden associeras med konstruktionen, inte tvärtom.

### 3. Ekvivalens i lexikografi och konstruktikografi

Den lexikografiska beskrivning som presenteras i en tvåspråkig ordbok bygger på en kontrastiv analys som syftar till att upprätta ekvivalensrelationer mellan enheter i de inblandade språken. Trots att ekvivalens i sig är ett komplicerat begrepp är det onekligen ekvivalensrelationer och ekvivalenter som utgör kärnan i den tvåspråkiga lexikografien. Framför allt är det lemmats semantiska innehåll och dess egenskaper på systemnivå respektive textnivå som har stötts och blötts i den lexikografiska och metalexikografiska litteraturen om ekvivalens (för en översikt se Adamska-Salaciak 2010).

#### 3.1. Fraseologi och ekvivalens i lexikografien

Hur de olika fraseologiska enheter som presenteras i en ordboksartikel skall behandlas i målspråket har inte rönt samma uppmärksamhet som de ekvivalensrelationer som redovisas mellan lemmat och ekvivalenterna (se dock Hannesdóttir & Jónsson 2001). I den tvåspråkiga e-ordboken kan hela ordbokstexten för båda språken göras sökbar och åtkomlig. Där kan lemmats kombinatoriska egenskaper redovisas på ett helt annat sätt än tidigare liksom de ekvivalensrelationer som upprättas mellan de båda språkens enheter. Trots det utgår den teoretiska diskussionen inom den tvåspråkiga lexikografien huvudsakligen från den tryckta, monodirektionella och monofunktionella ordboken, dvs. en ordbok där de båda språkens roller

är statiska som källspråk respektive målspråk och där ett av dessa språk är användarnas L1 och det andra L2 (Svensén 2009:202, jfr Hannesdóttir 2014).

Den traditionella uppdelningen i språkliga beskrivningsnivåer där den grammatiska och den lexikografiska nivån har hållits åtskilda är, i ett brukarperspektiv, inte längre lika relevant. I den tvåspråkiga e-ordboken där valensrelationer, selektionsrestriktioner och kollokationella mönster för vart och ett av språken kan redovisas och sökas aktualiseras frågan om vilka egenskaper det är som ska presenteras och hur. Förutom att spegla lemmats semantiska, kognitiva eller encyklopediska dimensioner har ekvivalensbegreppet utökats med strukturella, funktionella och kontextuella dimensioner.

### 3.2. Flerspråkig konstruktikografi

Till skillnad från tvåspråkig lexikografi är kontrastiv konstruktionsgrammatik vanligtvis inte inriktad på ekvivalensrelationer. I stället handlar det främst om detaljerade fallstudier där man undersöker likheter och skillnader mellan ett fåtal konstruktioner i två eller flera språk (se t.ex. bidragen i Boas 2010). Däremot går flerspråkig konstruktikografi i hög grad ut på att etablera korrespondenser mellan motsvarande konstruktioner i olika språk och har i det avseendet alltså mer gemensamt med lexikografi än med konstruktionsgrammatik.

Etablerandet av konstruktionsmotsvarigheter inbegriper alla de komplikationer som rör lexikal ekvivalens – betydelsenyanser, stil, distribution osv. – samt tillför ytterligare en dimension: form. Jämförelsen mellan en konstruktion i källspråket och eventuella motsvarigheter i målspråket måste beakta inte bara deras betydelse/funktion utan också deras grammatiska struktur. Matchningen mellan käll- och målkonstruktion omfattar därmed åtminstone fyra steg (Lyngfelt, Torrent m.fl. u.u.):<sup>3</sup>

- 1) *Motsvarighet*: Har målspråket en konstruktion, eller en grupp konstruktioner, som (mer eller mindre) kan sägas motsvara källkonstruktionen?
- 2) *Funktionell ekvivalens*: Är källkonstruktionen och dess närmaste motsvarighet i målspråket funktionellt ekvivalenta eller finns det funktionella skillnader mellan dem? Kan konstruktionerna matchas en-till-en eller delas den aktuella domänen in på olika sätt i de båda språken?
- 3) *Formell likhet*: Har käll- och målkonstruktion samma/likartad struktur eller uppträder formella skillnader mellan dem? (Här kan man i regel bortse från rent lexikala skillnader.)
- 4) *Egenart*: Är eventuella formella skillnader utmärkande för konstruk-

<sup>3</sup> På grund av denna ökade komplexitet används termen *ekvivalens* endast om den funktionella dimensionen (vilket i praktiken är vad som avses även inom lexikografi), *likhet* om formell struktur och *motsvarighet* om relationen som helhet.

tionerna i fråga eller härrör de från mer generella skillnader mellan språken? Strukturella drag som ordföljdsmonster, ev. artikelbruk, kongruensböjning m.m. kan slå igenom i ett stort antal konstruktioner utan att vara prominenta egenskaper hos mer än ett fåtal av dem. Exempelvis specificeras grundläggande nominalfrasstruktur i generella nominalfraskonstruktioner och behöver kanske inte specificeras i varje konstruktion som innehåller en nominalfras.

Steg 1–2 är relevanta även för tvåspråkig lexikografi, medan steg 3–4 normalt inte är det – åtminstone inte än så länge. Om och när man inom tvåspråkig lexikografi börjar länka inte bara lexikala enheter som sådana utan även deras valenskonstruktioner m.m. aktualiseras också den formella strukturen. Därmed väcks samtidigt frågan om hur denna struktur ska representeras. En formalism som är utvecklad för källspråket kanske inte är fullt ut tillämplig på målspråket, kategoribeteckningar kanske inte fullt likvärdiga osv. Ju större skillnad mellan språken, desto svårare att utveckla ett gemensamt – och helst användarvänligt – beskrivningsformat.

#### 4. Svenska konstruktioner och isländska

På grund av svenskans och isländskans nära släktskap är det rimligt att anta att det finns systematiska överensstämmelser i de båda språkens sätt att kodifiera ett visst semantiskt innehåll i vissa syntagmatiska strukturer eller konstruktioner. Det borde därför gå att identifiera isländska motsvarigheter till några av de närmare 400 svenska konstruktioner som beskrivs i *Svenskt konstruktikon*.

Redan vid en ytlig genomgång av konstruktikonet framträder konstruktioner som kan ha etablerade formella och funktionella motsvarigheter i isländskan. Bland dessa finns konstruktioner av alla de slag som beskrivs i avsnitt 2 ovan. Några exempel på konstruktioner med fasta lexikala led är: *ANTINGEN X ELLER Y* (*annað hvort X eða Y*), *BÅDE XP1 OCH XP2* (*bæði X og Y*), *DET ÄR SÅ X ATT Y* (*það er svo X að Y*) och *HA DET ADVP* (*hafa það AdvP*). Även möjliga isländska konstruktioner som innehåller variabla led med tydliga restriktioner är lätta att identifiera utifrån de svenska konstruktionerna: *REFLEXIV\_RESULTATIV* (*borða sig saddan*), *EN TIDS AKTIVITET* (*þriggja daga útilega; fimm vikna sumarfrí*).

Några av de möjliga isländska konstruktionsmotsvarigheterna uppvisar vid en närmare granskning vissa formella avvikelser från den svenska konstruktionen. Det gäller exempelvis *BARA.DESIDERATIV* (*bara jag hade mera pengar*) och den isländska varianten *bara að ég ætti meiri peninga*, där den syntaktiska strukturen skiljer sig från den svenska konstruktionens eftersom det inledande adverbet följs av subjunktionen *að*.

Trots en formell överensstämmelse mellan en svensk konstruktion och en isländsk struktur kan det finnas semantiska skillnader som gör att de snarare är fal-



ska vänner än konstruktionsmotsvarigheter. Ett sådant fall är konstruktionsparet I ADJ:ASTE *LAGET* (*i högsta laget*) och I ADJ:ASTA *LAGI* (*i hæsta lagi*). Substantivets form skiljer sig åt: bestämd form i svenskan och obestämd form i isländskan. Den avgörande skillnaden är dock semantisk och funktionell. Den svenska konstruktionen betyder att det som är ’i Adj:aste laget’ är på fel sida om en tänkt gräns – *i högsta laget* är för mycket, *i minsta laget* är för lite. Den isländska konstruktionen betyder däremot att det som är ’i Adj:asta lagi’ *inte* överskrider en viss gräns. Denna gräns kan inte, som i svenskan, underförstås utan måste uttryckas i form av ett preciserande komplement: *i hæsta lagi tíu metrar* betyder att det åsyftade är kortare än tio meter. Parallellen är ändå inte fullständig – den isländska frasens betydelse varierar beroende på det enskilda adjektivet och det finns även vissa semantiska restriktioner för vilka adjektiv som kan ingå i konstruktionen.

## 5. Konstruktioner i *Islex*

Hur ser det då ut i den flerspråkiga isländska nätordboken *Islex*? Där är isländskan primärspråk och de andra språken lexikografiskt sekundära i förhållande till den (Hannesdóttir 2014, 2015). *Islex* innehåller närmare 50 000 lemmor och omkring 14 000 isländska fraseologiska enheter (Úlfarsdóttir 2013). Dessa enheter utgörs av mer eller mindre lexikaliserade ordförbindelser av olika slag. De är avsedda att illustrera lemmors konstruktionspreferenser och selektionsrestriktioner, dess semantiska och syntaktiska valens och även metaforisk och annan oegentlig användning i form av idiom och ett och annat ordspråk. Förutom dessa fasta fraser innehåller ordboken omkring 30 000 språkprov. Såväl fraserna som språkproven är översatta till samtliga sekundärspråk, däribland svenska.<sup>4</sup>

De lexikografiskt motiverade enheterna i *Islex* är naturligtvis inte rakt av jämförbara med strukturer definierade genom konstruktionsgrammatik och den lexikografiska beskrivningen skiljer sig från den konstruktivlexikografiska. Eftersom åtkomststrukturen i *Islex* är mycket generös, är möjliga isländska och svenska konstruktioner åtkomliga genom fritextsökningar på de lexikala enheter som de består av. Med sökfunktionen ”avancerad sökning” i kategorin ”all artikeltext” på *Islex*-sidan kan fasta eller typiska innehållsord sökas både i det isländska materialet och det svenska.

Instansieringar av flera av de svenska konstruktioner som nämnts ovan påträffas i *Islex*, antingen i översättningar av språkprov eller som motsvarigheter till fraseologiska enheter av något slag. Förekomsterna är dock slumpmässiga. Exempelvis ger en sökning på strängen ”ha det” 33 träffar. Fyra av dem inleder instansieringar av konstruktionen HA DET ADVP: ”<ha det bra> på ålderns höst” s.v. *elli*, ”<ha det tråkigt> hela dagen lång” s.v. *dagur*, ”<ha det riktigt trevligt> på födelsedagskalaset” s.v. *afmæli* och ”hon bestämde sig för att ha det trevligt på kväl-

<sup>4</sup> I bidirektionella tvåspråkiga e-ordböcker är termen *sekundärspråk* lämpligare än *mål-språk* (se Hannesdóttir 2014).

len” s.v. *huggulegur*. Av de isländska enheter som genererar dessa svenska översättningar är det bara två som kan sägas innehålla en isländsk konstruktion HAFI ÞAÐ ADVP. Frasen ”<hafa það náðugt> í ellinni” är avsedd som en typisk kontext för lemmat *elli* och ”hún ákvað að hafa það huggulegt um kvöldið” ger exempel på adjektivet *huggulegur* i adverbial användning. En sökning på ”hafa það” ger nio träffar, varav fem är exempel på konstruktionen HAFI ÞAÐ ADVP. Fyra av dessa fem genererar konstruktionsekvivalenten HA DET ADVP medan frasen ”hafa það náðugt” istället för det formellt och funktionellt ekvivalenta ”ha det bra” översätts med det funktionellt ekvivalenta ”ta det lugnt, koppla av, softa”.

Även konstruktioner av typen EN TIDS AKTIVITET och REFLEXIV RESULTATIV kan sökas i *Islex* då uppsättningen möjliga innehållsord är relativt begränsad. En sökning på ”\_ dagars” ger fem träffar varav fyra är förekomster av EN TIDS AKTIVITET: ”tre dagars landssorg”, ”femton dagars friluftsliv”, ”tre dagars camping” och ”tio dagars vandring”. Samtliga dessa är översättningar av motsvarande isländsk fras som inkluderas i ett längre språkprov: ”þriggja daga þjóðarsorg”, ”fimmtán daga útivist”, ”þriggja daga útilegu” och ”tjú daga göngu”. Sökningar på genitivformen av andra tidsuttryck (*vikna*, *mánaða*, *ára*) ger liknande resultat. Också konstruktionen REFLEXIV RESULTATIV påträffas i ordboken, ”han drack sig svinfull” som översättning till just ”hann drakk sig svínfullan” och ”äta sig mätt” till ”borða sig saddan”.

Sökningar av detta slag efter flera konstruktionsmönster ger liknande resultat. De finns i *Islex*, än i det ena språket, än i det andra och i vissa fall i båda språken. Någon medveten strävan efter att redovisa en svensk konstruktionsekvivalent där så är möjligt avspeglas emellertid inte i ordboken.

## 6. Lexikografisk nytta?

Huruvida konstruktioner har en plats i en enspråkig ordbok har diskuterats i olika sammanhang (se Lyngfelt & Sköldberg 2013, Blensenius 2017). Men vad kan då konstruktionsgrammatik bidra med i den tvåspråkiga lexikografien?

Det finns ingen språklig nivå som inte är relevant i ett lexikografiskt perspektiv, vare sig det enspråkiga eller tvåspråkiga. Perspektivet på lexikografins centrala objekt, ordet och dess egenskaper, har vidgats till att omfatta egenskaper som tidigare har hänförs till språkets strukturella nivåer, syntaxen och grammatiken. Däremot är det naturligt att man begränsar sig till sådana egenskaper som är lexikaliskt relevanta, dvs. har med det aktuella lemmat att göra.

Den fraseologi som presenteras i en ordbok är lexikaliskt motiverad och det är frasens relation till uppslagsenheten som står i fokus – inte själva frasens inre struktur och semantiska egenskaper. Att ge en rättvisande karakterisering av konstruktionen som sådan är snarare en uppgift för ett konstruktikon. Men om ekvivalenta lexikala enheter i två språk förknippas med motsvarande konstruktioner kan det finnas ett värde i att systematiskt relatera dessa till varandra, och helst

belysa relevanta likheter och skillnader. I en elektronisk ordbok är detta inte bara önskvärt utan också praktiskt genomförbart.

I *Islex* finns det gott om exempel på att den lexikografiska analysen av en isländsk fras har resulterat i helt idiomatiska och adekvata men ändå arbiträra översättningar. Med tillgång till en samlad beskrivning av grammatiska konstruktioner, dvs ett konstruktikon, hade de ekvivalensrelationer som i *Islex* upprättats mellan isländska fraser och svenska i större utsträckning kunnat illustrera också konstruktionella överensstämmelser mellan språken.

## Litteratur

### Ordböcker och liknande

*Islex*. <<http://islex.se>>

Oxford English Dictionary. <<http://www.oed.com/>>

*Svenskt konstruktikon*. <<https://spraakbanken.gu.se/karp/konstruktikon>>

### Annan litteratur

Adamska-Salaciak, Arleta (2010): Examining Equivalence. I: *International Journal of Lexicography*, Volume 23:4, 387–409.

Blensenius, Kristian (2017): Valens- eller konstruktionsuppgifter: redovisningar av aspektuella konstruktioner i några svenska ordböcker och en dansk. I: *LexicoNordica* 24, 153–172.

Boas, Hans (red.) (2010): *Contrastive Studies in Construction Grammar*. Amsterdam: John Benjamins.

Fillmore, Charles J., Russell Lee-Goldman & Russell Rhomieux (2012): The FrameNet Constructicon. I: H. C. Boas & I. A. Sag (red.): *Sign-Based Construction Grammar*. Stanford: CSLI, 309–372.

Goldberg, Adele E. (2013): Constructionist approaches. I: Th. Hoffmann & G. Trousdale (red.): *The Oxford handbook of construction grammar*. Oxford & New York: Oxford University Press, 15–31.

Hannedóttir, Anna Helga (2014): Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi. I: *Nordiske studier i leksikografi* 12. Oslo: Novus forlag, 193–211.

Hannedóttir, Anna Helga (2015): What is a target language in an electronic dictionary? I: Kosem, I. 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.

- Hannesdóttir, Anna Helga & Jón Hilmar Jónsson (2001): Að hafa í sig og á. Is­ländsk fraseologi i ett isländskt-svenskt perspektiv. I: *LexicoNordica* 8, 67–91.
- Lyngfelt, Benjamin, Lars Borin, Linnéa Bäckström, Markus Forsberg, Leif-Jöran Olsson, Julia Prentice, Rudolf Rydstedt, Emma Sköldberg, Sofia Tingsell & Jonatan Uppström (2014): Ett svenskt konstruktikon. Grammatik möter lexikon. I: *Svenskans beskrivning* 33. Helsingfors, 268–279.
- Lyngfelt, Benjamin, Lars Borin, Kyoko Hirose Ohara & Tiago Timponi Torrent (red.) (u.u.): *Constructicography: Constructicon development across languages*. Amsterdam: John Benjamins.
- Lyngfelt, Benjamin, Tiago Timponi Torrent, Adrieli Laviola, Linnéa Bäckström, Anna Helga Hannesdóttir & Ely Edison da Silva Matos (u.u.): Aligning constructicons across languages: A trilingual comparison between English, Swedish, and Brazilian Portuguese. I: Lyngfelt, Borin m.fl. (u.u.).
- Lyngfelt, Benjamin & Emma Sköldberg (2013): Lexikon och konstruktikon – ett konstruktionsgrammatiskt perspektiv på lexikografi. I: *LexicoNordica* 20, 75–91.
- Stefanowitsch, Anatol (2013): Collostructional Analysis. I: Th. Hoffmann & G. Trousdale (red.), *The Oxford handbook of construction grammar*. Oxford & New York: Oxford University Press, 290–306.
- Svensén, Bo (2009): *A Handbook of Lexicography*. Cambridge: Cambridge University Press.
- Úlfarsdóttir, Þórdís (2013): ISLEX – norræn margmála orðabók. I: *Orð og tunga* 15, 41–71.

Anna Helga Hannesdóttir  
professor  
anna.hannesdottir@svenska.gu.se

Benjamin Lyngfelt  
professor  
benjamin.lyngfelt@svenska.gu.se

Göteborgs universitet  
Box 200  
SE-40530 Göteborg

# Ord i samtaler – findes de?

*Carsten Hansen & Jakob Steensig*

This is a presentation of the lexicographic part of *Projekt samtalesprog* (“Project Talk-in-interaction”). This project aims at describing lexemes in naturally occurring interactions, on the basis of how sound strings are actually produced and used in interaction. The authors have access to two corpora of Danish talk-in-interaction and the paper illustrates how a search for sound strings that are possible lexemes can be carried out using these corpora and the available search options. The search begins with an intuition, that there is a lexeme in spoken Danish, *vaffen*, meaning something like ‘which’, that does not occur in written corpora and has, consequently, not been described in written language based lexicography. A pilot study is devised, and this study provides evidence for two lexemes, *vaffen* and *vafnén*, which are used in different environments and with a possible difference in meaning. The paper argues that research that takes talk-in-interaction seriously as an object of study in its own right will be able to yield results, which may lead to a new understanding of lexeme boundaries and lexeme divisions. New search machines, based on sounds, not strings of characters, need to be developed in order for such an endeavour to gain ground.

## 1. Projekt samtalesprog

Vores baggrund er to projekter i Danmark der udforsker alment samtalesprog, ét med en grammatisk tilgang, *DanTIN* (Danish Talk-In-Interaction), og ét med en leksikografisk tilgang, *Ordbog over Dansk Talesprog* (ODT). Projekterne deler dogmet om samtaleelementernes interaktionelle forankring, og de har derudover teoretiske og metodiske ligheder såsom korpusfundering og multimodalitet. Begge vil bryde med den skriftsprogsbaserede sprogbeskrivelses hegemoni og via en nytænkning ikke bare af det leksikografiske lemmabegreb (ODT) men af selve det skriftsprogsbaserede leksembegreb bidrage til en konsekvent lydaseret tilgang til ytringens elementer. Metodisk rejser der sig mindst to spørgsmål som en leksikografi og en grammatik der vil tage samtalesproget alvorligt, må forholde sig til: (1) Hvor går grænserne for et leksem? Er fx *jojo* eller *nåja* ét eller to leksemer? Man kan ikke bruge mellemrum som kriterium, og af metodiske årsager løser det ikke problemet at spørge sprogbrugerne. (2) Er en streng af lyd udtalt med to forskellige intonationsmønstre det samme leksem, fx *okay* udtalt med henholdsvis stigende og faldende intonation? Da vi endnu ikke har svaret på disse spørgsmål, tager vi i vores undersøgelser udgangspunkt i *lydstreng*, dvs. sammenhængende udtalte lydsekvenser med mulig interaktionel forankring.

I artiklen demonstrerer vi med udgangspunkt i to eksempler fra henholdsvis

DanTIN's korpus (*AULing*) og fra ODT's korpus (*Lanchart*) hvordan vi teoretisk og metodisk nytænker den traditionelle opfattelse af arbejdsdelingen mellem grammatik og ordbog, hvor grammatikken fokuserer på det generelle og ordbogen det specielle.

Foredraget på konferencen og denne artikel er et udtryk for de igangværende bestræbelser på at integrere DanTIN og ODT i et nyt og innovativt samarbejde: *Projekt samtalesprog*.

## 2. Korpora

DanTIN's korpus består af ca. 50 timers audio- og videooptagelser af autentiske samtaler fra hverdagslivet forskellige steder i Danmark fra *AULing* (Aarhus Universitets samtalekorpus) og *Samtalebanken* (u.å.). Ca. 20 timer herfra er på nuværende tidspunkt udskrevet efter samme samtaleanalytiske konventioner (se nedenfor), som betoner nøjagtighed i gengivelsen af det lydlike, uden dog at være fonetiske (*Samtalegrammatik.dk* 2017). Samtaledeltagerne er tilfældigt udvalgt og ikke forsynet med sociolingvistiske oplysninger. Der findes heller ikke nogen opgørelse over antal tokens i dette korpus.

ODT's korpus består af ca. 1700 timers audiooptagelser fra *Lanchart*-korpuset som primært består af sociolingvistiske interviews og gruppesamtaler fra 1970'erne frem til i dag fra syv forskellige repræsentative områder i Danmark. Alle data er udskrevet ortografisk efter Retskrivningsordbogens regelsæt. Samtaledeltagerne er tilfældigt udvalgte interviewere og "informanter" – udvalgt efter de sociolingvistiske kriterier køn, alder, social baggrund, tid og sted. I udskrevet form tæller korpuset ca. 7 mio tokens, det bliver løbende udvidet med flere samtaler.

## 3. Et eksempel: *vaffen* eller "hvad for en"

Her vil vi vise hvordan en undersøgelse af en lydstreng i *Projekt samtalesprog* kan foregå. Der er tale om en indledende afsøgning som vi har foretaget i de korpora vi har redegjort for ovenfor.

### 3.1. En intuition gøres søgbar

Vi havde bemærket i vores egne erfaringer med samtaler at der findes en lydstreng som vi i vores indledende diskussioner refererede til som *vaffen* og som vi udtalte ['vafɸ]. Denne lydstreng var ikke beskrevet før, og det fandt vi interessant fordi der kunne være tale om et leksem som opførte sig anderledes end noget tilsvarende i skriftsproget.

Søgefaciliteterne til vores korpora giver indtil videre kun mulighed for at ope-

rere på skriftsprogsstrenge, og vi måtte derfor i vores søgning “oversætte” lydstrengen til skriftsprog. For søgninger i ODT’s korpus (se ovenfor) oversatte vi lydstrengen til *hvad for en*. For søgninger i samtalekorpus (AULing, se ovenfor) var det lidt sværere. Vi måtte begynde med en alt for bred søgning, nemlig på *hva*. Den gav brugbare resultater i form af *hva for en*, *hvar for en*,<sup>1</sup> *hvaffor en* (se Jørgensen 2015 om samtaleprogets forskellige ord for skriftsproget “hvad”), men den resulterede også i en overgenerering i form af en masse strenge vi ikke var interesseret i. Vores første søgning i de to korpora gav os ca. 600 forekomster som vi så gik i gang med at undersøge. Vi vil her redegøre for vores fremgangsmåde, gengive nogle af vores eksempler og bruge dem som illustration af vores metode.

### 3.2. Arbejdet med eksempler

Man kan ikke arbejde med samtaleprogseksempler uden at afspille dem (mange gange) og uden at udskrive dem. Heldigvis er udskrifterne af samtalerne i vores korpora linket til rådata på en måde der gør det nemt at afspille de lydstrengene vi havde fundet.

Da vi insisterer på at behandle en lydstreng i dens interaktionelle kontekst, er det ikke nok at udskrive strengen alene, vi skal også have dens samtalekontekst med. Derfor har vi lavet en ny udskrift af de lydstrengene vi har analyseret, hvor man kan se strengen i dens præcise samtalekontekst. Dertil har vi anvendt samtaleanalysens udskriftssystem (Jefferson 2004, *Samtalegrammatik.dk* 2017). Som udgangspunkt kan det være godt at vide at talen er gengivet “halv-ortografisk”, dvs. vi bruger almindelig retstavning men bruger en særlig stavning til strenge som vi vurderer findes i forskellige former. Kursiverede passager i udskrifterne er beskrivelser af nonverbale handlinger, og deres placering i forhold til talen vises med krøllede parenteser. Figur 1 viser et af de første eksempler vi undersøgte.

**Uddrag 1: ‘vaffor’n bunke’ [board-game-coffee3]**

```

01 SVE: det virker (jo) meget uover>skueligt det-< (0.3)
02 ka det oss godt være lige her i †st*arten*→
03 (0.5)
04 LIV: (okay okay)
05 SVE: men øh
06 {(1.4) }
   { Svend rejser sig }
07 SVE: {·hhhh vaffor’n bunke var nummer to du fik af mig>}
   {Svend peger på et punkt mellem to bunker lige foran TOR }
08 TOR: {den der> }
   {Tor peger på én bunke}
09 (2.8)
10 LIV: °o†ka|y→°
11 TOR: [a:h (m') der var en rækkefølge i det>
```

Figur 1: Uddrag 1: *vaffor ’n bunke*

<sup>1</sup> ‘hvar’ er DanTINs måde at udskrive [va] når det bruges som reparatur-indleder (Jørgensen 2015).

Dette eksempel vil vi nu bruge til at undersøge lydstrengen *vaffor 'n* i linje 7.

### 3.2.1. Et attributivt *vaffen* der udpeger én af to

Fire unge mennesker er i uddraget i figur 1 ved at spille et brætspil. Vi kommer ind i samtalen hvor der er blevet lagt nogle kort-bunker ud på bordet. Det er Svend der kender spillet i forvejen, og han er ved at forklare reglerne samtidig med at kort og figurer bliver delt ud. Foran Tor ligger der to bunker, og det er dem Svend forholder sig til i linje 6–7, hvor han stiller et spørgsmål til Tor: *vaffor 'n bunke var nummer to du fik af mig*∇. Tor behandler dette som en anmodning om at udpege én af de to bunker han har fået, det gør han i linje 8, hvilket Svend kvitterer for i linje 10, hvorved han også viser at Tor har gjort hvad han bad ham om.

Den lydstreng vi er interesseret i, *vaffor 'n*, udtales her [ˈvafɫm]. At den slutter på [m], skyldes assimilationen med [b] i *bunke*, og nasaleringen af [ɫ] skyldes påvirkning fra det efterfølgende [m]. Derudover kan vi konstatere at det har tryk (på første stavelse), at det udtales med ”mørkt a” [ɑ] og at der er en tydelig vokallyd [ɫ], der svarer til den vokallyd man finder i ordet *for* [fɫ].

Lydstrengen indleder et spørgsmål, den anmoder om at udpege én af to, og den står attributivt foran et substantiv. Vi fandt andre eksempler med lydstrengene der mindede om denne i både form og funktion. De blev brugt i spørgsmål der anmodede om at få udpeget én af to ting, og de stod attributivt. Nogle af dem havde en anden a-lyd [æ], nogle havde en svækket vokallyd i anden stavelse, men ellers mindede de i fonetisk form om vores [ˈvafɫm] fra første uddrag.

### 3.2.2. Et alene-stående *vafnén*

I vores søgen faldt vi over en anden udtale af *vaffen*, én vi ikke havde forudset at vi ville finde. Figur 2 viser et eksempel på denne.

**Uddrag 2 [odder2ny-06-KBJ\_vaf-nen]**

01	A:	sån lige <u>umiddelbart</u> .
02		(0.6)
03	?:	·mt=·hhh
04		(1.4)
05	B?:	°°(ha)°°=
06	B:	=v <u>a</u> =fen:: v <u>a</u> =fenen syn's du var p <u>æ</u> nest,
07		(1.1)
08	B:	°af de tre°
09		(0.4)
10:	A?:	°°h::°°

Figur 2: Uddrag 2: *va=fen:: va=fenen*



Uddraget kommer fra et af *Lancharts* sociolingvistiske interviews. Intervieweren (B) har afspillet nogle sprogrøver for A og beder ham udpege den af tre prøver han bedst kan lide. I uddrag 2 står lydstrengen *v<sub>a</sub>=f<sub>e</sub>nen* alene på sin syntaktiske plads. Vi har ikke taget svaret med her, men det fremgår at A forstår hvad B snakker om.

Udtalen af lydstrengen i linje 6 er ['væfne:ʔn], det har altså tryk på første stavelse, et ”lyst a” efterfulgt af en n-lyd, og så et tydeligt *en*, med forlængelse og stød, noget som normalt kendetegner talordet *en* (og ikke artiklen *en*). Før dette kommer strengen ['væfn:], altså noget der kan høres som begyndelsen på vores lydstreng, med en forlængelse af n-lyden. Forlængelsen fremstår som en søgen efter den rette formulering hvor lydstrengen ['væfne:ʔn] så er det resultat taleren når frem til.

Den lydlige forskel på denne variant og den vi så i uddrag 1, er for det første at der er en tydelig e-vokal med stød i den sidste del af ordet. Og, måske endnu mere bemærkelsesværdigt, at der kommer et [n] mellem *vaf* og *en*.

Vi fandt andre eksempler på det samme, altså alene-stående (ikke-attributive) lydstrengene der indeholder en n-lyd før “en”. Figur 3 viser endnu et eksempel, også fra *Lanchart*, men fra den lidt mere interaktionelle del da uddraget stammer fra en gruppesamtale.

**uddrag 3 [clarin3gl-10\_AOE\_176]**

01	A:	det der <u>det</u> der var °altså° ikk så godt
02	A:	f <u>det</u> for svæ <u>rt</u> .
03		(0.3)
04	A?:	(>det ve je' ikk<) j' kom b <u>are</u>
05	A?:	te' [å (snakke) ( )
06	B:	[ <u>va=fne</u> ɿ

Figur 3: Uddrag 3: *v<sub>a</sub>=f<sub>e</sub>nen*

A er ved at forklare noget som står uklart for B. Det får B til i linje 6 at igangsætte reparatur (Schegloff, Jefferson & Sacks 1977; Garly 2016), dvs. bede om at et element i A's ytring bliver specificeret. Igen har vi her en variant med et tydeligt [e], med stød, og med et [n] før dette, ['væfne:ʔn]. Og igen står lydstrengen for sig selv, den står her helt alene i sin tur hvor den udpeger en problemkilde i noget den forudgående taler har sagt.

### 3.3. Foreløbige konklusioner

Vi har nu undersøgt tilstrækkeligt til at vi kan formulere to hypoteser om lydstrengen “*vaff<sub>e</sub>n*” som vi satte os for at undersøge. Den videre søgen vil vise om hypoteserne holder.

Den første hypotese er at der findes to lydstrengene, med tydeligt forskellige realiseringer og tilhørende distributive og funktionelle egenskaber:

Den ene lydstreng er *vaffen*. Den kan have lyst eller mørkt a i første stavelse og enten [ʌ] eller ingen vokal (kun nasal) i anden stavelse. Lydstrengen bruges attributivt til at anmode om at få udpeget én af to. Den anden lydstreng, *vafnén*, indeholder en ekstra n-lyd, [ˈvæfn̩eːn], den har lyst a [æ] i første stavelse og tydelig [eː] i anden stavelse. Den bruges “substantivisk”, altså som alene-stående element (ikke attributivt), og den lader ikke til at være bundet til at udpege én af to.

Den anden hypotese er at det lader til at samtaledeltagerne behandler både *vaffen* og *vafnén* som ét leksem. Strengene *vaffen* og *vafnén* synes at fungere som andre leksemer der kan stå attributivt eller alene, hvilket jo er typisk for fx mange spørgeord. Men i samtalesproget findes der øjensynlig en formel specialisering hvor det ene ord, *vaffen*, er attributivt, og det andet, *vafnén*, er alene-stående.

I en leksikografi som vil tage samtalesproget alvorligt, skal vi altså kunne finde leksemerne *vaffen* og *vafnén*. Der er tale om leksemer, ikke om kollokationer af flere ord, som i “hvad for en” eller “hvad for en en”. Og der lader til at være tale om to forskellige leksemer, et der står attributivt og et der står alene.

#### 4. Diskussion og perspektiver

Da vi fremlagde ovenstående observationer på NFL i Reykjavík i juni 2017, var én af de kommentarer vi fik (efter hukommelsen) at det var udmærkede observationer, men at de da nemt bare kunne føjes til de beskrivelser som allerede findes. Kommentaren implicerer at det er unødvendigt at starte helt forfra, sådan som vi lægger op til. Vi kunne nøjes med at slå ned de steder hvor samtalesproget er anderledes end det skriftsprog hvis leksikon allerede findes beskrevet. Vi kunne kalde dette “tilføjelsesmetoden”: Grundlaget findes allerede i form af eksisterende (skriftsprogs-)ordbøger. Hertil kan bare tilføjes de ord vi finder i samtalesproget, som så kan tilføjes ordbogen med en bemærkning om at de er særligt “samtaleproglige”.

Der er efter vores mening ikke tale om et “enten-eller”. Det vil være fint at tilføje de samtaleproglige former til en ellers skriftsproglig leksikografi. Men tilføjelsesmetoden forsømmer at undersøge samtalesproget på dets egne betingelser. Den tager ikke højde for at samtalesproget indeholder lydstreng hvortil der ikke findes tilsvarende leksemer i skriftsproget. Tilføjelsesmetoden vil heller ikke kunne behandle samtalesprogets særlige principper for leksemgrænser og leksemopdeling. Og vi mener allerede at kunne konstatere at de kan være endog meget anderledes end dem vi kender fra skriftsprogsleksikografien.

Vores indledende undersøgelser har også vist at det ikke er ligetil at udforske samtalesprogets leksikografi på dets egne betingelser. Da vi ønskede at forfølge lydstrengen *vafnén*, måtte vi konstatere at det for nuværende kun kan lade sig gøre ved at gennemgå de 600 forekomster én for én, inklusiv at lytte grundigt til hver af dem. Det vil være et stort fremskridt for *Projekt samtalesprog* når der bliver lavet en søgemaskine der kan søge på lydstreng – uden om deres skriftlige form. Indtil da vil vi arbejde videre med det vi har.

Efter vores mening kan kun en leksikografi som tager udgangspunkt i de former der findes i samtaleproget, sådan som det *tales* i virkeligheden, yde samtaleproget retfærdighed. Det håber vi at kunne gøre med *Projekt samtaleprog*.

## Litteratur

- Garly, Katrine (2016): *Hvar – det danske huh*: En kollektionsanalyse af åbenklasse andeninitieret selvreparatur på dansk. I: *Skrifter om Samtalegrammatik* 3:3. <<http://samtalegrammatik.au.dk/skrifter-om-samtalegrammatik/aargang-3/>> (oktober 2017).
- Jefferson, Gail (2004): Glossary of transcript symbols with an introduction. I: Gene H. Lerner (red.): *Conversation Analysis: Studies from the first generation*. Amsterdam/Philadelphia: John Benjamins, 13–31.
- Jørgensen, Maria (2015): ”Hvad” i dansk talesprog. *Skrifter om Samtalegrammatik* 2:6. <<http://samtalegrammatik.au.dk/skrifter-om-samtalegrammatik/aargang-2/>> (oktober 2017).
- Samtalebank* (u.å.) <<http://samtalebank.talkbank.org/browser/>> (oktober 2017).
- Samtalegrammatik.dk* (2017): Udskriftssystem for samtaleeksempler. <<http://samtalegrammatik.au.dk/udskriftssystem-for-samtaleeksempler/>> (september 2017).
- Schegloff, Emanuel A., Gail Jefferson & Harvey Sacks (1977): The preference for self-correction in the organization of repair in conversation. I: *Language* 53, 361–82.

Carsten Hansen  
 Projektleder Ordbog over Dansk Tale-  
 sprog (ODT). Lektor emeritus  
 Institut for Nordiske Studier og Sprogvi-  
 denskab (NorS)  
 Njalsgade 136,5  
 Dk-2300 København S  
 carhan@hum.ku.dk

Jakob Steensig  
 Lektor, ph.d.  
 Lingvistik, Aarhus Universitet  
 Jens Chr. Skous Vej 2  
 DK-8000 Aarhus C  
 linjs@cc.au.dk

# Natur och digitala, tvåspråkiga ordböcker: En studie av finska som sökspråk i ISLEX

*Helga Hilmisdóttir & Marjakaisa Matthíasson*

This study deals with the Finnish version of the Icelandic-Scandinavian dictionary ISLEX. The source language of the dictionary is Icelandic and the lemmas that are included represent modern Icelandic language use. The Finnish language has a subordinate status in ISLEX and Finnish words are only entered into the dictionary if they can be used as equivalents for the Icelandic lemmas. However, this is complicated by the fact that ISLEX makes it possible to use both Icelandic and Finnish as search languages. This leads us to the question how representative the Finnish equivalents are for Finnish language use and whether the vocabulary is large enough to decode Finnish texts. This study focuses on words that have to do with nature: plants and animals, the weather and geological and geographical phenomena. The aim is to discuss to what degree Icelandic nature affects the lemma selection in ISLEX and what consequences that has for the user when Finnish is the search language.

## 1. Inledning

Den isländsk-skandinaviska ordboken ISLEX är en digital ordbok där isländska fungerar som källspråk och danska, finska, färiska, norska (bokmål och nynorska) och svenska fungerar som målspråk. Med andra ord har ordboken utarbetats utifrån det isländska språket och isländska förhållanden, och de skandinaviska språken har en underordnad status från redaktörernas sida (jfr Hannesdóttir 2015:240 om *main language* och *subordinate language*). Digitaliseringen av tvåspråkiga ordböcker har dock medfört nya utmaningar för lexikografer, och den nya tekniken har lett till att etablerade termer som *källspråk* och *målspråk* kan ifrågasättas (Hannesdóttir 2015:236). Digitaliseringen av ordböcker tillåter sökning i både käll- och målspråk, och en islänning kan t.ex. söka finska ord i ISLEX för att avkoda en text på finska. Ur användarens synvinkel fungerar finska då som *sökspråk* och isländska som ett *målspråk* (jfr Hannesdóttir 2015).

På grund av de olika sökmöjligheterna kan man säga att den digitala ordboken har ett bredare syfte än en tryckt ordbok. Ur användarens synvinkel fungerar nämligen den finska versionen av ISLEX både som isländsk-finsk och finsk-isländsk ordbok (jfr Hannesdóttir 2015:240). Bild 1 visar uppslagsordet *vogur* 'vik' när man använder sökspråket isländska och bild 2 för *lahti* om man väljer finska.

Bild 1: Uppslagsordet *vogur*.Bild 2: Det finska sökordet *lahti*.

När användarna söker isländska ord får de fram en artikel som ger information om ordets uttal, böjning och finska ekvivalenter. Som bild 1 visar har uppslagsordet *vogur* två finska ekvivalenter, *lahti* och *poukama*. Om de däremot söker på det finska ordet *lahti* får de fram en lista över alla isländska uppslagsord där sökordet anges som ekvivalent. Ordet förekommer under tre uppslagsord: *flói*, *vík* och *vogur*. Användarna kan sedan klicka sig fram till varje enskilt uppslagsord för att få fram mer information om eventuellt andra ekvivalenter och närmare förklaringar.

Den digitala utvecklingen ställer nya krav på det lexikografiska arbetet. Hanesdóttir (2012; 2015) har påpekat att ordböckernas inre struktur i framtiden kommer att ändras; i digitala ordböcker kommer båda språken att vara jämlika och ge lika bra sökresultat. Men även om ISLEX erbjuder sökning på både isländska och finska har isländskan en tydlig överordnad funktion. Uppslagsorden är valda utifrån isländska förhållanden och beskriver isländsk samhällsordning, kultur, traditioner och natur. Med tanke på detta kan man fråga sig i vilken utsträckning ISLEX fungerar när man använder finska som sökspråk. Ordboken omfattar två helt obesläktade språk som talas i länder som geografiskt sett är ganska olika varandra.

Syftet med denna studie är att undersöka hur bra det finska språket fungerar i ISLEX som sökspråk för naturord, dvs ord som har att göra med biologi, geologi, geografi och meteorologi. I artikeln kommer vi att dryfta följande tre frågor: 1) Finns det några allmänna tendenser angående valet av isländska uppslagsord? 2) Vilka problem kan uppstå när man söker ord som finns både i allmänspråket och som vetenskaplig terminologi? och 3) Hur ska man välja ekvivalenter för ord som har många synonymer, som t.ex. ord för 'storm'? För att svara på dessa frågor

jämför vi ISLEX med *Stora finsk-svenska ordboken* (2004) och gör två nedslag i ISLEX: det isländska uppslagsordet *refur* 'räv' och det finska sökordet *myrsky* 'storm'.

## 2. ISLEX

ISLEX är en isländsk-skandinavisk ordbok som har funnits på nätet sedan november 2011. Projektet är ett samarbete mellan Árni Magnússon institutet för isländska studier och universitet och forskningsinstitutioner i Norden. Arbetet med den finska versionen av ISLEX inleddes i januari 2013 och i januari 2018 hade 75 % av de isländska uppslagsorden översatts till finska (se Hilmisdóttir & Martola 2015 för en introduktion av den finska ordboken).

Det isländska materialet består av ca 50 000 uppslagsord, 30 000 språkprov och 14 000 fasta och halvfasta fraser (Úlfarsdóttir 2013). Uppslagsorden i ISLEX representerar modernt isländskt språkbruk men omfattar även en del vanlig terminologi och poetiska uttryck.

För redaktören är uppslagsorden i ISLEX grupperade enligt betydelsefält som t.ex. GEOGRAFI, GEOLOGI, DJUR och VÄXTER eller mer specifika grupper som VIND eller OVÄDER. En och samma ord kan ingå i flera olika betydelsefält som t.ex. *rok* 'storm' som ingår i grupperna VÄDER, VIND och VINDHASTIGHET. Som redaktör kan man ta fram dessa listor vilket ger intressanta möjligheter till jämförelser av uppslagsord i olika ordböcker.

## 3. Naturord i ISLEX

Om valet av uppslagsord i ISLEX kan man säga att det har en tydlig isländsk prägel och att redaktörerna har vinnlagt sig om att inkludera uppslagsord som beskriver landets natur. I detta kapitel diskuteras först valet av isländska uppslagsord i ISLEX och vilka betydelsefält man lagt särskild vikt vid (avsnitt 3.1.). Sedan diskuteras vi problem och utmaningar angående vetenskapliga och folkliga namn på djur och växter (3.2.) och avslutar med en diskussion om vind och oväder och möjligheten att överhuvudtaget hitta tillräckligt nyanserade ekvivalenter mellan två språk (3.3.).

### 3.1. Allmänna iakttagelser angående valet av isländska uppslagsord

I ISLEX har man inkluderat djur och växter som finns i isländsk natur. Många av dessa arter har spridning i hela Norden såsom fågelarterna *hettumávur/naurulokki* 'skrattmåsa' och *tjaldur/meriharakka* 'strandkata', och växter som *gleymmérei/peltolemmikki* 'åkerförgätmigej' och *sóley/niityleinikki* 'smörblomma'. Man kan

även hitta en del däggdjur som finns i Skandinavien men inte på Island såsom *ikorni/orava* 'ekorre', *elgur/hirvi* 'älg' och *hreysiköttur/kärppä* 'hermelin'.

Emellertid har ordboken en del luckor vad gäller fisk och skaldjur som inte finns på Island. Som exempel kan vi nämna uppslagsord som *silakka* 'strömning', *kuha* 'gös', *muikku* 'siklöja', *siika* 'sik', *rapu* 'kräfta' och *made* 'lake' som alla fångas i Finland och ingår i det finländska köket. Här vill vi dock påpeka att inte ens Nordiska rådets *Fiskordbog* anger isländska namn på dessa arter, så det är möjligt att det inte finns några etablerade ekvivalenter alls.<sup>1</sup> Vanliga finska bärsorter fattas också i ISLEX såsom *lakka* 'hjordron', *tyrni* 'havtorn' och *ruusunmarja* 'nypon'. Om det är redaktörernas mål att finska ska fungera som sökspråk för islänningar som vill förstå finska borde även dessa ord finnas i ordboken.

En tydlig tendens i ISLEX är att redaktörerna varit ambitiösa med att inkludera geologisk terminologi. En söking på GEOLOGI ger 333 isländska uppslagsord. Endast ett fåtal av dessa ord återfinns i *Stora finsk-svenska ordboken* (2004): t.ex. *jäätikkö* 'glaciär', *tulivuorenpurkaus* 'vulkanutbrott'. I ISLEX ingår däremot en rad geologiska termer som är viktiga i ett land som Island. Där kan man nämna ord som *öskulag/tuhkakerros* 'asklager', *dyngja/kilpitulivuori* 'sköldvulkan' och *skriðjökull/vuorenjuurijäätikkö* 'utloppsglaciär'. Andra betydelsefält som bedömts som viktiga i ISLEX är kategorierna VÄDER (390 ord), VIND (113 ord) och SNÖ (58 ord). En mera systematisk jämförelse mellan ISLEX och *Stora finsk-svenska ordboken* (2004) skulle sannolikt visa att det finns vissa betydelsefält som täckts in med mera ambition i den finska ordboken. Man kan t.ex. tänka sig att ord som har med skogen att göra har en större betydelse i en finsk kontext. I ISLEX är kategorin SKOG mycket liten och består av endast 12 ord.

### 3.2. Isländska och finska rävar

Som tidigare nämnts har redaktörerna för ISLEX strävat efter att inkludera de djur och växter som finns på Island. När det gäller fåglar, fiskar och växter har man tagit med både specifika arter som t.ex. *týsfjóra* 'ängsviol' och den mer allmänna beteckningen *fjóra* 'viol' som omfattar alla blommor av violfamiljen. Uppslagsorden är ofta försedda med bildmaterial, och efter artnamnen kommer den latinska termen inom parentes.

Eftersom ISLEX bygger på isländska förhållanden kan vissa djurbeteckningar tolkas på olika sätt beroende på om användaren är från Island eller Finland. Det bästa exemplet är uppslagsordet *refur* 'räv' som är ett ord i allmänspråket som kan syfta på flera arter. När islänningar använder ordet *refur* syftar de oftast på polarräv eftersom andra arter av rävsläktet inte lever vilt i isländsk natur. På finska heter polarräven (eller fjällräven) däremot *naali* och man gör en tydlig skillnad mellan *naali* 'polarräv' och *kettu* 'räv el. rödräv' som båda lever vilt i Finland. Eftersom användare på Island och Finland sannolikt har olika uppfattningar om

<sup>1</sup> Se <<http://docs.norden.org/faktaof/Tabel-med-latin.htm>> (hämtat i juli 2017).

hur en prototypisk räv ser ut har redaktörerna valt att inte visa bildmaterial för uppslagsordet *refur*.

### 3.3. Oväder

Isländskan har ett stort ordförråd för att beskriva väder och det kan vara svårt att beskriva alla betydelsenyanser. Dels syftar väderorden ofta på abstrakta företeelser som olika personer kan uppfatta på olika sätt, dels kan det finnas geografisk variation när det gäller vilka ord man använder och hur man definierar dem. Definitionen av väderord blir också mer komplicerad i och med att vissa ord kan användas både som meteorologiska termer och som vardagligt språk med ibland mindre precision. Ett exempel på detta är orden som används för att beskriva vindhastighet. Ordet *rok* är t.ex. mycket frekvent i allmännspråket och syftar på det som vi upplever som stark eller mycket stark vind. Som meteorologisk term används *rok* däremot för en vindstyrka som motsvarar 10 på Beaufortskalan, som betyder att *rok* är lite starkare än *stormur* (9 på Beaufort) och svagare än *ofsaveður* (11 på Beaufort). Därför kan man säga att det vardagliga ordet *rok* har en mycket snävare betydelse inom meteorologin. På grund av detta har man i ISLEX delat upp uppslagsord som *rok* i två olika betydelser: den allmänna och den terminologiska. Tabell 1 visar de meteorologiska termerna på isländska och finska med svensk översättning (se finska termer och tabell i Similä 1981:44–45).

Beaufort skala	Isländsk term	Finsk term och svensk översättning
0	logn	tyyni 'lugnt'
1	andvari	hyvin heikko tuuli 'mycket svag vind'
2	kul	heikko tuuli 'svag vind'
3	gola	kohtalainen tuuli 'måttlig vind'
4	stinningsgola	
5	kaldi	navakka tuuli 'frisk vind'
6	stinningskaldi	
7	allhvasst	kova tuuli 'hård vind'
8	hvassviðri	
9	stormur	myrsky 'storm'
10	rok	
11	ofsaveður	
12	färviðri	hirmumyrsky 'orkan'

Tabell 1: Beaufortskalan på isländska och finska.

Som framgår av tabell 1 så har den finska skalan färre vindord än den isländska. Medan den isländska skalan använder 12 olika substantiv och ett adjektiv/adverb för att precisera vindstyrka finns det endast 4 olika ord på finska: *tyyni* 'lugnt', *tuuli*



'vind', *myrsky* 'storm' och sammansättningen *hirmumyrsky* 'orkan'. Notera t.ex. att den korrekta termen för vindstyrka 9, 10 och 11 är *myrsky* på finska medan motsvarande vindstyrka beskrivs med tre olika termer i isländskan: *stormur* (9), *rok* (10) och *ofsaveður* (11). Det finska sökordet *myrsky* ger användaren således tre isländska (terminologiska) alternativ att välja mellan.

När de finska redaktörerna sökte finska ekvivalenter till isländska väderord valde de att fokusera mer på tydlighet än på att få fram en stor variation i ordförrådet. Detta återspeglas i de finska ekvivalenterna i ISLEX. När man söker på *\*tuuli* och *\*myrsky* får man fram 61 respektive 35 isländska uppslagsord. Tabell 2 visar sökresultaten för det finska sökordet *myrsky* där sökordet förekommer utan preciserande adjektiv eller förled (sökning utan asterisk). De isländska definitionerna är hämtade från den enspråkiga nätordboken *Íslensk nútímamálsorðabók* (=ÍNO) som bygger på samma databas som ISLEX.

Ekvivalent	Definition i ÍNO översatt till svenska
aftakarok	mycket stark vind
aftakaveður	väldigt stark vind
foráttuveður	stark vind och nederbörd, i sht hög vindhastighet
hávaðarok	mycket stark vind
hryssingur	storm med snö
ofsarok	väldigt stark vind
ofsaveður	väldigt stark vind
ofsi	svår storm
ofviðri	svår storm som orsakar skada på byggnadsverk
óveður	stark vind med hög vindhastighet, ibl. nederbörd
rok	stark vind
rosi	stark vind med regn, snöblandat regn eller snö
slagveður	storm och mycket regn
slagviðri	storm och mycket regn
stormur	mycket stark vind
veðurhamur	dåligt väder med hög vindhastighet
veðrahamur	dåligt väder med hög vindhastighet

Tabell 2: Uppslagsord som motsvarar det finska sökordet *myrsky* med definitioner hämtade från ÍNO i svensk översättning.

När man talar om storm och oväder påverkas valet av ord av många faktorer: Medför stormen nederbörd av något slag? Vilken årstid rör det sig om? Hur farligt är det att vara ute, hur påverkar stormen miljön och hur plötsligt börjar den?

Även om uppslagsorden i tabell 2 semantiskt sett är ganska lika finns det fina skillnader mellan dem. Man talar t.ex. om *slagviðri* om det både blåser och regnar, medan *rok* enbart syftar på vindhastigheten och *óveður* på oväder i allmänhet. Ordet *hávaðarok* leder däremot tankarna till vår hörupplevelse av mycket stark vind och *ofsi* betyder något extremt eller häftigt. En del ord används mycket frekvent

i isländskt tal- och skriftspråk, i synnerhet *rok* 'stark vind' och *stormur* 'mycket stark vind'. Andra ord som t.ex. *veðurhamur* och *veðrahamur* är mindre frekventa i modern isländska och visar mindre syntaktisk variation. En jämförelse av förekomsten av vindord i den isländska tidningsdatabasen *Tímarit.is* visar att *stormur* har 23 220 belägg medan *veðurhamur* har endast 107 belägg.

I nuläget ordnas listor på uppslagsord som innehåller ett finskt sökord i alfabetisk ordning. I de flesta fall spelar detta inte någon större roll eftersom sökorden oftast ger endast ett eller två uppslagord. När det gäller sökord som *myrsky* 'storm' eller *tuuli* 'vind' blir det dock mer komplicerat. För isländska användare spelar antalet ord inte någon större roll eftersom de snabbt kan gå igenom listorna och hitta de ord som passar kontexten, men för finska användare kan det här vara betydligt svårare. Idealt skulle man kunna programmera ordboken på så sätt att uppslagsorden rangordnas enligt frekvenslistor och därmed göra valet enklare för finskspråkiga användare.

## 4. Avslutning och diskussion

I denna artikel har vi diskuterat naturord i ISLEX för att ta fasta på vilka problem som kan uppstå när man har en digital ordbok som har två eller flera sökspråk. ISLEX har en stor målgrupp och utgångspunkten är att den ska fungera för både isländsk- och finsktalande användare. I analysen kom vi fram till att ISLEX har en tydlig tonvikt på isländska naturförhållanden. Det finns ovanligt många geologiska termer i ISLEX om man jämför ordboken med *Stora finsk-svenska ordboken* (2004). Även isländska djur och växter är mycket väl representerade. På så sätt kan man säga att ISLEX passar bra för finska användare som vill avkoda isländska texter. Om islänningar vill använda ISLEX för att avkoda finska texter är det risk att de stöter på vissa luckor i ordförrådet.

Som Hannesdóttir (2015) har påpekat har digitaliseringen av ordböcker lett till att användarna kan söka på målspråket. Om tanken är att tillmötesgå dessa nya användare och förbättra sökresultaten för sökspråket finska finns det en del saker som kunde förbättras. Man kunde t.ex. fokusera på ord som har med skogen att göra och tillägga en del fiskar, svampar, växter, bär och djur som har en central roll i finländsk kultur (både terminologi och ord ur allmänspråket). Sådana tillägg är enkla att göra och kräver inte mycket planering. Om man å andra sidan vill tillägga ord för abstrakta företeelser som väderförhållanden skulle det kräva en mer systematisk genomgång av det finska ordförrådet. Finska språket har precis som isländska ett stort och nyanserat ordförråd som beskriver väderförhållanden men eftersom ISLEX har utarbetats från källspråket isländska erbjuder ordboken få väderord på finska: De finska redaktörerna ansåg att det var viktigare att hitta tydliga och beskrivande ekvivalenter än att registrera alla nyanser inom vissa finska betydelsefält.

Slutligen kan man fråga sig om det går att redigera en existerande ordbok på så

sätt att man får lika bra sökresultat på det överordnade och underordnade språket. Hur får man t.ex. in finska sökord som inte har en ekvivalent på isländska? Vår studie visar att trots att man till viss mån kan använda finska som sökspråk i ISLEX kan ordboken inte ersätta en finsk-isländsk ordbok. För att få en webbordbok som ger lika bra sökresultat på båda språken är det viktigt att planera ordboken utifrån båda språken från början.

## Källförteckning

### Ordböcker

ISLEX. <[www.islex.is](http://www.islex.is)> (augusti 2017).

*Íslensk nútímamálsorðabók*. <[islenskordabok.arnastofnun.is](http://islenskordabok.arnastofnun.is)> (januari 2018, preliminär version).

Järvelä, Tuomas (2008): *Suomi-islanti-suomi sanakirja*. Helsinki: Gaudeamus.

*Stora finsk-svenska ordboken* (2004). Helsingfors: WSOY.

### Övrig litteratur

Hannedóttir, Anna Helga (2012): Ekvivalensrelationer i tvåspråkig lexikografi. I: *Lexico Nordica* 19, 39–58.

Hannedóttir, Anna Helga (2015): What is a Target Language in an Electronic Dictionary? I: *Electronic Lexicography in the 21st century: Linking Lexical Data in the Digital Age*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd, 236–249.

Hilmisdóttir, Helga & Nina Martola (2015): Flerspråkig lexikografi och obesläktade språk – erfarenheter från finska Islex. I: *Nordiske studier i leksikografi 13. Rapport fra 13. konference om leksikografi i Norden København 19.–22. maj 2015*, 143–154

Similä, Artturi & Toivo Vuorela (1981): *Artturi Similän sääkirja*. Porvoo: WSOY. *Timarit.is*. <[timarit.is](http://timarit.is)> (augusti 2017).

Úlfarsdóttir, Þórdís (2013): ISLEX – norræn margmála orðabók. I: *Orð og tunga*, 15, 41–71.

Úlfarsdóttir, Þórdís (2014): ISLEX – a Multilingual Web Dictionary. I: Calzolari et al (red.): *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation*. <<http://www.lrec-conf.org/proceedings/lrec2014/index.html>> (augusti 2017).

Helga Hilmisdóttir  
forskningslektor, fil.dr  
Árni Magnússon institútet för isländska  
studier  
Laugavegur 13  
101 Reykjavík  
helga.hilmisdottir@arnastofnun.is

Marjakaisa Matthíasson  
Redaktör, fil.mag.  
marjakaisa@internet.is

# Udflytning og nedarvning som formidlingsstrategi i ordbøger

*Henrik Hovmark*

Dictionaries are highly specialized texts containing many types of information and using a substantial amount of symbols and condensation, which constitutes a considerable communication challenge. In this article we discuss relocation as a means of condensation in (older) printed as well as (more recent) electronic dictionaries. We show that a simple hierarchical structure with clear indicators is paramount when using this type of condensation. We also show that a presentation that organizes information in fixed, visually recognizable positions has specific advantages. Information retrieval and decoding in dictionaries, however, still presuppose a high degree of user knowledge about the dictionary as a specialized text type and its typical component parts.

## 1. Indledning

I ordbøger er det almindeligt at oplysninger der gælder for flere enheder kun gives én gang, fx formelle oplysninger om ordklasse o.l. i artiklens begyndelse:

**spat** s amer. gruf  
**spat** s amer. dask  
> **spat** s amer. **1** gruff **2** dask

Figur 1: Eksempel på hierarkisk (markeret) udflytning (jf. Svensén 2004:112f.).

I eksemplet ovenfor (Figur 1) er informationerne ”s amer.” flyttet ud til et øvre overskriftsniveau hvis indhold gælder for både betydning 1 og 2 (gruf, dask). Udflytningen foregår i dette tilfælde i et eksplicit markeret hierarkisk system. I andre tilfælde nedarves information mellem sideordnede enheder, fx betydningsdefinitioner udfyldt med ”ds” (”det samme”, fx *Ømålsordbogen* (ØMO): ”*give sig an med ngn ds*” under *give* bet. 12), hvor informationen om betydningen skal hentes fra en foregående parallel enhed. Informationen kan også være efterstillet. I ØMO kan geografisk udbredelse fx udtrykkes: ”S, M, L-Fa(alm)”, hvor indholdet i parentes (”alm(indeligt)”) gælder for alle foranstillede forkortelser (”Sjælland, Møn, Lolland-Falster”). Brugeren skal altså hente information om ”S” og ”M” i den efterstillede parentes. Fænomenet er fx beskrevet som ”udflytning” (Svensén 2004:111-113; jf. Wiegand 1989:447ff.):

Uttflytning är en form av kondensering som huvudsakligen förekommer

inom mikrostrukturen. Den indebär att en angivelse placeras på ett sådant sätt att den kan ingå i flera bearbetningsenheter (dvs. ha flera adresser) [...], och den kondensering som blir resultatet av utflyttningen består i att angivelsen bara behöver ges en gång för alla och inte en gång för varje bearbetningsenhet. (Svensén 2004:112)

Som Svensén bemærker, er udflytning en form for kondensering eller komprimering (jf. fx forkortelser): En oplysning som gælder for flere enheder, nævnes kun én gang, dvs. at oplysningen ikke vil være anført eksplicit ved mindst én og muligvis flere enheder. Strategien kan altså skabe problemer i en formidlingsmæssig sammenhæng – i en teksttype som i forvejen er meget specialiseret, med mange oplysningstyper og en høj grad af kodificering (jf. fx Svensén 2004:96, Gerzymisch-Arbogast 1989). Ordbøger stiller store krav til brugeren, som ikke kun skal kunne finde frem til den relevante information, men også kunne afkode informationen (eller rekonstruere den, jf. Hausmann & Wiegand 1989:341) ud fra et kendskab til bestemte konventioner for hvordan informationen i netop ordbøger er kodificeret og/eller komprimeret. Man kan tænke på en ordbogsartikel som en tekst hvor et stort antal informationer er blevet fortættet og pakket omhyggeligt ned i et antal kasser, og hvor det er brugerens opgave at finde bestemte kasser og pakke den relevante information ud igen.

En ordbogsartikel præget af en tydelig organisering og entydige instrukser vil være en fordel når brugeren skal finde, udpakke og afkode en given information. En sådan organisering, præget af tydeligt hierarki og velafgrænsede enheder, er imidlertid ikke altid til stede, især ikke i de ældre, trykte ordbøger med udpræget lineær informationsorganisering og stærkt komprimeret information. Det kan være uklart for brugeren hvad en betydningsdefinition som ”ds” (’det samme’) refererer til, og hvor informationen helt præcist skal hentes.

Jeg vil i denne artikel diskutere udflytning og nedarvning som generel komprimeringstype og formidlingsstrategi i ordbøger (trykte såvel som digitale, med henholdsvis ØMO, *Ordbog over det danske Sprog* (ODS) og *Den Danske Ordbog* (DDO) som eksempler), og jeg vil sætte strategien i forhold til forskellige kontekstuelle forhold, fx den konkrete ordbogs funktion og format og den enkelte brugers forhåndskendskab til konventionaliserede typer af informationsorganisering.

## 2. Udflytning og nedarvning: hierarkisk og lineær

### 2.1. Hierarki, strukturindikatorer og afgrænsede informationsenheder

Når udflytning bl.a. beskrives ved at noget ”flyttes væk”, siges det også indirekte at oplysningerne i en ordbogsartikel 1) er koblet til hinanden på bestemte måder, dvs. indgår i bestemte relationer (jf. ordet ”adresserer” i Svenséns definition; Svensén 2004:100ff., Hausmann & Wiegand 1989:349ff.) og 2) er placeret på en bestemt

måde i forhold til hinanden. Helt minimalt kan man sige at en ordbog(sartikel) består af enheder (x) som der gives oplysninger om (y) (undertiden beskrevet ved hjælp af tema-remabegrebet, se fx Gerzymisch-Arbogast 1989). Informationen i teksten er organiseret i relationer hvor oplysning y hører til (siger noget om, peger på, adresserer) enhed x. Man vil i den sammenhæng som udgangspunkt forvente at oplysninger der står ved siden af hinanden, hører sammen: x y. I princippet behøver ”ved siden af” ikke at betyde ”lineært”, fx ikke i den elektroniske ordbog (jf. afsnit 3), men i realiteten vil det ofte være sådan. Ved siden af kan både være før og efter, jf. at udflytning kan være både højre- og venstrestillet (jf. Svensén 2004:113). I Figur 2 adresserer ”s” fx ”kæltringrak”, og citatet (”*der skulle være ...*”) har adresserende information både til venstre (redaktionel kommentar) og til højre (lokaliseringsangivelsen ”smst”) ( $x < > y$ ).

*kæltringrak s – (skældsord) dss kæltringpak nvS-*  
 (Gra): (folk snyder og bedrager) *der skulle være*  
*en hård justits med al det kæltringrak smst.*

Figur 2: Artiklen *kæltringrak* i ØMO (bind 11).

Eksemplet *kæltringrak* er simpelt, og x og y kan stå direkte ved siden af hinanden. Men ofte vil relationerne være mere komplekse: Nogle informationer kan gælde for flere dele af en artikel, fx oplysninger om køn, eller en enhed kan have tilknyttet flere informationer, som citatet i Figur 2. Under alle omstændigheder vil alle sammenhørende enheder ikke nødvendigvis stå ved siden af hinanden, dels fordi der ikke er plads, dels fordi man måske ønsker at organisere oplysningerne hierarkisk og/eller give dem bestemte, faste placeringer som kan markeres eksplicit, fx ved hjælp af udflytning. Det prototypiske eksempel på udflytning i Figur 1 foregår netop inden for et tydeligt hierarki: *dask* adresserer **spat**; *dask* (y) står ikke direkte ved siden af **spat** (x), men tilhørsforholdet er markeret med de hierarkiske markører 1 og 2. De to informationsenheder er adskilt af et klart defineret og markeret skel i en struktur: **1** og **2** ( $x < | > y$ ), og 1 og 2 har også indbyrdes en tydelig parallel struktur og en klar afgrænsning. De udflyttede elementers rækkevidde, dvs. det de gælder for (jf. Svensén 2004:102), er ligeledes tydelig: De står placeret på et overordnet niveau, som en slags overskrift, og gælder for både 1 og 2. Man kan også sige at de to undergrupperinger er selvbærende i den forstand at de giver fuld, eksplicit information i forhold til deres funktion, fx ved en søgning, nemlig ’betydning’. Og så har man sparet plads, dvs. komprimeret.

Brugeren vil med en hierarkisk, positionsbestemt organisering ikke nødvendigvis finde informationerne direkte ved siden af hinanden; til gengæld vil brugeren vænne sig til at bestemte oplysninger altid findes på bestemte pladser, og ofte vil oplysningstypen være markeret grafisk eller med en indleder (strukturindikator, jf. Svensén 2004:102). Undersøgelser inden for visuel processering og kognition giver et vist belæg for at antage at en sådan strategi har formidlingsmæssige for-

dele: Hjernen er trænet i at genkende enkle fremhævelser og klare afgrænsninger. I det øjeblik en repræsentation er præget af mange forskelligartede og komplekse elementer uden tydelige skel, eller af et for omfattende hierarki, vil hjernen til gengæld begynde at afkode elementer enkeltvist i en meget længere proces (jf. Treisman 1986:123f.) (jf. forskellen i det visuelle udtryk mellem DDO og ODS som app Figur 4).

## 2.2. Nedarvning, parallelle strukturer og tvetydigt hierarki

Svenséns definition og eksempler på udflytning bygger i høj grad på at informationerne i en ordbogsartikel er organiseret i velafgrænsede enheder i et tydeligt (markeret) hierarki (jf. Wolski 1989:965). Denne forudsætning er imidlertid ikke altid til stede, især ikke i de ældre ordbøger med udpræget brug af lineær informationsorganisering. Her udelades og/eller "(ned)arves" information ofte ud fra mere diffuse kriterier mellem sideordnede enheder ( $x < \dots > y$ ). For at afkode informationen korrekt skal brugeren bevæge sig tilbage (eller fremad) i en lineær opstilling og finde den relevante information "et sted i det umiddelbart foregående (eller efterfølgende)", ofte i en parallel struktur, men der vil ikke nødvendigvis være et tydeligt hierarki eller entydige strukturindikatorer. Jeg vil gennemgå et par eksempler i det følgende. Eksemplerne er fra ØMO, men skal ses som eksempler på en type som ses i mange tilsvarende ordbøger. Om man vil kategorisere typen som en variant af udflytning eller en særskilt form for komprimering, er en smagssag, men den har under alle omstændigheder det til fælles med udflytning at information (y) som kunne have stået eksplicit (ved siden af x), er rykket væk eller er blevet erstattet af en forkortelse (pladsholder).

Det første eksempel er brugen af "ds" (pladsholdertypen). I Figur 3 ses "ds" brugt som betydningsforklaring til flerordsforbindelsen *en fuldkommen kvind* (artiklen *kvinde*). Markøren "ds" fungerer som en pladsholder, dvs. at "ds" udfylder pladsen som definition, men uden at rumme selve definitionen.

vF(Dres), jf *for* III.21; i *forb* fuld kvind om  
fuldvoksen, dygtig tjenestepige øF(NSøby Molb.-  
S), vF(VPSeidelin.HistTopDres:261), jf *fuld* II.2;  
i *forb* en fuldkommen kvind ds sF(Horne  
CDalgas.SvenAmt:404), jf *fuldkommen* 1; hertil

Figur 3: Eksempel på nedarvning med pladsholder ("ds").

Teksten "ds" skal afkodes som en slags henvisning: Brugeren skal bevæge sig lineært bagud og finde betydningsforklaringen til den umiddelbart foregående flerordsforbindelse: *fuld kvind* – som er identisk med betydningsforklaringen til *en fuldkommen kvind*. Referencen er dobbelt og kompleks: "ds" adresserer ikke



kun *en fuldkommen kvind*, men også betydningsforklaringen til *fuld kvind*. Den første reference er tydelig nok i strukturen (x står ved siden af y), men til gengæld uden indhold. Den anden reference giver indholdet, men er til gengæld underspecificeret med hensyn til den strukturelle placering. En vellykket afkodning forudsætter at brugeren forstår at indholdet skal hentes i den foregående, parallelle struktur indledt med en flerordsforbindelse. Det er i dette tilfælde ikke så svært, da de to enheder er korte, har samme elementer og oven i købet sammenfald i den sproglige form ("fuld"). Dette er imidlertid ikke altid tilfældet, fx ikke i artiklen *kvovne* (bind 11). Informationsenheden "i pass form *kvovnes*" har her "ds" som pladsholderdefinition. Indholdet til "ds" skal imidlertid hentes hele otte linjer væk og i en struktur som ikke på samme måde spejler udgangsstrukturen ved "ds". Brugeren får heller ikke hjælp fra strukturindikatorerne, for de er ikke parallelle: "i pass form *kvovnes*" indledes af en dobbeltbjælke "||", mens den foregående enhed indledes af et betydningsnummer "-1)". Samtidig rummer indholdet ved betydningsnummeret både en definition, en flerordsforbindelse og en indledende grammatisk parentes: "(intrans) blive kvalt, kvæles (af mangel på frisk luft); især i forb *være ved at kvovne*". Brugeren kan altså også komme i tvivl om hvorvidt "ds" refererer til hele dette samlede indhold, eller kun dele af det.

Man kan også pege på det u hensigtsmæssige i at brugeren først skal bevæge sig fra højre mod venstre (for at hente indholdet til "ds"), og derefter tilbage igen fra venstre mod højre for at genoptage afkodningen af informationerne. Hvis man opfatter "sprogligt udtryk + betydningsforklaring" som en slags mindsteenhet som bør være selvbærende ved et opslag, føres læseren altså ud af denne enhed undervejs – enheden rummer ikke det nødvendige indhold for at kunne være selvbærende. Det er det som Grønvik (2016) peger på som problematisk ved omsætningen af (ældre) papirordbøger til databaser:

Leksikalske databaser skil seg frå papirordbøker ved at den einiskilde ord-artikkelen må kunna stå aleine. Han inngår i ei artikkelliste som blir vist fram, men artikkellista er bestemt av søket i databasen og varierer dermed frå gong til gong. Ein redaktør kan derfor ikkje bruka definisjonsplassen til å visa til føregåande artikkel, slik papirordbøker kan ha døme på. (Grønvik 2016:45)

Man kan udvide Grønviks argument og tilføje at søgninger i databaser ikke er bundet til at foregå via opslagsordet – en hvilken som helst oplysningstype kan i princippet danne udgangspunkt for en søgning. Det er derfor hensigtsmæssigt, undertiden nødvendigt at definere underenheder i en artikel som også bør være selvbærende, eller som i hvert fald ikke bør mangle bestemte oplysninger som eksplicit information, fx en definition ved et sprogligt udtryk (det være sig en ny (under)betydning eller en flerordsforbindelse). En isoleret søgning på flerordsforbindelsen *en fuldkommen kvind* vil blive utilfredsstillende og meningsløs da man vil få at vide at forbindelsen betyder "ds" uden umiddelbar tilgang til det indhold

som ”ds” refererer til. Man kan også tænke sig andre underenheder hvor bestemt information altid skal være eksplicit, fx lokaliseringsinformation ved citater (dette overholdes fx ikke i ØMO, hvor ”smst” anvendes som angivelse af lokalisering, dvs. ”smst” er pladsholder og peger på en lokaliseringsangivelse ved den umiddelbart foregående betydningsdefinition).

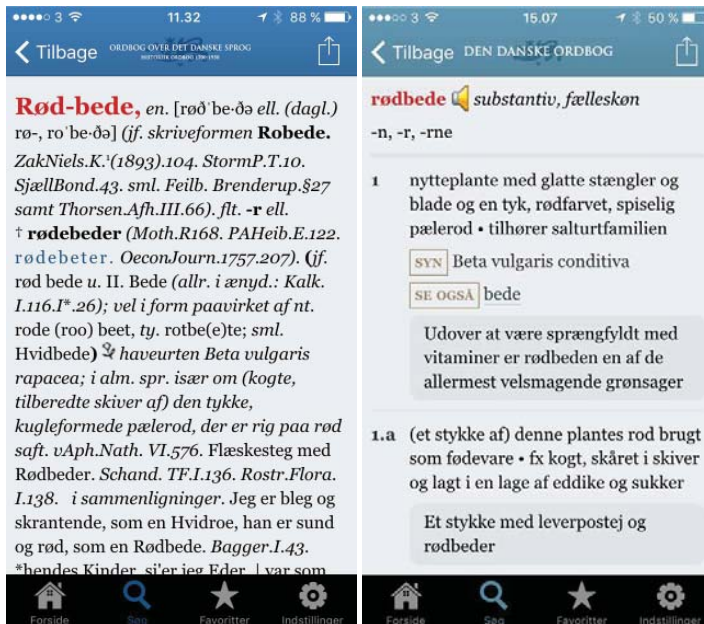
Til sidst skal jeg give et eksempel på nedarvning uden pladsholder, hvilket i højere grad ligner det indledende eksempel på udflytning (Figur 1). I ØMO angives oplysninger om geografisk udbredelse i meget komprimeret form ved hjælp af forkortelser, men man udnytter også en type af udflytning eller nedarvning baseret på en lineær informationsafkodning for at komprimere oplysningerne yderligere. I informationsenheden ”S, M, L-Fa(alm)” gælder indholdet i parentesens (”alm(indeligt)”) for alle foranstillede forkortelser (Sjælland, Møn, Lolland-Falster). Udtrykket svarer altså ukomprimeret til ”S(alm) og M(alm) og L-Fa(alm)”. Oplysningen ”alm” er rykket ud, men vel at mærke til højre – indholdet i parentesens fungerer som en opsamling for de foregående enheder, og der er tale om en slags ”oparvning” fra højre mod venstre. Det sidste giver en ekstra udfordring formidlingsmæssigt da det kolliderer med læseretningen venstre mod højre. Brugeren kan i princippet ikke vide at der mangler information ved ”S” (eller ”M”). Pudsigt nok aflæses informationen af ØMOs egen redaktion traditionelt i den logisk korrekte, omvendte rækkefølge, nemlig: ”almindeligt i dialektområderne Sjælland, Møn og Lolland-Falster”, dvs. ”alm(S, M, L-Fa)” (jf. ØMO.Tillæg:22).

### 3. Udflytning og nedarvning i kontekst og brug

Som vist kan nedarvning give problemer i forbindelse med afkodning af information, især i (ældre) udpræget lineært organiserede ordbøger. Det er imidlertid ikke ensbetydende med at udflytning og nedarvning pr. definition er en dårlig formidlingsstrategi. Manglende brug af udflytning e.l. vil i nogle tilfælde også resultere i en uoverskuelig fremstilling, fx hvis ordklasse skulle gentages som indledende information ved hver enkelt betydning (jf. Svensén 2004:112). Eventuel brug af udflytning og nedarvning er med andre ord et spørgsmål om omfang set i forhold til bestemte faktorer, og flere faktorer kan spille en rolle (funktion, størrelse, brugerens forhåndskendskab, jf. fx Bergenholtz & Svensén 1994).

På Figur 4 ses artiklen *rødbede* i appversionen af henholdsvis ODS og DDO. Forskellen mellem den lineært organiserede fremstilling i ODS og en fremstilling baseret på en tydeligere markering af et hierarki og en positionsbestemt formidlingsstrategi i DDO er markant. I ODS er brugerens afkodning bundet til en lineær aflæsning, og der er kun begrænset hjælp at hente i strukturindikatorer. Set på afstand fremstår teksten som én tekstenhed hvor forskellige skrifttyper og symboler blander sig med hinanden og uden tydelige pejlemærker. I DDO har bestemte oplysninger bestemte pladser som desuden er fremhævet ved hjælp af diskrete grafiske virkemidler. Set på afstand er teksten struktureret og overskuelig

med forskellige pejlemærker. Brugerens afkodningsstrategi er ikke bundet til en lineær aflæsning og kan i langt højere grad udnytte en evt. forhåndsviden om eller erfaring med at bestemte oplysninger kan findes på bestemte pladser, fx oplysningen om bøjning eller henvisningen til *bede*. Hierarkiet bruges til at skabe en visuel organisering af informationen, hvor et positionsskelet går igen sammen med visuelle grafiske virkemidler. Denne organisering af ordbogsartikler giver dermed også fornyet input til en række analyser af informationsstruktur m.v. som alle har været bundet til den trykte bog, fx artiklerne i Hausmann et al. (1989).



Figur 4: Artiklen *rødbede* i app-versionen af henholdsvis ODS og DDO.

Skemastrukturen bag ved de digitale ordbøger kræver ofte en strengere hierarkisk struktur og en mere konsekvent rubricering af oplysninger i veldefinerede kasser. Det er dog svært at sige om overgangen til databaseredigering ligefrem er den eneste eller direkte årsag til at formidlingsstrategien også har ændret sig i retning af langt tydeligere hierarkiske over-/underenheder og tydelig grafisk markering af oplysningstyper. Men i forhold til spørgsmålet om udflytning og nedarvning er det slående hvor lidt lineær informationsorganisering (bogstavelig talt) der bruges i *rødbede* i app-versionen af DDO i den forstand at næste oplysning i en række typisk anbringes på en ny linje. Det giver ringere basis for at udnytte nedarvning som komprimeringsstrategi. Til gengæld er mulighederne for at udnytte udflytning som strategi intakte, måske faktisk forøgede, jf. det forhold at en vellykket udflytning hænger tæt sammen med et tydeligt (markeret) hierarki og velafgrænsede (under)enheder (jf. afsnit 2). Som vist er det ikke sådan at den ældre line-

ært organiserede ordbog er blottet for hierarkier eller opdeling i (under)enheder, snarere tværtimod, men de er mere komplekse og måske først og fremmest ikke tydeligt markerede. Og den positionsbestemte informationsorganisering er heller ikke bundet til de digitale platforme web og app, men ses også på tryk, fx i den trykte version af DDO.

**rødbede** sb. fk.  
 -n, -r, -rne; [røð'be'də]  
 nytteplante med glatte stængler og blade og en tyk, rødfarvet, spiselig pælerod; **SVN** *Beta vulgaris conditiva*; **JF** *bede*; **ENC** tilhører salturtfamilien  
 □ *syltede rødbeder* □ *Udover at være sprængfyldt med vitaminer er rødbeden en af de allermest velsmagende grønsager* Hjemm92  
 • (et stykke af) denne plantes rod brugt som fødevarer · fx kogt, skåret i skiver og lagt i en lage af eddike og sukker □ *Et stykke med leverpostej og rødbeder* KBJens92  
**SMS** rødbedesaft, -salat, -skive,  
 -stunde

Figur 5: Artiklen *rødbede* i den trykte udgave af DDO (bind 5, 2005).

Man genkender tydeligt den positionsbestemte organisering og også brugen af markører af oplysningstyper i Figur 5. Man ser imidlertid også en del komprimering: Markørerne er enten forkortede eller omdannet til grafiske symboler som forudsætter et stort afkodningsarbejde hos brugeren (fx kvadraten som markør af eksempelmateriale; jf. Gerzymisch-Arbogast 1989:954), og koncentreret nedarvning ses i bøjningsoplysningerne og i orddannelserne (SMS). Nedarvning mellem parallelle strukturer i artiklen ses imidlertid ikke. Artiklen har både et ben i den databasebaserede formidlingslejr og i den lineære, trykte ordbog præget af et behov for at spare plads og komprimere.

Hjemmesiden som platform har til gengæld givet mulighed for at bryde med tidligere tiders pladsbesparende krav og skabe en langt mere brugervenlig formidling. Opstillingen af *rødbede*-artiklen udnytter en positionsbestemt informationsorganisering fuldt ud, og der anvendes hverken nedarvning eller forkortelser, selv sammensætningerne er skrevet ud. Og hvis man havde anvendt forkortelser, ville man have kunnet udnytte hjælpefunktioner som fx mouseover og udfoldning. Den eneste nedarvning der ses er opstillingen af bøjningsendelser med bindestreg, dvs. hvor ordets stamme i princippet nedarves, men da der netop er tale om bøjningsendelser, kan man samtidig hævde at der er tale om en indirekte eller ”medfødt” nedarvning der er indbygget pr. definition i oplysningstypens inventar.

**rødbede** substantiv, fælleskøn

BØJNING -n, -r, -rne

UDTALE [ˈʁøðˈbeːðə]

---

**Betydninger**

1. nytteplante med glatte stængler og blade og en tyk, rødfarvet, spiselig pælerod • tilhører salturfamilien

SYNONYM Beta vulgaris conditiva SE OGSÅ bede

BESLÆGTEDE ORD... ..vis

EKSEMPLER syltede rødbeder

Udover at være sprængfyldt med vitaminer er rødbeden en af de allermost velsmagende grønsager Hjemm 1992

1.a (et stykke af) denne plantes rod brugt som fødevarer – fx kogt, skåret i skiver og lagt i en luge af eddike og sukker

BESLÆGTEDE ORD... ..vis

Et stykke med leverpostej og rødbeder KB Jens92

---

**Orddannelser**

SAMMENSETNINGER rødbedesaft | rødbedesalat | rødbedeskive | rødbedesuppe

Figur 6: Artiklen *rødbede* i web-versionen af DDO (maj 2017).

Med app'en som platform vender kravene til pladsbesparelse til gengæld tilbage (jf. Figur 4), og man genkender træk fra den trykte udgave. Markørerne for de fleste oplysningstyper er sparet væk, og en af de to tilbageblevne er blevet forkortet ("SYN"). App-brugeren afkræves altså større forhåndsviden om oplysningernes art og om ordbogsartiklen som tekstgenre. Hertil kommer at hele oplysningskategorier er fjernet (fx orddannelser og etymologi). Det er dog også tydeligt at app-versionen er en koncentreret version af hjemmesiden, ikke af den oprindelige trykte udgave: Det hierarkiske skelet og det grafiske udtryk fra web-versionen er bevaret.

Generelt peger de valgte eksempler på at konteksten, i form af format og pladshensyn, stadig kan spille en rolle i den digitale ordbog, med deraf følgende behov for at forkorte, udelade og komprimere i bestemte situationer (fx ved hjælp af udflytning), og med deraf følgende øgede krav til brugerens forhåndskendskab til ordbogsgenren som tekst og de formidlingsstrategier der bruges. Men komprimeringen er generelt mindre omfattende og især er nedarvning stærkt minimeret, i hvert fald i disse eksempler.

#### 4. Diskussion og konklusion

Eksempelmaterialet og analysen peger på en række konkrete ting i forbindelse med brugen af udflytning og nedarvning som formidlingsstrategi i ordbøger, som fx at nedarvning er på retur i den nyere elektroniske ordbog, mens udflytning består, at det er vigtigt med et tydeligt hierarki og tydelige strukturindikatorer samt

en grafisk enkel repræsentation, og at man skal være varsom med komplekse relationstyper (fx pladsholder). Men der er også grund til at nævne et mere generelt forhold som forekommer underbelyst i diskussionen af udflytning og nedarvning, nemlig relationstypen mellem x og y, dvs. i hvor høj grad er det nødvendigt at hente en udflyttet information for at sikre en korrekt afkodning? Der vil ofte være tale om et kontinuum mellem relevant og nødvendig afhængig af brugerens behov eller interesser. I eksemplet ”S, M, L-Fa(alm)” er ”alm” decideret nødvendig information for en korrekt afkodning. I eksemplet ”**spat** s amer. **1** gruff **2** dask” er ”s” og ”amer.” ikke på samme måde nødvendig, men kan være relevant. I en formidlingsmæssig sammenhæng er det klart at en sådan faktor må tages i betragtning i det øjeblik man overvejer at benytte udflytning eller nedarvning i sin informationsorganisering: Jo mere nødvendig, jo større omhu med evt. komprimering i form af udflytning eller nedarvning.

Spørgsmålet er imidlertid om nedarvning pr. definition er en dårlig formidlingsstrategi? Strategien hvor en umiddelbart foregående oplysning udelades i en efterfølgende parallel struktur kan ikke siges at være ukendt, og en bruger kan med rimelighed antages at kunne genkende den. Nedarvningerne i ØMO er imidlertid komplekse. Ikke kun foregår de inden for et generelt utydeligt hierarki, men – nok så meget – i nogle former og med et indhold som er meget specifikt for netop ØMO eller lignende, ældre ordbøger, og som derfor ikke kan forudsættes bekendt (overhovedet?) af en bruger. En trænet bruger af ordbøger vil måske kunne genkende brugen af pladsholder-typen, dvs. ”ds” og ”smst”, fra ordbøger af samme generation og type som ØMO, i Danmark således i ODS. Men afkodningen af nedarvningen ved angivelse af geografisk udbredelse er helt specifik for ØMO, og afkodningen kræver et detaljeret kendskab til formidlings- og komprimeringskoden i netop ØMO.

Dette minder om at nok er overskrifter og punktopstillinger alment kendte og effektive, men de gør ikke en formidling i sig selv. Den vellykkede afkodning er knyttet til et specifikt, kontekstuel bundet behov og til tyding af et indhold (i dette tilfælde et sprogligt). Ordbogsartiklen som tekst kan ikke løbe fra en række konventioner som er bundet til det indhold som netop denne teksttype komprimerer. Brugeren skal stadig have et vist kendskab til hvad ”bøjningsoplysninger” er, og brugeren afkræves også et vist forhåndskendskab til at disse oplysninger typisk står på overskriftspladsen. Brugers evne til at genkende prototypiske træk ved teksttypen ordbogsartikel, og dermed også til det indhold der formidles i denne teksttype, er med andre ord lige så vigtig for en vellykket, dvs. kontekstuel og funktionel relevant udpakning og afkodning af information. En visuel god opstilling kan ikke give en afkodning i sig selv, men den kan sikre bedst mulige eller optimale betingelser for at brugeren kan fokusere på relevant information, og den kan minimere antallet af mulige forhindringer frem mod en afkodning.

## Litteratur

### Ordbøger

- DDO = *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab. Trykt udgave bind 1-6, Gyldendal 2003-2005. Web-version <ordnet.dk/ddo> (maj 2017). App-version (iPhone, maj 2017).
- ODS = *Ordbog over det danske Sprog*. København: Det Danske Sprog- og Litteraturselskab. App-version (iPhone, maj 2017).
- ØMO = *Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer 1–* (1992–). Tillægsbind 1992. København: Center for Dialektforskning, Københavns Universitet & Universitets-Jubilæets danske Samfund.

### Anden litteratur

- Bergenholtz, Henning & Bo Svensén (1994): Systematisk inledning till *Nordisk lexikografisk ordbok (NLO)*. I: *LexicoNordica* 1, 149–185.
- Gerzymisch-Arbogast, Heidrun (1989): Standardisierte Wörterbuchartikel des allgemeinen einsprachigen Wörterbuches als Texte: Probleme der Kohärenz und der Thema-Rhema-Struktur. I: Hausman et al. (Hrsg.), 946–956.
- Grønvik, Oddrun (2016): Vitskæpeligheit og samfunnsrelevans for store ordbogsverk. I: Asgerd Gudiksen & Henrik Hovmark (red.): *Nordiske Studier i Leksikografi* 13 (*Skrifter udgivet af Nordisk Forening for Leksikografi nr. 14*), 27–61.
- Hausman, Franz Josef et al. (Hrsg.) (1989): *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Erster Teilbd. (Handbücher zur Sprach- und Kommunikationswissenschaft (HSK) 5.1). Berlin/New York: de Gruyter.
- Hausman, Franz Josef & Herbert Ernst Wiegand (1989): Component Parts and Structures of General Monolingual Dictionaries: A Survey. I: Hausman et al. (Hrsg.), 328–360.
- Svensén, Bo (2004): *Handbok i leksikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Treisman, Anne (1986): Features and objects in visual processing. I: *Scientific American* 255 (5, Nov. 1986), 114–125.
- Wiegand, Herbert Ernst (1989): Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. I: Hausman et al. (Hrsg.), 409–462.
- Wolski, Werner (1989): Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch. I: Hausman et al. (Hrsg.), 956–967.

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



# Thaipiger, muskelhunde og fulde svenskere: nedsættende ord, stereotyper og ligestilling i *Den Danske Ordbog*

*Jonas Jensen, Henrik Lorentzen, Sanni Nimb, Mette-Marie Møller Svendsen & Lars Trap-Jensen*

This paper examines the ways in which the editors of The Danish Dictionary have sought to navigate the troubled waters of activism and descriptivism. On the one hand, an editor with an activist bent may feel an urge to right a few wrongs and omit certain linguistic examples even though they are in abundance in the corpus. On the other hand, the stern descriptivist may want to represent the harsh realities of the corpus – warts and all – whether or not it aligns with the editor’s personal convictions. Additionally, (unacknowledged) personal predispositions may have a bearing on an editor’s way of defining a word and choosing appropriate linguistic examples. One might inquire whether it is at all possible for editors to maintain an objective and neutral stance when it comes to issues known to prompt firm views or opinions, e.g. gender politics or racial epithets. This paper examines the principles of dealing with such issues in The Danish Dictionary. It focuses on lemma selection, definitional practice, usage labels, and choice of lexicographical examples, and it includes a discussion of the ways in which the composition of the corpus influences the depiction of reality presented in the dictionary. In conclusion, the paper casts a critical eye on the practice of The Danish Dictionary past and present, and outlines a number of potential improvements to the current method of working.

## 1. Indledning

I denne artikel beskæftiger vi os med den følelsesbelagte del af ordforrådet. Med vores sprog kan vi krænke andre mennesker gennem vores valg af ord. Nogle gange sker det bevidst, andre gange har det den virkning uden at vi selv var klar over det eller havde det til hensigt. Ligesom med den øvrige del af sproget er det ikke konstant og absolut: Hvad der opfattes som krænkende, ændrer sig hele tiden. Det der var neutral sprogbrug for 10 år siden, kan føles krænkende i dag – og omvendt. Ofte kommer vores valg af ord og sproglig udtryksmåde meget tæt på vores identitet, køn, etnicitet, alder eller overbevisning, og derfor reagerer mange mennesker spontant og følelsesmæssigt på den slags ord. Det er også noget *Den Danske Ordbogs* redaktion får mange, ofte passionerede henvendelser om fra brugerne. Med mere end 100.000 daglige brugere af *Den Danske Ordbog* kan det næsten ikke være anderledes.

Der er næppe nogen ordbogsredaktion der opfatter sig selv som racistisk eller sexistisk. Temaet er derfor snarere hvordan vi sikrer os at ordbogens lemmabe-

stand og beskrivelsen af den holdes opdateret og fri for krænkende, skæve og utidssvarende beskrivelser. Motivationen kommer fra forskellige kilder, og det kan være nyttigt at nævne de vigtigste da det samtidig tjener til afgrænsning af emnet og problematikken.

I følgende citat vises et eksempel hvor en bruger ønsker et ord fjernet fra ordbogen.

Hej Den Danske Ordbog.

Jeg vil gerne sende en klage til jer vedrørende jeres brug af det medieopfundne ord "muskelhund", der udelukkende bruges nedsættende om særlige hunderacer, som (gennem videnskabelige undersøgelser, som jeg glædeligt kan give kildehenvisninger til) er beviseligt ligeså harmløse, som alle andre hunde. Jeres beskrivelse af ordet er ganske enkelt grotesk og angribende mod en forsvarsløs minoritetsgruppe ...

Der er, skriver brugeren, tale om "et medieopfundet, nedsættende ord om bestemte hunderacer", og vedkommende fortsætter derefter længe og engageret, men hensigten er klar: Ordet hører ikke hjemme i en ordbog, for ingen hunde er født mere farlige og aggressive end andre, og derfor kan man ikke udnævne bestemte racer til at være muskelhunde. Argumentationen er subjektiv og puristisk: Der findes gode ord, og der findes dårlige ord, og de dårlige ord bør ikke stå i ordbogen – uanset om de findes og bruges i sproget.

Den næste diskussion går ikke på om bestemte ord hører hjemme i ordbogen, men hvordan de skal beskrives når de er kommet ind. Hvem har retten til at definere hvornår et ord er nedsættende, uformelt, om det bruges som skældsord eller helt bør undgås? Det er den velkendte diskussion af politisk korrekthed: Er det afsenderens intention der vejer tungest, eller den effekt som ordet har på modtageren? Et klassisk eksempel er *neger*. Redaktionen har gennem årene været i dialog med en række minoritetsgrupper, fx seksuelle minoriteter, især transpersoner (kan man bruge *transseksuel* om personer hvis kønsidentitet er i konflikt med deres biologiske køn?), religiøse mindretal (hvem er omfattet af ordet *hedning*?) og sociale mindretal (er man kriminel hvis man kører på motorcykel i lædertøj og har mærke på ryggen, jf. *biker*, *rocker*?).

Endelig er der en sidste og lidt anden form for korrekthed som har været en motivation for os. Ruth V. Fjeld (2015) har undersøgt den kønsmæssige ligestilling i nordiske ordbøger. Undersøgelsen rejser det metodiske spørgsmål hvilket forhold der er eller bør være mellem ordbogen og det sprog der beskrives. Er det fx rigtigt at et balanceret korpus giver en kønsmæssigt mere balanceret lemmasektion? Eller mere generelt formuleret: Hvis sproget er stereotyp, fordomsfyldt og ulige, bør en deskriptiv ordbog da stadig være et spejl af sproget og den samtid sproget bliver brugt i, eller er det i ligestillingens navn netop redaktørernes opgave at være aktivistiske og rette op på uligheder på trods af sprogbrugen?

I det følgende ser vi først på lemmaudvalg, definitioner, emneområder og va-

lørangivelser, derefter på kollokationer og citater. Den afsluttende konklusion giver en række konkrete forslag til fremtidige forbedringer af *Den Danske Ordbog* (herefter DDO) på dette område.

## 2. Lemmaselektion

DDO er korpusbaseret, og et ords frekvens i korpus er afgørende for om ordet kommer med i ordbogen. Hvad lemmaselektion angår, ligger det således redaktionens hjerte nær at lemmakandidater med en vis frekvens i sproget ikke udelukkes fra optagelse i ordbogen alene fordi ordene kan virke anstødelige eller bruges nedsættende. De seneste opdateringer af DDO byder af denne grund på nedsættende, men frekvente udtryk som fx *bøsserøv* (nedsættende udtryk for homoseksuel mand). Med denne type ord følger naturligvis en forpligtelse til dels at oplyse om at brugen af dem ikke er neutral, dels at udvælge passende citater, gerne af en type der problematiserer ordenes brug.

Så længe redaktørerne lever op til denne forpligtelse, bør de nye lemmaer man udvælger, ikke umiddelbart føre til øget stereotypisering, men det forholder sig anderledes med de lemmaer der ikke slipper gennem nåleøjet. I ældre ordbøger som fx *Ordbog over det danske Sprog* (ODS) (der udkom 1918–1956) kan det i dag virke påfaldende at man skal lede meget længe efter ord der omhandler homoseksualitet på trods af ordbogens store omfang. Det er svært at finde mere end ganske få ord i det oprindelige værk (*Homoseksualisme, Homoseksualist, Homo, homoseksuel, Svansspiller*) og enkelte underbetydninger (af *Jomfru* og *Kæreste-ri*). Efterfølgende har man i de supplementsbind der udkom 1992–2005, måttet tilføje mange ord med lignende betydninger som rent faktisk fandtes i ODS' periode, men enten blev overset eller bevidst fravalgt af redaktørerne. I en moderne korpus- og frekvensbaseret ordbog risikerer bestemte ord at gå under radaren fordi de tilhører domæner der kun sjældent behandles i mainstreamkulturen og den offentlige debat. Dermed opstår huller eller lemmalakuner i forbindelse med sprogbrugen hos bestemte subkulturer, religiøse eller seksuelle mindretal etc., hvilket kan anses for et demokratisk problem: Manglende registrering af bestemte ord og udtryk kan ses som en manglende anerkendelse af de mennesker der bruger dem.

Hvis ord af denne type skal sikres optagelse i ordbogen, kræver det en midlertidig tilsidesættelse af det strengt frekvensbaserede princip. Redaktøren må altså aktivt beslutte at prioritere fx *heteronormativ* højere end det langt mere frekvente *fastansat*. Er det en forsvarlig leksikografisk praksis? På baggrund af frekvens alene er svaret "nej", men det lavfrekvente ord er nyere (kendt fra 2001), sværere at afkode end *fastansat* og knytter an til en diskurs der er oppe i tiden. Der er grund til at formode at flere brugere vil efterspørge *heteronormativ* end *fastansat* i ordbogen. DDO har heldigvis haft resurserne til at tage begge ord med – og derfor kan vi nu se at antagelsen underbygges af søgetal fra Google Analytics, idet der i perioden 17. maj – 17. august 2017 er foretaget 301 søgninger på *heteronormativ* mod 144 på *fastansat*.

### 3. Definitioner, emnekoder og valøroplysninger

Når lemmaet er udvalgt til inklusion i ordbogen – uanset om det er sket for nylig eller ordet har været i ordbogen i 25 år – er det afgørende at det har en dækkende og tidssvarende definition. Indimellem får vi henvendelser fra brugere der ikke mener at det er tilfældet. Et sådant eksempel er *skamlæber*, der bl.a. optræder i definitionen til opslagsordet *omskæring*. Som Foreningen mod Pigeomskæring har gjort redaktionen opmærksom på, findes på dansk det udmærkede og umarkerede alternativ *kønslæber*, og det bør foretrækkes, for som foreningens repræsentant påpeger “[er der ikke] efter nutidige betragtninger forbundet nogen skam ved at være født med sådanne.”

Det skyldes typisk at det der oprindeligt har været opfattet som neutralt sprogbrug, nu ofte opfattes som fordomsfuldt. Det gælder fx brugen af *skamlæber* i artiklen **omskæring**, sb., *negroid* i **sort**, adj. og *sigøjner* i **romani**, sb. Når redaktøren er blevet opmærksom på problemet, kan han eller hun nemt fremsøge de enkelte ord og opdatere ordbogen.

Men hvordan finder vi frem til problemerne i tilfælde hvor vi ikke kan støtte os til bestemte søgeord eller henvendelser fra brugerne? En metode er at søge på *emnekoder*. Ordbogsartiklerne i DDO er for størstedelens vedkommende forsynet med et element kaldet <Sysfag>, som er synligt for redaktørerne, men ikke for ordbogens brugere. Det giver mulighed for at søge på emnekoder der forventes at rumme problemer, fx den selvforklarende kode <sex> og <fam>, der står for *familie*.

Som nævnt ovenfor bør redaktørerne være særligt opmærksomme på at *valøroplysninger* optræder ved anstødelige og nedsættende ord. Også her kan emnekoderne hjælpe os med at finde de pågældende ord, fx giver ovennævnte søgning på <sex> bl.a. adjektivet *billig* i betydningen ‘for nem at komme i seng med – især om kvinde’. Artiklen bør som minimum have tilføjet valøren “nedsættende”.

En søgning på selve betegnelsen “nedsættende” i elementet <Valør> vil typisk være ubrugelig, for problemerne knytter sig kun sjældent til artikler, hvor “nedsættende” allerede optræder. En interessant undtagelse er substantivet *bøsse*, som indtil oktober 2016 var forsynet med valøren “undertiden nedsættende”. Ordet er oplagt kandidat til en faktaboks (se nedenfor), idet det oprindeligt var et skældsord, men siden er blevet overtaget af de homoseksuelle selv og i dag er så neutralt et udtryk at det optræder i navnet på den danske interesseorganisation Landsforeningen for bøsser, lesbiske, biseksuelle og transpersoner. Ordet er altså gledet ind i hverdags sproget som umarkeret, men bruges dog i visse kredse fortsat som skældsord, hvilket kunne tale for en selvstændig betydning – og altså en faktaboks.

## 4. Faktabokse, datoangivelser, brugerinteraktion

Ord der tidligere har været opfattet som neutrale, men som i dag opfattes ned-sættende eller omvendt, kræver undertiden en mere omfattende behandling end et simpelt <Valør>-felt. Efter bl.a. svensk forbillede har DDO's redaktion derfor besluttet at gøre forsøg med faktabokse.

I figur 1 ses et eksempel på betaversionen af en faktaboks under substantivet *neger*, direkte inspireret af *Svensk Ordboks* ditto:

**Om ordet 'neger'**

Tidligere var ordet 'neger' en neutral betegnelse. Man kunne fx tale om 'borgerretsbevægelsen for negrenes frihed', som det fremgår af den eksempeltekst fra 1988 der optræder under opslagsordet. I dag opfattes betegnelsen af mange som groft nedsættende, og 'sort' (eller undertiden en sammensætning med 'afro-', fx 'afroamerikaner om amerikanske forhold) foretrækkes som neutral betegnelse.

Figur 1: Udkast til faktaboks ved ordet *neger* i DDO.

I tilgift til faktaboksene arbejder vi på at indføre datomærkning af ordbogens artikler, med to datoer pr. artikel: Én der angiver hvornår artiklen er skrevet, og én der angiver hvornår seneste revision har fundet sted. Dette tiltag ville ikke i sig selv udgøre en løsning, men datomærkningen ville dog fungere som en forklaring for brugere der støder på indhold de opfatter som utidssvarende. Denne model er allerede taget i brug i onlineudgaven af ODS.

Vedrørende brugerinteraktion overvejer DDO's redaktion desuden at gøre det nemmere at kommentere de enkelte opslagsord direkte via et link eller en formular på hjemmesiden eller i appen. Det forventes dog at en sådan invitation vil medføre et betragteligt merarbejde for redaktørerne, hvorfor den endelige form overvejes grundigt inden implementeringen.

## 5. Kollokationer

Når det gælder politisk korrekthed i en ordbog, er det også relevant at kigge nærmere på ordbogens kollokationer.

I DDO's nyhedsbrev fra 1999 skriver Kjeld Kristensen at "ordbogsredaktører ikke blot beskæftiger sig med sproget, men naturligvis også med den individuelle og sociale virkelighed som sproget i brug er i vekselvirkning med" (Kristensen 1999).

Kollokationerne i sproget er et felt der kan afsløre mange tidsbundne tendenser – på godt og ondt. Fx er en typisk ordforbindelse der optrådte i DDO, "palæstinen-

siske terrorister”. Ordforbindelsen er nu blevet fjernet fra opslagsordet **terrorist**, hvor den optrådte i ental, men kan fortsat findes under **palæstinensisk**. Denne kollokation har historisk set en klar udbredelse (især mellem 2006 og 2009 optræder kollokationen hyppigt), men med tiden er brugen af den kombination blevet mindre og mindre frekvent i sproget. En søgning i DDO’s opdaterede korpus giver kun et enkelt eksempel på “palæstinensiske terrorister” i 2017.

Også kollokationen “fulde svenskere” har været ret udbredt i det danske sprog. Den genuine brug af kollokationen er faldet drastisk de senere år. Dermed er der ikke nok belæg til at retfærdiggøre at den optræder i ordbogen, og samtidig sender den et uheldigt signal. Ordbogen skal afspejle virkelighedens sprogbrug, men der er ingen grund til at skævvride forskellen mellem forskellige befolkningsgrupper eller præsentere et stereotypet billede af en nation hvis der ikke længere er belæg til at understøtte den brug.

På redaktionen har vi diskuteret forskellige løsningsforslag i forhold til problemet med især forældede kollokationer der kan optræde i DDO. Tilføjelse af en markør der senere kan bruges til at tjekke de berørte kollokationer for deres aktualitet, kunne være et middel til at komme forældede kollokationer til livs.

En synlig datering af hvornår artiklen med den pågældende kollokation er blevet skrevet, og dertil også hvornår den sidst er blevet revideret, kunne som tidligere nævnt være en hjælp for brugeren på dette punkt.

## 6. Konnotative betydninger

De konnotationer som et ord kan have, medfører undertiden også problemer når man ønsker en korrekt, og samtidig tidssvarende, definition.

### Betydninger

kvinde eller (ung) pige fra Thailand – ofte med bibetydning af prostitueret kvinde

Figur 2: Uddrag af artiklen *thaipige* i DDO.

Et eksempel er ordet *thaipige*. *Thaipiges* denotation ville lyde ‘kvinde fra Thailand’, mens en udbredt konnotation er ‘prostitueret kvinde fra Thailand e.l.’. Den konnotative betydning af ordet er meget udbredt i korpus, og derfor stiller ordet leksikografen i et dilemma når det kommer til vægtningen af sproglig og politisk korrekthed. DDO’s løsningsforslag vises i figur 2. Definitionen er den sproglige tidsånd tro: Den konnotative betydning er medtaget i opslaget, og derfor bliver det nødvendigt at tjekke opslaget løbende for dets aktualitet.

## 7. Citater

I sin oversigt over kriterierne for det gode citat skriver Lorentzen (2001:215) bl.a.: “det må ikke indeholde forstyrrende elementer der tager opmærksomheden bort fra opslagsordet” og “det bør kun indeholde anstødelige udsagn hvis man gør udtrykkeligt opmærksom på stilværdien”.

Gode kriterier, men det er værd at overveje i hvilket omfang fordomsfuld sprogbrug og stereotype udsagn er forstyrrende elementer? Og hvor går grænsen for anstødelighed? Kan den have rykket sig i DDO’s levetid? Sidst, men ikke mindst: Hvordan afvejer man hensynet til frekvens i korpus i forhold til hensynet til at undgå stereotyper, fremme ligestilling etc.?

Om kønsstereotype citater skriver Fjeld (2015:35): “Kvinner og menn fremstår fortsatt svært ulikt i ordbøkene og derved fastholdes stereotype kjønnsforskjeller.” Denne tendens er beklagelig, men hvad gør vi som leksikografer hvis den er vidt udbredt i de korpora vi anvender? Antager vi at korpus repræsenterer den sproglige virkelighed og bliver på deskriptivismens smalle sti – eller sørger vi aktivt for at gå i rette med stereotyper? Hovedparten af citaterne i DDO stammer fra ordbogens oprindelige korpus (med tekster fra perioden 1983–92), og der er dermed en risiko for at visse udsagn forekommer gammeldags eller stereotype i dag.

Følgende citater har givet anledning til diskussion, internt i redaktionen eller med ordbogens brugere. Første eksempel udspringer af en brugerhenvendelse vedrørende substantiverne *karrierekvinde* og *karrieremand* med meget enslydende definitioner. *Karrierekvinde* er forsynet med citatet: “det er ikke rigtig hende at blive gift og få børn. Vores Inge er en karrierekvinde” fra ugebladet *Se og Hør*, 1983, hvorimod *karrieremand* ikke har noget citat. Brugeren skriver:

Jeg lagde mærke til at når man taler om kvinder er der fordomme om hvad det betyder ift. at man ikke vil have mand og børn, men det siger I ikke noget om for mænd. Dvs. hvis man er karrieremand så [er der ikke nogen] forventninger ift. familie. Det lyder ikke ligestillet for mig.

Bemærk formuleringen “det siger I ikke noget om”. Brugeren opfatter citatet som noget ordbogen “siger”, og med en vis ret: Citatet er udvalgt på bekostning af andre citater, det er artiklens eneste og står dermed uimodsagt, og det fremstår ekstra kraftfuldt ved sammenligning med artiklen *karrieremand*, som er helt uden citater.

Brugerens henvendelse peger dermed på at det kan være en god idé at sammenholde lemmaer der ender på *-mand/-kvinde*, *-dreng/-pige* etc. – også på citatniveau da eksemplerne undertiden kan bidrage til en uønsket kønsmæssig skævvridning eller stereotypi.

I den trykte udgave af DDO var der som udgangspunkt kun plads til ét citat ved hver betydning, men i den digitale udgave er der ændret praksis på dette punkt, og der bringes nu typisk to citater i nyskrevne artikler. Særligt når et citat peger entydigt i en bestemt retning, bør redaktører undgå kun at give ét citat. Endelig

er vores korpus inddelt i årgange, hvilket betyder at redaktørerne med fordel kan prioritere de årgange i korpus der ligger nærmest nutiden. De ældste korpustekster stammer fra 1983 – og der er sket meget med både sprog og samfund i de 35 år der er gået siden. Dette fremgår fx tydeligt når man genser det udvalg af citater fra DDO der præsenteres af Fjeld (2015:52), og som hun med rette mener afbilder en stereotyp opfattelse af kvinder og mænd i en forældet og kønsopdelt kultur. Alle eksempler stammer fra de oprindelige tekster fra ordbogens hovedperiode 1983–92, og det står klart at vi bør gennemgå en del af dette meget store materiale. I et forsøg på at afdække problemets omfang har vi i første omgang udtrukket alle citater der indeholdt strengen “kvind\*”, og læst dem igennem. Vi har derved fundet en lang række citater som nok bør tages op til genovervejelse med henblik på at supplere med ekstra citater fra nyere tekster og i værste fald, omend nødtigt, fjerne det ældre citat. I tilfældet *karrierekvinde* ovenfor valgte vi at supplere med et mere moderne citat, men lade det ældre og tidstypiske citat stå. Her et par eksempler som vi vil behandle på samme måde:

substantivet *buste*: “enhver kvinde kan altid finde tid på dagen til at gøre sine busteøvelser for at få en smuk og fast buste” (DRTV 1983)

adjektivet *modsigelsesfyldt*: “Med kvindeligt vægelsind bevæger hun sig ud i den ene modsigelsesfyldte argumentation efter den anden” (avisen Berlingske Tidende, 1992)

verbet *ævle*: “du godeste, hvor det kvindemenneske kunne ævle” (roman, 1991)

substantivet *vås*: “jeg [mener] at ligestilling er noget vås. Mænd og kvinder er forskellige og det skal vi blive ved med at være” (avisen Jyllands-Posten, 1992)

En tilsvarende søgning på strengen “mand\*/mænd\*” i citater afslører også stereotype forestillinger om mænd, men der er tale om færre tilfælde. Nedenfor ses nogle eksempler:

adjektivet *aggressiv*: “Mænd bliver lettere aggressive, de har sværere ved at finde sig i flabede bemærkninger end kvinder” (ugebladet Alt for Damerne, 1988)

substantivet *bæ*: “Jens skiftede ham og ordnede ham og tog den første sorte bæ, som ikke ret mange mænd har set” (debatbog, 1993)

præpositionen *bag*: “Der står en kvinde bag enhver mand” (skoleopgave, 1990)



substantivet *barn*: “De kontakthæmmede, pertentlige, udi egen opfattelse ufejlbarlige mænd er på bunden børn” (fagbog om litteratur, 1985)

Udover at de stereotype citater især er af lidt ældre dato, er vi klar over at der er et særligt problem i forbindelse med bestemte kilder. Som det også fremgår af Fjelds (2015:52) eksempler, er mange af citaterne fra dameblade (*Hjemmet, Ude og Hjemme, Familie Journalen, Hendes Verden, Ugebladet Søndag*) der velvilligt leverede tekster til DDO’s oprindelige korpus. Især *Familie Journalen* er kilde til mange kønsstereotype citater der handler om “stakkels kvinder der græder”. Dette skyldes de mange trivialnoveller i bladet. Skrivestilen i ugeblade overholder i høj grad formelle sproglige krav til “det gode citat” (kort, underholdende, idiomatisk etc.), men gengiver så desværre også i særlig høj grad stereotyper omkring køn og kønsroller, se fx følgende:

“Tårene, der silede ned ad hendes yndige ansigt, var glædestårer. Hun vidste, hvad hun kørte i møde. Gyldne timer. Gyldne dage. Gyldne år” (FamJour1983)

“hun [havde ikke] noget højere ønske end at kaste sig i hans arme og knuge sig ind til ham” (FamJour1985)

“Martin, hviskede hun, og en dyb rødme skyllede op i hendes kinder” (FamJour1984)

“Julie .. fortalte sin mand, hvordan landet lå. Hun ville forlade ham og tage børnene med” (FamJour1987)

Også erotiske magasiner (*Cupido, Ugens Rapport, Tidens Kvinder*) med en tilsvarende skrivestil var velvillige bidragydere til DDO’s korpus, men dermed også til mange stereotype citater i ordbogen. Man bør være varsom med at overforbruge disse kilder, og vi påtænker at gennemgå en del af det ældre materiale herfra, evt. indsnævret til citater med pronominerne *hun/hendes* for at gøre et gennemsyn mere overkommeligt.

## 8. Konklusion

DDO – og en hvilken som helst anden betydningsordbog – bør revideres og opdateres jævnlige hvis den skal fremstå tidssvarende. Det gælder både ordbogens artikler om teknologi og artiklerne om andre emneområder under udvikling i samfundet, fx kønsroller.

Som denne afsluttende liste over mulige tiltag illustrerer, er en sådan opdatering omfattende, men ikke en umulig opgave. Således kan redaktionen bl.a.:

- Justere kriterier for udvælgelse af citater. (Er visse kilder decideret uønskede? Bør de seneste årgange prioriteres højest? Bør potentielt anstødeligt indhold i højere grad undgås i tilfælde hvor det potentielt anstødelige ikke udgøres af selve lemmaet?).
- Udstyre artikler der revideres i denne omgang, med mærket ‘tjek aktualitet’ så de er garanteret et eftersyn om nogle år.
- Finde og omformulere utidssvarende definitioner (fx ved søgning på emnekode, bestemte ord, “-kvinde/-mand” etc.).
- Tilføje faktabokse (efter bl.a. svensk forbillede) der oplyser om et ords historiske udvikling og aktuelle status (*neger* og *bøsse* er to oplagte kandidater).
- Tilføje synlig datering med én dato for *skrevet* og én for *senest revideret* (ikke en egentlig løsning, men dog en forklaring for slutbrugeren).
- Tilføje eller revidere valøroplysninger.
- Give mulighed for mere direkte brugerinddragelse (fx mulighed for at indberette/kommentere den enkelte ordbogsartikel med et enkelt klik).

I processen skal vi imidlertid huske at:

- Tage hensyn til det tidsbillede vi erstatter; tidligere versioner bør gemmes af hensyn til eftertiden og den historiske interesse (dette er en ny problemstilling i den digitale tidsalder).
- Sondre (i det omfang det er muligt) mellem det prototypiske og det stereotype.
- Søge en gylden middelvej mellem deskriptivismen og den aktive modstand mod stereotyper og fordomsfuld sprogbrug (fx ved at bringe ét citat fra hver (syns)vinkel).

## Litteratur

### Ordbøger

DDO = *Den Danske Ordbog* 1–6 (2003–05). Online: <ordnet.dk/ddo>.

ODS = *Ordbog over det danske Sprog*. 1–28 (1918–56). Online: <ordnet.dk/ods>.

*Svensk ordbok* (2009). Utgiven af Svenska Akademien. Norstedts (i distribution).

Online: <svenska.se>.

## Anden litteratur

- Fjeld, Ruth Vatvedt (2015): Om ordbokseksempler og stereotypisering av kjønn i noen nordiske ordbøker. I: Caroline Sandström, Ilse Cantell, Eija-Riitta Grönros, Pirkko Nuolijärvi & Eivor Sommardahl (red.): *Perspektiv på lexikografi, grammatik och språkpolitik i Norden*. Helsingfors: Institutet för de inhemska språken, 35–65.
- Kristensen, Kjeld (1999): Politisk (u)korrekthet i Den Danske Ordbog. I: *Nyhedsbrev fra Den Danske Ordbog* 11.
- Lorentzen, Henrik (2001): Jagten på det gode citat. Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus. I: Martin Gellerstam, Kristinn Jóhansson, Bo Ralph & Lena Rogström (red.): *Nordiska studier i leksikografi 5. Rapport från Konferens om Lexikografi i Norden*, Göteborg 27–29. maj 1999. Göteborg, 202–216.

Jonas Jensen  
redaktør  
jj@dsl.dk

Henrik Lorentzen  
seniorredaktør  
hl@dsl.dk

Sanni Nimb  
seniorredaktør  
sn@dsl.dk

Mette-Marie Møller Svendsen  
assisterende redaktør  
mette.svendsen@dsl.dk

Lars Trap-Jensen  
ledende redaktør  
ltj@dsl.dk

Det Danske Sprog- og Litteraturselskab  
Christians Brygge 1  
DK-1219 København K

# Middelaldertekster som sproglig ressource

*Ellert Þór Jóhannsson & Simonetta Battista*

The topic of this article is the medieval source material for *A Dictionary of Old Norse Prose*. We discuss the dictionary's approach to the text sources and how different types of text make up the body of citations and what information about the source material is registered. Furthermore we account for the distribution of citations between different text genres as well as time periods and geographical provenance.

## 1. Indledning og baggrund

I denne artikel diskuterer vi middelaldertekster som sproglig ressource med udgangspunkt i *Ordbog over det norrøne prosasprog* (ONP) eller *A Dictionary of Old Norse Prose*, som er et ordbogsprojekt ved Københavns Universitet. Vi redegør for arten af ONPs kildemateriale og hvilke informationer der er registreret om de enkelte middelaldertekster.

Projektet blev grundlagt i 1939 med det oprindelige navn *Den Arnamagnæanske Kommissions Ordbog*. Formålet med ONP var at registrere og belyse ordforrådet i de norrøne prosatekster, som er overleveret i norske og islandske middelalderhåndskrifter, med et værk som skulle være mere moderne, videnskabeligt og omfattende end de tidligere ordbøger fra slutningen af 1800-tallet, fx Johan Fritzners *Ordbog over Det gamle norske Sprog* (Fritzner 1886, 1891, 1896), og *An Icelandic-English Dictionary* af Cleasby & Vigfusson (1874). For et detaljeret historisk overblik cf. Battista & Jóhannsson (2014).

En oprindelig planlagt trykt publikation på 12 bind blev indstillet i 2004, efter at 3. bind var udkommet, til fordel for en digital udgave. I 2010 blev såvel udgivet som uudgivet ordbogsmateriale gjort tilgængeligt på ONP online (se nærmere Jóhannsson & Battista 2016 for redegørelse af online udgaven). Siden da er ordbogsmaterialet løbende blevet udvidet med nye ordbogsartikler udgivet på nettet, og digitaludgaven er under konstant udvikling.

## 2. Kildemateriale

ONPs kildemateriale er i sin natur begrænset og ufuldkommet, idet mange håndskrifter er gået tabt, og derfor er det kun et vilkårligt antal tekster, som er bevaret fra middelalderen. Kildematerialet kan afgrænses under tre hovedkriterier: teksttyper, geografisk oprindelse (proveniens) og tidsperiode.

## 2.1. Teksttyper

Ordbogen omfatter alle genrer af norrøne prosatekster: sagaer, jura, oversættelser osv. Finnur Jónssons reviderede udgave af *Lexicon Poeticum* fra 1931 lige var blevet udgivet, da ordbogsprojektet startede. Derfor besluttede man at fokusere på prosasproget.

## 2.2. Geografisk oprindelse (proveniens)

Den geografiske oprindelse for teksterne er Island og Norge. Sproget er vestnordisk, og håndskrifterne stammer både fra Norge og Island. Der er nogle diplomer, som muligvis stammer fra Færøerne, og som er grupperet sammen med norske tekster.

## 2.3. Tidsperiode

Tidsperioden strækker sig fra ca. 1150 indtil slutningen af middelalderen. Når det gælder tekster fra Norge går grænsen ved 1370–1400. I 1300-tallet går den sproglige udvikling i Norge hurtigt og efter 1370 afviger sproget i en sådan grad, at man ikke længere kan tale om et fælles sprog for Norge og Island. Ordbogen registrerer valgte eksempler fra norske kilder i perioden 1370–1400, men markerer dem som yngre norsk.

For Islands vedkommende strækker perioden sig til 1540. Selvom håndskriftstraditionen fortsætter i Island uafbrudt indtil 19. og 20. århundrede, begrænser ONP sig til tidsperioden fra 12. århundrede frem til 1540, dvs. året når den første islandske bog ”Det nye testamente” blev trykt. ONP har også nogle eksempler efter den tid, men de er markeret som yngre islandsk.

## 3. Principper og procedurer

Oprindeligt var ONP tænkt som en slags supplement til Fritzners store oldnordiske ordbog, *Ordbog over Det gamle norske Sprog* (Fritzner 1886, 1891, 1896), men hurtigt udvikledes det til et selvstændigt projekt. I bestemmelserne for projektet stod der, at den nye ordbog skulle forsøge at give et bredt syn på hele periodens ordforråd med repræsentative eksempler fra alle bevarede tekstgenrer og eksempler på alle kendte ord ”med særlig hensyntagen til ældste forekomst af såvel sproglige som syntaktiske fænomener og til kulturhistorisk interessant ordforråd.” (Widding 1964:21).

Ifølge ONPs redaktionelle principper skal ordbogens materiale genspejle originalkilden så nøjagtigt som muligt, hvilket betyder, at citaterne ikke er normaliserede og tit gengivet med kildetekstens ortografi. I de første mange år bestod

ordbogsarbejdet hovedsagelig i excerpering af tekster, primært fra diplomatariske udgaver. Man har også excerperet direkte fra håndskrifter, i de tilfælde hvor den pågældende tekst ikke er blevet udgivet, eller man har betragtet en udgave som upålidelig i sin gengivelse af kildeteksten.

Excerperingsarbejdet foregik med papir og blyant. Mindst to redaktører læste en tekstudgave og understregede de ord, som de syntes belyste ordets brug eller betydning og også ordformer af særlig interesse. Den indstregede udgave med markerede meningsfulde citater blev så givet videre til opskrivning på sedler, som siden blev ordnet i en fysisk seddelsamling i alfabetisk rækkefølge efter opslagsord. På den måde voksede ordbogens arkiv til nogle hundrede tusind håndskrevne sedler.

Faren ved denne arbejdsgang er, at resultatet ikke repræsenterer sproget, men bliver en samling af underligheder, mærkelige former og sjældne ord. Efter at man havde excerperet tekster i nogle år, opdagede man, at der manglede flere eksempler på almindelige ord, konjunktioner, præpositioner, pronomener og talord, men også de mest almindelige ord i sproget, som fx verberne *hafa* "have" og *vera* "være" etc. Derfor blev citatsamlingen udvidet med en komplet excerpering af alle ord fra nogle udvalgte nøgleværker, som repræsenterer forskellige genrer fra særligt relevante håndskrifter. Det drejer sig om: *Snorra Edda* (GKS 2367 4°), et fragment af *Egils saga Skallagrímssonar* (AM 162 A 0 fol), *Heiðarvíga saga* (Holm perg 18 4°, Lbs fig 1), *Íslendingabók* (AM 113 fol<sup>s</sup>) og *Ólafs saga Tryggvasonar* (Holm perg 18 4°).

Resultatet blev en stor samling citater, som opfyldte de kriterier, som blev vedtaget ved projektets begyndelse. Da citatsamlingen havde nået en tilstrækkelig mængde sedler (ca. 750.000), kunne man gå i gang med at ordne citaterne og strukturere dem i ordbogsartikler (nærmere om redigering cf. Jóhannsson & Battista 2016).

## 4. ONPs data og registrering af oplysninger

Materialet, som danner grundlag for redigeringsarbejdet, består primært af selve citaterne, mange slags forskellige supplerende informationer om hvert enkelt citat og adskillige andre oplysninger om kildeteksterne. Disse informationer, som tidligere kun eksisterede på papirsedler, er nu organiseret i en database, og en stor del af dem bliver benyttet i onlineudgaven af ONP (ONP Online). ONP har ca. 65.000 opslagsord og hvert enkelt citat er organiseret under det opslagsord, det hører til, i alt ca. 795.000 citater, inkl. de citater der løbende er blevet tilføjet.

### 4.1. Værker

Kildeteksterne består af 428 værker, som varierer i størrelse. Alle citater er hentet fra disse værker. Et værk kan meget rudimentært beskrives som en enhed af

beslægtede tekster. Mange værker er klart definerede, som fx *Bójarlög (Bl)* (*Byloven*) (cf. Figur 1). Her drejer det sig om den samme basistekst, som sandsynligvis har én oprindelse, dog overleveret i forskellige håndskrifter med mindre variationer. På samme måde er fx *Njáls saga* og mange andre sagaer, som stammer fra én forfatter, ét værk.

Men betegnelsen værk er ikke altid så klar som i disse tilfælde. ONP betragter andre tekster som samme værk, selvom de ikke er en helhed på samme måde. Fx har ONP registreret *Encyclopædica (Enc)* som ét værk, selvom det ikke er et værk af en enkelt herkomst eller af samme forfatter. Betegnelsen dækker over en vifte af encyklopædiske tekster, som har mange fællestræk. Man skal derfor være opmærksom på, at et værk i ONP kan betegne forskellige slags grupperinger af beslægtede tekster.

## 4.2. Signaturer

Alle citater hører til et bestemt værk og er markeret med en unik signatur. En signatur består typisk af værkforkortelse plus nærmere indikatorer, ofte håndskriftsforkortelser, som peger på detaljer vedr. tekstens oprindelse, efterfulgt af side og linje (cf. Figur 1). Hvert værk kan indeholde mange forskellige signaturer.

## 4.3. Register

Oplysningerne om hvilket håndskrift der ligger til grund for en bestemt signatur, finder man i ordbogens første og måske vigtigste bind, registerbindet, som udkom i 1989, og som oplister alle anvendte tekstudgaver og håndskrifter i ONPs citatsamling. Registerbindet indeholder for alle værker de tilhørende signaturer, og hvilke håndskrifter de henviser til, oplysninger om udgaver, evt. udenlandske paralleller osv. Registerne er også tilgængelige i onlineudgaven (ONP Online).

Bl	Bójarlög (Bjarkeyjarréttr hinn nýi) & réttarbótr	
<i>Bl</i>	Keyser & Munch 1848 [INGL 2] 185-274 <sup>2</sup> , 287-288 ("Ba")	AM 323 fol 34-109, 121-123 »34-123« (c1325-1350) (cf. var...: AM 323 fol)
<i>BlApp</i>	ed. cit., 289 <sup>3-17</sup> ("Na")	GKS 3262 4 <sup>o</sup> 97v-98r »1r-98r« (1576) (cf. var...: GKS 3262 4 <sup>o</sup> )
<i>BlApp</i> <sup>x</sup>	ed. cit., 289 <sup>19-26</sup>	AM 85 4 <sup>ox</sup> 244v (1543 & postea)
<i>BlFar</i>	ed. cit., 274 <sup>3</sup> -286 ("Bb")	AM 60 4 <sup>o</sup> 93vb-98va »81va-98va« (c1320) (cf. var...: AM 60 4 <sup>o</sup> )
<i>BlFarApp</i>	ed. cit., 289 <sup>28</sup> -290 ("Bh")	AM 61 a 4 <sup>o</sup> 67v-68r »52v-71v« (c1350-1370) (cf. var...: AM 61 a 4 <sup>o</sup> )
var...: AM 304 fol	ed. cit., var.app. ("Bd")	AM 304 fol »44rb-62ra« (c1320)
var...: AM 305 fol	ed. cit., var.app. ("Ob")	AM 305 fol »85ra-100vb« (c1450)
var...: AM 307 fol	ed. cit., var.app. ("Xc")	AM 307 fol »64va-72vb« (c1350)
var...: AM 309 fol	ed. cit., var.app. ("Oa")	AM 309 fol »72r-88v« (c1325)
var...: AM 315 h fol	ed. cit., var.app. ("Bo")	AM 315 h fol »1ra-1vb, 4ra-4vb« (c1300-1350)

Figur 1: Fra registerbind. Oplysninger om værket *Bójarlög* (*Byloven*) med værkforkortelsen *Bl*. I venstre spalte vises signaturene sammen med varianthensvisninger. Man kan fx se, at signaturen *BlFar* er brugt til citater fra s. 274–286 i udgaven, hvor teksten er baseret på håndskriftet AM 60 4<sup>o</sup>, som er dateret til ca. 1320 (ONP Registre: 36).

Registret er i sig selv et vigtigt værktøj og et selvstændigt opslagsværk for alle, der beskæftiger sig med oldnordiske tekster.

## 5. Genre, periode og oprindelsessted

### 5.1. Gruppering af teksttyper

Værkerne er markeret med genre eller teksttype i ONPs database. Denne opdeling bestemmes af ONPs redaktører, men følger stort set den traditionelle opdeling af vestnordisk prosalitteratur fra middelalderen. Nogle af værkerne er dog lidt svære at placere inden for en bestemt genre, så muligvis bliver genreopdelingen revideret. På nuværende tidspunkt er de 428 værker delt op i 11 genregupper:

- **Islændingesagaer** — fortællinger om islændinge (fx *Njáls saga*, *Egils saga*)
- **Totter** — korte fortællinger om islændinge
- **Legendariske sagaer** — heltesagn fra den oldgermanske fortid
- **Samtidssagaer** — (fx *Sturlunga saga* og *Byskupa sögur*)
- **Høvisk litteratur** — primært riddersagaer, dvs. høviske tekster



oversat fra oldfransk og oversættelser og originale værker inspireret af international litteratur

- **Historiske tekster** — fortællinger om de norske konger (fx *Heimskringla*)
- **Lærde tekster** — *Snorra-Edda* og diverse tekster med videnskabeligt indhold (fx *Encyclopædica*)
- **Pseudo-historisk litteratur** — Oversættelser af fortællinger om historiske begivenheder i oldtiden (fx *Rómverja saga*, *Trójumanna saga*)
- **Religiøse tekster** — hagiografiske tekster og legendelitteratur primært oversat fra latin (fx *Hómiljubók*, *Heilagra manna sögur* og *Postula sögur*)
- **Juridiske tekster** — (fx *Grágás*, *Norges gamle love*)
- **Diplomer** — breve og officielle dokumenter som belyser mange aspekter af dagliglivet i middelalderen og det relaterede ordforråd

## 5.2. Genrer, værker og citater

Der er stor forskel på, hvordan disse tekstgenrer er repræsenterede i ONPs materiale. Det er også meget forskelligt, hvor mange tekster og værker der hører til hver gruppe.

Genrer	Værker		Citater	
Religiøse tekster	122	28,5%	184.781	23,2%
Historiske tekster	32	7,5%	137.586	17,3%
Islændingesagaer	41	9,6%	95.528	12,0%
Juridiske tekster	49	11,4%	75.034	9,4%
Høvisk litteratur	48	11,2%	70.662	8,9%
Diplomer	4	0,9%	61.163	7,7%
Samtidssagaer	11	2,6%	58.085	7,3%
Pseudo-historisk litteratur	6	1,4%	20.392	2,6%
Legendariske sagaer	29	6,8%	33.759	4,2%
Lærde tekster	21	4,9%	40.950	5,1%
Totter	65	15,2%	17.333	2,2%
I ALT	428	100%	795.273	100%

Tabel 1: Proportionerne mellem forskellige genrer, opdelt både efter værker og citater. Tabellen er ordnet efter antal citater.

Det fremgår, at den største gruppe er ”Religiøse tekster”, både når det gælder antal værker og antal citater. Til denne gruppe hører 122 ud af 428 værker, som udgør 28,5% af alle definerede værker i ONPs register og 23,2% af citaterne.

Andre grupper, som springer i øjnene, er grupperne ”Totter” og ”Diplomer”. Man kan se, at hele 65 værker, men få citater, hører til gruppen ”Totter”. Det skyldes primært totternes karakteristiske korte fortællestruktur. Det omvendte gælder for gruppen ”Diplomer”, som indeholder mange forskellige slags dokumenter: fx breve, testamente, kirkeinventarer, kontrakter og vidnesbyrd. Hvert diplom udgør en tekstenhed for sig selv, men betragtes ikke som et individuelt værk. Diplomerne er opdelt i fire værker baseret på geografiske og tidsmæssige kriterier, og derfor er det svært at sammenligne dem med de andre grupper, da hvert værk består af mange hundrede eller tusinde forskellige dokumenter som er skrevet i et stort område i løbet af mange århundreder.

Det er lidt overraskende, at gruppen ”Islændingesagaer”, som repræsenterer den mest berømte genre inden for norrøn litteratur, ligger på en tredjeplads og kun udgør 12% af alle citaterne. Den beslægtede genre ’Totter’, som også indeholder fortællinger om islændinge, kunne måske også tælles med, men alligevel står islandske sagaer og totter kun for 14% af materialet (dog ca. en fjerdedel af alle værker, eller 106 ud af 428). Det viser, at selvom fokus tit er på Islændingesagaer, når det gælder middelalderlitteratur fra Norden, omfatter norrøn prosa mange forskellige typer tekster. Dette forhold genspejles i ONPs materiale.

Inden for hver af disse 11 genregrupper er det meget forskelligt, hvor meget der er excerperet fra hvert værk. Her vises detaljerne for en af grupperne, ”Samtidssagaer”, som indeholder 11 værker, og som har bidraget med 58.085 citater.

Værk	Værkbetegnelse	Citater
Arons saga	Ar	359
Árna saga byskups	ÁBp	4.935
Guðmundar saga byskups	GBp	9.805
Hrafns saga Sveinbjarnarsonar	Hrafn	1.671
Hungrvaka	Hungrv	727
Jóns saga Helga	JBp	3.489
Kristni saga	Kristni	529
Lárentíuss saga byskups	LBp	2.328
Páls saga byskups	PBp	633
Sturlunga saga	Stu	28.238
Þorláks saga Helga	ÞBp	5.371
I ALT:		58.085

Tabel 2: Samtidssagaers 11 værker og antal citater tilhørende til hvert værk.

Det viser sig faktisk, at det allerhyppigst excerperede værk i ONP hører til her, nemlig *Sturlunga saga*, hvor hele 28.000 (ca. 3,5% af alle citater i ONP) er taget fra. Det er næsten halvdelen af alle citater fra denne gruppe. Andre værker bidrager med kun nogle få hundrede, fx *Kristni saga* og *Páls saga byskups*. Det illustrerer den store forskel, som findes mellem individuelle værker, og hvor meget de

vægter i ONPs citatsamling. Der er nogle få store værker, som har bidraget med mange tusind citater, ligesom *Sturlunga saga*, men der findes også adskillige små værker, som kun har bidraget med fem eller færre citater, fx *Markúss saga postula* (2 citater), som hører til gruppen ”Religiøs litteratur”, og som er uudgivet og kun bevaret i et enkelt håndskrift.

### 5.3. Tidsperioder og citater

ONP registrerer også oplysninger om datering af kildeteksterne. Her er der stor forskel mellem århundrederne.

Periode	Citater	
1100-tallet	2.880	0,4%
1200-tallet	126.165	15,9%
1300-tallet	357.395	45,0%
1400-tallet	115.227	14,5%
1500-tallet	50.961	6,4%
Yngre materiale	142.645	17,9%
I ALT:	795.273	100%

Tabel 3: Citaternes distribution over tid.

Der er få citater fra 1100-tallet. Det er også kun nogle få håndskrifter, som er dateret til denne periode. Antallet af citater øges markant i 1200-tallet, men 45% af alle citater i ONPs citatsamling kommer fra 1300-tallet. Det er i takt med den generelle produktion af norrøne tekster, som når et højdepunkt i denne periode. Det fortsætter frem til 1400-tallet, men efter 1370 stammer citaterne kun fra det islandske materiale. I 1500-tallet strækker ONPs periode sig kun frem til 1540, derfor det tilsvarende lave tal.

Hele 18% af materialet betragtes som yngre materiale. Det drejer sig primært om tekster, som er dateret til middelalderen, men som kun er bevaret i yngre papirhåndskrifter. Et meget berømt eksempel er *Íslendingabók*, som sandsynligvis er en af de ældste bevarede tekster, affattet ca. 1200, men som kun er bevaret i papirhåndskrifter fra anden halvdel af 1600-tallet, og er derfor markeret som yngre i ONPs database.

### 5.4. Oprindelse (proveniens) og citater

Opdeling af citaterne efter oprindelsessted (Island eller Norge) viser at størstedelen af citaterne stammer fra islandske kilder. Der findes islandsk materiale fra alle 11 genrer, mens det norske materiale repræsenterer færre genrer: primært religiøse tekster, juridiske tekster og diplomer, men også historiske tekster, høvisk litteratur og legendariske sagaer.

Proveniensi	Citater	
Island	567.496	71,4%
Norge	49.631	6,2%
Island/Norge	47.912	6,0%
Umarkeret	130.613	16,4%
I ALT:	795.273	100%

Tabel 4: Citaternes opdeling efter oprindelse.

Kategorien Island/Norge betegner tekster med blandet oprindelse, mest norske tekster bevarede i islandske håndskrifter.

## 6. Konklusion

Som omtalt ovenfor er der mange forskellige slags supplerende informationer om kildeteksterne, som findes ved hvert enkelt citat i ordbogsmaterialet. Videreudviklingen af ONP sigter mod at gøre disse informationer tilgængelige og søgbare. Udover at give adgang til middelaldertekster kan ordbogens citatsamling bruges som sproglig ressource til gavn for alle der interesserer sig for nordisk sprog og kultur i middelalderen.

## Litteratur

### Ordbøger

- Cleasby, Richard & Gudbrand Vigfusson (1874): *An Icelandic-English Dictionary*. Oxford: Clarendon Press.
- Finnur Jónsson (1931): *Lexicon poeticum antiquæ linguæ Septentrionalis / Ordbog over det norsk-islandske skjaldesprog*. Opr. udg. af Sveinbjörn Egilsson. 2. udg. København: Kongelige nordiske oldskriftselskab.
- Fritzner, Johan (1886, 1891, 1896): *Ordbog over Det gamle norske Sprog* 1–3, rev. udg. Kristiania: Den norske forlagsforening.
- Fritzner Online. Johan Fritzners *Ordbog over Det gamle norske Sprog* <[www.edd.uio.no/perl/search/search.cgi?appid=86&tabid=1275](http://www.edd.uio.no/perl/search/search.cgi?appid=86&tabid=1275)> (maj 2017)
- 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 (1994). 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](http://onp.ku.dk)> (maj 2017)

## Anden litteratur

- Battista, Simonetta & Ellert Þór Jóhannsson (2014): "Ordbog over det norrøne prosasprog 2004–2014 — fra trykt udgave til netversion", *LEDA-Nyt* 57, 6–23.
- 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.
- Widding, Ole (1964): *Den Arnemagnæanske Kommissions Ordbog, 1939–1964: Rapport og plan*, Copenhagen: G.E.C.GADS Forlag.

Ellert Þór Jóhannsson  
Ordbogsredaktør, ph.d.  
nkv950@hum.ku.dk

Simonetta Battista  
Ordbogsredaktør, ph.d.  
sb@hum.ku.dk

Københavns Universitet  
Njalsgade 136  
DK-2300 København S

# Redigering af adjektiver i ONP Online

*Ellert Þór Jóhannsson, Simonetta Battista, Þorbjörg Helgadóttir & Aldís Sigurðardóttir*

In this article we discuss the editing of adjectives in the online version of *A Dictionary of Old Norse Prose*. After a short account of the editing procedure, where the focus is on individual word classes or groups, we discuss some features of the entry structure and show the entry for the adjective *langr* ‘long’.

## 1. Indledning

*Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose* (ONP) er et ordbogsprojekt ved Københavns Universitet. ONP registrerer ordforrådet i de norrøne prosatekster, som er overleveret i norske og islandske håndskrifter fra ca. 1150 til slutningen af middelalderen. Fra 1989 til 2004 udkom 4 af 12 planlagte trykte ordbogsbind (ONP 1989–2004). I 2005 blev den trykte udgave indstillet til fordel for en digital udgave. ONP Online, der blev lanceret i 2010, indeholdt det allerede trykte materiale (a–em) samt lemmaliste og uredigerede citater i resten af alfabetet (cf. Battista & Jóhannsson 2014). Siden da er redigeringsarbejdet fortsat, og ONP publicerer løbende nye artikler på websiden <onp.ku.dk>.

I denne artikel giver vi et overblik over redigeringsprincipper og artikelstruktur med særlig fokus på redigering af adjektiver. Vi redegør for nogle vigtige elementer i en ordbogsartikel med fokus på adjektivernes særlige karakteristika, og endelig viser vi en færdigstruktureret ordbogsartikel med adjektivet *langr* ‘lang’.

## 2. Redigeringsprocedure og ordgrupper

I forbindelse med digitaliseringen (ONP Online) blev der foretaget nogle ændringer i redigeringsproceduren (cf. Battista & Jóhannsson 2014). Fx blev den alfabetiske redigering af opslagsordene, som i den trykte udgave, forladt til fordel for redigering af 12 ordgrupper baseret på ordklasser og morfologisk kompleksitet. Disse 12 grupper består af simpleks substantiver (med flere end 10 citater), simpleks substantiver (med færre end 10 citater), sammensatte substantiver, verber, simpleks adjektiver, sammensatte adjektiver, simpleks adverbier, sammensatte adverbier, pronomener, talord, konjunktioner og præpositioner. Denne opdeling i grupper har vist sig at være gavnlig både med hensyn til rationalisering af redaktionsarbejdet og produktivitet.

Desuden opereres der med to redigeringsniveauer, dvs. semantisk og struk-

turelt (se nærmere fx Jóhannsson & Battista 2016:168–169). Vi er nu færdige med redigering af alle simpleks substantiver (semantisk og strukturelt), verber (strukturelt) og præpositioner (hovedsagelig strukturelt). I indeværende periode fokuserer vi på gruppen simpleks adjektiver.

### 3. Artikelstruktur

En typisk artikel i ONP indeholder mange forskellige oplysninger (cf. Battista & Jóhannsson 2014:12–13). Desuden giver onlineudgaven adgang til mange supplerende oplysninger om kildetekster og håndskrifter (cf. nærmere diskussion om onlineudgaven i Jóhannsson & Battista 2016). I dette afsnit kommenterer vi nogle aspekter, som især er relevante i forbindelse med redigering af adjektiverne. Der henvises til artiklen *langr* ‘lang’, som er gengivet i afsnit 4.

#### 3.1. Bøjningsmønster

Efter opslagsordet følger en bøjningsklamme, som indeholder relevante morfologiske oplysninger baseret på de bøjningsformer, som findes i ONPs citatsamling. Ved adjektiver angives komparativ og superlativ, eller der henvises til et andet opslagsord ved suppletiv bøjning, fx *langr* adj. [compar. *lengri*, superl. *lengstr*] og *gamall* adj. [compar. & superl. → *ellri* adj.]. Også andre træk vises, fx acc. sg. masc. i positiv ved bortfald af vokal: *gamall* adj. [*gamlan*] og *j* eller *v* foran visse endelser: *rikr* adj. [*rikjan*].

#### 3.2. Definitioner

ONP angiver definitioner på dansk og engelsk. Den primære betydning står først, og definitionerne nummereres (1, 2, 3, ...), men ikke i en hierarkisk opstilling. Alle citater, som ONP har optegnet, og som ligger under det pågældende opslagsord, bliver placeret under den relevante betydning. Til citater, som er tvetydige, kan der tilføjes en kommentar med forslag til en anden tolkningsmulighed og dermed et andet definitionsnummer.

#### 3.3. Citater og kilder

Citaterne vises i kronologisk rækkefølge under hver definition. En del af citaterne er forsynet med udenlandske paralleltekster, varianter fra andre håndskrifter og/eller redaktionelle bemærkninger. Citatsignaturen henviser til udgave og det tilgrundliggende håndskrift. Signaturen kan slås op i ONPs register, hvor der findes oplysninger om udgaven, de forskellige håndskrifter, deres datering, eventuelle udenlandske kilder osv. (cf. ONP Registre el. ONP Online).

### 3.4. Fraser

En ordforbindelse kan gengives som såkaldt frase efter relativt frie kriterier. Det kan fx være en ordforbindelse, der optræder hyppigt, en kollokation eller en frase, der kræver en selvstændig forklaring. I artikelstrukturen i 4. afsnit vises fraserne med en indledende bullet (●).

### 3.5. Afdelinger og sektioner

Hvor materialet kræver det, bruger ONP en overordnet inddeling efter syntaktiske og andre grammatiske eller funktionelle kriterier. Afdelinger betegnes ved romertal (I, II, III osv.) og sektioner ved store bogstaver (A, B, C osv.).

Afdelingerne (I, II etc.) omfatter almindelig brug, tilnavne (cogn.), fx *Bergsveinn langi* ‘Bergsvein den lange’, benævnelser (*proprial.*), fx skibsnavne, yngre islandske kilder (*y. isl.*), yngre norske kilder (*y. no.*) og topografiske betegnelser (*propr. topogr.*), fx stednavnet *Langidalr* ‘Langedal’. Kun afdeling I defineres.

Inddeling i sektioner sker i de tilfælde, hvor opslagsordet fremviser en særlig brug, som fx i forbindelse med adjektivet *langr* de forskellige former for adverbial brug: B. **lǫngu** (*dat. sg. = adv.*), C. **lǫngum** (*dat. pl. = adv.*) D. **lengrum** (*compar. dat. pl. = adv.*)

### 3.6. Konstruktionseksempler

Den syntaktiske struktur i fraserne vises ved hjælp af såkaldte konstruktionseksempler, som noteres i en klamme, og som fx kan bestå af et forkortet pronomen, et præpositionsled, type af ledsagende sætning [!] eller infinitivkonstruktion [■], fx i def. 10: **þykkir langt** (*n. sg. = adv.*) [e–m] [■] [með e–jum] ‘(ngn) synes det er kedeligt (*at + inf.*)’.

### 3.7. Præfikser

Ifølge ONPs redigeringsprincipper bliver adjektiver, der består af et præfiks + adjektiv, redigeret under simpleksadjektiver. Det drejer sig fx om præfikserne *al-*, *all-*, *jafn-* og *of-* som modificerer adjektivets betydning, fx i def. 1: *oflangr* ‘for lang’.

### 3.8. Comp., Gloss. og Litt.

Nederst i artiklen finder man nogle supplerende oplysninger om opslagsordets brug og betydning.



- Under *Comp.* finder man en liste over de sammensatte ord i ONP, hvor opslagsordet udgør sidste led.
- Under *Gloss.* henvises der til andre ordbøgers og glossarers relevante opslagsord. ONP Online har link til en af de tidligere ordbøger (Fritzner Online).
- Under *Litt.* finder man registreret litteratur om opslagsordet (hovedsagelig vedr. betydningen).

#### 4. Ordbogsartikel: langr, adj.

Som illustration på de fremlagte aspekter i 3. afsnit følger her artiklen med adjektivet *langr* ‘lang’, dog med et stærkt reduceret udvalg af artiklens ca. 650 citater. Alle henvisninger og forkortelser findes i ONPs registre og bibliografi (cf. ONP Online). Den komplette artikel bliver offentliggjort på hjemmesiden som en del af ONP Online ved næste opdatering i 2018.

**langr** adj. [*compar.* lengri, *superl.* lengstr]

I. A.

1) [e-u] (*om længde el. højde*) lang, høj

þeir hófu long suerð ok buklara *Fornk 17<sup>21</sup>*, [c1300]

gulr æ hær. langr vaxste *JBpB(2003) 59<sup>2</sup>* [c1350-1365]

oflangr var hans bukr; hann var .VIII. alna upp fra lindastad *ElisD 44<sup>33</sup>* [c1450-1475]

● **á langan** *på den lange led*

Ec vil isinn lata hocva ... scolu fiorir menn hocva <sup>f</sup>faðm a langann [*var. fadm a leingd Sv-Flat 689*; *var. fadms langt Sv81 235*] oc .iiii. a breidd *Sv 180<sup>35</sup>* [c1300]

● **langt** (*n. sg., substantiv.*) *i længden*

ver ... skipadum herbudir lx. milna langt ok breitt *AlexBr226 162<sup>26</sup>* (~ *lat. in longum Epi-stola 29<sup>9</sup>*) [c1350-1360]

2) *som når langt, langtrækkende*

Þa taka þeir aptr anat skeið, ok þa er Hvgi \*kemr til skeiðs enda ok hann snytz aptr, þa var langt kolfskot til Þialfa *SnE 55<sup>10</sup>* [c1300-1350]

● **langt** (*n. sg. = adv.*)

kom hann fjarri niðr hestinum, en sverðit hraut langt á völlinn *Rém 289<sup>7</sup>*

3) (*om afstand*) lang

þa skal þa fara þo at su (*þjóðgata*) se lengri *GulI 44<sup>18</sup>* [c1250-1300]

● **long leið**

● **langr vegr**

● **um langan veg, um langa leið** *på lang afstand*

● **it lengsta**

● **langt** (*n. sg. = adv.*) *langt, langvejs, langvejsfra*

skip morg sigldo sunnan. ok atto eigi langt til eyiarinnar *HkrFris 80<sup>7</sup>* [c1300-1325]

menn, er hingat hafa lengst sótt *Laxd 168<sup>20</sup>* [c1330-1370]

● langt af [!] *på ingen måde, så sandelig ikke*

● **langt (í) frá** [e-u] *på ingen måde, så sandelig ikke*

● **langt um** ●1 [!] *på lang afstand, langt væk, langvejsfra*

●2 [!] *i vid udstrækning*

4) [(burt) frá e-m] *som ligger (ngn) fjernt*

● **langt** (*n. sg. = adv.*)

5) *forskellig*

6) *vidtstrakt, stor*

7) (*om tid*) *lang, som varer længe, langvarig, længevarende*

● **føstudagr hinn langi**

● **gera jafn langt** [e-t ok e-t]

● **of langa leið** *en tid lang*

● **langt líf**

● **lõng stund**

● **fyrir lõngu (áðr)** *for lang tid siden, for længe siden*

● **it lengsta** *allerlængst*

● **langt** (*n. sg. = adv.*) *længe*

● **langt eftir** [!] [e-s]

8) (*om tale/udtalelse*) *ordrig, detaljeret, omstændelig, vidtløftig*

● **langt** (*n. sg. = adv.*)

● **gera langt** [um e-t]

● **hvat (...) langt** [af e-u / um !] *det korte af det lange*

● **tala/... (...) langt** [af e-m/■(e-u)] [um e-t]

● **hér þarf eigi langt um** [!]

● **hvat þarf hér/þar langt um** [!]

9) (*gramm.*) (*om bogstavs el. stavelses længde*) *lang*

10) *langtrukken, kedelig, kedsommelig*

● **þykkir langt** (*n. sg. = adv.*) [e-m] [■] [með e-jum]

11) ● **er langt** [e-m] [at e-m] (*ngn*) *længes (efter ngn)*

12) *som udgør en stor del*

● **langt** (*n. sg. = adv.*) (*intens.*) *meget*

● **langt um** [!] *meget*

13) ● **vera langt um** [e-n] [at e-u] *overgå (ngn) (mht. ngt)*

I. B. **lõngu** (*dat. sg. = adv.*)

*for længe siden, for længst*

● **lõngu (...) áðr** *længe før*

● **lõngu fyrr/fyrir** [!] *længe forud*

● **lõngu næst**

● **lõngu síðarr** *lang tid senere*

I. C. **lõngum** (*dat. pl. = adv.*)

*i lang tid, længe, ofte, jævnligt*

I. D. **lengrum** (*compar. dat. pl. = adv.*)

1) *i længere tid, længere, oftest, for det meste*

2) (*om afstand*) *længere*

I. E. **lengstum** (*superl. dat. pl. = adv.*)

*det meste af tiden, for det meste, oftest*

II. (*cogn.*)

Bergsveinn langi merkis-maðr Sverris konungs Sv 156<sup>4</sup> [c1300]

III. (*proprial.*)

(om skib)

[**var.** + nu skal skipít taka nafn af ormi þessum ok skal kalla orm en sakir uaxstar skipsíns ok léingdar þiki mer hann hæita mega ormr hinn lange] *var.* ÓT<sup>II</sup> 173<sup>6</sup>: GKS 1005 fol “D<sup>2</sup>” [c1387-1395]

IV. (*proprial.*)

(om sværd)

gyrdr suerði þui er hann kallaði Láng mikít vápn ok gott *EgM(2001)* 82<sup>6</sup> [c1330-1370]

V. (*proprial.*)

(om rod)

Hinn fyrsta (*holurt*) hefir langa rót ok þvi heitir hun long *MedMisc* 57 [c1475-1500]

VI. (*y. isl.*)

huad uier hefðum laungu fyrir sied vorum undersatum til goda *DF<sup>s</sup>* 117<sup>25</sup> (*y. isl.* 1654-1657)

VII. (*y. no.*)

skall Þorder Gunnaldzs ægha ... æin klefua iamlangher stofuone *DN<sup>r</sup>* 335<sup>3</sup> (*y. no.* 1377)

VIII. (*propr. topogr.*)

j. langa dal *JarðKálf* 19b<sup>17</sup> [c1360]

**Comp.:** af-; aftan-; alnar-; ár-; at-; austr-; dag-; endi-; end-; erma-; faðms-; ferða-; fingr-; fót-; fram-; full-; furu-; hala-; hals-; haust-; hendi-; hús-; hæri-; hæru-; hoku-; jarð-; kjaft-; kveld-; langfingrs-; lög-; mat-; mið-; nátt-; nef-; náetr-; rastar-; snoppu-; spannar-; stór-; sumar-; vár-; vetrar-; vetr-.

**Gloss.:** *Clv* item (all-) alllangr, (jafn) jafnlangr; *Suppl2*; *Fr* item alllangr, langt *adv.*, lengra *adv.*, lengrum *adv.*, lengstum *adv.*, lǫngu *adv.*, lǫngum *adv.*; *LL* jafnlangr (*cf. corr.* p. 399); *Hertz* item jafnlangr, langt *adv.*; *Hertz* *Till* jafnlangr; *Suppl4* jafnlangr; *Bin* item Lang-, Langa-, Langi, -langi, -langr; *RĒm*; *LP* item jafnlangr, oflangr; *Med*; *AH* item iamlangr, longum *adv.*; *deVr*; *Fr4* item jafnlangr, lǫngu *adv.*, oflangr; *NO* item jafnlangr, langr *sb.*, langt *adv.*, lengra *adv.*, lengrum *adv.*, lengstum *adv.*, lǫngu, lǫngum, oflangr; *ÁBLM*; (*Bl* item alllangur)

**Litt.:** Hægstad 1899 36; Falk 1914 (Waff) 54; Lind 1920-1921 [NB 8] 126-127; Björn K. Þórólfsson 1926 [ANF 42] 79; Grøtvedt 1939 104; Lindblad 1952 179; Seip 1955 25, 28, 110, 149, 170, 334; Iversen 1957 [KLNLM 2] 148; Heidermanns 1993 360-362.

## Litteratur

### Ordbøger

Fritzner Online = Johan Fritzners ordbog. (*Ordbog over Det gamle norske Sprog*) <www.edd.uio.no/perl/search/search.cgi?appid=86&tabid=1275> (maj 2017).

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 (1994). ONP 2: ban–ða (2000). ONP 3: de–em (2004). København: Den Arnamagnæanske Kommission.

ONP Online = *Ordbog over det norrøne prosasprog Online*. <onp.ku.dk> (maj 2017).

## Anden litteratur

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

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.

Ellert Þór Jóhannsson  
Ordbogsredaktør, ph.d.  
nk950@hum.ku.dk

Simonetta Battista  
Ordbogsredaktør, ph.d.  
sb@hum.ku.dk

Þorbjörg Helgadóttir  
Ordbogsredaktør, cand. mag.  
tobba@hum.ku.dk

Aldís Sigurðardóttir  
Ordbogsredaktør, mag. art.  
cws301@hum.ku.dk

Københavns Universitet  
Njalsgade 136  
DK-2300 København S

# Får flickor vara vildbasare? Om könstillhörighet och genusangivelser i SAOL och hos Dalin

*Lennart Larsson*

Starting with the two nouns *vildbasare* and *vildhjärna*, with similar meaning ('madcap') and referring to humans, this article discusses the interplay between grammatical gender and sex in A. F. Dalin's glossary from the 1850's, and different editions of the Swedish Academy Glossary (SAOL) from 1874 and onwards. The background is the transition in the Swedish language from a three-gender system of masculine, feminine and neuter to the present two-gender system with a non-neuter and neuter, first adopted in the 1923 eighth edition of SAOL. The first six SAOL editions continued to use the archaic three-gender system, while in the seventh edition (1900), a semantic gender was provided parallel to the three-gender system. As shown in this article, Dalin's glossary also demonstrates, although somewhat hesitantly and unsystematically, certain efforts to provide semantic gender.

## 1. Vildbasare och vildhjärnor

Bland sammansättningarna till adjektivet *vild* i *Svenska Akademiens ordbok* (SAOB) finns de två personbeteckningarna *vildbasare* och *vildhjärna*, där förleden anknyter till den andra huvudbetydelsen av ordet, 'ostyrig, oregerlig, okontrollerad'. Båda de här orden betecknar alltså en person som inte riktigt låter sig styras eller regeras eller kontrolleras, som på ett eller annat sätt bryter mot normer eller regler. Dock är de inte synonyma. Deras ostyrlighet eller oregerlighet är av lite olika slag, och de typiska referenterna skiljer sig också i viss mån åt: medan *vildbasare*, åtminstone numera, framför allt används om barn, så kan det i fallet *vildhjärna* lika gärna handla om vuxna.

Men hur är det med referenternas könstillhörighet? Det är den fråga som ska behandlas här, hur det ser ut dels i det faktiska språkbruket, dels i ordböckerna, närmare bestämt Dalins *Ordbok öfver svenska språket* från början av 1850-talet samt de olika utgåvorna av *Svenska Akademiens ordlista* (SAOL) från 1874 och framåt.

## 2. Könstillhörigheten i det faktiska språkbruket

Oftast är det nog så att man ser en pojke eller en man framför sig när man föreställer sig en vildbasare eller en vildhjärna. Och det är inget konstigt i det: det hör ju

till en djupt förankrad patriarkal ordning att flickor är lydiga, fogliga och snälla. Den traditionella kvinnorollen lämnar inte särskilt mycket utrymme för att vara ostyrig och oregerlig, och att bryta mot normer eller regler kan nog sägas av hävd ha uppfattats som pojknarnas och männens privilegium – även om de kanske inte gör det ostraffat så betraktas det i alla fall med visst överseende. Det är också lätt att konstatera att majoriteten av vildbasarna och vildhjärnorna i det faktiska språkbruket är av mankön. Men å andra sidan behöver man inte leta länge i källorna för att hitta kvinnliga referenter, vare sig i nutida texter eller i språkprov från tidigare århundraden. (Exemplen nedan är hämtade från dels SAOB:s samlingar av språkprov (SAOBs), dels databaserna *Svenska dagstidningar* och *Mediearkivet*.)

Till de tidigaste exemplen på kvinnliga vildbasare hör två språkprov från Aftonbladet 1842 respektive 1845 (*Svenska dagstidningar*) – alltså något decennium innan Dalins ordbok utkom. I det första, hämtat från en följetongsroman, finns repliken ”Du lilla vildbasare!” riktad till en ung kvinna (som inte uppträder riktigt så undergivet som hennes manliga uppvaktare önskar), i det andra möter vi frasen ”qvinliga vildbasare” i en operarecension (där den kvinnliga huvudpersonen till recensentens besvikelse visar sig vara mindre ädelmodig än det från början såg ut).

Bland 1900-talsexemplen märks de följande, båda excerperade från romanöversättningar där *vildbasare* står som predikativ med kvinnliga referenter som predikationsbas:

Det är mycket hälsosamt, förstås, men det gör henne allt till ett stycke vildbasare. (SAOBs 1912).

Fastän Bernadette aldrig varit någon vildbasare så lekte hon (*osv.*) (SAOBs 1942).

Som exempel från nutida tidningstext (i *Mediearkivet*) kan anföras dels Svenska Dagbladet 2009 där skribenten omnämner ”de moderlösa rödhåriga vildbasarna Anne Shirley och Pippi Långstrump”, dels Göteborgsposten 2016 där det talas om ”två vildbasare till tvillingsystrar”.

Också i fallet *vildhjärna* uppträder kvinnliga referenter redan på 1800-talet. Så t.ex. i teaterstycket *Mjölmar-fröken* av Johan Jolin från 1865 (i SAOBs) där en baron kallar sin dotter för ”Lilla yrhätta, vildhjärna!” och i ett tidningsklipp från Svenska Dagbladet 1899 (i *Svenska dagstidningar*) där vi har en beskrivning av ”hjältinnan och vildhjärnan” i en rolltolkning. Och från modernare tidningstexter (i *Mediearkivet*) kan anföras Expressen 1991 där en viss Lena kallas ”kreatören, vildhjärnan”, liksom Året runt 2015 där ”Filmaktuella Martina Haag” påstår sig ha varit ”en vildhjärna som ung”.

### 3. Könstillhörigheten i ordböckerna

Går vi så till ordböckerna får vi en delvis annorlunda bild. (Upplagorna 2–5 av SAOL har här – liksom i SAOLhist – förbigåtts eftersom de ”är i stort sett identiska” (Johannisson 1974:4) med förstaupplagan.) För *vildbasare* ser det ut så här:

Dalin: **VILDBASARE**, m. 5. (fam.) Odygdig, yr, oregerlig pojke; tygellös, utsväfvande yngling.

SAOL 1 (1874): **basare** (pl. lika), s. m.

SAOL 6 (1889): **basare** (pl. =) m.

SAOL 7 (1900): **basare** (~n) mf.

SAOL 8 (1923): **basare** (~n, äv. *-basarn*; pl. =) s.

SAOL 9 (1950): **basare** *-basar[e]n* pl. = s.

SAOL 10 (1973): **basare** s. **base** ~n *basar* s. bråkstake, ostyrig pojke

SAOL 11 (1986): **basare** s. ostyrig pojke, rustibus o. d. **base** ~n *basar* s. vildbasare

SAOL 12 (1998): **vildbasare** el. **vildbase** s. n; pl. =, best. pl. *vildbasarna* ostyrig pojke o. d.

SAOL 13 (2006): **vildbasare** el. **vildbase** s. n; pl. =, best. pl. *vildbasarna* ostyrig pojke

SAOL 14 (2015): **vildbasare** s. n; pl. ~, best. pl. *-basarna* el. **vildbase** n *-basar* ostyrig pojke

Genom att använda *pojke* och *yngling* som överbegrepp i sin definition begränsar alltså Dalin uttryckligen vildbasarnas referenter till personer av mankön. Vad gäller SAOL är det ju av tradition ingen definitionsordbok, och i de nio första upplagorna saknar ordet helt betydelseförklaringar. Men när *vildbasare* fr.o.m. den tionde upplagan 1973 ges en betydelseangivelse<sup>1</sup> så får vi också här *pojke* som genomgående överbegrepp; också i SAOL-redaktörernas ögon tycks alltså vildbasaren till sin natur vara en mansperson.

Emellertid finns det en upplaga av SAOL som på ett diskret men otvetydigt sätt öppnar för kvinnliga vildbasare. Det handlar om SAOL 7 från 1900, som med sin genusangivelse ”mf.” anger att vildbasaren kan vara inte bara *han* utan också *hon*.

Bakgrunden till de olika genusangivelserna i SAOL har att göra med svenskans övergång från ett ursprungligt grammatiskt tregenussystem med maskulinum, femininum och neutrum till vårt nutida tvågenussystem med utrum och neutrum. Redan i tidig nysvenska började de grammatiska gränslinjerna mellan maskuli-

<sup>1</sup> Tydligt ansåg redaktören Gösta Mattsson (1974:104) att *vildbasare* tillhörde de ”stilistiskt särpräglade och i vardagens språk numera ganska ovanliga ord” som behövde förklaras.

num och femininum suddas ut, och under det senare 1800-talet var det gamla tregenussystemet redan klart föråldrat; fränsett dialekterna så var det i praktiken tvågenussystemet som gällde (för en initierad beskrivning av svenskans genusproblematik hänvisas till Tegnér 1892 [1962]). Men i de sex första utgåvorna av SAOL höll man fortfarande hårdnackat fast vid det gamla. Icke-neutrala substantiv klassificerades genomgående som maskuliner eller femininer, ibland på (något som tydligt framgår hos Tegnér 1892 [1962]) ganska så osäkra eller tvivelaktiga grunder.

Att försöka upprätthålla den här föråldrade klassificeringen innebar flera problem, och ett av dem var att det inte sällan blev en konflikt mellan å ena sidan ordens grammatiska och å den andra deras semantiska eller sexuella genus. Ett exempel som Tegnér (1892 [1962:17 f.]) framhåller är *flickunge* som, trots att ordet givetvis betecknar en person av kvinnokön, ändå fick genusangivelsen ”m.” eftersom efterleden *unge* ju är ett traditionellt svagt maskulinum; det är lätt att föreställa sig att en sådan klassificering kan ha framstått som lite märklig när den inte längre hade något stöd i språkkänslan.

I fallet *flickunge* förbigår alltså de tidiga SAOL-redaktörerna konflikten mellan grammatiskt och sexuellt genus med tystnad. Men i andra fall har de uppenbarligen sett sig tvungna att förklara sig. Ett exempel är det traditionella femininet *vakt* där man i de tidiga upplagorna tillfogat att ”i bet. *vaktkarl*, *vaktande soldat* o. d. användes ordet som m.”. (En intressant sak här – och det är någonting som i sig kan ses som ett uttryck för den rådande könsmaktsordningen – är att den här typen av klargörande bara tycks förekomma när det gäller grammatiska femininer som används om manliga individer; något motsvarande när det gäller grammatiska maskuliner har jag inte lyckats hitta.)

Så var alltså fallet i de sex första utgåvorna av SAOL. Och det var först i och med SAOL 8 1923 som man hittade fram till en enkel lösning på problemet: man nöjde sig helt enkelt med att ange genus i form av bestämdhetsartikeln (se ovan). I SAOL 7 1900 valde man däremot en betydligt mer komplicerad variant, ett försök att kombinera det obsoleta tregenussystemet med den nya tidens krav; detta genom att dels införa en fjärde genuskategori reale, dels kombinera de tre beteckningarna ”m.”, ”f.” och ”r.” (dvs. ”reale”) med varandra för att därigenom också ange semantiskt eller sexuellt genus vid sidan av traditionellt grammatiskt. För personbeteckningarnas del beskrivs det så här i förordet till SAOL 7 (s. VIII):

I fråga om naturligt kön användas fortfarande beteckningarna *m.* (maskulinum) och *f.* (femininum). Vid sådana ord som *släkting*, *artist*, om hvilka kan sägas både *han* och *hon*, skrives *mf.*

På det här sättet fick alltså genusangivelserna en semantisk funktion vid sidan av den grammatiska. Och här har vi då förklaringen till genusangivelsen för *vildbasare*: ordets referenter kan vara både män och kvinnor.

Att närmare undersöka vilka personbeteckningar i SAOL 7 som fått angivel-



sen ”mf.” – eller i några enstaka fall ”fm.” med femininum placerat först – och vilka som bara har ”m.” eller ”f.” är intressant av flera skäl (för en mera ingående redovisning hänvisas till Larsson 2017). Dels därför att det ger en intressant bild av kvinnans ställning i samhället kring förra sekelskiftet (eller mer exakt: ordlisteredaktörens bild av hennes ställning), dels därför att man får en inblick i ett lexikografiskt arbete av synnerligen komplicerat och mödosamt slag. I princip var alltså redaktören tvungen att för varje enskild personbeteckning ta ställning till referenternas könstillhörighet, tvungen att tänka efter om det handlade om män eller kvinnor eller bådadera. Och det kan inte ha varit lätt, inte minst med tanke på att det handlade om en tid av stora samhälleliga förändringar, en tid då ”[d]efinitionerna av gränserna för manligt och kvinnligt var skiftande och ständigt disputerade, fluktuerande” (Wikander 2006:80). Det är också uppenbart att SAOL 7 innehåller ett stort antal inkonsekvenser och förbiseenden i det här avseendet. Oftast handlar det om att redaktören oreflekterat övertagit de tidigare utgåvornas grammatiskt grundade genusangivelser, framför allt att personbeteckningarna har ett ”m.” där det utifrån den explicit uttryckta avsikten i SAOL 7 tveklöst borde ha varit ”mf.”. Men motsatsen förekommer också, att vi har ett ”f.” där det rimligtvis borde ha varit en kombinerad genusangivelse. Och till den kategorin måste man räkna *vildhjärna*, som ser ut så här i ordböckerna:

Dalin: **VILLHJERNA**, f. 1. Svärmare, fantast.

SAOL 1 (1874): **-hjerna** (pl. *-or*), s. f.

SAOL 6 (1889): **-hjärna** (*-or*) f.

SAOL 7 (1900): **-hjärna** f.

SAOL 8 (1923): **-hjärna** vard.

SAOL 9 (1950): **-hjärna** vard.

SAOL 10 (1973): **-hjärna** vard.

SAOL 11 (1986): **-hjärna** vard.

SAOL 12 (1998): **vildhjärna** s. (vard.)

SAOL 13 (2006): **vildhjärna** s. (vard.)

SAOL 14 (2015): **vildhjärna** s. *-n -hjärnor* (vard.) vild person; till *vild 2*

Här hittar vi alltså betydelseangivelser i två av ordböckerna, Dalin och SAOL 14, och båda gångerna – till skillnad från vad som var fallet med *vildbasare* – utan att angivelserna begränsar referenternas könstillhörighet. Och när det gäller SAOL 7 så har vi ett enkelt ”f.” där det enligt de principer som finns uttryckta i förordet – ”vid sådana ord [...] om hvilka kan sägas både *han* och *hon*, skrives *mf*” – en dubblerad genusangivelse utan tvekan hade varit på sin plats.

#### 4. Dalin och genusproblematiken

Även om det traditionella tregenussystemet är förhärskande också i den äldsta av ordböckerna, Dalins, så gäller det inte helt och hållet: som Sven-Göran Malmgren (1988:201) påpekat finns det ansatser till en semantiskt grundad genusklassificering även hos Dalin. Problematiken berörs i ordbokens förord där Dalin skriver som följer (s. 17):

För öfrigt bör nämnas, att i öfverensstämmelse med flera språkkännare, bland andra den berömde Rask, för flera substantiver antagits ett genus commune (tvekön), d. v. s. på en gång både maskulin och feminin, i ordboken betecknad med *c*. Att sådana till betydligt antal finnas i språket, läser icke kunna bestridas, och detta förhållande framträder helt uppenbart i dagligt tal, i det vi aldrig utmärka sådana substantiver med det personella pronomen *han*, *hon*, utan alltid med det demonstrativa *den*. Häruti ligger på en gång en ovisshet och ett dunkelt erkännande af tvekönets tillvaro i ordet.

Vad Dalin egentligen menar med ”ett genus commune (tvekön), d. v. s. på en gång både maskulin och feminin” är inte helt lätt att säga. Läser man vidare tycks det nämligen inte handla om ord som är både maskulina och feminina utan snarast varken eller, dvs. sådana som vi ”utmärka” inte med ”det personella pronomen *han*, *hon*, utan alltid med det demonstrativa *den*” och alltså motsvarande det som i SAOL 7 gavs den nya genusbeteckningen reale.

Så även om Dalin gör vissa ansatser till att bryta upp från det traditionella tregenussystemet präglas det helt klart av den ovisshet och dunkelhet som han nämner i sitt förord. Vad gäller beteckningen ”*c*” för genus commune har jag vid en fritextsökning på Dalin hittat sammanlagt 195 sådana. Till största delen handlar det om sakbetecknande substantiv, men i drygt en fjärdedel av fallen, närmare bestämt 53, är det frågan om personbeteckningar, från *baccant* och *bannlyst* till *vän* och *älskling*. Och här rör det sig alltså om ord där man *inte* skulle använda *den* som pronomen utan *han* eller *hon* beroende av kön. Dessutom finns det i själva lemmalistan ytterligare en variant för ”erkännande af tvekönets tillvaro” som inte nämns i förordet, nämligen en kombinerad genusangivelse ”*m. o. f.*”. Och i de fall det inte rör sig om pronomen eller räkneord o.d. utan om substantiv handlar det nästan uteslutande om personbeteckningar, sammanlagt 20 stycken.

Att de här angivelserna av sexuellt genus hos Dalin bara handlar om trevande och osystematiska ansatser framgår på flera sätt. För det första är det sammanlagda antalet personbeteckningar som fått angivelserna ”*c*” eller ”*m. o. f.*” bara 73, vilket är mindre än en tiondedel av de med ”*mf.*” eller ”*fm.*” i SAOL 7 – trots att det också i SAOL 7 finns ytterligare ett stort antal personbeteckningar som utan tvekan borde ha fått en sådan dubblerad angivelse om förordets ambition hade tillämpats konsekvent (se vidare Larsson 2017).

För det andra framgår det av de här angivelsernas fördelning i alfabetet. Påfal-

lande när det gäller personbeteckningarna med "c." är att det inte alls finns några beteckningar som börjar på *a*, medan de på *b* däremot omfattar så många som elva; därefter är de förhållandevis jämnt fördelade över alfabetet, som mest sex på *s* och fem vardera på *l* och *t*.

Ser man däremot på angivelsen "m. o. f." så börjar tolv av de 20 angivelserna, från *allierad* till *avliden*, på *a*, fem på *d* och så en vardera på *b*, *e* och *f*. Här anar man alltså en ambition som fanns inledningsvis men under *b* övergavs till förmån för angivelsen "c.", sedan blommade upp lite under bokstaven *d* men därefter snabbt utslocknade, medan angivelsen "c." hade sin blomstring under *b* och sedan mer sporadiskt kom att användas fortsättningsvis.

Än tydligare blir Dalins tvehågsenhet och inkonsekvens när det gäller personbeteckningarnas genus om man närstuderar vissa ord. Ett exempel är det ovan nämnda *vän* som alltså har angivelsen "c.". Ser man till de 27 sammansättningarna med *vän* som efterled så har *blomstervän* "c." och *dansvän* "m. o. f.", medan de övriga 25, t.ex. *bokvän*, *frihetsvän*, *konstvän*, *människovän* och *ovän*, däremot bara har "m." I ett fall finns det emellertid ett inslag i definitionen som närmast ter sig som en komplettering av genusangivelsen. Det gäller *barnvän* vars första betydelsemoment lyder så här:

En, som älskar, hyllar barn. Ordet brukas äfv. om kvinnor.

I stället för tillägget att ordet även brukas om kvinnor kan man ju tycka att det hade varit enklare att ha genusangivelsen "m. o. f.".

Ett annat exempel är ordet *unge* som är klassificerat som "m." trots att det ju helt klart också inbegriper flickor och i ett av betydelsemomenten rentav förklaras just som "Flicka". Av de 16 personbetecknande sammansättningarna med *unge* finns två med angivelsen "c.", *sötunge* (vars första moment förklaras som "Säges om vackra flickor" och det tredje har "Glädjeflicka" som synonymdefinition) samt (kanske lite oväntat) *tjuvunge*. Annars är det nästan genomgående "m.", också i ett fall som *smekunge* vilket definieras som "Barn ell. flicka, som man håller mycket af och ständigt smeker". Men det finns ett undantag och det är det redan nämnda *flickunge* som ser ut som följer:

FLICKUNGE, 2. pl. – *ungar*. (fam.) Liten flicka.

Här har alltså Dalin utelämnat genusangivelsen, något som annars knappast alls tycks förekomma i hans ordbok.<sup>2</sup> Det kan förstås vara en slump, ett rent misstag eller förbiseende. Men det skulle också kunna vara ett medvetet utelämnande för att undslippa den för Dalin – med tanke på hans gryende medvetenhet om det gamla genusystemets ohållbarhet – nog ganska svårsmälta angivelsen "m.". För

<sup>2</sup> Att *flickunge* är placerat i ordbokens supplement ("Tillägg och förbättringar") saknar betydelse i sammanhanget; också där är Dalin i övrigt mycket noggrann med att ange ordens genus.

eftersom referenten i det här fallet *måste* vara av kvinnokön hade ju ”c.” eller ”m. o. f.” varit malplacerade här; och ett ”f.” hade kanske på ett alltför flagrant sätt brutit mot det etablerade grammatiska tregenussystemet.

## Litteratur

### Ordböcker

- Dalin, A. F. (1850–55): *Ordbok öfver svenska språket*. Stockholm: Svenska bokförlaget/Norstedts. På nätet: <<http://spraakdata.gu.se/saolhist/>>.
- SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien* (Svenska Akademiens ordbok) (1893–). Lund. På nätet: <<http://saob.se>>.
- SAOL = *Svenska Akademiens ordlista 1–14* (1874–2015). På nätet: SAOLhist. <<http://spraakdata.gu.se/saolhist/>>.
- SAOL 7 = *Ordlista öfver svenska språket utgiven af Svenska Akademien* (Svenska Akademiens ordlista) 7 uppl. (1900). Stockholm: P. A. Norstedt & söners förlag. På nätet: <<http://runeberg.org/saol/7/>>.

### Övrig litteratur

- Johannisson, Ture (1974): Ordlistans första nio upplagor – en historik. I: Ture Johannisson & Gösta Mattsson: *Svenska Akademiens ordlista under 100 år*. (Skrifter utgivna av Svenska språknämnden 55.) Stockholm: Esselte studium, 1–54.
- Larsson, Lennart (2017): Vilken plats hade kvinnan? Om grammatiskt och sexuellt genus i SAOL:s sjunde upplaga 1900. I: *Språk och stil*, ny följd 27, 49–66.
- Malmgren, Sven-Göran (1988): Almqvist, Dalin och den svenska definitionsordbokens födelse. I: Red. Gertrud Pettersson: *Studier i svensk språkhistoria*. (Lundastudier i nordisk språkvetenskap A 41.). Lund University Press, 195–213.
- Mattsson, Gösta (1974): Ordlistans tionde upplaga – riktlinjer och kommentarer. I: Ture Johannisson & Gösta Mattsson: *Svenska Akademiens ordlista under 100 år*. (Skrifter utgivna av Svenska språknämnden 55.) Stockholm: Esselte studium, 55–116.
- Mediearkivet (Retriever Research). <<http://web.retriever-info.com/services/archive/>>.
- SAOBs = Samlingar till Svenska Akademiens ordbok. Lund.
- Svenska dagstidningar*. Sök bland svenska dagstidningar. Kungl. Biblioteket. <<http://tidningar.kb.se/>>.
- Tegnér, Esaias (1892 [1962]): Om genus i svenskan. I: *Svenska Akademiens handlingar 1891*. [Nytryck 1962.] Stockholm.

Wikander, Ulla (2006): *Kvinnoarbete i Europa 1789–1950. Genus, makt och arbetsdelning*. Stockholm: Atlas Akademi.

Lennart Larsson

docent i nordiska språk, ordboksredaktör

SAOB, Institutionen för nordiska språk vid Uppsala universitet

[lennart.larsson@svenskaakademien.se](mailto:lennart.larsson@svenskaakademien.se), [lennart.larsson@nordiska.uu.se](mailto:lennart.larsson@nordiska.uu.se)

# Finska språkets lånordsförråd i ljuset av Martti Rapolas ordsamling

*Petri Lauerma*

Martti Rapola (1891–1972) was a respected scholar of Finnish dialects and Old Literary Finnish. He also studied 19th century literary Finnish and compiled a large lexical corpus of this period. This paper deals with mainly Swedish loan words in the Rapola collection. Most of the loan words in the collection concern areas such as administration, literature, technology, philosophy, mathematics, economy, natural sciences, music and politics. Due to the somewhat puristic language planning politics in Finland, a mere half of these loan words remained in use in modern Finnish, especially in the areas of philosophy, politics, physics and chemistry. In the late 20th century, however, many loan words came back into use.

## 1. Martti Rapola och hans ordsamling

Martti Rapola (1891–1972) var på sin tid en av de ledande forskarna i finska dialekter och äldre finskt skriftspråk. Rapola forskade också på det finska ordförrådets utveckling på 1800-talet. Han publicerade många artiklar (t.ex. Rapola 1941–1946) och ett urval av första belägg på finska ord (Rapola 1960), men hans mest ambitiösa forskningsprojekt om denna period blev inte slutfört. På 1930- och 1940-talet samlade Rapola material om olika begrepps utveckling på 1800-talet, en period då ordförrådet utvecklades mycket snabbt. Rapolas excerpter är baserade på olika slag av skönlitteratur, sakprosa och ordböcker. Hans källor omfattar över 900 publikationer. Hela samlingen, som finns bevarad vid Institutet för de inhemska språken, omfattar ungefär 44 000 ordsedlar. Största delen av dessa är ordnade enligt betydelse: också synonymer är placerade tillsammans, vanligen i ett kuvert.

Utifrån detta material har jag planerat och utarbetat en databas på cirka 5600 ord, där tyngdpunkten ligger på begrepp som har fått sina nuvarande betydelse och form på 1800-talet. De flesta begreppen motsvaras av helt finska avledningar (41 % av materialet), sammansättningar (45 %), ordfogningar (7 %) eller arvord (2 %), men 5 procent av materialet är lånord.

## 2. Lånordens identifiering, anpassning och ursprung

De flesta lånorden i materialet är så unga, att de är lätta att identifiera utgående

från ofinska fonetiska särdrag (cf. Karlsson 1976:15). Finskan har t.ex. inga initiala konsonantförbindelser eller tonande klusiler, men många lånord börjar med sådana (t.ex. *barrikadi* 'barrikad', *protokolla* 'protokoll'). Även inuti ord är klusilerna *b*, *g* – och till största delen också *d* – främmande för den finska fonotaxen. Bristen på vokalharmoni hör likaså till lånordens typiska särdrag (t. ex. *kasööri* 'kassör'). Några ord har bibehållit originalspråkets kvantitativa drag (t.ex. *graadi* 'grad', *kopiia* 'kopia') eller ortografiska konvention (t.ex. *c*, *q* och *ph*: *cassa* 'kassa', *equationi* 'ekvation', *photographia* 'fotografi'), som inte längre var typiska för 1800-talets finska i allmänhet (cf. Lauerma 2007:224, 331–334).

Äldre lånord har anpassats till finska språkets fonotaktiska system på så sätt att de har förlorat sina initiala konsonantförbindelser, tonande klusiler eller andra främmande särdrag (Karlsson 1976:22–23). Man kan se en sådan utveckling även i några av 1800-talets lånord (t.ex. *pankrutti* 'bankrutt', *trottningi* 'drottning', *asvaltti* 'asfalt', *komelia* 'komedi'). Men många lånord, som i originalspråket slutar på konsonant har fått slutvokal enligt den kanoniska formen för finska ord. Den allra vanligaste slutvokalen är *i*. I äldre lånord var *a* en typisk slutvokal, men i 1800-talets finska är *a* redan tämligen sällsynt. Lånord med *a* som slutvokal har ofta varianter som slutar med *i* (jämför t. ex. *arkiiva* – *arkiivi* 'arkiv', *biblioteka* – *biblioteekki* 'bibliotek'). I dessa fall pekar *i* på svenska, *a* på ryska som ursprungsspråk (< архив, genitiv -a; библиотека). Det finns också några ord med ovanliga varianter som slutar på *o* (*materialismo* 'materialism', *realismo* 'realism', *sosialismo* 'socialism'), men dessa ord är trots det knappast lånade från italienska.

De flesta lånord har kommit från svenskan, direkt eller indirekt. Svenskan har också förmedlat en stor andel av ordförrådet från andra europeiska språk. Det finns ju några ryska lånord i materialet (t.ex. *kanava* < канава 'kanal', *kapakka* < кабак 'krog', *kansellista* < канцелярист 'kanslist', *leima* < клеймо 'stämpel', *tyrmä* < тюрьма 'fängelsehåla'). Ryska ord var vanligare i 1800-talets skriftspråk än i dagens finska, eftersom Finland var en del av det ryska riket på 1800-talet (om ryska lånord se Plöger 1973). Det finns i materialet också några ord som kan ha kommit direkt från latinet (t.ex. *docens* < *docens* 'docent', *mytologia* < *mythologia* 'mytologi').

### 3. Semantiska grupper av lånorden

Det semantiska spektret av 1800-talets lånord domineras av vissa livsområden och vetenskapsgrenar. En del av dessa lånord är fortfarande i bruk, en del har blivit ersatta med finska avledningar eller sammansättningar och en del har försvunnit.

Semantisk grupp av lånord	Frekv. i 1800-talets finska	Frekv. i modern finska	Semantisk grupp av lånord	Frekv. i 1800-talets finska	Frekv. i modern finska
Administration	42	19	Akademiska livet	9	4
Litteratur	33	13	Språk	8	2
Teknik	29	15	Religion	7	3
Filosofi	20	17	Psykologi	6	3
Matematik	14	5	Medicin	6	1
Ekonomi	12	6	Pedagogik	6	4
Fysik och kemi	11	8	Botanik och zoologi	6	4
Geografi och geologi	11	8	Restaurangliv	6	4
Musik	10	6	Bildkonst	5	3
Politik	10	9	<b>Tillsammans</b>	<b>251</b>	<b>134</b>

Tabel 1: Lånordens grupper och frekvenser i materialet.

Den största gruppen består av olika slags administrativa termer. Dessa är benämningar på olika sammanslutningar (t.ex. *directioni* 'direktion', *instruktioni* 'instruktion', *programmi* 'program'), personer som arbetar inom dem (t. ex. *roodmanni* ~ *raatmanni* 'rådman', *skriivari* 'skrivare'), olika slags av kollektiv (t.ex. *komiteetti* ~ *kommittee* 'kommitté') och produkter av arbete (*arkiiva* ~ *arkkiivi* 'arkiv', *stempeli* 'stämpel'); det finns också några militära termer (*artilleria* 'artilleri', *fästninki* 'fästning'). Administration kom till Finland från andra länder. Nästan hälften av alla administrativa lånord är fortfarande i bruk (t.ex. *protokolla* 'protokoll', *rekisteri* 'register'; *amanuessi* 'amanuens', *kanslisti* 'kanslist', *kenraali* 'general', *ministeri* 'minister'; *parlamentti* 'parlament', *raati* 'råd', *senaat* 'senat'; *leima* 'stämpel').

I den näst största gruppen finns litteraturvetenskapliga lånord (t.ex. *litteratur* 'litteratur', *roman* 'roman', *stroofi* 'strof') samt ord som har att göra med publikationsprocesser av tryckalster (*korehtuuri* 'korrektur', *prenumeratio* 'prenumeration') och teater (*aktöri* 'aktör', *deklamationi* 'deklamation', *komediantti* 'komediant'). Det stora antalet av sådana lånord är inte överraskande, om man beaktar att hela den finskspråkiga skönlitterära kulturen utvecklades först på 1800-talet. Mer än tredjedel av alla dessa lånord är fortfarande i bruk, särskilt många litteraturvetenskapliga genrebenedämningarna (*balladi* 'ballad', *draama* 'drama', *komedia* 'komedi', *lyriikka* 'lyrik', *novelli* 'novell', *romaani* 'roman').

Den tredje största gruppen av lånord består av tekniska termer (t.ex. *daguerreotyp* 'daguerrotyp', *fotografeerata* 'fotografera', *telegrammi* 'telegram', *velosipedi* 'velociped'). Tekniska företeelser blev allt vanligare under 1800-talet och många tekniska lånord har blivit ersatta med finska ord (sådana är t.ex. sammansättningen *valo/kuva* 'fotografi' och avledningen *sähke* 'telegram'), men hälften



av lånorden finns kvar (t.ex. *barometri* 'barometer', *barrikadi* 'barrikad', *kiikari* 'kikare', *ruuvi* 'skruv', *sirkkeli* 'cirkelsåg').

Den fjärde största gruppen består av de filosofiska lånorden. De flesta av dessa ord är ännu i bruk (t.ex. *idealisti* 'idealist', *logiikka* 'logik' och sådana grundläggande termer som *metodi* 'metod', *teoria* 'teori' och *termi* 'term'). Det är bara *diskussioni* 'diskussion', *individi* 'individ' och *opinion* 'opinion' som inte hör till det nufinska ordförrådet.

Också matematiska lånord var allmänna i 1800-talets finska (t.ex. *addeerata* 'addera', *multiplikationi* 'multiplikation', *tärninki* 'tärning'). Men bara *ellipsi* 'el-lips', *geometria* 'geometri', *nolla* 'noll', *statistiikka* 'statistik' och *trigonometria* 'trigonometri' är fortfarande i dag matematiska termer i finskan.

Det finns också lånord som har att göra med ekonomi och handel (t.ex. *auction* ~ *auksjoni* ~ *avisjoni* 'auktion', *pankrutti* 'bankrutt', *puukhollari* 'bokhållare'). Hälften av dessa lånord används i dagens finska (*kassa* 'kassa', *krediitti* 'kredit', *konkurrenssi* 'konkurrens', *konkurssi* 'konkurs', *piljetti* 'biljett' och *puoti* 'bod').

Finskan har också lånat många ord inom vetenskaper som fysik och kemi. Den största delen av dessa ord är ännu i bruk (t.ex. *alkoholi* 'alkohol', *fysiikka* 'fysik'). Numera kan också *elementti* 'element' och *skaala* 'skala' igen användas i finskan, men *graadi* 'grad', *elektrisiteetti* 'elektricitet' och *syyrä* 'syre' har blivit ersatta med finska ord.

Av de geografiska och geologiska lånorden finns också den största delen kvar (t.ex. *laava* 'lava', *horisontti* 'horisont'). Som undantag kan nämnas orden *juveeli* 'juvel', *globi* 'glob', jordklot' och *madjaari* 'madjar, ungrare', medan *geografia* 'geografi', *ekvaattori* 'ekvator' och *kolonia* 'koloni' har kommit tillbaka.

Också olika slags musikerter lånades (t.ex. *kompositöri* 'kompositör', *tuoni* 'ton, melodi'). Av dessa finns bara *piano* 'piano', *konsertti* 'konsert', *nuotti* 'not' och *rytmi* 'rytm' kvar, och ord som *akordi* 'ackord' och särskilt *melodia* 'melodi' har igen blivit populära.

Politiska termer, särskilt benämningar på nya politiska ideer och deras anhängare, är ofta lån (t.ex. *autonomia* 'autonomi', *konservatiivi* 'konservativ', *liberalismi* 'liberalism'). Nästan alla dessa lånord har bevarats. Bara *republiikki* 'republik' är numera ersatt av sammansättningen *tasa/valta*.

På samma sätt finns det många lånord bland akademiska termer (t.ex. *disputationi* 'disputation', *studentti* 'student'). Man använder fortfarande ord som *dosentti* 'docent', *kandidaatti* 'kandidat' och *lisensiaatti* 'licentiat' på finska, även *akatemia* 'akademi' i vissa kontexter.

Det finns också språk(vetenskap)liga och grammatiska lånord i materialet (t.ex. *grammatiikka* 'grammatik', *puntti* 'punkt'). Finska uttryck har ersatt nästan alla dessa ord, utom orden *ortografia* 'ortografi' och *vokaali* 'vokal'.

Det religiösa och teologiska ordförrådet utvecklades till största delen redan i det äldre skriftspråket, så i materialet finns ganska få sådana ord (t.ex. *katolikki* 'katolik' och *missionäri* 'missionär'). *Muslimi* 'muslim', *mytologia* 'mytologi' och *reformaatio* 'reformation' tillhör fortfarande det finska ordförrådet.

Kännetecknande för den ifrågavarande perioden är att det i materialet finns nästan lika många lånord inom det religiösa området som inom det psykologiska området (t.ex. *miini* 'min', numera *ilme*; *passiooni* 'passion, mani', numera *into/himo*). Av de psykologiska lånorden blev *psykologia* 'psykologi' i slutet av 1800-talet *sielu/tiede* (ordagrant "självetenskap"). Numera talar man åter bara om *psykologia* och använder lånord som *intelligentti* 'intelligent' och *intresantti* 'intressant'.

Det finns också några medicinska lånord i materialet (t.ex. *hospitaali* 'sjukhus', *vaccinationi* 'vaccination', *vaksineerata* 'vaccinera'). Bara ett gammalt uttryck för hospital, *lasaretti* < *lasarett*, är ännu i dag möjligt som ett humoristiskt ord.

Av de pedagogiska lånorden i materialet är benämningar på olika skolor fortfarande i bruk som (skol)historiska termer. Sådanan ord är *kymnaasi* (i materialet också *kymnasiumi* ~ *kymnasio* ~ *kimnasia*) 'gymnasium', *lyseo* (~ *lyseumi* ~ *lykeiona* ~ *lysea* ~ *lykeio* ~ *lysei*) 'lyceum' och *seminaari* (~ *seminariumi*) 'seminarium'. *Gymnastiikki* 'gymnastik' förekommer inte längre, men *sportti* 'sport' har kommit tillbaka.

Det finns också några botaniska och ett zoologiskt lånord i materialet. Av dessa ord är *ananassi* 'ananas' numera också i finskan bara *ananas*, *kryyti* 'krydda' är bara *ryyti* och *puumuli* 'bomull' *pumpuli*. *Basti* 'bast' och *träkolsmestari* 'träkolsmästare' har försvunnit, men *zoologia* 'zoologi' har kommit tillbaka.

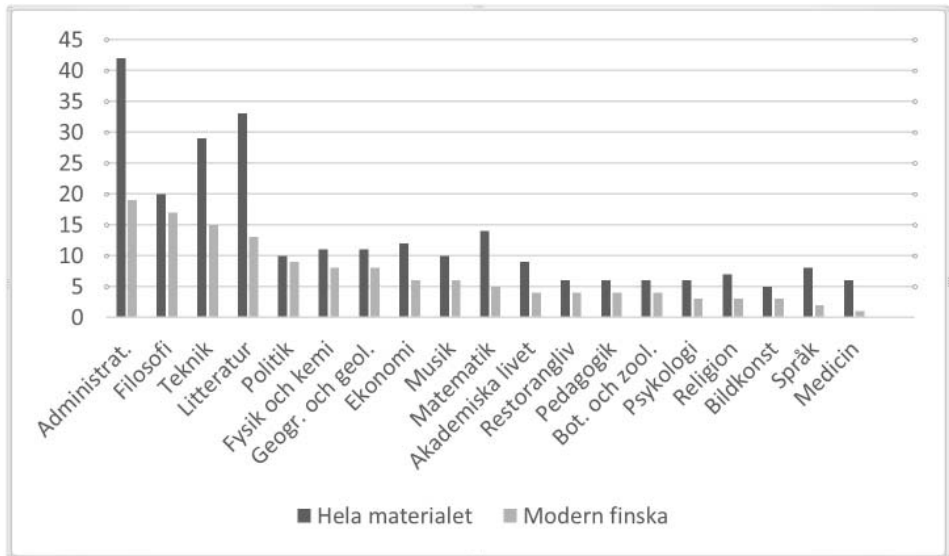
Mera vardagliga är ord som hör till matlagning och restaurangliv. Av dessa har *kryydätä* 'krydda' har blivit ersatt med avledningen *maustaa* och *restaurationsi* 'restaurang' med avledningen *ravintola*. Andra benämningar på olika slags restauranger är ännu i bruk: 'krog' är på finska ännu *krouvi* eller ett ryskt lån *kapakka*. Ordet för vin, *viini*, med sitt finala *i* är från 1800-talet; tidigare drack finnarna bara *viina*, som betydde både 'vin' och 'sprit'. Ordet för 'kök', *kyökki* (numera vanligen *keittiö*), är ett äldre lån från svenska.

Inom bildkonsten finns det färre lånord i materialet. Alla dessa ord blev senare ersatta med finska motsvarigheter (*artisti* 'artist' och *konstnären* 'konstnär' > *taitelija*; *bildhugari* 'bildhuggare' > *kuvan/veistäjä*; *portretti* 'porträtt' > *muoto/kuva*; *reliefi* 'relief(bild)' > *korko/kuva*), men *reliefi*, *portretti* och *artisti* har kommit tillbaka.

#### 4. 1800-talets lånord och modern finska

Enligt den nufinska ordboken *Kielitoimiston sanakirja* finns mer än hälften av lånorden (53%) i Rapolas material kvar, men många av dessa ord har blivit vardagliga eller humoristiska (t.ex. *aviisi* 'avis', *piletti* 'biljett', *potretti* 'porträtt'). Svenska lånord var typiska för de stora tvåspråkiga städernas talspråk, särskilt i språkbruk mellan tjänarinnor och deras matmödrar. Sådant språk (som man började kalla *kyökki/suomi* 'kökfinska', Lehtinen 1990) var inte respekterat och lånorden har därför senare fått humoristiska konnotationer.

Lånorden har inte heller bevarats lika bra i alla semantiska grupper, men de har bevarats bättre i vissa vetenskapsgrenar. De svarta staplarna i figur 1 visar absoluta frekvenser för lånorden i databasen. De grå staplarna visar, hur många av dessa ord som fortfarande kan användas i finskan.



Figur 1: Lånorden i hela materialet och i modern finska.

Mera än tre fjärdedelar av de filosofiska lånorden finns fortfarande kvar. De filosofiska termernas betydelse är ofta så abstrakt att det kanske har varit svårt att hitta goda finska ekvivalenter. De viktiga etymologiska konnotationerna hos många filosofiska termer gör det lättare att använda originalord i stället för översättningar (Arppe 2012). Men man bör också notera att i 1800-talets Finland skrevs inga filosofiska läroböcker för små skolbarn eller för allmänheten, och därför behövde man inte översätta alla filosofiska termer till finska, på samma sätt som inom mer praktiska discipliner.

Det finns fortfarande kvar många administrativa lånord. Men nästan hälften av dessa ersattes med finska ord när den officiella finskan utvecklades i slutet av 1800-talet. Orden för internationella politiska ismer och deras anhängare har kvarstått som internationella lånord. Det är också lätt att förstå, att ordförrådet i fysik och kemi fortfarande innehåller många lån. Även benämningarna på centrala litteraturvetenskapliga begrepp är fortfarande lånord, trots att de ord som hör till områden som tryckning och publicering av böcker har förfinskats, liksom de flesta tekniska orden i materialet. Andelen lånord har minskat mest inom praktiska områden.

Å andra sidan finns nästan inga av de medicinska och bara få av de botaniska och språkvetenskapliga lånorden i materialet kvar. Den grundläggande finska ter-

minologin för medicin började utvecklas på 1830-talet och för botanik på 1840-talet. En stor del av dessa termer är ännu i bruk. (Koivusalo 1975:10–11, 15; Pitkänen 2008:273, 280.) Man utvecklade finsk terminologi också för språkvetenskap, men denna blev bestående bara i skolundervisningen, inte i forskningen. Finska termer blev mer allmänna i skolvärlden än i akademiska kretsar. Inom matematik var finskspråkiga dissertationer ovanliga.

Antalet lånord har minskat inom alla discipliner i materialet. Men många lånord, som på 1800-talet hade blivit ersatta med finska avledningar och sammansättningar har kommit tillbaka på 1900-talet, nu mestadels som en följd av det engelska språkets inflytande. Man kan se denna process även i många vetenskapsbenämningar (cf. Latikka 2013:199; Koukkunen 1992). Det finska ordförrådet utvecklas åter i en mer internationell riktning. Detta betyder också att en stor del av de finska ord som introducerades på 1800-talet har marginaliserats eller försvunnit.

## Litteratur

### Ordböcker och ordsamlingar

*Kielitoimiston sanakirja*. Nätpublikation. Kotimaisten kielten keskuksen verkkojulkaisuja 35. Senaste uppdatering 28.2.2017.  
<<http://www.kielitoimistonsanakirja.fi/>> (använt 6.9.2017).

Martti Rapolas samling av 1800-talets finska ord. Institutet för de inhemska språk, Helsingfors.

### Annan litteratur

Arppe, Tiina (2012): De la traduction de la philosophie. I: *Traduire* 227, 29–34.  
<<http://traduire.revues.org/469>>.

Karlsson, Fred (1976): *Finskans struktur*. Lund: LiberLäromedel.

Koivusalo, Esko (1975): Lääketieteen suomen kieli aatteiden ja asenteiden muokkauksessa. I: *Duodecim*, 16–30. Helsinki: Kustannus Oy Duodecim.

Koukkunen, Kalevi (1992): Suuretieteestä matematiikkaan. [Summary: From "quantity science" to mathematics]. I: *Arkhimedes* n:o 3, 245–252, n:o 4, 295–305. Helsinki: Suomen fyysikkoseura. & Suomen matemaattinen yhdistys.

Latikka, Anne-Maria (2013): Psykologinen kirjallisuus. I: K. H. Riikonen & Outi Paloposki (ed.), *Suomennetun tietokirjallisuuden historia 1800-luvulta 2000-luvulle*, 196–213. Suomalaisen Kirjallisuuden Seuran Toimituksia 1189. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Lauerma, Petri (2007): Vieraiden grafeemien väistyminen varhaisnykysuomesta etenkin Jakob Johan Malmbergin tuotannon valossa (Summary: The abandonment of foreign graphemes in Early Modern Finnish with special reference to Jakob Johan Malmberg's works). *Virittäjä*, 322–345. Helsinki: Kotikielen seura.
- Lehtinen, Raija (1990): Kelpo kyökkisuomea. I: Maija Länsimäki (ed.), *Sana kiertää: Kirjoituksia suomen kielestä*, 336–340. Kotimaisten kielten tutkimuskeskuksen julkaisuja 80. Helsinki: Kotimaisten kielten keskus.
- Pitkänen, Kaarina (2008): *Suomi kasvitieteen kieleksi. Elias Lönnrot termistön kehittäjänä* [Abstract: Development of botanical Finnish. Elias Lönnrot as the creator of new terminology]. Suomalaisen Kirjallisuuden Seuran Toimituksia 1164. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Plöger, Angela (1973): *Die russischen Lehnwörter der finnischen Schriftsprache*. Veröffentlichungen der Societas Uralo-Altaica 8. Wiesbaden: Otto Harrassowitz.
- Rapola, Martti (1941–1946): Lisiä kirjasuomen sanaston historiaan 1–11 [Résumés: Contributions à l'histoire du vocabulaire de la langue finnoise littéraire]. *Virittäjä*. Helsinki: Kotikielen Seura.
- Rapola, Martti (1960): *Sanojemme ensiesiintymiä Agricolaista Yrjö-Koskiseen*. Tietolipas 22. Helsinki: Suomalaisen Kirjallisuuden Seura.

Petri Lauerma  
specialexpert, fil. dr., docent i finska språket  
Institutet för de inhemska språken  
Hagnäskajen 6  
FI-00530 Helsingfors  
Finland  
petri.lauerma@kotus.fi

# Sproressourcer kontra ordbøger: Hvad er bedst?

*Patrick Leroyer & Henrik Køhler Simonsen*

This article discusses whether language resources can replace dictionaries and whether there is a measurable quality of the texts translated and produced by means of dictionaries or language resources, respectively. We carried out a controlled experiment involving twenty test subjects divided into two groups. One group had access to four specially selected online dictionaries and the other group had access to online language resources. On the basis of the analysis of the screen recordings and the texts we were not able to identify significant differences in the quality of the texts translated and produced, and we conclude that the use of dictionaries over language resources does not seem to result in higher text quality.

## 1. Indledning og problem

Leksikografien som teori og praksis er under voldsom transformation, og som det fremgår af f.eks. Simonsen (2016), står leksikografien som videnskab og praksis i en disruptiv tid.

Der er især tre grunde til, at ordbøger synes at være under pres. For det første påvirker en række trends brugen og opfattelsen af ordbøger. Som det fremgår af f.eks. Kernerman (2013), går ordbøger i stigende grad fra håndgribelige til virtuelle artefakter, fra menneske til maskine, fra indhold til teknologi, fra ord til sprog og fra leksikografi til leksikografikologi. Kernerman (2013:7) går endda så vidt til at sige med reference til en af Frank Zappas gamle sange, at "one could say dictionaries are not dead, they just smell funny".

For det andet synes der at være et stort videnskabeligt underskud i relation til brugerundersøgelser. Vi har ganske vist været vidne til en række nye brugerundersøgelser, som det fremgår af f.eks. ENeL (2017), som er en survey om brugen af monolingvale ordbøger, og af f.eks. Simonsen (2011) og (2014), som handler om datatilgang på mobile enheder. Men der er indtil nu ganske få empiriske studier i ordbogsbrug i forbindelse med især oversættelse til og fra et fremmedsprog eller tekstproduktion på fremmedsprog, se også Lew (2016), og der er os bekendt heller ikke blevet lavet kontrastive undersøgelser af forskellige hjælpemidler, ligesom der heller ikke er blevet lavet undersøgelser, hvor brugerne forinden er blevet instrueret i ordbogsbrug. Vi har heller ikke kunnet finde resultater fra eksperimenter, hvor man har gennemført studier af udvalgte værktøjer, som matcher den påtænkte brugersituation. Samtidig synes der at være en vis modstand mod undersøgelser af denne type, som det f.eks. fremgår af Tarp (2009).

For det tredje synes der at være et fald i brugen af ordbøger inklusive virk-

somhedsordbøger og termbaser, som det fremgår af Bundgaard (2017). Bundgaard påviser i en helt ny undersøgelse, at brugere, der arbejder professionelt med fagoversættelse i stigende grad synes at fravælge ordbøger, inklusive virksomhedsordbøger, og vores egen undersøgelse her påviser da også, at selv studerende synes at bruge Google som det foretrukne værktøj, fordi Google straks giver kontekstuelle oplysninger om ordenes brug. Sammenfattende synes der således at være en række tungtvejende grunde til at øge forskningen på dette felt.

I dette bidrag undersøger vi således, hvorvidt sproglige ressourcer (f.eks. automatisk stavekontrol, Google, Google Translate etc.) kan erstatte ordbøger, og hvorvidt man kan måle forskelle i kvaliteten af tekster produceret med henholdsvis ordbøger og sproressourcer. Vi diskuterer på basis af resultaterne fra et kontrolleret pilotforsøg, hvorvidt ordbogen er ved at miste sin betydning på grund af faktorer som ændrede behov, nye teknologier, eller fordi folk ikke længere har brug (eller føler de ikke har brug) for ordbøger. De to forskningsspørgsmål, som vi adresserer er:

- A. I hvilket omfang kan sproressourcer erstatte online-ordbøger i forbindelse med oversættelse fra modersmål til fremmedsprog og i forbindelse med tekstproduktion på fremmedsprog?
- B. I hvilket omfang kan man måle kvaliteten af oversættelser fra modersmål til fremmedsprog og i forbindelse med tekstproduktion på fremmedsproget med hhv. online-ordbøger og sproressourcer?

## 2. Metode

Vi designede et kontrolleret forsøg med to grupper forsøgspersoner. Forsøgspersonerne blev rekrutteret fra bacheloruddannelsen *English and Organizational Communication* på CBS i efteråret 2016.

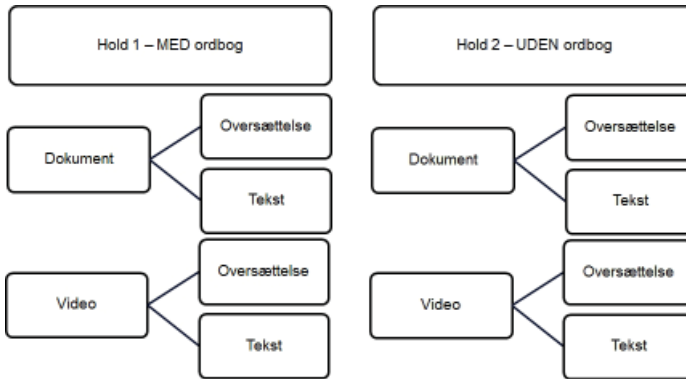
Det ene hold fik forudgående instruktion i anvendelsen af fire ordbøger, der passede nøje til den undersøgte brugersituation og fik dermed adgang til disse ordbøger, se også V&B, FAG, ØKON og OED. Det andet hold måtte kun bruge sproressourcer på nettet.

Der var 10 forsøgspersoner på hvert hold, men vi vil i dette bidrag kun inddrage data fra 10 tilfældigt udvalgte forsøgspersoner. På basis af de erfaringer vi fik før, under og efter undersøgelsen gennemfører vi i efteråret 2017 en større og forbedret undersøgelse på CBS. Her vil vi også udvide undersøgelsen af testpersonernes skrive- og søgeprocesser.

Både Hold 1 (MED ordbog) og Hold 2 (UDEN ordbog), se også Figur 1 herunder, fik til opgave at løse udvalgte dele af en tidligere (ukendt) eksamensopgave, og forsøgspersonerne blev bedt om at:

- A. Oversætte 7–8 linjer fra dansk til engelsk
- B. Producere en e-mail på engelsk på basis af dansk oplæg
- C. Aflevere oversat og produceret tekst
- D. Optage og aflevere videofilm af arbejdsprocessen vha. skærmoptager.

En oversigt over artiklens empiriske data ses i Figur 1.



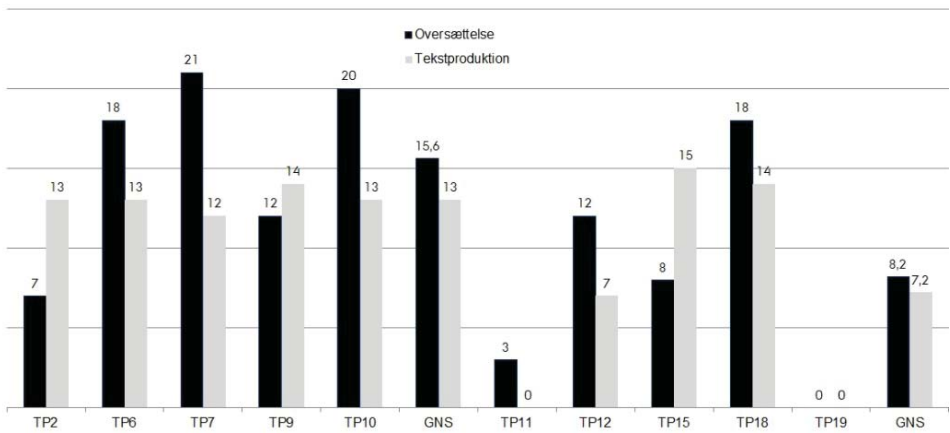
Figur 1: Oversigt over empiriske data.

Vi definerede en række målepunkter og har i dette pilotforsøg især interesseret os for hhv. processen, dvs. ”hvordan har forsøgspersonen arbejdet” og for produktet, dvs. ”hvad har forsøgspersonen leveret”. Vi har kunnet måle på hhv. ”hvordan” og ”hvad”, idet vi for hver af de 20 forsøgspersoner har kunnet analysere en video (dvs. vi har kunnet analysere ”hvordan”) og et produkt (dvs. vi har kunnet analysere ”hvad”) på basis af en foruddefineret rettevejledning. For hver af de to hold definerede vi 10 målepunkter, men vi vil her kun behandle fire målepunkter nemlig ”proces antal opslag/aktiveringer”, ”proces tid forbrugt”, ”produkt grammatisk korrekthed” og ”produkt ækvivalens”. Vi vil nedenfor også diskutere korrelationer ml. det anvendte værktøj, opgave, antal opslag, tid forbrugt, grammatisk korrekthed og ækvivalens, ligesom vi i analysen og diskussionen vil inddrage kvalitative kommentarer på basis af analysen af både videoer og tekster.

### 3. Resultater og analyse

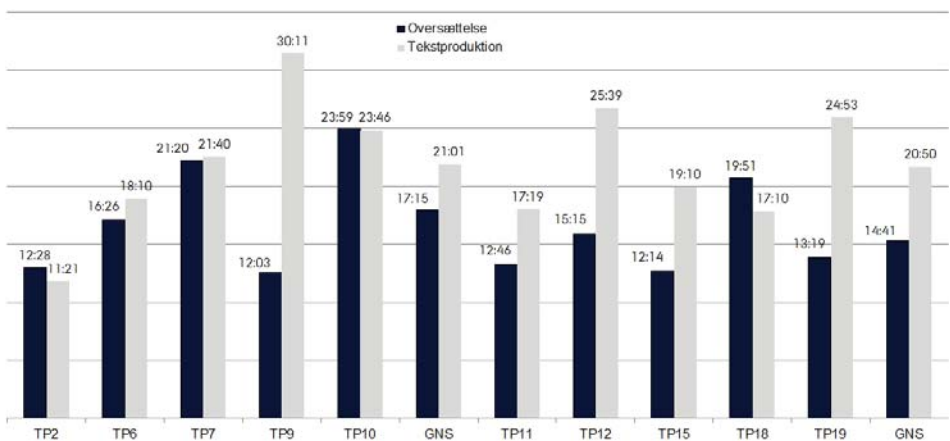
Analysen af både processen og produktet vha. ti tilfældigt udvalgte videoer og tekster syntes at pege på en række interessante resultater. Testperson 2, 6, 7, 9, og 10 udgør Hold 1 (MED ordbøger) og testperson 11, 12, 15, 18 og 19 udgør Hold 2 (UDEN ordbøger). Forsøgspersonerne er forkortet TP (Test Persons) og vises på X-aksen. Den mørke søjle er oversættelse og den lyse søjle er tekstproduktion. Figur 2 herunder viser antal opslag/aktiveringer ved både oversættelse og tekstproduktion og fokuserer på data fra processen.





Figur 2: Antal opslag/aktiveringer – proces begge grupper.

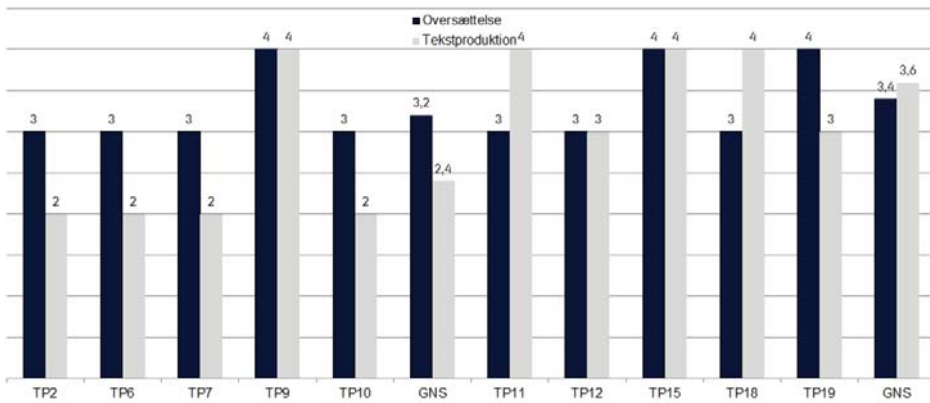
Som det fremgår af Figur 2 foretager Hold 1 (MED ordbog) omtrent dobbelt så mange opslag som Hold 2 (UDEN ordbog) i forbindelse med både oversættelse og tekstproduktion, og det ses, at Hold 2 (UDEN ordbog) generelt foretager ganske få opslag/aktiveringer. Tallene er 15,6 opslag vs. 8,2 opslag ved oversættelse og 13 vs. 7,2 opslag ved tekstproduktion. Forskellen kan formentlig skyldes, at forsøgspersonerne på Hold 1 (MED ordbog) i højere grad end forsøgspersonerne på Hold 2 (UDEN ordbog) er opmærksomme på, og måske mere tilbøjelige til at slå ord op i ordbogen, og dermed er der en potentiel bias i forsøget, fordi forsøgspersonerne ligesom er blevet præpareret til at skulle anvende ordbøger. Et andet perspektiv fremgår af Figur 3 herunder, som viser tid forbrugt ved både oversættelse og tekstproduktion, og igen fokuseres der her på data fra processen.



Figur 3: Tid forbrugt – proces begge grupper.

Som Figur 3 viser, har forsøgspersonerne brugt totalt set næsten lige meget tid. Dog har Hold 1 (MED ordbog) brugt marginalt længere tid på oversættelse end Hold 2 (UDEN ordbog) – 17:15 sekunder vs. 14:41 sekunder. Tallene for tekstproduktion er næsten identiske – 21:01 vs. 20:50 sekunder. Dermed kan vi på basis af målinger af tid heller ikke se den store forskel på, hvorvidt man bruger ordbog eller sprogresstype.

Det interessante er imidlertid at korrelere vores observationer fra processen med vores observationer fra produktet. Figur 4 herunder viser en flig af de indsigter, som vi fik i forbindelse med analysen af forsøgspersonernes oversættelser og tekster. Figur 4 viser nemlig en måling af den grammatiske korrekthed i de afleverede tekster vurderet på basis af en foruddefineret rettevejledning og over 20 års erfaring med undervisning i engelsk grammatik samt oversættelse og tekstproduktion på engelsk.

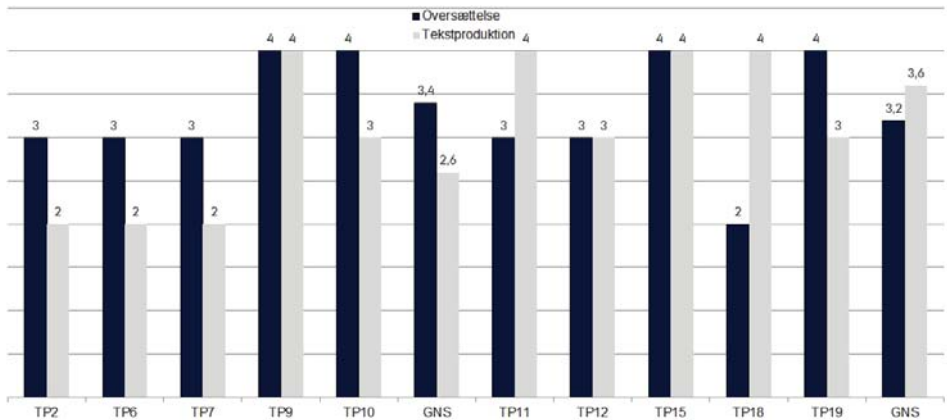


Figur 4: Grammatisk korrekthed – produkt begge grupper.

Som det fremgår af Figur 4 har Hold 2 (UDEN ordbøger) en marginalt højere, dvs. en lidt bedre, grammatisk korrekthed i deres tekster end Hold 1 (MED ordbøger), hvilket i sig selv er interessant, fordi man kunne forvente, at teksterne fra især Hold 1 (MED ordbog) burde have haft en langt højere grammatisk korrekthed, idet de jo har haft adgang til grammatiske oplysninger i ordbøgerne. Den største forskel ses i forbindelse med tekstproduktion, hvor den grammatiske korrekthed er størst, dvs. 2,4 for Hold 1 (MED ordbog) vs. 3,6 for Hold 2 (UDEN ordbog). Dette resultat synes derfor at bidrage til diskussionen af grammatiske oplysninger i ordbøger, og der synes også at være behov for mere forskning på dette punkt.

Denne observation understøttes af målinger af ækvivalenskvaliteten i testpersonernes tekster, hvilket vises i Figur 5 herunder. Nu er diskussionen af ækvivalens i sig selv omdiskuteret, særligt også i leksikografien, men i vores vurdering har indgået både leksikalsk, syntaktisk og tekstuel ækvivalens. Som det fremgår af Figur 5 har Hold 1 (MED ordbøger) haft en marginalt bedre ækvivalens i deres

oversættelser i forhold til Hold 2 (UDEN ordbøger), dvs. 3,4 for Hold 1 (MED ordbog) vs. 3,2 for Hold 2 (UDEN ordbøger). Billedet er dog markant anderledes for de producerede tekster, hvor Hold 2 (UDEN ordbøger) har en noget bedre ækvivalens end Hold 1 (MED ordbog), dvs. 2,6 for Hold 1 (MED ordbog) vs. 3,6 (UDEN ordbog). At de testpersoner, som har brugt sprogressourcer synes at have en bedre ækvivalens kan formentlig skyldes, at sprogressourcerne indeholder konkrete og situationsafhængige, kontekstuelle oplysninger, som testpersonerne har kunnet anvende i tekstproduktionssituationen.



Figur 5: Ækvivalens – produkt begge grupper.

Analysen af processen viste også, at Hold 2 (UDEN ordbog) for både oversættelse og tekstproduktion anvendte flere værktøjer end Hold 1 (MED ordbog), hvilket kan forklares med at Hold 1 (MED ordbog) kun havde lov til at bruge fire specielt udvalgte ordbøger.

## 4. Diskussion

Analysen synes at pege på to overordnede konklusioner. For det første at antal og type af værktøjer ikke synes at have indflydelse på kvaliteten af oversættelser lavet af Hold 1 (MED ordbog) eller oversættelser lavet af Hold 2 (UDEN ordbog), og for det andet at antal og type af værktøjer heller ikke synes at have indflydelse på kvaliteten af tekster produceret af hverken Hold 1 (MED ordbog) eller tekster produceret af Hold 2 (UDEN ordbog).

Undersøgelsen påviste således, at ordbøger i dette eksperiment ikke gjorde den store forskel og at anvendelsen af ordbøger frem for sprogressourcer således IKKE synes at kunne ses mærkbart på kvaliteten af hverken oversættelse eller tekstproduktion.

Undersøgelsen viste også, at de generelle IT-kompetencer synes at have indfly-

delse på hvor hurtigt begge hold forsøgspersoner løste opgaver, og de mange videooptagelser viste endvidere, at testpersonerne på begge hold i høj grad anvendte de integrerede stavekontroller og synonymordbøger i Microsoft Word, og at de i mange tilfælde simpelthen undlod at foretage opslag og/eller aktivere søgninger i sprogressourcer.

Analysen af de mange skærmoptagelser viste også, at forsøgspersonerne på Hold 2 (UDEN ordbog) i stor udstrækning anvendte sprogressourcerne til både at svare på kommunikations- og vidensorienterede spørgsmål - også selvom det i udgangspunktet kun var et kommunikationsorienteret spørgsmål. Flere af testpersonerne startede endda med at stille Google vidensorienterede spørgsmål som "Hvad er en væge" i forbindelse med oversættelsen af fagtermen *væge*. Alt dette er på flere områder særdeles interessant og stiller spørgsmålstejn ved den måde leksikografien traditionelt laver ordbøger.

Skærmoptagelserne viste også, at testpersonerne på Hold 2 (UDEN ordbog) især brugte Google Translate og derefter Google til at validere oversættelsesforslag, hvilket for så vidt er en fornuftig fremgangsmåde.

Teknologien synes dog ikke at få folk til at koble hjernen fra, som man ellers kunne have frygtet. Man kunne ligeledes frygte, at de studerendes indlæring muligvis påvirkes i negativ retning, eftersom brugen af online sproglige ressourcer, der fejlagtigt tildeles autoritet, i et vist omfang anvendes ukritisk. Det gælder i særdeleshed for automatiske oversættelsesressourcer, der kræver en hel del (post-)editing, som det fremgår af Bundgaard (2017).

Samme frygt kan i øvrigt næres i relation til ureflekteret og ukritisk brug af ordbøger, for så vidt de studerende tit og ofte ikke har fået den nødvendige træning i professionel ordbogsbrug. Flere af forsøgspersonerne søgte endvidere også i sprogressourcer uden hensyn til sprogretning og skiftede mellem sprogene for at finde ækvivalente svar.

Analysen viste desuden også en lang række andre resultater af mere kvalitativ art. Det mest interessante var nok at se, hvordan testpersonerne helt konkret arbejdede med at tilgå, udvælge, tjekke og anvende leksikografiske data. Brugerundersøgelser af denne art, hvor man helt konkret kan se, hvordan brugerne arbejder, synes at være særligt værdifulde for forståelsen af brugersituationen og dermed også for udviklingen af nye typer værktøjer.

## 5. Konklusion

På basis af dette lille kontrollerede eksperimentet var det ikke muligt at identificere nogen større forskel i kvaliteten på de oversættelser eller tekster, som forsøgspersonerne producerede med ordbøger eller sprogressourcer. Undersøgelsen viste således, at sprogressourcer umiddelbart sagtens kan erstatte online-ordbøger.

Når det er sagt er det dog vigtigt at anføre, at der skal meget mere forskning til på dette felt. Der er således stor brug for mere forskning i brugen af – og i ikke-

brugen af – ordbøger, og der er brug for mere procesforskning i tekstproduktion – både med og uden leksikografiske hjælpemidler med henblik på at få en større forståelse for brugssituationen. Derfor gennemfører vi i efteråret 2017 en større og forbedret undersøgelse på CBS, hvor vi ser endnu nærmere på testpersonernes skrive- og søgeprocesser.

Eksperimentet her synes også at pege på behovet for forskning i, hvordan leksikografiske data af høj kvalitet med fordel kan inkorporeres i forskellige typer af applikationer og AI-systemer, herunder writing assistants.

Noget tyder dog på, at brugerne ikke tager teknologien blindt til sig, men bevarer en kritisk distance. Det lover godt for fremtiden.

## Litteratur

### Ordbøger og andre leksikografiske ressourcer

ENeL = European Network of e-Lexicography. <<http://www.elexicography.eu/>> (august 2017).

FAG = Thomas Arentoft Nielsen, Charlotte Langkilde, Jørgen Høedt: *Gyldendals Røde Ordbøger – Fagordbog*, København: Gyldendal. <<http://ordbog.gyldendal.dk.esc-web.lib.cbs.dk>> (august 2017).

V&B = Vinterberg & Bodelsen: *Gyldendals Røde Ordbøger – Vinterberg & Bodelsen*, København: Gyldendal. <<http://ordbog.gyldendal.dk.esc-web.lib.cbs.dk>> (august 2017).

ØKON = Annemette Lyng Svensson: *Økonomisk Ordbog Engelsk-Dansk/ Dansk-Engelsk*. Samfundslitteratur. <<http://ordbog.gyldendal.dk.esc-web.lib.cbs.dk>> (august 2017).

OED = OED Oxford English Dictionary. Oxford University Press: OED <<http://www.oed.com.esc-web.lib.cbs.dk/>> (august 2017).

### Anden litteratur

Bundgaard, K. (2017): (Post-editing) *A workplace study of translator-computer interaction at Textminded Danmark A/S*. PhD dissertation. Aarhus: Aarhus Universitet, Department of Management.

Kernerman, I. (2013). *Kernerman News*, Number 21, July 2013, 1–2.

Lew, R. (2016): Can a Dictionary Help you Write Better? A user Study of an Active Bilingual Dictionary for Polish Learners of English. I: *International Journal of Lexicography* 29 (3), 353–366.

Simonsen, H. K. (2011): User Consultation Behaviour in Internet Dictionaries: An Eye-Tracking Study. I: *Hermes – Journal of Language and Communication Studies* 46(2011), 75–101.

- Simonsen, H. K. (2014): Mobile Lexicography: A Survey of the Mobile User Situation. I: Abel, Andrea, Chiara Vettori & Natascia Ralli (eds.): *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 15–19 July 2014, Bolzano/Bozen. Bolzano/Bozen: EURAC research. Volume I, 249–261.
- Simonsen, H. K. (2016): *Hvor er forretningsmodellen? En analyse af de forretningsmæssige udfordringer i forlags- og informationsindustrien med særlig fokus på opslagsværker*. MBA-afhandling. Institut for Økonomi og Ledelse. Aalborg Universitet.
- Tarp, S. (2009): Reflections on lexicographical user research. I: *Lexikos* 19, 275–296.

Patrick Leroyer  
Lektor, PhD  
Aarhus University  
School of Communication and Culture  
Jens Chr. Skous Vej 4  
DK-8000 Aarhus C  
pl@cc.au.dk

Henrik Køhler Simonsen  
Ekstern lektor, PhD, MBA  
Copenhagen Business School  
Management, Society & Communication  
Dalgas Have 15  
DK-2000 Frederiksberg  
hks.msc@cbs.dk

# Bildliga betydelser i SAOB

*Pär Nilsson*

In this study, I investigate what a figurative sense is in the Swedish Academy Dictionary (SAOB). The description of figurative senses in the dictionary is analyzed in terms of word classes and the use of specific labels in the definitions, such as *bildl.* (figurative), is mapped out. From a cognitive semantic perspective, I investigate what the figurative uses of the present cases consist of, in a cognitive sense. What semantic process is at hand (metaphor, metonymy etc.)? And to what extent does a figurative sense imply polysemy? The study is a part of a licentiate project and the aim of this article is to present some preliminary results of that project.

## 1. Inledning

I föreliggande studie anlägger jag ett teoretiskt perspektiv på lexikografisk semantisk analys och undersöker vad en bildlig betydelse i *Ordbok över svenska språket* utgiven av Svenska Akademien (SAOB) är för något. I fokus står relationen mellan kognitiv semantik och lexikografi och beskrivningen av metaforiska och metonymiska processer.

Studien ingår som en del i ett licentiandprojekt vid Lunds universitet. Syftet med föreliggande studie är att presentera några preliminära resultat från projektet och att skissera utgångspunkter för vidare studier. Projektets övergripande syfte är att bedöma hur relevant och konsekvent SAOB:s analysmetod är idag, rent teoretiskt. I vilken utsträckning är beskrivningen av betydelseutvecklingsprocesser förenlig med samtida teoretiska ideal?

I det följande diskuteras beskrivningen av bildlighet i SAOB med utgångspunkt i ordklass, position i ordbokens hierarkiska struktur, definitionsformler som anger bildlighet och vilken semantisk process som betecknas med dessa formler. Närmast följer dock en allmän teoretisk bakgrund.

## 2. Teoretisk bakgrund

### 2.1. Kognitiv semantik, metafor och metonymi

Utgångspunkten för den kognitiva semantiken är iakttagelsen att människans sätt att skapa och utveckla språkliga betydelser står i direkt samband med sättet att kategorisera och strukturera information om omvärlden i övrigt. Språklig begrepps-bildning beror på eller står i nära samband med våra upplevelser av den fysiska

världen. Abstrakta idéer och begrepp utvecklas metaforiskt, genom att erfarenheter i den fysiska verkligheten översätts till nya kontexter (jfr t.ex. Evans & Green 2006, Geeraerts 2010:182–183 m. fl.).

Sedan Lakoff & Johnsons *Metaphors we live by* (1980) är metafor-teori en central del av den kognitiva semantiken. Lakoff och Johnson (1980:5) menar att ”the essence of metaphor is understanding and experiencing one kind of thing in terms of another”, vilket innebär att det inte bara är våra språkhandlingar som får metaforisk gestaltning – utan vi tänker och handlar också strukturellt i dessa termer. Vidare hävdar författarna att metaforer låter oss förstå en *erfarenhetsdomän* i termer av en annan (Lakoff & Johnson 1980:117). I den konceptuella metaforen TIME IS MONEY förstår vi begreppet tid och tidens domän i termer av pengar (vilket leder till att vi kan tänka och uttrycka oss i termer som: ”du slösar min tid”). Förståelse och varseblivning kopplas således inte till isolerade språkliga uttryck, utan till hela domäner. Med domän avses här en strukturerad helhet, en sammanhängande organisation av erfarenheter (som bl. a. strukturerar naturliga dimensioner och relationer såsom delar, nivåer, orsaker m. m.). Domäner anses som naturliga i så måtto att de grundas på erfarenheter som har upplevts genom interaktion med andra personer och genom våra kroppar (Lakoff & Johnson 1980:117–118).

När vi ifråga om metaforer alltså överför erfarenheter *mellan* olika domäner (”domain mapping”), gäller för metonymi en relation *inom* domänen: ”we are using one entity to refer to another that is related to it” (Lakoff & Johnson 1980:35). Vid exempel som ”He likes to read the Marquis de Sade” refererar författaren i uttrycket till författarens verk. Både författare och verk befinner sig inom samma domän.

Den här artikeln tar sikte på relationen mellan kognitiv semantik och praktisk lexikografi. Att det finns beröringspunkter mellan de slutsatser man dragit inom den teoretiska litteraturen och de analyser man gjort inom den praktiskt orienterade ordboksverksamheten är dock inte en ny upptäckt. Denna iakttagelse har uppmärksammats tidigare och tillämpning av teorin på praktiken har bl. a. diskuterats av Geeraerts (t.ex. 1990), Atkins & Rundell (2008) och Ostermann (2015) för att endast nämna några.

## 2.2. Bildligt språk och metaforer i ordböcker

Metaforik är också centralt i ordböcker och det har inom den praktiska lexikografin utvecklats konventioner för att beskriva fenomenet. Åtminstone gäller detta definitionsordböcker som beskriver betydelseutveckling. En sådan konvention är bruket av särskilda och återkommande termer eller formler (i fortsättningen används just beteckningen *formler* för detta fenomen) i definitionerna som används för att beskriva den aktuella processen. I de germanska ordböckerna stöter man på formler som *billedlig*, *übertragen*, *figurative*, *bildlich* osv. Dessa tar fasta på dels upplevelsen av att framkalla en mental bild, dels hur själva överföringen mellan bild och den faktiska företeelsen ser ut. Formeln *metaforisk*, eller *metafor*, är dä-



remot sällsynt i lexikografiska sammanhang. Detta hänger ihop med att termen är färgad av den traditionella och snävare innebörden av begreppet, såsom stilfigur (trop). Termen är alltså traditionellt sätt snarare retorisk än språkvetenskaplig.

I SAOB:s fall förekommer också den här typen av definitionsformler och flera är i omlopp. Således kan man stöta på formuleringar som *bildl.*, *i överförd anv.* och (i enstaka fall även) *figurlig* och *i andlig bet.* Formeln *bidl.* (*i bildl. anv.*, *med bildl. bet.* osv.) är av dessa den klart mest frekventa. En sökning i nätversionen av ordboken på söksträngen *bidl\** genererar över 40 000 träffar mellan initialbokstäverna A och T.

Men trots att formeln används så flitigt i ordboken finns inga närmare instruktioner för redaktörerna om hur den semantiska analysen ska genomföras och hur man ska bedöma vad som konstituerar en bildlig användning. Det finns heller ingen anvisning till läsaren hur definitionsformeln ska tolkas. Det antas helt enkelt var något som man intuitivt ska förstå eller känna på sig.

Mot bakgrund av detta anser jag det motiverat att undersöka vad en bildlig betydelse egentligen är för något.

### 3. Material och metod

Ett sätt att bedöma SAOB:s behandling av metaforiska processer är att studera förekomsterna av formeln *bidl.* i ordboken. Vad döljer sig egentligen bakom formeln? Mitt teoretiska perspektiv är alltså kognitiv semantik och jag frågar mig således vad bilden innebär rent kognitivt, vad det handlar om för slag av semantisk process och hur bilden relateras till grundbetydelsen.

Metoden vilar på två ben. Dels undersöks vad som konstituerar en bildlig användning rent lexikografiskt, dels undersöks hur en sådan beskrivs utifrån perspektivet kognitiv semantik. I det följande redogörs för hur undersökningsmaterialet har valts ut och hur analysmallen är komponerad.

#### 3.1. Urval

Som nämndes ovan är *bidl.* en högfrekvent fras i ordboken. För att ha möjlighet att utföra en kvalitativ studie har undersökningen begränsats till att omfatta 300 förekomster av den aktuella formeln. De förekomster som har studerats representerar olika delar av alfabetet och olika perioder under ordbokens tillkomsttid. I SAOB anges för varje ordboksartikel aktuellt tryckår. Materialinsamlandet har tagit hänsyn till initialbokstav och tryckår för varje artikel där definitionsformeln *bidl.* förekommer.

Urvalet har skett på följande vis: en sökning på söksträngen *bidl\** (trunkerad sökning) har gjorts i den digitala versionen av SAOB. För att få spridning har materialet fördelats jämnt på de fyra olika perioder i SAOB:s tillkomsthistoria som Larsson (2014) diskuterar och som i sig representerar olika omfångs- och

ambitionsnivåer genom tiderna. Den första perioden sträcker sig mellan åren 1898–1920, den andra mellan 1920–1960, den tredje från 1960–2002 och den fjärde från 2002 och framåt.<sup>1</sup>

Följaktligen har 75 träffar från var och en av dessa perioder analyserats. Varje träff har närlästs och analyserats i detalj genom studium av den aktuella artikel som formeln förekommer i. Analysen har baserats på en mall som har upprättats i syfte att på ett så systematiskt sätt som möjligt fånga teoretiskt relevanta aspekter hos den semantiska relation och process som kan kopplas till formeln *bildlig* i varje enskilt fall. I det följande presenteras delar av denna mall.

### 3.2. Analysmall

Mallen som ligger till grund för analysen tar hänsyn till en rad faktorer som är eller kan vara relevanta vid bedömningen av bildlighet. Bl. a. undersöks *ordklassstillhörighet*. I exemplen i litteraturen representeras metaforer och bildliga användningar ofta av konkreta substantiv, men hur pass ordklassstyrt är fenomenet egentligen? En del av analysen består i att undersöka hur beskrivningen i SAOB ser ut i det här avseendet.

För att bedöma *graden av autonomi*, dvs. hur pass självständig den bildliga betydelsen är i förhållande till grundbetydelsen,<sup>2</sup> har den bildliga betydelsens position i SAOB:s artikelstruktur noterats. SAOB-artiklarna är hierarkiskt strukturerade med betydelsemoment på flera olika nivåer: överordnade huvudmoment (markerat med siffra), undermoment (kursiv bokstav), undermoment till undermoment (grekisk kursiv bokstav) osv. Var någonstans i strukturen som formeln *bildlig* påträffas skulle alltså kunna säga något om relationen till grundbetydelsen – ju längre ifrån denna, desto självständigare betydelse. Norén & Linell (2006) talar om semantisk variation som en skala från ren homonymi till semantiska aspekter. Metaforer och metonymier betraktas här som prototypisk polysemi, och ungefär i mitten på skalan (2006:11).<sup>3</sup> Metaforik betraktas alltså protoypiskt som en process som ger upphov till betydelser med relativ självständig prägel. Mot bakgrund av

<sup>1</sup> Larsson urskiljer två perioder (period 1 och 3) där man på ordboksredaktionen snarast strävat efter fullständighet i fråga om ord- och betydelsebeskrivning, med långa och utförliga ordboksartiklar och utvikningar i form av anmärkningar som följd. Som en reaktion följer på var och en av dessa en period där hållningen är den motsatta: åtstramningar och explicit formulerade begränsningsregler införs under period 2 resp. 4.

<sup>2</sup> I artikeln används i enlighet med Ekberg (2004) termen *grundbetydelse* för betydelse ifrån vilken en metafor eller bildlig betydelse har sin (bokstavliga) grund. Parallellt och i den här artikeln utan klar åtskillnad används, när tanken är riktad mot relation mellan domäner, termerna *källa* och *mål* (*källdomän*, *måldomän* osv.) i enlighet med Lakoffs *source* och *target* (se t.ex. Lakoff 1987).

<sup>3</sup> Med tanke på ordklassfrågan ovan är det för övrigt intressant att notera att 4 av de 5 exempel som författarna ger på metaforer och metonymier just utgörs av (konkreta) substantiv.

detta borde man i SAOB följaktligen hitta många av de bildliga användningarna i egna huvudmoment i artikelstrukturen.

Vidare tar analysmallen fasta på vad som med Grady, Oakley & Coulsons (1999) termer kan kallas för *metaforicitet* ("metaforicity"): i vilken mån är det fråga om en metaforisk process (och inte t. ex. metonymi)?<sup>4</sup> Bildlighet förknippas vanligen med metaforik (i ovan angiven bemärkelse) och hittills i artikeln har heller ingen klar gräns dragits mellan begreppen bild och metafor. Men frågan är förstas om termen *bildlig* i SAOB enbart täcker processen metafor eller om formeln även rymmer andra processer. Metaforicitetsfrågan närmar jag mig ifrån två olika håll.

För det första tas hänsyn till redaktörernas etikettering av de aktuella semantiska processerna i definitionen. Vad som är en bildlig användning eller inte är inte alltid ett klart fall. Dels kan det utan teoretiska ramar vara svårt vid den semantiska analysen att bedöma vilken process det är fråga om, dels verkar det (som nämnts ovan) vara en gradfråga. I SAOB kombineras eller modifieras definitionsformlerna ofta just för att hantera detta faktum. I ordbokens spalter kan man stöta på sådana formuleringar som: *mer l. mindre bildl., i utvidgad l. bildl anv., oeg. l. bildl.* osv. osv. Med analysmallen har noterats hur ofta och i vilka fall dessa modifierade definitionsformler uppträder.

För det andra har en semantisk analys gjorts vid samtliga förekomster av formeln *bildlig* i materialet för att ur ett teoretiskt perspektiv (med utgångspunkt i de ovan angivna definitionerna av konceptuella metaforer och metonymier) bedöma vilken semantisk process det är fråga om. Denna analys baseras bl. a. på hur relationen mellan käll- och måldomän ser ut och tar sin utgångspunkt i de definitioner som ges i respektive artikel och de språkexempel som illustrerar ordanvändningen.

#### 4. Resultat och analys

Mot bakgrund av metodbeskrivningen ovan kommer i det följande några resultat och iakttagelser från studien att presenteras i kortfattad form. I tabell 1 presenteras fördelningen av bildliga användningar per ordklass i det undersökta materialet.

---

<sup>4</sup> Svanlund (2001) skiljer mellan *bildlighetsgrad* och *metaforicitet*, och distinktionen är även relevant för min avhandling. Skillnaden mellan begreppen diskuteras dock inte vidare i den här artikeln.

	Formeln <i>bildl.</i> per ordklass i SAOB	Ordklassfördelning totalt i SAOB
Substantiv	64% (192)	62% (31 583)
Verb	21% (64)	21% (10 671)
Adjektiv	13% (39)	14% (7170)
Preposition	1 % (3)	0,5 % (310)
Adverb	0,5% (1)	2% (1095)
Pronomen	0,5% (1)	0,5% (59)
Summa:	100% (300)	100% (50 888)

Tabell 1: Ordklassfördelning för SAOB:s definitionsformel *bildl.*

Av tabellen framgår att bildliga användningar i materialet är klart mest frekvent hos substantiv. Men ser vi till hur ordklasserna fördelar sig totalt i SAOB och jämför de sex ordklasser som förekommer i undersökningens material med dessa ordklassers totala förekomst i ordboken kan vi konstatera att bildlighet här inte är ett substantivfenomen, utan istället något som verkar helt oberoende av ordklass. Proportionerna mellan de bildliga användningarnas ordklassstillhörighet och ordklassernas fördelning i stort är i det närmaste identiska.<sup>5</sup> Resultatet är inget häpnadsväckande i sig, men väl intressant att konstatera. Det stödjer den konceptuella metaforteoris tanke om att metaforik (och metonymi) inte är knutet till formen hos enskilda ord och uttryck, utan till större, bakomliggande begrepp. Således anspelar både adjektivet *uddlös* i ”En uddlös qvickhet” (språkexempel från period 4) och substantivet *gadd* i ”Satyrens gadd” (period 2) på en bakomliggande konceptuell metafor som kan formuleras exempelvis på följande sätt: SPRÅKET ÄR ETT (SÄRANDE) VAPEN.

I tabell 2 redovisas var någonstans i ordbokens hierarkiska struktur som de bildliga användningarna i materialet har påträffats.

Formeln <i>bildl.</i> :s position i artikelstrukturen	%
I huvudmoment ihop med grundbetydelsen	37,4
I eget undermoment som en särskild betydelse	43,1
I självständigt huvudmoment	19,5
I självständig artikel	0
Summa:	100

Tabell 2: Position i SAOB:s artikelstruktur för formeln *bildl.*

Tolkat som ett uttryck för graden av autonomi kan följande iakttagelser göras. Bildlig användning i SAOB är inte homonymi. I det undersökta materialet påträff-

<sup>5</sup> Uppgifterna är hämtade 2016 med hjälp av Erik Bäckerud, it-ansvarig på SAOB-redaktionen.

fas aldrig formeln *bildl.* i en från grundbetydelsen separat artikel. Ibland, men totalt sett relativt sällan, står bildlighet för det som i SAOB normalt klassas som prototypisk polysemi, nämligen inom artikeln som en självständig huvudbetydelse (markerat med momentsiffra). Oftare påträffas formeln inom grundbetydelsens moment i direkt anslutning till denna och allra oftast förekommer den som en separat presenterad underordnad betydelse (i undermoment markerat med kursiv bokstav). Resultatet står i kontrast till Norén och Linells beskrivning (jfr avsnitt 3.2. ovan), vilket är intressant och bl. a. väcker frågan om bildlighetsgrad: står bildlighet i SAOB för ett särskilt slag av metaforik (och/eller metonymi)? En stark koppling (kort avstånd) mellan källa och mål innebär i regel upplevelsen av en färgstark metafor – är det just sådana som klassificeras som bildliga i SAOB?

Som nämndes i föregående avsnitt finns en tendens hos redaktörerna att gardera sig vid beskrivningen av bildligt språkbruk, med sådana modifierade definitionsformler som *mer l. mindre bildl.*, *i utvidgad l. bildl. anv.* osv. Vid en genomgång visar det sig att sådana kombinerade eller modifierade definitionsformler föreligger vid drygt en tredjedel av fallen i materialet (39 %). Företeelsen verkar också öka över tid – det är dubbelt så vanligt i period 4 jämfört med period 1.

På innehållsplanet har de aktuella bildliga användningarna analyserats med avsikten att bedöma vilken semantisk process som döljer sig bakom definitionsformeln (metafor, metonymi, liknelse eller annat), och som väntat är det i nästan samtliga fall fråga om metaforer. I 96 % av fallen handlar det om att tala om en företeelse i termer av en annan. Resterande fall utgörs av vad som skulle kunna betraktas som tveksamma eller rent felaktiga bedömningar av redaktörerna (icke-metaforiska generaliseringar t.ex.) eller också av ren metonymi. Relationen mellan metafor och metonymi är emellertid ofta komplex, vilket jag strax kommer att återkomma till.

Nedan ges exempel från materialet där SAOB:s etikettering *bildl.* tolkats som (i tur och ordning) metafor och metonymi. Exempelen inleds av uppslagsform, följt av definitionstext och språkexempel.

**PACKA**, v.1

5 b) /.../ om lokal o. d.: fullproppad, proppfull; särsk. i uttr. *packad med l. av* (folk o. d.); äv. bildl.

Hans minne är packadt med bara årtal. Sjöberg *FörsLSAOB* (1815).

**VADMALSGRÅ**, adj.

som har den grå färg som vadmal förr ofta hade; äv. oeg. l. bildl., om bonde (med tanke på att vadmalsslagg förr användes av i sht bondebefolkningen). /.../

På marknadsängen trängs främmande köpmän med vadmalsgrå mälardalsbönder. *TurÅ* 1962, s. 231.

Ovan visades att bildlighet i undersökningsmaterialet i nästan samtliga fall innebär metaforisk användning. Metaforerna i materialet kombineras emellertid ofta

med metonymier på olika sätt, ett exempel på samverkan mellan metafor och metonymi i materialet ges nedan i uttrycket *kräla* eller *krypa i stoftet*:

**STOFT**, sbst.5

1) ss. ämnesnamn: anhopning l. mängd av ytterst finfördelade (mikroskopiskt små), torra partiklar av fast ämne (i sht av de lösa jordlagren) l. fasta ämnen (som lätt virvlas upp l. kan hålla sig svävande i luften); damm

*d)* i sådana uttr. som */.../ kräla l. krypa l. ligga l. rulla sig i stoftet* (för ngn l. ngt), företrädesvis hyperboliskt l. bildl. (jfr 2), betecknande att ngn fullständigt underkastar sig (ngn l. ngt) l. ödmjucar sig (för ngn l. ngt) l. är överdrivet ödmjuk l. servil l. inställsam l. lismande (inför ngn l. ngt) o. dyl. l. visar den djupaste vördnad l. artighet o. d. */.../*

Borde inte en sådan som han krypa i stoftet för de uslaste af mina dikter? Söderberg *MBirck* 155 (1901). Denne Feodoroff, som var upprörande orättvis mot underlydande, men krälade i stoftet för överordnade. Ackté-Jalander *Jalander* 19 (1932).

Krypandet eller krälandet står här i metonymisk relation till att någon utför handlingen som ett tecken på underkastelse (del för helhet). Men i de fall då uttrycket används och ingen i själva verket verkligen kryper på marken för någon, har vi att göra med en metaforisk användning av metonymin.

I litteraturen har den här gränslinjen eller kopplingen mellan metafor och metonymi uppmärksamrats i olika sammanhang, bl. a. av Barcelona (2000), som har gett ut en hel antologi om fenomenet. Flera av författarna där talar just om skalor mellan å ena sidan entydiga metaforer via liknelser, hyperboler till rena metonymier. Däremellan hittar man metonymiska metaforer eller metaforer inom metonymier. Goossens (1990) har för detta fenomen myntat begreppet *metaphonymy*, översatt till svenska: *metaftonymi*.

I SAOB:s fall verkar metaftonymi vara svårbeskrivet för redaktörernas del och fenomenet utgör inget klart fall av bildlighet. Ovan nämndes att SAOB-redaktörerna ofta anger dubbla etiketter för den semantiska process som ska beskrivas (*oeg. l. bildl. anv. osv.*) och att detta bruk också verkar ha ökat över tid. Om man tittar på de fall i materialet där metafor och metonymi samspelar, och om man fokuserar på den andra halvan av materialet (period 3 och 4), där dubbla etiketter är som mest frekvent, visar det sig att hela 23 av 29 metaftonymiska fall (79 %) just är markerade med dubbel etikett. Jag menar därför att det på en rent pragmatisk väg kan tolkas som att man på ordboksredaktionen har utvecklat en metod att beskriva denna företeelse, nämligen genom att gardera sig vid etiketteringen, eller nyansera beskrivningen om man vill, med modifieringar av formeln *bildl.: mer l. mindre bildl., i bildl. l. utvidgad anv. osv.*

## 5. Sammanfattande avslutning

De ovan presenterade resultaten från undersökningsmaterialet pekar på att bildlighet i SAOB inte är något unikt fenomen för substantiv. Vidare verkar bildlighet i SAOB inte inbegripa homonymi; oftare rör det sig om underbetydelser (i Norén och Linells termer (2006:11)) än om ren polysemi. Bildlighet är inte heller ett fenomen som är antingen svart eller vitt. Det finns en gråskala som bl.a. utgörs av *metafonymi* och denna gråskala uppmärksammas i ordbokstexten genom en kombination eller modifiering av definitionsformler.

Med detta hoppas jag ha nyanserat bilden lite av vad metaforer och bildlighet är, och hur fenomenen beskrivs, i SAOB. När vi diskuterar betydelseutveckling och semantiska förändringsprocesser behöver vi kanske inte börja från noll varje gång på det teoretiska planet – jag menar att det går att diskutera teori också utifrån vad som står i ordböckerna.

## Litteratur

- Atkins, B. T. Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Barcelona, Antonio (red.) (2000): *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. Berlin/New York: Mouton de Gruyter.
- Ekberg, Lena. (2004): *Grundbetydelse och förändringsprinciper hos relationella lexem. Exemplet följa*. Nordlund 24. Småskriften från Institutionen för nordiska språk i Lund.
- Evans, Vyvayan & Melanie Green (2006): *Cognitive Linguistics: an introduction*. Edinburgh: Edinburgh University Press.
- Geeraerts, Dirk (1990): The lexicographical treatment of prototypical polysemy. I: Savas L. Tsohatzidis (red.): *Meanings and Prototypes. Studies in linguistic categorization*. London: Routledge, 195–210.
- Geeraerts, Dirk (2010): *Theories of lexical semantics*. Oxford: Oxford University Press.
- Goossens, Louis (1990): Metaphtonymy: the interaction of metaphor and metonymy in expressions for linguistic action. I: *Cognitive Linguistics* 1–3:323–340.
- Grady, J., T. Oakley & S. Coulson (1999): Blending and metaphor. I: G. Steen & R. Gibbs (red.): *Metaphor in cognitive linguistics*. Amsterdam/Philadelphia: John Benjamins, 101–124.
- Lakoff, George (1987): *Women, fire and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, George & Mark Johnson (1980): *Metaphors We Live By*. Chicago/London: The University of Chicago Press.
- Larsson, Lennart (2014): En ”mer l. mindre stor” stor ordbok – om variationerna i SAOB:s omfång och ambitionsnivå. I: *LexicoNordica* 21, 61–80.

- Norén, Kerstin & Per Linell (2006): *Meningspotentialer i den språkliga praktiken*. MISS 54. Göteborg: Institutionen för svenska språket.
- Ostermann, Carolin. (2015): *Cognitive Lexicography. A new approach to lexicography making use of cognitive semantics*. Berlin/ Boston: Walter de Gruyter.
- Rosch, Eleanor (1978): Principles of Categorization. I: Eleanor Rosch & Barbara B. Loyd (red.): *Cognition and Categorization*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 27–48.
- SAOB (1898–) = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1–. Lund.
- Svanlund, Jan (2001): *Metaforen som konvention. Graden av bildlighet i svenskans vikt- och tyngdmetaforer*. Stockholm: Almqvist och Wiksell.

Pär Nilsson  
redaktör  
Svenska Akademiens ordboksredaktion  
Dalbyvägen 3  
S-224 60 Lund  
Par.nilsson@svenskaakademien.se



# Orddannelsestypen konversjon i leksikografien

*Carina Nilstun*

Descriptions of word formation seem to have a gap to fill when it comes to conversion of inflected forms. Conversion is said to be a process of word formation by changing part of speech. The examples given tend to be a base form that is converted to a new base form of a different part of speech. As a lexicographer you experience that this picture is deficient. There are several types of conversion where an inflected form is converted into a base form (accompanied or not by a change of part of speech). This field is inexhaustively explored within the linguistic field – a field that lexicography tends to turn to in matters concerning grammar. The lack of an academic foundation has led to an unsystematically lexicographic handling of these words, at least in the Nordic languages. In this article I will argue that conversion of inflected forms, as a very productive type of word formation, should be treated as a type of derivation with the lexicographic consequence that their use will be described in separate entries instead of in the dictionary entry of their derivation base, as is often the case in today's lexicographic practice.

## 1. Innledning

Konferansen i Reykjavik innbød til å snakke om «Ordbøker og sproglige resurser». Dette kan forstås både som ordbøker og andre språklige ressurser brukt supplerende til ordbøker, eller som ordbøker og de språklige ressurser de selv bygger på. Dette igjen kan forstås som korpuser og deres råmateriale, men også som lingvistikken og grammatikken som faglige kilder for det leksikografiske arbeidet. Leksikografiske beskrivelser bygger ofte på disse, slik som inndelingen av ordklasser og syntaktisk analyse for funksjonsordene. Også når det gjelder orddannelse vender vi oss til lingvistikken og grammatikken. Det kan for eksempel gjelde grensedragningen mellom affiks og sammensetningsledd.

Et praktisk problem vi har felles i nordisk leksikografi knytter seg til leksikalisering av bøyde former. Mest vanlig er presens og perfektum partisipp av verb som brukes som adjektiv, og spørsmålet er når disse skal regnes som selvstendige enheter i språkets ordforråd til forskjell fra å betraktes som mer tilfeldig bruk, slik som gjennomslittige sammensetninger (Bußmann 2002:405). Når slike bøyde former er leksikalisert, skal de også beskrives under egne oppslagsord i ordboken. Jeg ønsker å sette dette inn i den faglige rammen av orddannelse, nærmere bestemt konversjon, og med det som grunnlag gjøre en vurdering av hvordan dette kan behandles leksikografisk. Ut fra konkrete eksempler vil jeg presentere noen ulike måter å angi denne ordtypen på.

## 2. Konversjon i faglitteraturen

### 2.1. Hva er konversjon?

Konversjon refererer til orddannelse ved hjelp av ordklasseskifte, men uten bruk av affikser (Bußmann 2002:380; Askedal et al. 2017:164). Konversjon blir i faglitteraturen behandlet som en type avledning, og ofte kalles den *nullavledning* eller *kortavledning*. Et typisk eksempel er *rangle* (verb) > *rangle* (substantiv). De fleste ordklasser er involvert i konversjon, og retningen er ikke fast. Noen andre eksempler er *ja* (interjeksjon) > *ja* (substantiv), *sykkel* (substantiv) > *sykle* (verb), *lam* (adjektiv) > *lamme* (verb). Avledningsbasen og det konverterte ordet kan enten ha helt identisk form, eller det konverterte ordet har endelse etter sin ordklasse.

Hos Fjeld & Vikør (2008:44–45) behandles denne ordtypen under avledninger, og termen *kortavledning* brukes. *Norsk grammatikk* (Askedal et al. 2017:164) bruker termen *ordstammeavledning*, mens *Norsk referansegrammatikk* bruker termene *konversjon* (om orddannelsesmåten) og *avledningsbase* (om utgangspunktet for dannelsen) (NRG 1997:124–125). Heretter vil jeg følge terminologien fra NRG.

### 2.2. Konversjon av bøyde former

I ulike beskrivelser av konversjon er det påfallende at eksemplene som gis har en grunnform som avledningsbase. Beskrivelsene er dessuten langt mer overfladiske enn for andre orddannelsestyper og ikke intendert som uttømmende. Hos Bußmann (2002:380) og Fjeld & Vikør (2008:44–45) nevnes ikke konversjon av bøyde former overhodet. I NRG finner vi underkapittelet «Adjektiver konvertert fra perfektum partisipp». De illustreres med setningene *Meningene er delte*, *Hans dager er talte* og *Skoene er utslitte*, og begrunnes med kongruensbøyningen som kjennetegner adjektiviske predikativer (1997:124–125). I *Norsk grammatikk* (Askedal et al. 2017:187) står det å lese: «Noen faktiske preposisjoner er oppstått ved konversjon av verbalt perfektum eller presens partisipp: *alt unntatt* (eldre: *unntagen*) *sannheten*, *forutsatt finansiering*, *inkludert bidrag til folketrygden*, *fratrasket moms*, *uansett alder* og *ansiennitet*; *vedrørende dette*, *angående våre felles problemer*.» Begge grammatikkene behandler altså orddannelsestypen på linje med andre orddannelsestyper, men går ikke inn for en fullblåst beskrivelse.

Definisjonen av konversjon som orddannelse ved hjelp av ordklasseskifte er upresis og dels misvisende. Ofte omfatter prosessen et ordklasseskifte, men det finnes unntak, slik som for orddannelsen *høre* > *høres*. Det samme forhold finner man noen ganger også i definisjonen av suffiksavledninger. Fjeld & Vikør (2008:43) skriver at «avleiingsaffiksa i norsk og nærskyldte språk har to hoved-

funksjonar: Den eine er å *flytte eit rotord over i ei anna ordklasse.*»<sup>1</sup> I avledningen *søt* > *søtlig* finner ikke et ordklaseskifte sted.

Et ordklaseskifte står altså ikke sentralt, hverken for konversjon eller for suffiksavledning. NRG nevner at konversjon også kan finne sted uten ordklaseskifte, og anfører parverbene<sup>2</sup> som eksempel (1997:123).

### 2.2.1. Ulike typer konversjon av bøyde former

Nedenfor har jeg forsøkt å sette opp en uttømmende liste over denne orddannelsesstypen. Listen bygger på egen erfaring med leksikografisk arbeid.

- Passiver (*høres*, passiv av *høre* OG ny grunnform, verb)
- Partisipper (*spent*, perfektum partisipp av *spenne* OG ny grunnform, adjektiv; *spennende*, presens partisipp av *spenne* OG ny grunnform, adjektiv)
- Partisipper (*ansatt*, perfektum partisipp av *ansette* OG ny grunnform, substantiv)
- Nøytrumsadjektiver (*snart*, nøytrum av adjektivet *snar* OG ny grunnform, adverb)
- Genitiver (*damaskes*, genitiv av *damask* OG ny grunnform, adjektiv)
- Gradbøyde former (*ytterligere*, komparativ av *ytterlig* (adverb) OG ny grunnform, adjektiv)
- Imperativer (*blås*, imperativ av *blåse* OG ny grunnform, intereksjon)

### 2.3. Konversjon i et større perspektiv

At denne orddannelsesstypen er såpass lite beskrevet i språkvitenskapen og også så lite konvensjonalisert i leksikografisk praksis virker underlig av flere grunner. En ting er at den er hyppig i nåtidsspråket. En annen er at man ikke løfter den frem i sammenhenger der man er bekymret for alle nye lån som kommer inn i språket. Disse avledningene er både hjemlige og produktive. Den siste grunnen er at denne orddannelsesstypen er godt etablert, også utenfor norsk og bakenfor vår egen tid. En rekke ord av latinsk opprinnelse er dannet ved konversjon av bøyde

<sup>1</sup> Den andre er at enkelte suffiks også tilfører semantisk innhold, jf. *-aktig*, *-inne* m.fl. Det er dessuten upresist å her snakke om avledningsaffiks, det gjelder kun suffiksene.

<sup>2</sup> *Brenne* (sterkt) – *brenne* (svakt), *bryte* – *brøyte*, *ligge* – *legge*, *sitte* – *sette* o.l.

former. Disse lånene oppfatter vi i dagens norsk som aldeles fullverdige ord, og ikke lenger knyttet til sin avledningsbase. Det viser at denne orddannelsen er vel så kraftfull som andre.<sup>3</sup>

**attributt** fra latin *attributum*, grunnbetydning 'det tildelte', substantivering av adjektivisk perfektum partisipp i nøytrum av *attribuere*

**senior** fra latin *senior* 'eldre', komparativ av *senex* 'gammel'

**voksen** av norrønt *vaxinn*, perfektum partisipp av *vaxa* 'vokse'

## 2.4. En ny definisjon av konversjon

Den mangelfulle beskrivelsen av leksikalisering av bøyde former i lingvistisk litteratur skyldes nok at spørsmålet der ikke oppfattes som så relevant. I leksikografien oppfattes spørsmålet derimot presserende. Hvorfor? Fordi grunnformen er vår arbeidsenhet. Erfaringene fra leksikografisk arbeid viser også at vi kan definere konversjon og avledninger for øvrig enda mer presist. Tradisjonelt har man sagt at suffiksavledning og konversjon er orddannelse ved ordklasseskifte. Men det finnes eksempler på suffiksavledning og konversjon uten ordklasseskifte,<sup>4</sup> det er altså ikke ordklasseskiftet som er cruxet. *Ved konversjon (og all annen avledning) dannes en ny grunnform på grunnlag av en annen grunnform eller en bøyd form.*

Til den leksikografiske behandlingen må vi tenke på konversjon på samme måte som andre avledninger, og klare å skille de to (ofte identiske) formene (avledningsbase og ny grunnform) som homografer. Vi vurderer aldri en avlednings frekvens mot avledningsbasens frekvens når vi skal vurdere den tatt inn i ordboken. De to formene eksisterer parallelt.

## 3. Konversjon i leksikografien

### 3.1. Orddannelsesinformasjon i NAOB

Orddannelsestype fremgår i NAOB eksplisitt eller implisitt i etymologiledet. Der angivelsen er eksplisitt er det brukt et sett av standardfraser, eksempelvis

sammensatt av x og y  
avledet av x med suffikset y  
sammentrukket form av x

<sup>3</sup> Samtlige etymologier i eksemplene er hentet fra *Det Norske Akademis ordbok* (NAOB), som ligger gratis tilgjengelig på <www.naob.no> fra ultimo desember 2017.

<sup>4</sup> Eksempler er *søt* (adjektiv) > *søtlig* (adjektiv) og *høre* (verb) > *høres* (verb).

For orddannelsestypen konversjon har vi på grunn av fraværet av affiks valgt å ikke bruke standardfrasen «avledet av ... med ...», men i stedet brukt preposisjonen «til».

**komme** substantiv et  
substantivering av [komme](#) (verb)

For konversjon av bøyde former, fremkommer dette i etymologiledet, men ikke med en eksplisitt angivelse av orddannelsesmåten eller et eventuelt ordklasseskifte. Generelt har vi av pedagogiske hensyn unngått fagspesifikke termer som konversjon, samdannelse o.a.

**visst** adverb  
nøytrum av [viss](#)

I *Norsk Riksmålsordbok* (NRO 1937–95), seksbindsverket som NAOB bygger på, har man ikke betraktet disse tilfellene som homografer, slik jeg foreslår i denne artikkelen. I verboppslag finner man stadig separate betydninger med teksten «særlig brukt i adjektivisk perfektum partisipp». Man er altså ikke villig til å erkjenne et ordklasseskifte. Man holdt fast ved formen og oppgav en ny funksjon, fremfor å ta utgangspunkt i den nye funksjonen og oppgi formen. Frasen «brukt adjektivisk» er mer tilbakeholden enn «brukt som adjektiv». I revisjonsarbeidet flytter vi stadig slike enheter ut av verboppslaget og gjør dem til egne oppslag. Verbets bøyingsparadigme forblir det samme – uavhengig av om den enkelte form finnes i praktisk bruk eller ei. Uten at det finnes mye sikker kunnskap om dette, kan vi observere (også i egen språkforståelse) hvordan disse ordene entydig oppfattes som egne ord i det mentale leksikon. Adverbet *snart* (nøytrumsform av adjektivet *snar*) er et talende eksempel: det kom sent inn som eget oppslag i norske ordbøker, men ingen ville finne det naturlig å søke det opp under *snar* (adjektiv). Hva så med adverbene *visst* og *snaut*?

### 3.2. Lemmaseleksjon og lemmastatus

Når man undersøker nordiske ettspråklige allmennordbøker for å se hvordan konverterte bøyde former behandles, finner man raskt at det mangler en enhetlig behandling, også innen det enkelte verk. Det virker nokså tilfeldig hvilke former som er gitt lemmastatus og hvilke som behandles under avledningsbasen. Jeg tror dette skyldes at vi mangler en faglig forståelse av orddannelsestypen – både innen lingvistikken og innen leksikografien. Jeg vil her forsøke å finne frem til noen innfallsvinkler til hvordan vi kan forstå og behandle disse ordene.

I *Redigeringshandboka til Norsk Ordbok 2014* (2016:80; jf. kapittel 2.3.2) beskrives først praksisen for oppføring av oppslagsord på belegg i seddelarkivet: «Ordformer med form som regulære bøyingsformer (usamansette og samansette

partisipp og ing-avleiingar) vart som fast praksis førde under grunnordet, av di det ut frå einskildbelegg ikkje er mogleg å seia om ei bøyingsform er leksikalisert.» Når redaktøren så skal lage ordbokartikler ut fra seddelarkivet, er det satt opp følgende tre vurderingspunkter for lemmatisering (ibid.):

- om ordforma har tyding eller bruksmåte som ikkje er dekt av opphavsortet (sjå dømet **farga** nedanfor);
- om ordforma er polysem, dvs om det finst fleire opplysningar som spesielt gjeld ordforma, spreidde ut over fleire tydingsbolkar i grunnordsartikkelen, slik at det er vanskeleg for lesaren å finna dei att;
- om ordforma har eit stort (dominerande) beleggssomfang i høve til dokumentasjonen av grunnordet (det hender t d at mest alle setlane på eit gjevi verb syner adjektivisk bruk, jfr **frustrera – frustrert**).

De to første punktene, jeg kaller dem **egen semantikk** og **polysemi**, anser jeg som sentrale. Om det derimot er relevant å bruke frekvensfordelingen mellom de to formene som argument, vil jeg komme tilbake til.

Svensén (2004:128–130) skriver mot slutten av kapittelet «Lemmats grammatiska form» at enkelte partisipper nærmest har mistet sin verbale betydning og blitt adjektiver (som *slående* og *uttalad*) og at verbet i andre tilfeller nærmest har gått ut av bruk (som *påfallande* og *uthungrad*). I disse tilfellene anbefaler han at partisippformen får lemmastatus. Svensén ser dog ikke ut til å se disse tilfellene som orddannelse, idet han tematiserer dem sammen med tilfeller der lemmaformen avviker fra konvensjonen om hvilken form som er lemmaform for den enkelte ordklasse, så som at et substantiv får lemmaform i bestemt form eller i flertallsform da ubestemt form entall ikke forekommer i praktisk språkbruk. Hans beskrivelse utover det – med vekt på semantikk og frekvens er verdt å ta med seg videre.

### 3.3. Eksempler og vurderinger

Jeg vil ut fra konkrete (norske) eksempler forsøke å identifisere noen ulike vurderingspunkter.

Syntaksen som analyseverktøy for ordklasseidentifisering spiller åpenbart en rolle, men står den seg som kriterium alene? Vi kan ta *skrikende* som eksempel: *skrikende barn, måker*. Syntaktisk er *skrikende* et adjektiv, men trenger vi adjektivoppslaget *skrikende* for å beskrive disse kollokasjonene? Betydningen vil jeg si er verbal ut fra følgende parafrase: ‘barn, måker som skriker’.

*Skrikende forskjeller, skrikende farger og skrikende barn*. Syntaktisk er de alle adjektiver, men semantisk oppfatter jeg altså det siste verbalt: ‘barn som skriker’. De andre ville jeg parafrasert med andre adjektiver: *skrikende forskjeller* > ‘store/gapende forskjeller’ og *skrikende farger* > ‘grelle farger’. **Parafrasen** kan si oss noe om **semantikken**.

Ingen nordiske ordbøker<sup>5</sup> jeg har konsultert har adjektivoppslag for *svelgende*, brukt i den noenlunde faste forbindelsen *svelgende dyp* med betydning ‘uoverstigelig forskjell; avgrunn’. Hadde det derimot hett *svelgemessig dyp* ville vi jo hatt oppslaget *svelgemessig*. *Svelgende dyp* er ikke særlig frekvent, men det er altså frekvent nok til å beskrives i ordbøkene, spørsmålet er ene og alene knyttet til lemmatiseringen. For ordboksføring av affiksavledningene vurderes aldri deres **frekvens** mot avledningsbasens frekvens.

Informasjonen vi helst skulle hatt, var hva språkbrukerne oppfatter som ord og hvordan de søker. Det vet vi for lite om. Det vil uansett gi en **pedagogisk** gevinst å skille ut adjektivisk bruk osv. fordi artiklene blir mer oversiktlige. Utflytningen, mener jeg, må basere seg på i hvilken grad formen oppfattes som leksikalisert, som vi best kan si noe om ved å parafasere. En annen måte å vurdere leksikalisering kan være koordineringen av ordpar eller ordgrupper som utgjør et paradigme og dermed må gis samme funksjonsanalyse. Eksempelvis *ny(tt)/brukt, få/mange, gift/forlovet/skilt/separert*.

NRG velger å analysere *låst* i *Døra er låst* som et adjektiv selv om man der ikke kan støtte seg på kongruens. De setter det som motsetning til *Døra er åpen*, uten eksplisitt av nevne denne parrelasjonen *åpen – lukket* som argument i seg selv (NRG 1997:125).

### 3.4. Forslag til løsninger

Basert på den forutgående analysen av konverterte bøyde former ville jeg for eksemplene i dette avsnittet foreslått følgende løsninger: Man burde ha eget adjektivoppslag for *skrikende*. Her ville *skrikende forskjeller* og *skrikende farger* hørt hjemme, etter sitt adjektiviske innhold. *Skrikende barn* og *skrikende måker* hører hjemme i verboppslaget *skrike*, etter sitt verbale innhold. Dette viser tydelig den parallelle eksistensen av de to formene. Verbparadigmet er uforandret selv om det har avfødt et nytt ord (med eget bøyingsparadigme). Vanskelighetene bunner i at de to formene er homografer og semantisk beslektet. De krever altså en analyse som ikke vektlegger form. Også *svelgende* skal etter min analyse være et eget oppslag. *Svelgende* i verbal bruk har muligens ingen typiske kollokasjoner og nevnes muligens overhodet ikke i oppslaget *svelge*, men er altså like fullt del av verbets bøyingsparadigme. At *svelgende dyp* ikke bør stå under *svelge*, er dels at vi ikke har noen tilsvarende verbal betydning (jf. Svensén 2004:130), dels den semantiske analysen eller parafrasen: uoverstigelig forskjell.

<sup>5</sup> *Bokmålsordboka* (2005), *Nynorskordboka* (2006), *Norsk Ordbok 2014* (1966–2015), *Norsk ordbok med 1000 illustrasjoner* (2008), *Den Danske Ordbog* (2003–05), *Svensk ordbok* og SAOB.

## 4. Konklusjon

Analysen av konversjon som en type avledning (eller i alle fall en type orddannelse) – er ikke kontroversiell (jf. kapittel 1.1). Flere kilder gir også eksempler hvor bøyde former er konvertert, uten at dette beskrives uttømmende. Denne analysen – altså at dette er en type orddannelse og ikke bare en mer eller mindre avvikende funksjon for en bøyd form – bør ligge til grunn når disse ordene skal beskrives leksikografisk. Lemmaseleksjonen må da basere seg på semantikk og leksikaliseringsgrad, og frekvenskravene bør ikke være strengere enn for andre avledninger.

## Litteratur

### Ordbøker

- Bokmålsordboka* (2005). Boye Wangensteen (red.). 3. utgave. Oslo: Kunnskapsforlaget.
- Den Danske Ordbog* (2003–05). Bind 1–6. Ebba Hjorth & Kjeld Kristensen (hovedred.). København: Det Danske Sprog- og Litteraturselskab og Gyldendal.
- NLO (1997) = Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- NAOB = *Det Norske Akademis ordbok* (2017): [www.naob.no](http://www.naob.no).
- Norsk ordbok med 1000 illustrasjoner* (2008). Tor Guttu (red.). 2. utgave. Oslo: Kunnskapsforlaget.
- Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet* (1966–2015). Oslo: Det Norske Samlaget.
- NRO (1937–95) = Trygve Knudsen & Alf Sommerfelt (red.): *Norsk Riksmålsordbok I–VI*. Oslo: Kunnskapsforlaget.
- Nynorskordboka* (2006). Marit Hovdenak et al. (red.). 4. utgåva. Oslo: Det Norske Samlaget.
- SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien* (1893–). Lund: Gleerups förlag.
- Svensk ordbok = *Svensk ordbok utgiven av Svenska Akademien* (2009). Stockholm: Norstedts Akademiska förlag.

### Annen litteratur

- Askedal, John Ole et al. (2017): *Norsk grammatikk*. 2. utgave. Oslo: Kunnskapsforlaget.
- Bußmann, Hadumod (2002): *Lexikon der Sprachwissenschaft*. Stuttgart: Alfred Kröner Verlag.



- Fjeld, Ruth Vatvedt & Lars S. Vikør (2008): *Ord og ordbøker*. Kristiansand: Høyskoleforlaget.
- NRG = Jan Terje Faarlund, Kjell Ivar Vannebo & Svein Lie (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Redigeringshandbok for Norsk Ordbok 2014* (2016): <http://no2014.uio.no/eNo/tekst/redigeringshandboka/redigeringshandboka.pdf> (november 2017)
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska förlag.

Carina Nilstun  
leksikograf, MA  
Kunnskapsforlaget  
Gullhaug torg 1  
Postboks 4432 Nydalen  
N-0403 Oslo  
carina.nilstun@gmail.com

# Gammalt projekt och nya resurser – SAOB:s hantering av nytillkomna elektroniska textmängder

*Bodil Rosqvist & Bo-A. Wendt*

As all the big historical dictionary projects launched in the 19th century, *Svenska Akademiens ordbok* (SAOB) is based on paper excerpts manually selected from authentic texts. Today entirely new possibilities have emerged when it comes to anchoring lexical analysis and description in authentic language use. In this article we report on how the editorial staff of the SAOB handles the fact that nowadays our source material can so easily be complemented by electronic sources. We also discuss the upcoming revision of the SAOB and the question of what material(s) to use in the future.

## 1. SAOB och dess excerptmaterial

*Ordbok över svenska språket* utgiven av Svenska Akademien (SAOB) är en historisk ordbok som beskriver det svenska språket från 1521 fram till idag. Projektet inleddes redan på 1880-talet, och det material som ännu är ryggraden i projektet skapades utifrån sin samtids förutsättningar, i en tid när man knappast ens kunde drömma om de möjligheter som datorteknikens textdatabaser nu har givit oss. Materialet insamlades genom selektiv excerptering, där man valde ut i genomsnitt ett par, tre belägg per sida i de valda källorna och gjorde excerpter av dessa samman med det excerpterade ordets närkontext.<sup>1</sup> Sådan excerptering genomfördes i ett mycket stort antal mycket mångskiftande slags verk. Resultatet har därmed blivit ett brett men inte så djupgående urval av språkbelägg. Relativt sett missgynnas språkets allra vanligaste ord i materialet, men å andra sidan fångas väldigt många olika ord upp, också ovanligare ord och sådana vars bruk inskränker sig till vissa genrer eller vissa ämnesområden.

Denna stora bredd, både genremässigt och inte minst viktigt för en historisk ordbok kronologiskt, är den verkligt stora fördelen med SAOB:s samlingar (SAOBArkiv), liksom omvänt den relativa hanterligheten i omfång för språkets mer frekventa ord vid det praktiska analysarbetet. Det blir tydligt om man jämför med nutidens allt flera och allt större elektroniska korpusar, som per definition är total Konkordanser och därmed har ett djup som SAOBArkiv inte kommer ens i

---

<sup>1</sup> Lundbladh (1996:95) framhåller att denna excerptering oriktigt har beskrivits som slumpmässig. Tvärtom är urvalet högst medvetet från excerptistens sida. Ibland har detta helt visst gynnat det anmärkningsvärda eller rent av konstiga, men givna instruktioner har understrukt vikten av att också excerptera det typiska och vardagsgrå. Ätminstone idealt kan urvalet därför beskrivas som stratifierat.

närheten av men som föreligger till priset av antingen – och vanligen – att korpusen omfattar mycket färre olika källor eller att den har enastående stort och därigenom vid detaljanalys svårhanterligt omfång. Och inte ens de allra största korpusarna har samma bredd i genrer och tid som SAOBArkiv. En bredd av detta slag är förstås mycket viktig för en historisk ordbok där man på förhand inte vet vilka ord som över huvud taget funnits eller vilka användningar ett i och för sig känt ord en gång i tiden har kunnat ha. Å andra sidan är förstås det motsvarande mindre djupet en uppenbar nackdel när det kommer till att med bästa möjliga precision hitta det äldsta belägget på en viss betydelse eller användning. (Jämför med diskussionen hos Svensén (2004:56–57) om det han kallar beläggsmetoden kontra korpusmetoden.)

Skillnaderna mellan SAOBArkiv och moderna digitala korpusar kan tydliggöras med följande sifferförhållanden. SAOBArkiv omfattar numera uppemot 9 miljoner excerpter (där varje excerpt har ett stickord, som blir den enda lexikala ingången till excerpten).<sup>2</sup> SAOBArkiv källförteckning omfattar idag 21 460 poster. Man får då komma ihåg att många av dessa källor endast lämnat enstaka belägg, ibland bara ett enda; vi torde kunna räkna med att endast omkring 11 500 av källorna är genomexcerperade på det selektiva sätt som beskrivits ovan (Lundblad 1996:94). Därmed skulle man alltså få ett genomsnitt på omkring 750 belägg per genomexcerperad källa. Detta kan så jämföras med Fornsvenska textbankens nysvenska annex som omfattar 786 397 ord fördelade på endast 19 källor<sup>3</sup>, en 600-del av antalet källor i SAOBArkiv. Det ger ett genomsnitt på 41 389 ord/källa, mer än 50 gånger högre än snittet för SAOBArkiv. Ett material av innehållsligt helt annat slag är den av historiker i Uppsala sammanställda Gender and Work-databasen, som bygger på handskrivet material från 1500–1700-talet, främst domböcker. Denna omfattar 498 336 ord från 37 olika källor (ofta i urval, i form av nedslag för vissa år). Båda dessa textdatabaser har med nutida ögon sett ändå ett beskedligt omfång. Så mycket större är Korps historiska material, som med borträknande av fornsvenska källor omfattar hisnande 1,33 miljarder ord fördelade på 114 korpusar, i sig förstås överlag samlingskorpusar. Bland dessa samlingskorpusar återfinns Litteraturbanken, som idag i egen rätt består av 117 425 315 ord fördelade på 2 143 verk.

Så långt alltså bredd och djup översatta till nakna siffror, men intressantare blir det om vi ser på ett exempel på en följd av dessa förhållanden. Exemplet vi mer eller mindre godtyckligt har valt avser det alfabetiska spannet *yl–ymta* (främst för att det inte ger upphov till så mycken störande stavningsvariation vid korpussökningarna). Detta är ännu endast förberedande bearbetat vid SAOBArkiv redakti-

<sup>2</sup> Också det är ju för övrigt ett utslag av att SAOBArkiv är ett barn av en annan tid att vi faktiskt inte vet exakt hur många materialets excerpter är. Tidigare har ofta siffran 7,5 miljoner excerpter anförts (Hast 1983:160), men denna hänför sig till en tid då inte redaktionens egna kompletteringar av materialet, inte minst ur andra ordböcker, sparades efter redigering, vilket görs numera.

<sup>3</sup> Av dessa är 17 tillika SAOBArkiv-källor.

on. Av tabell 1 framgår vilka och hur många olika ord som SAOBArkiv och några elektroniska korpusar samt Dahlgrens Glossarium (som redovisar nysvenska ord som var föråldrade vid dess utgivning 1914) lyckas fånga upp i spannet. Som framgår härav återfinns i SAOBArkiv 16 olika ord, förutom ytterligare 10 hapax-ord som inte tagits upp i tabellen och rimligen inte heller skall beskrivas i ordboken. Av dessa 16 ord har Dahlgren med nio, vilket är rätt imponerande, i synnerhet om man tar i beaktande att han säkerligen uteslutit några av de andra orden för att de inte bedömts som föråldrade. Fornsvenska textbankens nysvenska annex har fem av orden, och samma antal återfinns i Gender and Work-databasen. Tillsammans får dessa båda databaser ihop sju ord i spannet. Lika många, men inte samma, är automatiskt identifierade i Korps historiska material. Poängen med denna jämförelse är å ena sidan att SAOBArkiv fångar upp betydligt fler olika ord än de två i huvudsak äldre nysvenska textdatabaserna tillsammans – och med beläggsantal för vardera ordet som någotsånär möjliggör en kvalificerad analys av dem. Å andra sidan har SAOBArkiv i jämförelse med Korp inte ohanterligt många belägg per ord. Så känns ju till exempel 15 000 belägg för *ymnig* lätt övermåttigt för en lexikograf som i sin historiska analys arbetar förutsättningslöst, nedifrån och upp.

	SAOBArkiv	Dahlgren	NsvA	GaW	KorpH
YL s. 'ylande'	16	–	–	–	5 591
YLA, v.	143	x	1	–	1 996
YLLE, s.	885	–	1	6	21 854
YLLEN, adj.	143	x	1	1	307
YLONS, s. '(slags) öl'	3	–	–	–	–
YLVA, s. 'varghona'	3	–	–	–	–
YLVA, v. 'yla?'	2	–	1	–	–
YMLA, v. 'mumla'	7	x	–	–	–
YMMEL, s. 'mummel'	8	x	–	1	–
YMN(E), s. 'ymnighet'	7	x	–	–	–
YMNA, v. 'göra fruktbar'	8	x	–	–	–
YMNELIGA, adv.	5	x	–	–	–
YMNIG, adj.	788	x	25	2	15 343
YMP, s.	75	–	–	2	178
YMPA, v.	803	–	–	–	1 339
YMTA, v. 'antyda'	9	(x)	–	–	–
<b>Σ ord</b>	<b>16 ord</b>	<b>9 ord</b>	<b>5 ord</b>	<b>5 ord</b>	<b>7 ord</b>
<b>Σ belägg</b>	<b>2905</b>		<b>29</b>	<b>12</b>	<b>46 608</b>

Tabell 1: Ord inom det alfabetiska spannet *yl–ymta* i SAOBArkiv, Dahlgrens *Glossarium*, Fornsvenska textbankens nysvenska annex (NsvA), Gender and Work-databasen (GaW) och Korps historiska material (ej fornsvenska källor, endast automatiskt identifierade ord; KorpH).

## 2. Elektroniska resurser i SAOB-arbetet idag

Digitala korpusar som Fornsvenska textbanken, Korps historiska material, Litteraturbanken och Gender and Work-databasen nämns ovan. Andra välkända elektroniska resurser är Projekt Runeberg och, inte minst, Google Books. Alla dessa resurser används ibland i arbetet med SAOB, exempelvis för att hitta äldre (eller yngre) belägg än de som finns i SAOBArkiv eller språkexempel som illustrerar en viss betydelse på ett bra sätt.

Databaser med äldre material är dock fortfarande förhållandevis små och/eller svåra och tidskrävande att söka i, ibland är de också osäkra till sitt innehåll. Det sistnämnda gäller i synnerhet Google Books. Procentuellt sett täcker SAOBArkiv också det äldre språket bättre än det nyare. Samtidigt som antalet publicerade texter har ökat lavinartat under 1900-talet har excerperingen på SAOB:s redaktion tvärtom minskat.

Vid arbetet med SAOB görs därför mest omfattande kompletterande sökningar i material från de senaste 150 åren, dvs. ”senare material” ur vårt historiska perspektiv. Vidare används främst tidningsdatabaser. Detta eftersom de är både väldefinierade till sitt innehåll och förhållandevis lättillgängliga. De tidningsdatabaser som används på redaktionen är framför allt:

- Mediearkivet
- KB dagstidningar (Svenska dagstidningar, Kungliga biblioteket)
- Finlandssvenska dagstidningar (Nationalbiblioteket i Finlands digitala samlingar)
- DN-arkivet (Dagens Nyheters arkiv)
- SvD-arkivet (Svenska Dagbladets arkiv)

Mediearkivet, KB dagstidningar och Finlandssvenska dagstidningar utgör samlingsarkiv för en stor del av den svenska dagspressen. Mediearkivet innehåller ca 70 olika tidningar från 1980-talet och framåt. Denna resurs används flitigt, för att hitta sena illustrativa belägg om sådana saknas i SAOBArkiv, för att bekräfta betydelseanalysen eller för att kontrollera brukligheten hos ord eller användningar.

KB dagstidningar, som tillhandahåller tidningar som publicerats från 1700-talet och framåt, och Finlandssvenska dagstidningar, som täcker perioden 1771–1920, används framför allt för att hitta äldre första belägg än de som finns i SAOBArkiv.

Ibland begränsas visningen av texter i samlingsarkiven p.g.a. lagen om upphovsrätt. Därför används också några enskilda tidningars arkiv som är tillgängliga för prenumeranter, exempelvis DN:s och SvD:s arkiv.

### 2.1. Några specifika exempel från ordboksarbetet

Just nu är ord som börjar på *vre-vä* under redigering på SAOBArkiv redaktion, bl.a. sammansättningarna till ordet *värld*. För att visa vad tidningsdatabaserna kan

tillföra och vilket merarbete som hanteringen av dem medför i det enskilda fallet kommer några av dessa sammansättningar att diskuteras här.

I SAOBArkiv finns 9 belägg på ordet *världsartikel*, det äldsta från *Nordisk familjebok* (1920, 30:767). En tentativ definition, baserad på detta material lyder ”i sht *handel.* artikel (se d. o. III 3) som säljs i stora delar av världen”. En sökning i KB dagstidningar gör att årtalet för första belägg kan flyttas tillbaka 27 år, till 1893. Det tidigaste exemplet i Finlandssvenska dagstidningar är från samma år. (DN:s och SvD:s arkiv behöver inte undersökas eftersom de ingår i samlingsarkivet KB dagstidningar.) Under den aktuella perioden har ordet *värld* haft stavningsvarianter med *w-* och *-e-*. Således bör även dessa varianter beaktas. En sökning på strängen ”verldsartik\*” visar sig ge det tidigaste belägget, från DN (19/11 1866, s. 2):

Äfven vår stad har lemnat ett bidrag till den stora verldsexpositionen i Paris. Härifrån har neml. afsändts biskop Corts tofflor [...]. Som dessa tofflor ej synas ega någon annan märkvärdighet än den att hafva tillhört nämnde biskop föreställa vi oss, att de bland de öfriga verldsartiklarne komma att intaga en blygsam plats.

Som framgår av citatet ovan avser ordet *världsartikel* här inte en ’vara som säljs i stora delar av världen’. Artikeln i fråga är ett par tofflor som tillhört en biskop i Strängnäs och som ska ställas ut på världsutställningen i Paris. Den tidigare formulerade definitionen behöver alltså utökas för att inbegripa detta nytillkomna belägg. Den nya versionen lyder ”i sht *handel.* artikel (se d. o. III 3) som säljs i stora delar av världen; förr äv. om sak eller persedel av världsintresse”.

Av de ca 200 sammansättningar till *värld* som hittills är skrivna har tidigare belägg hittats för ca 25 %, bl.a. för orden *världsarena* (1906→1865), *världsatlats* (1924→1888), *världsberömmelse* (1929→1900), *världsess* (1963→1936) och *världsledande* (1974→1926). De kompletterande sökningarna i tidningsdatabaserna har alltså höjt kvaliteten på ordboksartiklarna, såväl vad gäller tidsfästning som definitioner.

Å andra sidan är kompletteringsarbetet tidskrävande. Även om tidningsdatabaserna kan ses som lättillgängliga kräver de ett omfattande merarbete. Äldre tidningar har ofta upp till sex spalter tättskriven text, och det är inte alltid som det aktuella textstället är markerat på sidan. Det kan ta lång tid att hitta fram till det eftersökta ordet. Vad gäller yngre tidningstexter, som skyddas av upphovsrättslagen, visas eftersökta textställen inte i sin helhet (KB dagstidningar begränsar exempelvis texter fr.o.m. 1900). Det kan därför vara omöjligt att citera texten utan att gå till originalkällan. Det kan också vara svårt eller omöjligt att med utgångspunkt i det lilla textutdrag som visas säkert avgöra det aktuella ordets betydelse.

## 2.2. Kvalitet kontra arbetsbörda och tidsåtgång

Det har framgått ovan att ett systematiskt och grundligt kompletteringsarbete kan

höja kvaliteten på ordboksartiklarna, och att kvalitetshöjningen har en prislapp i form av ökad arbetsbörda. Så som redaktionen arbetar för närvarande görs därför inte systematiska sökningar på varje enskilt ord. Det är upp till den aktuella redaktören att bedöma vad som är en ”rimlig insats” i det enskilda fallet.

Något som talar för en ambitiös komplettering av materialet är att angivandet av äldsta belägg på ord- och betydelsenivå är en grundläggande uppgift för SAOB. Ett större material betyder också att betydelseanalyserna blir säkrare och mer representativa. På minussidan kommer dock att arbetet tar betydligt längre tid om SAOB:s uppgift utökas från att beskriva det material som finns i SAOBArkiv till att innefatta allt som, mer eller mindre enkelt, går att hitta ”någonstans i världen”. Dessutom skulle ett systematiskt utnyttjande av elektroniska resurser medföra en obalans i verket SAOB, eftersom det bara är under de senaste åren som dessa möjligheter till kompletteringar stått till buds. Det skulle också bli obalans i den kronologiska beskrivningen av språket, eftersom det i nuläget främst är den senare delen av den nysvenska perioden som kan förbättras. Såsom läget är just nu blir det snedfördelning även vad gäller genre; tidningstexter blir överrepresenterade p.g.a. sin lättillgänglighet. Det viktigaste argumentet mot en systematisk och grundlig komplettering är kanske ändå att det blir obalans mellan olika ordtyper. Sammansättningar och monosema ord gynnas medan homografer och polysema ord missgynnas.

### 3. SAOB:s material i framtiden

Färdigställandet av SAOB närmar sig, och inom några få år är det dags att börja uppdatera den tryckta upplagan. En stor fråga blir då vilket material som ska beaktas. En önskedröm skulle vara att ha en mycket stor elektroniskt sökbar, lemmatiserad och taggad korpus som grund för hela ordbosinnehållet, från A till Ö. Korpusen skulle vara balanserad med avseende på kronologi, ämnesområde, genre, regionalitet och genus och innehålla ett brett urval texter som föreligger i sin helhet och som täcker hela beskrivningsperioden. Helst skulle den också vara specialanpassad för ökad hanterlighet och innehålla genvägar som gör det möjligt för lexikografen att på ett enkelt sätt välja ut och citera textställen. Att åstadkomma en sådan korpus skulle kräva specialkunskaper på en rad olika områden och enorma insatser vad gäller arbete, tid och pengar. En representativ och i alla avseende balanserad korpus måste nog dessutom betraktas som en utopi. Det förhåller sig ju onekligen så att vi lexikografer, som Atkins & Rundell uttrycker det (2008:74), ”are sampling from a population whose nature is unknowable and whose extent is unlimited”. En möjlig lösning som vi ser framför oss just nu är en fortsatt kombination av olika resurser.

För det första har vi ”SAOBArkiv 2” till vårt förfogande, dvs. den del av SAOBArkiv som har kommit in ”för sent” för varje enskilt ordboksband. Det rör sig om ca 500 000 excerpter, insamlade från 1890-talet och framåt. För det andra

tror vi oss ha goda möjligheter att skapa en mindre egen digital korpus, anpassad efter våra behov och utformad med en medveten strävan mot balansering och representativitet. Grunden till en sådan korpus skulle vi kunna få via vetenskapligt noggranna projekt som t.ex. de tidigare nämnda Litteraturbanken och Gender and Work. För det tredje tänker vi oss att vi även fortsättningsvis kommer att använda fritt tillgängliga textdatabaser. De olika materialen skulle kunna användas på olika sätt för olika syften. Det är ju nämligen så att den kommande uppdateringen rör högst olika aspekter med delvis olika materialmässiga implikationer.

En första viktig del av denna uppdatering är givetvis en komplettering med de ord som nu saknas i SAOB, det vill säga framförallt under 1900-talet i språket inkomna ord. Här kan man sannolikt komma rätt långt med den enklare lösning vi skisserade ovan. En annan del gäller komplettering av nya, återigen i huvudsak 1900-talsnya, användningar hos redan beskrivna ord. Detta innebär förstås lexikografiskt en större utmaning, eftersom det nu handlar om att gå in och ändra i befintliga artiklar. Materialmässigt bör det emellertid kunna klaras med samma eller liknande material som för kompletteringen av helt nya ord. Ett tredje slags komplettering gäller nya äldsta belägg hos redan beskrivna ord och användningar, ibland för den delen också yngsta belägg för mindre brukliga sådana. Just här, där det snarast handlar om material äldre än 1900-talet, kan man faktiskt tänka sig att korpusbalansering trots allt inte är en så tvingande nödvändighet. Eftersom man här kan utgå från SAOB:s befintliga beskrivning kan man i detta hänseende föreställa sig punktvisa sökningar i varjehanda äldre material, små som stora, gener specifika som mer allmänna, utifrån devisen att varje nytt äldstabelägg (eller yngstabelägg) är en vinst i sig. Malmgren (1999:201) framhåller att äldstabelägg måste ses som approximationer, särskilt för äldre tid, men att kravet på exakthet samtidigt ökar med tillgången till stora korpusar.

Ett gott material är naturligtvis en grundförutsättning för en god ordbok, och också om vi här skisserat möjliga vägar för SAOB:s materialmässiga framtid, är det tveklöst så att mycket ännu återstår att begrunda och överväga i detta hänseende, innan arbetet med SAOB:s andra upplaga kan inledas på allvar.

## Litteratur

### Ordböcker, textsamlingar och textdatabaser

- Dahlgren, F. A. (1914–1916): *Glossarium öfver föråldrade eller ovanliga ord och talesätt i svenska språket från och med 1500-talets andra årtionde*. Lund: C.W.K. Gleerups förlag.
- DN-arkivet: <arkivet.dn.se> (maj 2017).
- Finlandssvenska dagstidningar: <digi.kansalliskirjasto.fi/sanomalehti> (maj 2017).



Fornsvenska textbankens nysvenska annex: <project2.sol.lu.se/fornsvenska> (maj 2017).  
Gender and Work-databasen (GaW): <gaw.hist.uu.se> (maj 2017).  
KB dagstidningar: <tidningar.kb.se> (maj 2017).  
Korps historiska material: <spraakbanken.gu.se/korp/?mode=all\_hist> (maj 2017).  
Litteraturbanken: <litteraturbanken.se> (maj 2017).  
Mediearkivet: <retriever.se> (maj 2017).  
SAOB = *Ordbok över svenska språket*, utgiven av Svenska Akademien (1893–).  
Lund.  
SAOBArkiv = Samlingar til SAOB. Lund.  
SvD-arkivet: <svd.se/arkiv> (maj 2017).

### Annan litteratur

Atkins, Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.  
Hast, Sture (1983): Om SAOB:s material. I: *Nysvenska studier* 63, 159–192.  
Lundbladh, Carl-Erik (1996): Värdering av SAOB:s korpus och språkprovssamling. I: *LexicoNordica* 3, 91–103.  
Malmgren, Sven-Göran (2000): Korpusar i ordboksarbete och grammatikforskning. I: Gunilla Byrman, Hans Lindqvist & Magnus Levin: *Korpusar i forskning och undervisning. Rapport från ASLA:s höstsymposium, Växjö, 11-12 november 1999*. Uppsala: Association Suedoise de Linguistique Appliquée, 192–208.  
Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.

Bodil Rosqvist  
Biträdande huvudredaktör för SAOB  
Bodil.Rosqvist@svenskaakademien.se

Bo-A. Wendt  
Huvudredaktör för SAOB  
Bo.Wendt@svenskaakademien.se

Svenska Akademiens ordboksredaktion  
Dalbyvägen 3  
S-224 60 Lund

# Perfektum participium og ordklassen adjektiver

*Jørgen Schack*

This paper discusses the word form *perfect participle* from a word class perspective. In Danish, a formal distinction is made between participles with verbal meaning and participles with adjectival (property-ascribing) meaning. When used attributively, no such distinction is made. However, when used following the auxiliary verb *være*, the former word forms occur in supine constructions, where there is no inflectional agreement in number between the subject and the supine verb form. The latter word forms, on the other hand, occur in predicative constructions, where the participle agrees in number with the subject. Participles formed from some verbal stems only occur in verbal contexts, while participles formed from other verbal stems occur in both verbal and adjectival contexts. The formal distinction between verbal and adjectival participles can be used in order to decide which perfect participles could or should be listed in dictionaries as lemmas, i.e. adjectives.

## 1. Ordklassekriterier

I nordisk grammatisk tradition inddeles ordene i ordklasser efter tre kriterier: *betydning*, *bøjning* og *syntaktisk funktion*. Det fremgår bl.a. af de tre store referencegrammatikker Grammatik over det Danske Sprog (2011, herefter GDS), Norsk referansegrammatikk (1997, herefter NRG) og Svenska Akademiens Grammatik (1999, herefter SAG).

I GDS tages der ikke eksplicit stilling til inddelingskriteriernes rangorden. Det gør der til gengæld i NRG og SAG. Ifølge de to sidstnævnte grammatikker er hovedkriteriet bøjning, dernæst kommer syntaktisk funktion, og til sidst kommer betydning (NRG:21, SAG:7).

Ved ubøjelige ord må man gribe til det syntaktiske kriterium. Det gælder fx præsens participium, som i NRG henføres til adjektiverne fordi deres hovedanvendelse er adjektivisk (NRG:472). Det tredje og sidste kriterium, betydningskriteriet, har først og fremmest en kompletterende funktion i forhold til de to overordnede kriterier (NRG:21, SAG:7).

Det er ikke umiddelbart indlysende hvilken ordklasse ordformen perfektum participium tilhører. I de grammatiske fremstillinger karakteriseres ordformen ofte som adjektivisk (Nielsen 2014:181) fordi den har samme syntaktiske funktion som adjektiverne. Ligesom adjektiverne kan perfektum participium optræde som attribut i et nominalsyntaxme, og participiet bøjes i denne stilling i numerus og bestemthed, nøjagtig ligesom adjektiverne, fx *et aflyst fly – det aflyste fly – de af-*

*lyste fly*. Genusbøjning forekommer kun ved participier som er dannet af en række stærktbøjede verber, men jeg vil se bort fra genus i denne sammenhæng.

Efter inddelingskriterium 1 og 2, bøjning og syntaktisk funktion, er participiet *aflyst* altså en adjektivisk ordform snarere end en infinitiv form af verbet *aflyse*. Men er disse adjektiviske træk tilstrækkelige til at vi kan klassificere participiet som et adjektiv? *Betydningskriteriet* skal vel også være opfyldt, uanset hvordan det rangeres i grammatikkerne. En ordform er ikke et rigtigt adjektiv hvis den ikke har adjektivisk betydning.

Adjektivernes overordnede semantiske kendetegn er ifølge grammatikkerne at de udtrykker egenskaber ved referenter (se fx NRG:346). Adjektiver har *egenskabstilskrivende* betydning. I en attributivforbindelse som *et rødt hus* tilskriver adjektivet *rødt* referenten *hus* en egenskab. Spørgsmålet er nu om participier som fx *aflyst* i *et aflyst fly* og *skuffet* i *en skuffet tilhører* på samme måde som et adjektiv tilskriver referenterne 'fly' og 'tilhører' en egenskab, eller om de blot angiver at referenterne har været udsat for den handling som verberne *aflyse* og *skuffe* betegner.

I attributiv stilling er der ikke noget i formen der signalerer om et participium har egenskabstilskrivende eller verbal betydning. Et sådant signal har vi derimod i stillingen efter hjælpeverbet *være*. I dansk er det nemlig sådan at nogle participier typisk kongruerer i numerus med subjektet i stillingen efter *være*. Det gælder ikke mindst participier der er dannet af de såkaldte psykologiske verber, altså participier som fx *skuffet* og *chokeret*. Hvis man laver  $\chi^2$ -test på fordelingen af bøjningsformerne, viser det sig at tendensen til kongruensbøjning her er så stærk at vi får p-værdier der nærmer sig 0<sup>1</sup>, fx *subjekt pl. være skuffet/skuffede* ( $p < 0,0001$ ); *subjekt pl. være chokeret/chokerede* ( $p < 0,0001$ ). Det er m.a.o. helt usandsynligt at fordelingen af bøjningsformerne skyldes en tilfældighed.

Kongruensbøjning i stillingen efter *være* anses normalt som signal for at participiet udtrykker en egenskab (Becker-Christensen 2001:132, Hansen 1993:54), og hvis den karakteristik af bøjningsformens funktion er rigtig, så opfylder participier som de netop nævnte alle kriterier for at kunne klassificeres som adjektiver.

Andre participier synes derimod nærmest at insistere på ubøjet form i stillingen efter *være*. Det gælder fx participier som *aflyst* og *anholdt*, fx *subjekt pl. VÆRE aflyst/aflyste* ( $p < 0,0001$ ); *subjekt pl. VÆRE anholdt/anholdte* ( $p < 0,0001$ ). Den ubøjede form er signal for at *aflyst*, *anholdt* og *solgt* har verbal betydning. Ordformerne optræder her som hovedverbum i en perifrastisk passiv og har altså ingen adjektiviske træk. De bør derfor i denne stilling kaldes supinum snarere end perfektum participium.

Men mange participier vælger ikke side hvad angår spørgsmålet om verbal eller adjektivisk betydning; de optræder utvungent i begge kontekster. I stillingen efter *være* kan de have verbal betydning og optræder da typisk i ubøjet form, altså som supinum. Et eksempel er ordformen *trykt*:

<sup>1</sup> Mine frekvensoplysninger bygger på søgninger i et korpus med tekster fra landsdækkende aviser, regionsaviser og lokalaviser 2004 ff. Korpusset indeholder ca. 792 mio. løbende ord.

Valgplakaterne er *trykt*, så tiden er ikke til fordybelse og fælles løsninger.  
(Berlingske Tidende 30.8.2010, min fremhævelse)

Men *trykt* kan også have en klart egenskabsstilskrivende betydning, og da optræder den typisk i bøjet form:

På officielle kort over Libanon – hvad enten de er *trykte* eller digitale – bliver flygtningelejrene ofte vist som grå klatter, uden gader tegnet ind.  
(Arkitekten 3.6.2016, min fremhævelse)

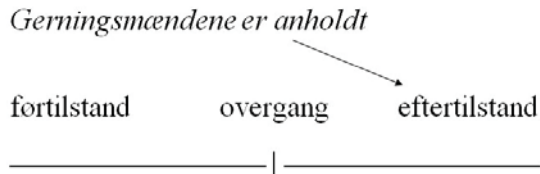
## 2. Participiernes betydningspotentiale

De forskelle i bøjningspotentiale som man kan påvise ved de danske participier, afspejler participiernes betydningspotentiale. Nogle participier synes nærmest kun at optræde med verbal betydning, mens andre participier *både* kan have verbal og adjektivisk betydning – og nogle af de sidstnævnte optræder som nævnt ovenfor altovervejende i adjektiviske kontekster. Disse forskelle ser ud til at hænge sammen med den semantiske rolle som de tilsvarende transitive verber tildeler et direkte objekt.

Den amerikanske grammatiker Van Valin (2001:24) opdeler den semantiske rolle som traditionelt benævnes *patients* ('den referent der udsættes for verbalhandlingen'), i to roller, nemlig *patient* og *theme*. At en objektsreferent har rollen *patient*, betyder hos Van Valin at den verbalhandling som objektsreferenten udsættes for, implicerer en forandring af referentens (interne) egenskaber eller beskaffenhed: "A patient argument is either in a state or undergoes a change of state or condition" (ibid.). En objektsreferent med rollen *theme* er derimod karakteriseret ved at verbalhandlingen ikke har nogen indvirkning på dens egenskaber eller beskaffenhed (ibid.). Som eksempler på prototypiske *theme*-argumenter nævner Van Valin objektsreferenter der som følge af verbalhandlingen ændrer lokation eller skifter ejer, fx *børnene* i *Jeg henter børnene* og *malerierne* i *Vi har solgt malerierne*.

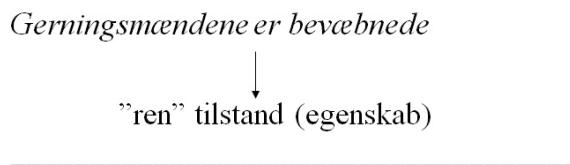
I en konstruktion som *Det svenske politi har anholdt gerningsmændene* har objektsreferenten en semantisk rolle der svarer til den netop nævnte karakteristik af rollen *theme*. Verbalhandlingen 'anholde' synes nemlig ikke at implicere en forandring af objektsreferentens egenskaber. Det ses tydeligere i en tilsvarende *være*-passiv, hvor udtrykssubjektet er indholdsobjekt for supinum (Nielsen 2014:184). En passivkonstruktion som *Gerningsmændene er anholdt* informerer ikke om en egenskab ved indholdsobjektets referent. Den informerer snarere om en *situation*, dvs. de omstændigheder som referenten er underlagt efter at den har været udsat for verbalhandlingen: "Ved kombination med et overgangsverbum [...] udtrykker *være*-passiv den tilstand der følger efter den overgang som verbet benævner, dvs. der er et element af førtidighed i forhold til et senere referencetidspunkt." (Nielsen

2012:185). En sådan konstruktion har m.a.o. en heterogen *tidskontur* eller *tidsprofil* (se fx GDS:192 ff.). Den udtrykker en eftertilstand, og dermed præsupponerer den en førtilstand (jf. figur 1).



Figur 1: Passivkonstruktion

Hvis vi i stedet for *anholdt* indsætter participiet *bevæbnet*, får vi derimod en konstruktion som informerer om en *egenskab* ved subjektsreferenten. Participiet tilskriver subjektsreferenten en egenskab, nøjagtig ligesom et egentligt adjektiv, og vi kan her uden videre sideordne participiet med et adjektiv: *Gerningsmændene er bevæbnede og farlige*. Konstruktionen er en prædikativforbindelse, og sådanne konstruktioner har en homogen tidsprofil. Der udtrykkes ingen overgang fra en førtilstand til en eftertilstand. Den eneste tid der er involveret, er den som kopula-verbet bidrager med (jf. figur 2).



Figur 2: Prædikativkonstruktion

Udover den homogene tidsprofil er prædikativforbindelsen karakteriseret ved at forestillingen om en agens er helt fraværende, og man vil da heller ikke kunne tilføje et agensadverbial: \**Gerningsmændene er bevæbnede af NN*. Forestillingen om en agens ligger derimod lige for i passivkonstruktionen, og her vil man uden videre kunne tilføje et agensadverbial: *Gerningsmændene er anholdt af det svenske politi*.

De transitive overgangsverber kan overordnet inddeles i to typer, som jeg her for nemheds skyld omtaler som type A og type B.

Verber af *type A* der styrer et objekt med en rolle der er beslægtet med Van Valins *theme*-rolle, i den forstand at verbalhandlingen ikke implicerer en forandring af objektsreferentens egenskaber. Participer der dannes af sådanne verber, har verbal betydning. Når de optræder efter *være*, har vi en passivkonstruktion, dvs. en konstruktion med supinum, fx *Brevet er sendt*; *Malerierne er solgt*; *Gids-*

*lerne er frigivet*. Når de optræder i attributiv stilling, altså som participium, har de samme passiviske betydning som i stillingen efter *være*, fx *det sendte brev* 'brevet som er blevet sendt (af nogen)', *de solgte møbler* 'møblerne som er blevet solgt (af nogen)', *de frigivne gidsler* 'gidslerne som er blevet frigivet (af nogen)'.

Type A-verber kan indordnes i forskellige semantiske felter. Her følger nogle eksempler på forskellige "type A-felter": *annoncere, bekendtgøre, meddele, offentliggøre – anholde, pågribe, tilbageholde; frigive, løslade; sigte, dømme – af-dække, afsløre, opdage, opklare – indfri, opfylde; nå, opnå – indlede, starte; af-slutte, stoppe – aflevere, sende; hente, modtage*.

Verber af type B styrer objekter hvis referenter forandrer egenskaber som følge af verbalhandlingen. De tilsvarende participier kan, afhængigt af konteksten, have verbal eller adjektivisk betydning, fx *Dækkene er slidt op* : *Dækkene er slidte*; *Gulvene er lakeret to gange* : *Gulvene er lakerede*. Nogle participier i denne gruppe optræder dog som nævnt typisk med adjektivisk betydning. Når sådanne participier optræder i stillingen efter *være*, vil vi som default tolke konstruktionen som en prædikativforbindelse snarere end som en passivkonstruktion, fx *Eleverne er udmattede*; *Bestyrelsesmedlemmerne var chokerede*; *Møblerne er ramponerede*. Forestillingen om en agentiv eller kausativ størrelse er her typisk helt fraværende (jf. Hansen 1967:58).

Også type B-verberne kan indordnes i forskellige semantiske felter, fx *fernisere, lakere, male, ludbehandle – bruge, ramponere, skramme, slide – afkræfte, nedslide, svække, udmatte – marinere, røge, salte, sylte – hækle, strikke, sy, væve – demokratisere, marginalisere, radikaliserer, sekularisere – chokere, forundre, overraske, skræmme*.

### 3. Perfektum participium som opslagsord i ordbøgerne

I Retskrivningsordbogens "Vejledning i ordbogens brug", der bl.a. oplyser om ordbogens ordforråd, står der følgende om participierne:

Participier er normalt ikke opført som selvstændige opslagsord. Ord(former) som fx *renset* og *syngende* er således i ordbogen regnet for bøjningsformer af verberne *rense* og *syng*. Participier er dog medtaget som selvstændige adjektiver hvis deres betydning eller dannelsesmåde afviger fra de tilsvarende verber, fx *frossen, henrykt, sleben, spændende*. (Retskrivningsordbogen:12)

Den praksis der er beskrevet her, svarer vistnok til den gængse praksis i tilsvarende ordbøger.

Men Retskrivningsordbogen indeholder, ligesom mange andre ordbøger, en del participialadjektiver der hverken afviger mht. betydning eller dannelsesmåde. Det gælder bl.a. participier som er dannet af psykologiske verber. Både *Retskriv-*

*ningsordbogen* og *Den Danske Ordbog* medtager fx participiet *bekymret* som et selvstændigt adjektiv, selvom participiets betydning kan udledes direkte af verbets betydning (se opslagene *bekymre* og *bekymret* i *Den Danske ordbog*).

Antallet af psykologiske verber er ganske stort. Der er over 100 i dansk, og jeg gætter på at der er et nogenlunde tilsvarende antal i svensk og norsk. Her er nogle eksempler fra dansk: *bekymre*, *chokere*, *forbavse*, *forbløffe*, *forundre*, *forvirre*, *imponere*, *irritere*, *krænke*, *ophidse*, *overraske*, *ryste*, *skuffe*, *stresse*, *såre*, *traumatisere*, *tænde*. Ved de understregede verber er de tilsvarende participier medtaget som selvstændige adjektiver i *Retskrivningsordbogen* og/eller *Den Danske Ordbog*.

Ordbøgernes udvalg af sådanne adjektiver ser ud til at være lidt tilfældigt, for participier som fx *chokeret*, *skuffet* og *såret* er nøjagtig lige så adjektiviske som fx *bekymret*, *forbavset* og *stressed*. Her har vi altså et af de semantiske felter som ordbogsredaktørerne kunne se lidt nærmere på hvad angår udvalget af adjektiver.

Indenfor mange semantiske felter er det sådan at jo længere vi bevæger os ned ad abstraktionsstigen, desto større er sandsynligheden for at et participium optræder med adjektivisk betydning. Verbalstammer med et forholdsvis højt abstraktionsniveau synes nemlig ofte at have et ringere adjektivisk potentiale end deres hyponymer, dvs. verbalstammer med en mere specifik betydning. Det gælder fx de verber der styrer såkaldte efficerede objekter, dvs. objekter der opstår ('bliver til') som følge af verbalhandlingen, fx *producere biler*, *bygge et hus*, *væve et tæppe*. I hyponymier som fx *bygge* > *mure* > *fuldmure* og *fremstille* > *væve* > *maskinvæve* optræder de participier der dannes af verbalstammerne på det øverste trin, normalt kun med verbal betydning, fx *Husene er bygget (i 1960'erne)*; *Tæpperne er fremstillet (i hånden)*. Participierne på mellemtrinnet kan optræde med såvel verbal som adjektivisk betydning, fx *Tæpperne er vævet af en kunsthåndværker*; *Tæpperne er vævede*. Også participierne på det nederste trin, *fuldmuret* og *maskinvævet*, kan optræde med begge betydninger, men i praksis optræder de helt overvejende med adjektivisk betydning. Verbalstammerne på det nederste trin optræder nemlig forholdsvis sjældent som fuldblods verber, altså i finit form eller som infinitiv. Mange sammensatte verber af denne type er – eller opleves som – tilbagedannelser. Om de historisk set faktisk ér tilbagedannet fra et participium, er ofte meget svært at afgøre.

Som ordbogsredaktør kan man have svært ved at afgøre om den verbale stamme i tilfælde som de ovennævnte skal repræsenteres af et verbum eller et adjektiv, eller måske begge dele. Det kan illustreres med et eksempel fra det semantiske felt 'tilberedning af fødevarer'. I hyponymien *tilberede* > *stege* > *sprødstege* – *tørstege* er de to kohyponyme verbers stammer, *tørsteg-* og *sprødsteg-*, i *Retskrivningsordbogen* repræsenteret af henholdsvis et verbum, *tørstege*, og et adjektiv, *sprødstegt*. Jeg kan ikke pege på nogen tungtvejende grunde til at vi redaktører har truffet netop dét valg, og det er nok heller ikke indlysende for ordbogsbrugerne. Der kan m.a.o. være grund til at se nærmere på hvordan man repræsenterer verbalstammer af denne type i sin ordbog.

#### 4. Perfektum participium i svensk og norsk

I *svensk* kongruensbøjes perfektum participium på samme måde som adjektiverne (SAG:620). I stillingen efter *vara* og *bli* er participiet altså kongruensbøjet, uanset om det har verbal eller adjektivisk betydning, fx *Alla flyg är inställda* (jf. dansk: *Alle fly er aflyst*); *Gärningsmännen är beväpnade*.

I *nynorsk* benyttes som hovedregel fuld kongruensbøjning, på omtrent samme måde som i svensk. Men man kan også følge et system hvor kun de stærke participier kongruensbøjes – eller et system hvor ingen af participierne kongruensbøjes (NRG:518). Uanset hvilket system man følger, gør bøjningen altså heller ikke her nogen forskel på adjektiviske og verbale participier.

I *bokmål* bliver ”adjektivet i princippet kongruensbøyd, mens partisippet ikke får slik bøyning” efter *være* og *bli* (NRG:345). I praksis ser det ud til at participierne normalt kun bøjes hvis deres betydning afviger fra de tilsvarende verber, eller hvis de ikke har et ægte verbalt modsvar og derfor kun etymologisk set er participier. Heller ikke de participier som i dansk altovervejende optræder i bøjet form efter *være*, bøjes i bokmål. Det gælder fx participier dannet af psykologiske verber. Søgninger i *Norsk aviskorpus*<sup>2</sup> viser at sådanne participier i de fleste tilfælde optræder i ubøjet form (med p-værdier der nærmer sig 0), også selvom de er gradsbestemte og dermed tydeligt markeret som adjektiviske, fx

Vi er *forbauset* og alvorlig *bekymret*. (Aftenposten 11.2.2000, mine fremhævelser).

#### 5. Afsluttende bemærkninger

Både dansk, svensk og norsk har oprindeligt haft fuld kongruensbøjning af perfektum participium. Denne tilstand er bevaret i svensk og til dels i nynorsk. Bokmål har derimod som hovedregel ubøjet form efter *være*. Dansk befinder sig i en mellemposition, idet man efter *være* skelner mellem ordformer med henholdsvis verbal og adjektivisk betydning. De førstnævnte bøjes ikke; de sidstnævnte kongruerer med subjektet i numerus. Mange verbalstammer danner kun participier med verbal betydning. Sådanne participier optræder udelukkende i attributiv stilling. I stillingen efter *være* optræder de tilsvarende ordformer som supinum, typisk som hovedverbum i en perifrastisk passiv.

I dansk vil man altså kunne benytte bøjningskriteriet i forbindelse med ordklasseinddelingen af participierne. Og som det er fremgået, er det participiets betydning der afgør om det kongruensbøjes eller ej, så her er det i en vis forstand betydningskriteriet der får det sidste ord.

<sup>2</sup> Jf. <<http://avis.uib.no/avis>> (oktober 2017).



## Litteratur

### Ordbøger

*Den Danske Ordbog*. Udgivet af Det Danske Sprog- og Litteraturselskab. <<http://ordnet.dk/ddo>> (maj 2017).

Retskrivningsordbogen (2012) = A.Å. Jervelund, J. Schack, J.N. Jensen, M.H. Andersen: *Retskrivningsordbogen*. 4. udgave. København: Alinea.

### Anden litteratur

Becker-Christensen, Christian (2001): Den grædte pige og lignende historier. Om perfektum participium efter *blive* og *være* og som adled. I: Pia Jarvad, Frans Gregersen, Lars Heltoft, Jørn Lund og Ole Tøgeby (red.): *Sproglige Åbninger: E som Erik, H som halvfjerds*. Festskrift til Erik Hansen 18. september 2001. København: Hans Reitzels Forlag, 119–139.

GDS = Erik Hansen og Lars Heltoft (2011): *Grammatik over det Danske Sprog*. Bind 1–3. København: Det Danske Sprog- og Litteraturselskab.

Hansen, Aage (1967): *Moderne Dansk*. Bind 3. København: Grafisk Forlag.

Hansen, Erik (1997): *Rigtigt Dansk*. 2. udgave. København: Hans Reitzels Forlag.

Nielsen, Peter Juul (2012): Supinum i dansk og svensk. *Ny Forskning i Grammatik* 19, 2012, 181–198.

Nielsen, Peter Juul (2014): Supinum versus perfektum participium i dansk Indholdsdifferentiering, mulige bøjningsvalg og andenordensnul. *Ny Forskning i Grammatik* 21, 177–192.

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

SAG = Ulf Teleman, Staffan Hellberg & Erik Andersson (1999): *Svenska Akademiens Grammatik*. Bind 2. Stockholm: Norstedts Ordbok.

Van Valin, Robert D. (2001): *An Introduction to Syntax*. Cambridge: Cambridge University Press.

Jørgen Schack  
seniorforsker, cand.mag.  
Dansk Sprognævn  
Worsaaesvej 19, 4. sal  
DK-1972 Frederiksberg  
schack@dsn.dk

# Nye islandske retskrivningsregler og den islandske retskrivningsordbog

*Jóhannes B. Sigtryggsson*

In 2016 new Icelandic official spelling rules by the Icelandic Language Council were published by the Education ministry. Also, in 2016, a new edition of the official Icelandic spelling dictionary, *Stafsetningarorðabókin*, was published on the web portal *málið.is*. In this article, the genesis of these works is discussed, in particular the emphasis that was put on their integration by using numerous references to the new spelling rules in the dictionary.

## 1. Indledning

Året 2016 var på mange måder et vigtigt år for islandsk retskrivning: nye retskrivningsregler var udgivne af Islandsk sprognævn og kulturministeriet og en ny udgave (2. udg.) af den offentlige islandske retskrivningsordbog, *Stafsetningarorðabókin*, blev lanceret på webportalen *málið.is*. Jeg vil gøre kort rede for arbejdet ved disse forskellige projekter og hvordan de har påvirket hinanden. Jeg vil også diskutere gavnet af det at have et nært forhold mellem en retskrivningsordbog og retskrivningsregler og gå lidt ind på hvordan man kan styrke ortografiske ordbøgers rolle eller formål i nutiden med at synliggøre mere end før denne forbindelse.

Disse to projekter er udgivne af forskellige aktører: retskrivningsreglerne af Islandsk sprognævn men *Stafsetningarorðabókin* af sprogrøgtsavdelingen i Árni Magnússon-instituttet for islandske studier.<sup>1</sup>

## 2. Nye islandske retskrivningsregler (2016)

### 2.1. Islandsk retskrivning

Islandsk retskrivning er ret konservativ og bygger på et traditionelt skriftsprog som har en lang historie.<sup>2</sup> Før nye reglerne blev godtaget var de sidste officielle retskrivningsregler fra 1974, med småændringer 1977.<sup>3</sup> Kulturministeriet udgav

<sup>1</sup> Jeg deltog i begge disse projekter, var hovedredaktør for den nye udgave af ordbogen og tog også del i at lave de nye retskrivningsregler som medlem af en arbejdsgruppe.

<sup>2</sup> God oversigt over retskrivningshistorien er hos Stefán Karlsson (2004).

<sup>3</sup> *Auglýsing um íslenska stafsetningu nr. 132/1974* [Meddelelse om islandsk retskrivning nr. 132/1974] og *Auglýsing um breyting á aughýsingu nr. 132/1974 um íslenska stafsetningu nr. 261/1977* [Meddelelse nr. 261/1977 om ændring på meddelelse nr. 132/1974 om islandsk retskrivning] (se *Auglýsing um íslenska stafsetningu*).

reglerne fra 1974 (1977). Disse regler var på mange måder vellykkede men også forældede på visse områder og gik ikke nok i detaljerne.

Her skal også nævnes at i første udgaven af *Stafsetningarorðabókin* (2006) var også publicerede retskrivningsregler. De havde ikke et selvstændig officiel værdi, men byggede på reglerne fra 1974.<sup>4</sup> Disse regler fra 2006 var et forsøg til at uddybe reglerne fra 1974, bl.a. med langt flere eksempler end før og var et stort fremskridt på og dannede grundlaget i mange måder for de nye retskrivningsregler som var udgivne 2016.

## 2.2. Islandsk sprognævn og nye retskrivningsregler

Ifølge lovene om islandsk sprog og islandsk tegnsprog, nr. 61 fra 2011 (paragraf 6), og ældre lov om Árni Magnússon-instituttet for islandske studier fra 2006 fik Islandsk sprognævn den opgave at affatte retskrivningsregler som gælder bl.a. for undervisning af retskrivning i skoler og kulturministeren udgiver. Grundlæggende ændringer af reglerne skal godkendes af kulturministeren.<sup>5</sup>

Formålet med at affatte nye retskrivningsregler var blandt andet at de skulle få et selvstændigt værdi og basis i lovene og ville ikke kun bygges på reglerne fra 1974. Fremstillingen skulle også gøres tydeligere end i ældre regler og antallet af eksempler forøges.

Islandsk sprognævn stiftede en arbejdsgruppe 2009 som arbejdede med pauser til 2015 på at affatte nye regler. Sprognævnet afleverede et udkast til nye retskrivningsregler til kulturministeriet 6. april 2016 og ministeriet udgav dem 6. juni samme år.<sup>6</sup> Reglerne blev dernæst også udgivet på sprognævnets webside (Íslensk málnefnd 2016).

Disse nye retskrivningsregler er derfor første reglerne som sprognævnet udgiver efter at nævnet fik denne klare opgave i lovene. Udgivelsen af de nye regler er derfor et vis vendepunkt i den islandske retskrivningshistorie.

## 2.3. Ændringer i de nye retskrivningsregler

Hvis man kigger på ændringer i de nye retskrivningsregler fra ældre regler er de ikke ret store. Man kan helst se dem i kapitler som handler om store og små bogstaver og et eller flere ord. Største ændringerne er hellere i fremstilling af emnet og antal eksempler (se Jóhannes B. Sigtryggsson 2016).

Et eksempel om en ændring er at i reglerne fra 1974 står at alle dyre- og plantnavne skal skrives med en lille bogstav endog første led i dem er oprindelig et

<sup>4</sup> *Ritreglur í samræmi við auglýsingar menntamálaráðuneytis nr. 132/1974, 133/1974, 184/1974 og 261/1977* [Retkskrivningsregler i overensstemmelse med meddelelser nr. 132/1974, 133/1974, 184/1974 og 261/1977] (*Stafsetningarorðabókin* 2006:673–736).

<sup>5</sup> *Lög um stöðu íslenskrar tungu og íslensks tákn máls nr. 61/2011*.

<sup>6</sup> *Auglýsing um setningu íslenskra ritreglna nr. 695/2016. 6. júní 2016*.

proprium. Et eksempel på det er ordet *óðinshani* ‘odinshane’ hvor den første led er oprindeligt navnet *Óðinn*. I de nye retskrivningsregler har denne regel været ekspanderet til flere grupper ord, specielt ord om mad og medicinord. De skal nu også altid skrives med små bogstaver, f.eks. *pekingönd* ‘pekingand’ og *alzheimer sveiki* ‘Alzheimers sygdom’.

Anden ændring som kan nævnes er at man har formindsket antallet af sideformer. Ofte har det med et eller flere ord at gøre. For eksempel er den eneste stavemåde som nævnes i nye reglerne for ordforbindelsen *alls konar*, som betyder ‘mange slags’, at skrive hvert ord separat. I ældre regler kunne man derimod enten skrive dem separat eller i et ord (*allskonar*, *alls konar*). Det er vores erfaring i sprogrøgtsavdelingen hos Árni Magnússon-instituttet i islandske studier at almenheden vil hellere have færre sideformer og varianter i retskrivningen end flere.

### 3. *Stafsetningarorðabókin* – den officielle islandske retskrivningsordbog

#### 3.1. Retskrivning i islandske ordbøger og databaser

Standardordbogen over moderne islandsk nævnes *Íslensk orðabók*. Første udgave af den udkom 1964 men den sidste store revidering af den er fra 2002 (*Íslensk orðabók* 2002). Denne ordbog er ikke så normativ som *Stafsetningarorðabókin* og retskrivningsreglerne og der forekommer flere varianter i retskrivningen end i dem.

Islandsk sprognævn udgav i 1989 *Réttritunarorðabók handa grunnskólum*, (retskrivningsordbog for grundskoler). Den blev ret populær og er endnu brugt i grundskolen i Island. Største problemet med den er at den er blevet forældet på mange måder. Den har ikke været opdateret fra året 1989 og følger ikke helt retskrivningsreglerne fra 2006 og selvfølgelig ikke heller de nye regler fra 2016. Ordforrådet i den består også kun af cirka 15.000 ord.

De sidste cirka femten år er den såkaldte *Beygingarlýsing íslensks nútímamáls* (bøjningsdatabase for islandsk nutidssprog) (se <<http://bin.arnastofnun.is>>), som Árni Magnússon-instituttet udgiver, blevet den de facto sted hvor man går til at finde oplysninger om bøjning af næsten 280.000 islandske ord og den er blevet meget mere kendt og populær end *Stafsetningarorðabókin*. Oprindeligt var denne database udviklet for sprogteknologi og har altid kun ligget på nettet. Den er deskriptiv og viser derfor ret mange lemmaformer som ikke er anerkendte i formel skriftsprog, f.eks. findes der såvel *tölva* ‘computer’ som sideformet *talva*. Den første varianten er den anerkendte form i skriftsproget men *talva* er hyppigt brugt i talesproget. Største problemet med denne imponerende database er at almenheden bruger den nu ofte som en normativ kilde. Det er derfor vigtigt at fundere på den islandske retskrivningsordbogs eksistensberettigelse, specielt fordi den samme

institution, Árni Magnússon-instituttet i islandske studier, udgiver begge projekterne, bøjningsdatabasen og *Stafsetningarorðabókin*. Det var blandt andet med denne bøjningsdatabase i tanke som forholdet mellem *Stafsetningarorðabókin*, og retskrivningsreglerne er i 2. udg. af ordbogen blevet tættere end før, som diskuteres her i kapitel 3.3.

### 3.2. *Stafsetningarorðabókin*

*Réttitunarorðabók handa grunnskólum* dannede grundlaget for *Stafsetningarorðabókin* (ca. 65.000 lemmaer) som udkom for første gang 2006.<sup>7</sup> Sekretariatet for Islandsk sprogævn lavede den og sprogævnet udgav denne første officielle retskrivningsordbog i Island. Den blev populær men var kritiseret blandt andet for mangel af henvisninger mellem sideformer. I den nye udgave af denne ordbog fra sidste år har der været sat ind mange henvisninger til sideformer. Et eksempel på det er ordet *skritinn* som betyder ‘underlig, mærkelig, besynderlig’ og kan enten skrives med *i* eller *y*. I den nye udgave af ordbogen henvises der til sideformene, som ikke var gjort i første udgaven af ordbogen:

*Stafsetningarorðabókin* (2006; 1. udg.):

**skritinn** skritin; skritið STIGB -nari, -nastur

**skrýtinn** skrýtin; skrýtið STIGB -nari, -nastur

*Stafsetningarorðabókin* (2016; 2. udg.):

**skritinn** (einnig **skrýtinn**) skritin; skritið STIGB -nari, -nastur

**skrýtinn** (einnig **skritinn**) skrýtin; skrýtið STIGB -nari, -nastur

En vigtig grund til at have en retskrivningsordbog er at vise sådan ligeværdige sideformer. Det er en del af det normative formål med således ordbøger. Retskrivningsordbøger er også vigtige til at sprede ud viden om overvejet retskrivning af f.eks. låneord. Vigtigste formålet med retskrivningsordbøger er at være en kanal til at fremføre normativ praksis, jf. definitionen for *rettskrivningsordbok* i NLO (1997:226): ”ordbok der formålet er å angi ordenes rettskrivning ifølge en viss norm.”<sup>8</sup>

Revidering af emnet i ordbogen har taget nogle år og tidligt i dette arbejde var bestemt af redaktionen at den nye udgave ikke skulle trykkes men kun publiceres på nettet. Det er i takt med udviklingen i Island og andre lande hvor meget færre ordbøger er trykte end før. Med lanceringen af den nye webportal *málið.is* i november 2016 fik *Stafsetningarorðabókin* også et nyt hjem og den nye version var publiceret der. Endnu er det den eneste sted hvor den ligger men det diskuteres nu om der skal også oprettes som en selvstændig webside. Det som jeg håber også

<sup>7</sup> Om historien bag den se Ari Páll Kristinsson (2008).

<sup>8</sup> Diskussion om særpræg og formål ved ortografiske ordbøger er f.eks. hos Nerius (1990).

at gøre i den nærmeste fremtid er at sætte html-link til retskrivningsreglerne fra opslagsord i *Stafsetningarorðabókin* så at man flyttes direkte til den relevante paragraf i reglerne når man taster på paragrafnummeret.

### 3.3. Henvisninger til nye retskrivningsregler i *Stafsetningarorðabókin*

En af største nyhederne i denne nye udgave af retskrivningsordbogen er systematiske henvisninger i lemmaer til relevante paragrafer i de nye officielle retskrivningsregler.

I første udgave af *Stafsetningarorðabókin* fra 2006 var det kun få henvisninger til daværende retskrivningsregler. Tanken fra begyndelsen af arbejdet med den anden udgave af ordbogen har været at den og de nye retskrivningsregler skulle styrke hver anden og kobles mere sammen end før. Forudsætningen for at den nye udgave af retskrivningsordbogen kunne færdiggøres var derfor at de nye retskrivningsregler ville være færdige. Disse to projekter måtte derfor gå hånd i hånd.

Hensigten var at henvisninger i ordbogen til paragraffer i retskrivningsreglerne ville styrke ordbogens rolle som den officielle retskrivningsordbog og skelne den fra andre islandsk-islandske ordbøger og databaser. Det ville også forhåbentlig hjælpe med at sprede viden om retskrivningsreglerne.

Nogle eksempler om henvisninger til paragrafer i reglerne kan nævnes her: *týsdagur* 'tirsdag' er det gamle islandske navn på dagen *tirsdag* men som nu nævnes i islandsk *priðjudagur*. Første led i den er navnet *Týr*. Her henvises til artikel 1.3.2 c i de nye retskrivningsregler hvor der står at navne på dage og måneder skal skrives med et lille bogstav. Ordets retskrivning har ændret sig fra 2006 hvor den begyndte på en stor bogstav:

*Stafsetningarorðabókin* (2006): **Týsdagur** -inn -dags; -dagar

*Stafsetningarorðabókin* (2016): **týsdagur** -inn -dags; -dagar (sjá §1.3.2 c í Ritreglum).

Et andet eksempel kan nævnes. Det er almindeligt i islandsk at sammensatte ord som oprindeligt har haft et proprium som første del har tabt den direkte henvisning til det og er blevet et fællesnavn, for eksempel *akkillesarhæll* 'akilleshæl', som er ledt af det græske navn *Akkilles*. Her henvises i den nye udgave af *Stafsetningarorðabókin* til paragraf 1.3.3.2 g i retskrivningsreglerne hvor således ord er behandlede:

*Stafsetningarorðabókin* (2006): **akkillesarhæll** -inn -hæls

*Stafsetningarorðabókin* (2016): **akkillesarhæll** -inn -hæls (sjá §1.3.3.2 g í Ritreglum)

Et forbillede i dette arbejde med henvisningerne og fremstillingen af dem har været den danske retskrivningsordbog, *Retskrivningsordbogen*, og specielt 4.

udgaven fra 2013. Antal henvisninger i *Stafsetningarorðabókin* er derimod højere end i den.

#### 4. Konklusion

Nu har jeg diskuteret historien bag disse projekter. Man kan måske drage lære af dette om retskrivningsordbøger og retskrivningsregler: Retskrivningsordbøger må finde en eksistensberettigelse med at skelne sig fra andre typer af ordbøger/databaser. Ortografiske ordbøger er vigtige til at sprede ud viden om overvejet retskrivning af f.eks. låneord og de er en god vej/kanal til at fremføre normativ praksis. Tæt (synlig) forbindelse mellem en retskrivningsordbog og retskrivningsregler er vigtig.

Forhåbentligt har alt dette arbejde ved nye retskrivningsregler og den nye udgave af retskrivningsordbogen defineret klarere end før deres plads og rolle i islandsk retskrivning.

#### Litteratur

##### Ordbøger

*Beygingarlýsing íslensks nútímamáls*. Kristín Bjarnadóttir (red.). Stofnun Árna Magnússonar í íslenskum fræðum. <<http://bin.arnastofnun.is/>>, hentet 11.10.2017.

*Íslensk orðabók* (2002). 3. udg. Mördur Árnason (red.). Reykjavík: Edda.

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

*Retskrivningsordbogen*. (2013). 4. udg. Dansk Sprognævn.

*Réttitunarorðabók handa grunnskólum*. (1989). Reykjavík: Námsgagnastofnun og Íslensk málnefnd.

*Stafsetningarorðabókin*. (2006). 1. udg. Dóra Hafsteinsdóttir (red.). Reykjavík: Íslensk málnefnd / JPV.

*Stafsetningarorðabókin*. (2016). 2. udg. Stofnun Árna Magnússonar í íslenskum fræðum. <<http://málið.is/>>, hentet 11.10.2017.

## Anden litteratur

- Ari Páll Kristinsson (2008): Den nye islandske rettskrivningsordboka. I: Ásta Svavarsdóttir et al. (red.): *Nordiske Studier i Leksikografi. 9. Rapport fra konference om leksikografi i Norden. Akureyri 22.–26. maj 2007*. Reykjavík: Nordisk Forening for Leksikografi, 19–30.
- Auglýsing um íslenska stafsetningu. [Þetta er auglýsing nr. 132/1974, með innfelldum breytingum skv. auglýsingu nr. 261/1977.]* <<https://notendur.hi.is/eirikur/stafsreg.htm>>, hentet 11.10.2017.
- Auglýsing um setningu íslenskra ritreglna nr. 695/2016. 6. júní 2016.* <<https://www.stjornarradid.is/media/menntamalaraduneyti-media/media/fretta-tengt2016/Auglysing-um-islenskar-ritreglur.pdf>>, hentet 4.4.2018.
- Íslensk málnefnd (2016). *Ritreglur. Auglýsing mennta- og menningarmálaráðuneytis nr. 695/2016 með leiðréttingum.* <<http://islenskan.is/images/ritreglur-IM-2016.pdf>>, hentet 11.10.2017.
- Jóhannes B. Sigtryggsson (2016): Greinargerð um helstu breytingar á ritreglum. I: *Skíma* 39, 40–41.
- Lög um stöðu íslenskrar tungu og íslensks tákn máls nr. 61/2011.* <<https://www.althingi.is/altext/stjt/2011.061.html>>, hentet 11.10.2017.
- Nerius, Dieter (1990). *Das Orthographiewörterbuch. I: Franz Josef Hausman et al. (red.): Wörterbücher. Ein internationales Handbuch zur Lexikographie.* Bind 5.2. Berlin / New York: De Gruyter, 1297–1304.
- Stefán Karlsson (2004). *The Icelandic language.* Translated by Rory McTurk. Viking Society for Northern Research. University College London.

Jóhannes B. Sigtryggsson  
 forskningslektor  
 Árni Magnússon-instituttet for islandske studier  
 Laugavegur 13  
 101 Reykjavík  
 johannes.b.sigtryggsson@arnastofnun.is



# Amatörlexikografiska insatser på sajten *Folkmun.se*

*Emma Sköldberg & Lena Wenner*

This paper describes the results of a study of the user-generated web dictionary *Folkmun.se*. The platform was launched in 2007 and contains (December 2016) about 4 600 headwords (lemmas). We analyse how the content of *Folkmun.se* has developed over time with particular focus on ten contributor aliases and how they have participated in the development of the dictionary. Our study displays some recurrent pattern in how different aliases work. We observe similarities in what kind of words they contribute with and the manner in which they define words. Many lemmas have in common that they are not described in traditional lexicons. Depending on contributor the articles differ in quality regarding both form and content.

## 1. Inledning

Föreliggande artikel redovisar resultaten av en kvalitativ, diakron studie av den användargenererade ordboken *Folkmun.se*. Med ”användargenererad” menas att sajten består av ord och uttryck som skickats in av allmänheten. Vårt fokus ligger på enskilda bidrag och bidragsgivare och deras roll i sajten utveckling. Vilka bidrar till ordboken, hur tycks dessa personer arbeta och resonera? Går det att skönja mönster i bidragsgivandet? Artikeln inleds med en forskningsbakgrund och därefter presenteras själva ordbokssajten. Därpå redogör vi för förutsättningarna för studien, vårt material och våra resultat. Artikeln avslutas med en sammanfattande diskussion.

## 2. Kort om användargenererade sajter

Användargenererade sajter är per definition svåra att undersöka eftersom de snarare är pågående processer än statiska produkter. Vidare är de flyktiga; de har en tendens att byta webbadress eller helt enkelt försvinna efter några år (Törnqvist 2015:155). Trots denna (inneboende) problematik har det de senaste åren forskats allt mer på resurser av detta slag. Exempelvis har det gjorts kvantitativa analyser av *Wiktionary* för att studera hur sajten utvecklas. Viktiga frågeställningar är hur *många* personer som ligger bakom en viss sajt, *när* artiklarna revideras och *hur* de revideras (se t.ex. Wolfer & Müller-Spitzer 2016). Tidigare forskningsinsatser till trots vet vi fortfarande mycket lite om själva bidragsgivarna.

Det råder delade meningar om de användargenererade sajternas värde och kvalitet (se t.ex. Hanks 2012:77–82 och Meyer & Gurevych 2012). Exempelvis

vinner *Wiktionary* terräng på en rad områden (såväl lingvistiska som språkteknologiska), och användargenererade ordböcker har potential att bli allvarliga konkurrenter till både förlagsbaserade och akademiska ordböcker (Wolfer & Müller-Spitzer 2016:347). Innehållet i dem fyller också en kompletterande funktion i förhållande till traditionella ordböcker t.ex. när det gäller dokumentation av helt nya, och kanske lite mer tillfälliga, ord. Den kritik som framförts har framförallt riktats mot grundantagandet att samarbete mellan många bidragsgivare leder till förbättrat innehåll (se t.ex. Mattus 2009:183). En annan kritisk synpunkt mot användargenererade sajter är att bidragsgivarna, i stället för att samarbeta i syfte att uppnå mer fullständiga beskrivningar, ibland mest tycks vilja lufta sina fördomar (Lew 2011). På många användargenererade sajter som t.ex. Wikipedia förekommer också vandalism och s.k. *edit-wars* i och med att man kan ändra och t.o.m. radera andras artiklar (Mattus 2009:192). Att bidragsgivare kan förhålla sig till befintliga artiklar och komplettera dem är dock en viktig drivkraft. Aaltonen & Seiler (2016) visar exempelvis att bidragsgivare är mer motiverade att bygga vidare på artiklar än att skapa helt nya.

### 3. Kort om *Folkmun.se*

Sajten *Folkmun.se* drivs av privatpersonen David Eriksson, som startade den år 2007 i syfte att ”dokumentera dialektala uttryck (men också slang och nyord)” (se vidare Sköldberg & Wenner 2017 med anförda referenser). Som Sköldberg (2015:232–233) visat finns det redaktionella riktlinjer till dem som vill föra in ett nytt ord. Bidragsgivarna uppmanas bl.a. att förklara ordet och lägga in språkexempel.

Eriksson har gett oss tillgång till en databas där allt material som inkommit till *Folkmun.se* från starten till december 2016 finns samlat.<sup>1</sup> Totalt innehåller databasen 5 523 bidrag. Av dessa är 4 641 poster synliga på webbsidan och de har skickats in av 1 664 unika bidragsgivare. Eriksson är ensam moderator och bestämmer vilka bidrag som publiceras på webben. Han går igenom alla insända och rapporterade ord, men är restriktiv med att stryka eller censurera material. Det motiverar han med att om ett nedsättande ord existerar, är det bra att språkbrukarna kan googla sig fram till definitionen och sedan förhoppningsvis välja att inte använda det. Det händer att synnerligen nedsättande bidrag inte görs publika, men de vanligaste skälen till utebliven publicering är dels att bidragsgivaren inte

<sup>1</sup> Stort tack till David Eriksson som så frikostigt delat med sig av sitt material och svarat på våra frågor. Avseende studiens forskningsetiska aspekter så har bidragsgivarna inte informerats om sin medverkan i studien. Erhållen data är anonymiserad så att den inte går att koppla till fysiska personer. Däremot går det att utifrån anonymiserade IP-uppgifter till exempel utläsa om flera användaridentiteter använder samma dator-uppkoppling. Därigenom menar vi att vi har beaktat befintliga forskningsetiska riktlinjer om informanternas anonymitet och skydd av informanter.

gett sitt medgivande till att artikeln publiceras på sajten (något som görs via mejl), dels att hen av misstag skickat in ett och samma bidrag flera gånger.

Statistik via Google Analytics visar att antalet inskickade ord med något undantag har minskat för varje år sedan sajten startades, men ser man till antalet sajtbesökare har siffran stadigt ökat. Under oktober månad 2007 hade *Folkmun.se* 3 500 besökare. Det kan jämföras med oktober månad 2016 då 12 000 personer besökte sidan.

## 4. I tio bidragsgivares fotspår

I den här studien har vi valt att närma oss bidragsgivarna genom att undersöka de spår de lämnat efter sig i materialet. Man kan se tydliga likheter mellan den valda metoden och studier som har gjorts av ordboksanvändares sökmönster (jfr t.ex. Almind 2008 och Hult 2017).

I avgränsningssyfte har vi närstuderat artiklar som härrör från tio olika bidragsgivare (se vidare nedan). Just deras signaturer (dvs. bidragsgivarnamn) valdes ut eftersom deras bidrag skiljer sig åt till antal och omfång m.m. De aktuella bidragsgivarna får, inom ramen för vår studie, representera olika typer av användare som är med och utvecklar *Folkmun.se*. Mer specifikt har vi funderat kring deras val av signatur och undersökt hur många bidrag de skickat in, när de skickats in och hur många av dessa som är publika på webbsidan. Naturligtvis har vi också granskat vad det rör sig om för ord och uttryck. Finns det återkommande informationskategorier i artiklarna? Och, inte minst, vilken lexikografisk kvalitet har artiklarna? Notera att den statistik som vi presenterar nedan inbegriper allt material som lagts in av användarna, även sådant som av olika anledningar inte gjorts publikt via sajten.

### 4.1. Dimbo-Klas

Den första bidragsgivare som vi följt kallar sig Dimbo-Klas. Möjligen är detta en man som bor eller kommer från småorten Dimbo i Västra Götalands län. Om så är fallet kan hen vilja signalera någon form av geografisk hemvist genom sin signatur. Men detta kan vi, med tanke på aktuell metod, förstås inte veta. Dimbo-Klas har endast lagt in fyra artiklar, tre simplex och ett flerordsuttryck, och det har skett under loppet av ca fem minuter en kväll i september 2013. Samtliga artiklar rör etablerade dialektord från ett specifikt angivet område. Ett typiskt exempel är substantivet *glisker* (se figur 1).

The screenshot shows the Folkmun.se website interface. At the top left is the logo 'folkmun.se' with a sun icon. To the right is an alphabetical menu 'ABCDEFGHIJKLMNOPQRSTUVWXYZÅÖ' and a search bar with the text 'Sök på folkmun.se'. Below the search bar are links for 'Senaste', 'Lägg till ord', 'Kontakt', and 'Blogg'. The main heading is 'Glisker' in green. Below it are navigation links: '« glirv', 'Överraska mig!', and 'Gloglas »'. The article text reads: 'Glisker', '"Vår hund har blivit gammal, han är inte så glisker längre"', 'Västgötska ( Redväg ) för "på alerten", "på hugget"', and 'av Dimbo-Klas 2013-09-18 - Rapportera'. At the bottom is a link 'Lägg till din egen definition av Glisker'.

Figur 1: Artikeln *glisker* i Folkmun.se (något beskuren; hämtad 30.11.2017).

Ordboksartikeln innehåller ett lemma, ett språkprov, en dialektangivelse, en geografisk uppgift (Redväg) samt en betydelseangivelse i form av två synonyma uttryck.

#### 4.2. Go4it

Signatur nummer två har valt att kalla sig Go4it. Hen har skickat in totalt fem bidrag en fredagsförmiddag 2013 och samtliga artiklar är publika på sajten. Det rör sig om fyra enkla ord och ett uttryck (*fläng* 'dum', *tölit* 'tråkigt', *töla* 'prata; pratglad kvinna', *Urban* 'namn; person som är på dåligt humör' och *ge-rää* 'ge dig, sluta'). De kan alla sägas vara relativt etablerade, talspråkliga uttryckssätt och enligt Go4it hör de till Karlstadsdialekt. I artiklarna finns förutom lemma och dialektangivelse många andra informationskategorier representerade, t.ex. uppgifter om uttal, betydelsevariation, konnotationer och exempel. Go4it:s bidrag är på så vis relativt rika på information, men vissa av artiklarna rymmer samtidigt ett antal skrivfel och brott mot svenska skrivregler.

#### 4.3. Benson

I materialet återkommer signaturen Benson 61 gånger. Bland Bensons artiklar finns många dubletter; totalt rör det sig om 47 unika bidrag. Typiskt har hen lagt in 3–4 bidrag åt gången, i regel med cirka två minuters mellanrum. Artiklarna, som rör ord som *linka* 'halta el. knäa' och *pyngla* 'liten börda', ger ett synnerligen

enhetligt intryck. Först återfinns ett uppslagsord (simplex) och därefter en enklare betydelseangivelse, oftast i form av en eller två synonymer. Benson använder också konsekvent vissa etablerade förkortningar. De allra flesta orden karakteriseras dessutom som ”blekingska”. Sammantaget vittnar bidragen om såväl goda dialektkunskaper som ett systematiskt arbetssätt från just denne bidragsgivares sida.

#### 4.4. Luscinia

Signaturen Luscinia – ett möjligt kvinnligt förnamn, men också det vetenskapliga namnet på ett släkte trastfåglar – har skickat in ca 50 bidrag mellan 2010 och 2011. Så gott som samtliga av dessa är publika. Typiskt har denna bidragsgivare lagt in ett par ord, gärna på kvällen, för att sedan tystna igen. De inmatade orden tycks, om man studerar dem lite snabbt, ha få gemensamma nämnare. Exempelvis har Luscinia den 16/1 2010 lagt in ord som *snigeltåg* ’långsamt tåg’, *golbög* ’den som samarbetar med polisen’ och *zlatanist* ’anhängare av Zlatan’. Men om man granskar samtliga bidrag framträder ett mönster. Ett flertal av orden anges vara skånska, t.ex. *piggatrå(d)* ’taggtråd’ och *cocabonnar* ’glasögon med tjocka glas’. Andra uppslagsord är semantiskt besläktade med varandra, t.ex. *småglin*, *glin* och *glytt* (som alla betyder ’små barn’). Värt att notera är att hen också bidragit med flera slangord som kopplas till fågelskådning, och valet av alias är därmed kanske inte så märkligt.

#### 4.5. Hetrometrus

Hetrometrus (en signatur som sammanfaller med en skorpionbeteckning) är med i den absoluta starten av sajten under 2007, och är ytterst aktiv under tre månaders tid. Hen lägger in 208 bidrag varav så gott som alla är publika. Artiklarna har överlag enkel struktur, t.ex.: ”Hossor Strumpor Sydsåknskt”.

Inledningsvis dokumenterar Hetrometrus mer allmänt kända, vardagliga ord. Därefter går hen över till att i följd lägga in dryga trettioalet uppslagsord som karakteriseras som ”göingska”. (Göinge ligger i nordöstra Skåne.) Därpå matar hen in mer allmänt skånska ord – över 70 uppslagsord efter varandra har etiketten ”skånska”. Avslutningsvis återgår Hetrometrus till mer allmänna svenska slangord och slanguttryck. En närmare granskning av de renodlat skånska orden ger vid handen att de faktiskt är alfabetiskt ordnade. Vidare ger artiklarna ett mycket enhetligt intryck och de har dessutom matats in i mycket snabb takt. Mycket tyder alltså på att Hetrometrus utgått från någon förlaga, men vi har inte kunnat utröna vilken denna i sådana fall skulle kunna vara.

#### 4.6. Mads & Preben

Vidare har vi granskat bidrag skrivna av Mads & Preben. Av vårt material framgår att dessa två män (?) i princip endast varit aktiva vid ett par mycket sena lördagskvällar i september 2007. Sammanlagt ligger de bakom 31 bidrag, men inget av dessa syns på sajten. Rent innehållsmässigt uppvisar deras ordboksartiklar stora likheter; nästan uteslutande rör det sig om grova beteckningar på könsorgan med obscena betydelsebeskrivningar (under lemman som *drängakuk* och *klitta*) eller rasistiska inlägg (under lemman som *habibber*, *blåneger* och *zigenar arsle*). Med tanke på den sena timmen kan det kanske vara två vänner som, i deras tycke, roat sig några ”blöta” kvällar. Man kan dock betrakta deras insatser som en form av vandalisering på webbsidan (se avsnitt 2), i synnerhet som de i ett fall använt sig av Hetrometrus namn och lagt in en allmänt kränkande artikel under denna signatur.

#### 4.7. Richard (J)

Signaturen Richard, alternativt Richard J, har under ca ett års tid skickat in 41 bidrag till *Folkmun.se*. Typiskt har hen matat in 3–4 bidrag i taget och detta på kvällstid. Bland uppslagsorden återfinns *pröjsa* ’betala’, *bull-TV* ’mikrovågsugn’, *dumknollad* ’person som är dum/bortkommen’, *syntetneger* ’person som ligger i solarium’, *lakritstomte* ’afrikan’, *pervo* ’pervers person’, *gambiaskalle* ’negerboll’, *brödkvarn* ’mun’, *PLO-burgare* ’kebab’, *skinkprins* ’homosexuell man’ och *dadeltrampare* ’person från mellanöstern’. Som synes rör det sig mest om (etablerade) slangord och enkla förklaringar, inte sällan med rasistiska och/eller homofoba inslag. Till skillnad från Mads & Prebens bidrag är dessa artiklar synliga på sajten. Richard (J) har bekräftat att hen vill att de ska synas och sajtägaren har också godkänt dem.

#### 4.8. Chooniii

Bidragsgivare nummer åtta har valt att kalla sig Chooniii. Signaturen kan möjligen kopplas till det multietniska ordet *shono/shonne* med betydelsen ’snubbe’ och med tanke på bidragens karaktär är detta inte helt orimligt. Signaturen kan kopplas till åtta bidrag som alla är publika. Uppslagsorden lades in under loppet av en kvart, sent en fredagskväll i juni 2009. Bland dessa återfinns *göt*, *katt* och *yani* med betydelserna ’rumpa’, ’snygg’ och ’typ’ samt verbet *klippa*, vilket är försett med språkexempel som ska illustrera de tre olika betydelser som anförs, ’stjäla’, ’slå’ och ’åka fast’. Vad gäller innehållet i artiklarna kan vi se att det efter hand blir fylligare – bidragsgivaren tycks bli lite varmare i kläderna.

#### 4.9. Tjo fräzzen

Signaturen Tjo fräzzen har lagt in tio bidrag en eftermiddag i juli år 2009 och åtta av dessa visas på sajten. Hen kan sägas stå bakom en ny typ av gruppspråk i ordboken i förhållande till de signaturer som vi behandlat hittills. Bland bidragen finns nämligen flera engelskklingande uttryck och förkortningar varav vissa anges komma från dataspelsvärlden, särskilt *World of Warcraft*. För att visa hur hen kan ha resonerat i samband med bidragsgivandet ges följande exempel: klockan 14.14.04 lägger hen in uppslagsordet *inv ffs* med betydelsen 'när du vill att nån ska invita dig jätte mkt till ett party'. 39 sekunder senare läggs artikeln *ffs* in som anges betyda 'for fuck sake'. Tjo fräzzen verkar alltså ha kommit på att också *ffs* kan tjäna som uppslagsord i *Folkmun.se* och att det kan behöva förklaras.

#### 4.10. Tyke

Slutligen har vi valt att titta närmare på signaturen Tyke som står bakom 49 bidrag (varav 46 är publika). Under loppet av ca 3,5 års tid har hen återkommit med några månaders mellanrum och då matat in ca 10 bidrag åt gången. Bland bidragen finns flera slangartade ord men också ett antal flerordsuttryck. Betydelsebeskrivningarna är ofta långa och inte sällan anför språkprov. Det stora antalet bidrag över en relativt lång tidsperiod gör att det går att urskilja ett mönster. Användaren har främst lagt in semantiskt besläktade ord och uttryck som kan föras till tre övergripande teman. Dessa är: 1) sex och könsorgan (t.ex. *aktertankare*, *polera hjälmen*, *åderpåk*, 2) ointelligens (t.ex. *felknullad*, *bombad*, *larmet går men brandkåren sover* och 3) meningslös, icke framgångsrik, aktivitet (*slakrunka*, *torrjucka* och *fastna i vinkelvolten*). Det är förstås svårt att säga varför just dessa teman återkommer, men kanske har hen associerat sig fram till mer eller mindre synonyma ord och uttryck till dem som hen redan bidragit med.

### 5. Sammanfattande diskussion

I den här artikeln har vi redogjort för en studie som utgår från drygt 5 500 bidrag som matats in *Folkmun.se*'s webbformulär 2007–2016. Materialet är omfattande och här har vi bara haft möjlighet att göra några nedslag. I avgränsningssyfte närstuderas tio signaturer som representerar olika typer av bidragsgivare. Personerna ifråga har lagt in allt från 4 till 208 artiklar var. Sammanlagt omfattar vår studie närmare 470 inlägg.

Vår undersökning visar att många bidragsgivare tycks vilja uttrycka något genom sitt val av signatur. Vidare kan vi skönja mönster i deras arbetssätt (hur många ord de matar in åt gången, vilket tempo de har) och vad de bidrar med (simplex/flerordsuttryck, semantiskt innehåll, varietet och stilnivå).

Artiklarna i *Folkmun.se* är, både till form och innehåll, av skiftande kvalitet. Bidragsgivarna har i varierande grad följt riktlinjerna. Tydligt är också att både rasistiska och homofoba bidrag är publika på webbsidan. Det bidrar till att sajten innehåller ett flertal beskrivningar av ord och uttryck som inte är beskrivna i mer konventionella ordböcker. Sajten kan därför potentiellt ha en kodifierande verkan jämförbar med mer etablerade ordböcker.

I tidigare forskning framhålls ofta motsättningen mellan utbildade lexikografer och amatörer respektive traditionellt utarbetade ordböcker och användargenrerade. Också i vår titel antyds att insatserna på *Folkmun.se* görs av amatörlexikografer, men utifrån vår studie kan man fråga sig om det verkligen bara är amatörer som bidrar. Många bidrag är systematiskt uppbyggda och visar på kännedom om grammatik, lexikografi, dialekter osv.

Avslutningsvis kan man fundera kring möjliga motiv till att skicka in ordboksartiklar. Några förefaller vilja testa sajten (se t.ex. signaturerna Dimbo-Klas, Go4it och Tjo fräzzen). Andra verkar primärt vilja sprida kunskaper om eller bevara en dialekt (Benson och Hetrometrus) eller sociolekt (Choooniii). Åter andra tycks vilja lyfta fram mer eller mindre spridda slangord (Luscinia, Tyke) eller tänja på gränserna genom att lägga in slangord som uppenbarligen kan väcka anstöt (Richard (J)). Slutligen tycks några mest vilja orsaka förtretligheter för andra eller förstöra (Mads & Preben). Mer omfattande vandalism förekommer dock inte vilket kan bero på att man på *Folkmun.se* inte redigerar varandras artiklar (jfr t.ex. Wikipedia). Vill man förtydliga innehållet i en annans artikel kan man endast lägga in en ny post.

För att se hur typiska de tio bidragsgivare som studerats här är, skulle vi i framtiden vilja göra fler närstudier. Vi skulle också vilja testa andra undersökningsmetoder, exempelvis lägga upp en enkät på ordbokssajten. Därigenom kan vi få kontakt med personer som bidrar med ord och få mer säkra svar på frågor som inte går att besvara enbart utifrån nuvarande material. Förutom att vi får veta mer om personen bakom en signatur (ålder, geografisk bakgrund m.m.) kan vi få mer kunskap om vad som driver en person att publicera ord på *Folkmun.se* och vilket syfte hen har med sina bidrag.

## Litteratur

### Ordböcker

*Folkmun.se*. <<http://www.folkmun.se>> [september 2017].

*Wiktionary* (svensk version). <<http://sv.wiktionary.org/wiki/Wiktionary>> [september 2017].



## Annan litteratur

- Aaltonen, Alexi & Stephan Seiler (2016): Cumulative growth in user-generated content production: Evidence from Wikipedia. I: *Management Science* 62(7), 2054–2069.
- Almind, Richard (2008): Søgemønstre i logfiler. I: *LexicoNordica* 15, 33–55.
- Hanks, Patrick (2012): Corpus evidence and electronic lexicography. I: Granger, Sylviane & Magali Paquot (eds.): *Electronic Lexicography*. Oxford: Oxford University Press, 57–82.
- Hult, Ann-Kristin (2017): *Ordboksanvändning på nätet. En undersökning av användningen av Lexins svenska lexikon*. (Göteborgsstudier i nordisk språkvetenskap 27.) Göteborg.
- Lew, Robert (2011): Online Dictionaries of English. I: Fuertes-Olivera, Pedro A. & Henning Bergenholtz (eds.): *E-lexicography. The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum, 230–250.
- Mattus, Maria (2009): Wikipedia – Free and Reliable? Aspects of a Collaboratively Shaped Encyclopaedia. I: *Nordicom Review* 30, 183–199.
- Meyer, Christian, M. & Iryna Gurevych (2012): Wiktionary: a new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. I: Granger, Sylviane & Magali Paquot (eds.): *Electronic Lexicography*, Oxford: Oxford University Press, 259–291.
- Sköldberg, Emma (2015): Bland runda ord och dialektala uttryck. Om den användargenererade sajten Folkmun.se. I: Sandström, Caroline, Cantell, Ilse, Grönros, Eija-Riitta, Nuolijärvi, Pirkko & Eivor Sommar Dahl (red.): *Perspektiv på lexikografi, grammatik och språkpolitik*. (Instituten för de inhemska språken 39.) Helsingfors, 229–241.
- Sköldberg, Emma & Lena Wenner (2017): Folkmun och Luxikon. En jämförelse mellan två användargenererade ordbokssajter. I: *Språk och stil. Tidskrift för svensk språkforskning* 27, 21–48.
- Törnqvist, Lars (2015): Nordiska dialekt- och slangordböcker på Internet. I: *LexicoNordica* 22, 151–169.
- Wolfer, Sascha & Carolin Müller-Spitzer (2016): How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions. I: *Lexikos* 26, 347–37.

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

Lena Wenner  
forskningsarkivarie, fil. dr  
Instituet för språk och folkminnen  
Vallgatan 22  
411 16 Göteborg  
lena.wenner@spraakochfolkminnen.se

# An Icelandic Gigaword Corpus

*Steinþór Steingrímsson, Sigrún Helgadóttir & Eiríkur Rögnvaldsson*

The paper describes work in progress to compile an Icelandic Gigaword Corpus (IGC). The initial aim of the project was to compile a large corpus of contemporary texts with at least a billion running words, with the minimum amount of work and resources. Thus we focussed on material not protected by copyright and sources which could provide us with large chunks of text for each cleared permission. The two main sources considered were therefore official texts and texts from news media. Only digitally available texts are included in the corpus, and formats that can pose problems are not processed. The corpus texts are morpho-syntactically tagged and provided with metadata. Processes have been set up for continuous text collection, cleaning and annotation. The corpus will be made available for search and download with permissive licenses. The first version of the corpus will be released by the end of 2017. Texts will be added continually and a new version published every year.

## 1. Introduction

The lack of a very large Icelandic text corpus has been evident for some time. The compilation of such a corpus has therefore been considered a top priority in order to further Language Technology (LT) in Iceland (Anna Björk Nikulásdóttir et al. 2017). Large text corpora are e.g. necessary for the design of language models that are used in building a variety of LT tools such as speech recognizers, spell and grammar checkers and automatic machine translation. With the increased importance of machine learning methods such as neural networks in LT, the importance of large text corpora and other textual resources has increased considerably.

The aim of the corpus project is to compile as large a corpus as possible with the minimum amount of work and resources. We want the corpus to be attractive for use in LT projects as well as for other research and study. In planning the project it was decided to aim for the following goals:

- The IGC will contain more than a billion running words, morpho-syntactically tagged and lemmatized and provided with metadata.
- Only digitally available texts will be included in the IGC. Formats that may pose a difficulty will not be processed.
- The IGC will be open and constantly expanding.
- A closed version will be published every year.
- The IGC will be accessible through an online concordance search tool.

- Trend data from the IGC will be searchable in an n-gram viewer.
- The IGC will be made available for download with a permissive license.

In Section 2 the compilation of the MIM corpus (Sigrún Helgadóttir et al. 2012) is described where the intention was to create a “balanced” and a “representative” text collection. In order to achieve representativity and balance text was sampled from many genres and often a very small chunk of text was acquired for each license. However, there are several problems connected with trying to achieve representativeness in a corpus. For the first, what should it be representative of? And because it can be hard to determine where a variety of language ends and another begins, any corpus is ‘virtually by definition biased to a greater or a lesser extent’ (Nelson 2010).

One of the design goals for the IGC is for it to be open, that it will be constantly expanding, but closed versions will be published every year to make it possible for researchers to verify others’ results. Furthermore, in order to accomplish our goal of more than a billion words we need to build a collection of texts from sources who have available material that is not protected by copyright or where it is possible to get big chunks of text for each license secured. The two main sources considered are therefore official texts and texts from news media. Only digitally available texts will be included in the corpus and formats that are difficult to process, like pdf documents, will not be used. This design makes it even harder to consider representativeness. The corpus will therefore be biased towards journalistic and official texts, but more detailed description of the corpus texts is given in section 3.2.

The corpus texts are morphosyntactically tagged and provided with metadata. Processing pipelines are set up for continuous text collection, text cleaning and annotation where the processing tools will be continually updated.

This paper is structured as follows. In Section 2 we describe briefly existing Icelandic corpora. In Section 3 an account is given of the creation of the IGC. Availability of the corpus is discussed in Section 4 and in Section 5 we sum up and conclude the paper.

## 2. Icelandic Corpora

In this section existing Icelandic corpora are listed and described briefly, to explain their shortcomings and hence the need for a new corpus.

A small corpus was compiled at the Institute of Lexicography<sup>1</sup> for the making of the Icelandic Frequency Dictionary (IFD), Íslensk orðtíðnibók, published in 1991 (Jörgen Pind et al. 1991). The IFD corpus<sup>2</sup> consists of just over half a mil-

<sup>1</sup> Now a part of the Árni Magnússon Institute for Icelandic Studies.

<sup>2</sup> Available at <<http://www.malfong.is>>.

lion running words. The corpus has a heavy literary bias as about 80% of the texts stem from fiction. The corpus is annotated with morphosyntactic tags and lemmata. Tagging and lemmatization was manually corrected and hence the corpus has been used as a gold standard for training part-of-speech (PoS) taggers, lemmatizers and parsers. It can be stated that the IFD corpus has laid the ground for most work on PoS tagging, lemmatization and parsing that has been performed on Icelandic during the last 15 years.

The Tagged Icelandic Corpus (MIM) was released in the spring of 2013, both for search<sup>3</sup> and download.<sup>4</sup> This corpus contains 25 million running words from various genres dating from the first decade of the 21<sup>st</sup> century (Sigrún Helgadóttir et al. 2012). The corpus was intended for use in LT projects and for linguistic research. About 86% of the texts are protected by copyright, the remainder being official text (parliamentary speeches, legal text, adjudications and text from government websites). The largest proportion of the text, just less than 24%, comes from published books containing both fiction and non-fiction. The second largest portion, about 22%, derives from newspapers, mostly printed newspapers. The corpus is annotated with morphosyntactic tags and lemmata.

To enable the use of the corpus in LT projects it was considered important to secure copyright clearance for the texts to be used. All owners of copyrighted text signed a special declaration and agreed that their material may be used free of licensing charges.

MIM-GOLD is a corpus of about 1 million running words which was sampled from the MIM corpus (Hrafn Loftsson et al. 2010; Sigrún Helgadóttir et al. 2012; Steinþór Steingrímsson et al. 2015). The corpus is intended as a reliable standard for the development of LT tools. Tagging of this subcorpus has been manually corrected. MIM-GOLD will augment the IFD corpus for training statistical taggers and developing LT tools. The MIM-GOLD corpus is nearly twice the size of the IFD corpus and the texts are more varied, less than 25% of the texts in MIM-GOLD are literary texts compared to about 80% of the texts in the IFD corpus.

Training and testing using the Average Perceptron Tagger *Stagger* (Östling 2012) on MIM-GOLD after two correction phases has already been described (Steinþór Steingrímsson et al. 2015). The result showed that there were still errors in the tagging that needed to be corrected. Work on locating and correcting these errors was completed in the fall of 2017.

*The Icelandic Parsed Historical Corpus (IcePaHC)*<sup>5</sup> is a diachronic treebank that contains about one million running words from every century between the 12th and the 21st centuries, inclusive (Eiríkur Rögnvaldsson et al. 2011). The texts are annotated for phrase structure, PoS-tagged and lemmatized. The corpus is designed to serve both as an LT tool and a syntactic research tool. The corpus is completely free and open since most of the texts are no longer in copyright.

<sup>3</sup> *Mörkuð íslensk málheild*: <<http://mim.arnastofnun.is>>.

<sup>4</sup> At <<http://www.malfong.is>>.

<sup>5</sup> <[http://www.linguist.is/icelandic\\_treebank/](http://www.linguist.is/icelandic_treebank/)>.

Íslenskur orðasjóður<sup>6</sup> is an Icelandic corpus of more than 550 million running words collected from all domains ending in *.is* in 2005 and 2010 (approx. 33 million sentences). Moreover, additional newspaper texts (2 million sentences) and the Icelandic Wikipedia are included. The web texts were cleaned substantially before their inclusion in the corpus.

Although the corpora mentioned in this section have been useful in LT and language research they do not fulfill the requirements that present day LT makes to language resources as regards size and quality. Therefore it was considered necessary to embark on the project of compiling the IGC.

### 3. Creating the corpus

In Section 1 the aims of the corpus project were described, the primary aim being to compile as large a corpus as possible, at least a billion words, with the minimum amount of work and resources. In this section we will give an account of permissions clearance, collecting the texts and the cleaning and annotation process.

#### 3.1 Permission clearance and licensing

One of the design considerations for the IGC was to make the corpus available with a permissive license, such as a Creative Commons license.<sup>7</sup> Work on permission clearance for the first version of the corpus concluded in early 2017. We cleared permission from 19 content providers but found that Creative Commons licensing is not widely known in Iceland so eventually it was necessary to use the license used for the compilation of the MIM corpus for a substantial part of the texts. Although some of the copyright protected texts in the IGC will be made available with a CC license a great part will be tied to the special license developed for the MIM corpus. Together with text not protected by copyright we have access to more than 40 different text sources. The texts include general and local news from print and the web, transcribed television and radio news, commentary on politics and current affairs and texts on scientific matters. Furthermore, we collect parliamentary speeches, adjudications from courts and a selection of recent fiction and non-fiction from The Árni Magnússon Institute's text collection.

#### 3.2 Collecting texts

A pragmatic approach to text collection was adopted. Texts requiring a minimum of cleaning and processing and texts accompanied by relevant metadata are preferred. This applies to texts obtained from databases of text owners and text har-

<sup>6</sup> <[http://wortschatz.uni-leipzig.de/ws\\_ice/](http://wortschatz.uni-leipzig.de/ws_ice/)>.

<sup>7</sup> Cf. <<https://creativecommons.org/>>.

vested from the web. Texts in MS Word document format, in Excel spreadsheets or in XML format have also been accepted.

Texts not protected by copyright will be collected from official sources, the biggest of which is the Icelandic parliament, providing parliamentary speeches dating back to 1940 in XML format, containing all relevant metadata. The speeches are transcribed at Alþingi and have been extensively proofread. We also harvest legal text and adjudications from official websites.

Text has been acquired from all the largest newspaper publishers in Iceland, and a number of smaller ones have given permission for use of their text both from online and printed sources. The corpus collection includes the Icelandic Wikipedia, the University of Iceland's Science Web, The Árni Magnússon Institute's text collection (fiction and non-fiction, from recent decades), translations of EEA documents and other smaller sources.

Text genre	Sources	Word count
Newspaper articles	<i>Morgunblaðið</i> , <i>Visir</i> , <i>DV</i> and various other smaller news sources	745,708,958
Parliamentary speeches	Alþingi	210,580,253
Adjudications	Supreme court and district courts of Iceland	88,351,996
Transcribed radio/television news	RÚV and 365	54,129,051
Sports news	Fótbolti.net and 433.is	45,992,991
Current affair blogs	Jónas.is, Andríki.is and other smaller sources.	13,030,217
Informational articles	Wikipedia and Science Web	10,738,060
Gossip/entertainment	Bleikt.is	5,316,675
Total		1,173,848,201

Table 1: Retrieved texts as of August 2017.

Table 1 lists text genres and word count for texts that have been retrieved in August 2017. At that point the majority of texts in the first version of the IGC have been processed. Unprocessed sources are listed in table 2.

Text source	Estimated word count
EEA translations	20,000,000
Newspaper articles (6 smaller news sources)	30,000,000
Legal text	5,000,000
The Árni Magnússon Institute's text collection	70,000,000
<b>Total</b>	<b>125,000,000</b>

Table 2: Texts to be included in the first version, not retrieved in August 2017.

### 3.3 Text cleaning and annotation

Texts in the corpus can be divided into written texts and transcribed spoken text. Transcribed spoken text includes parliamentary speeches and news from the main radio and television stations in Iceland.

Procedures have been devised for automatic editing and cleaning of the text, annotation and metadata extraction. There is no manual post-editing.

The annotation phase consists of sentence segmentation, tokenization, morphosyntactic tagging and lemmatization. After morphosyntactic tagging and lemmatization, the texts, together with the relevant metadata, are transferred into TEI-conformant XML format (TEI Consortium 2017). N-grams (n up to 5) are also created for use with the n-gram viewer and for distribution.

Sentence segmentation and tokenization is performed with the same procedures as were used for the MIM corpus (Sigrún Helgadóttir et al. 2012). *IceStagger* (Hrafn Loftsson & Östling 2013) is used for tagging the IGC, initially trained on the IFD corpus but will be retrained and rerun when MIM-GOLD is available.

A new tool is currently being developed for lemmatizing Icelandic text. This tool will be used for lemmatizing the IGC and first results indicate a great improvement over the tool used to lemmatize the MIM corpus. A thorough analysis and comparison of the two systems remains to be done.

A pipeline for harvesting, cleaning and annotating the corpus texts has been developed. Individual tools in the pipeline will be continually updated to produce a more precise and reliable annotation with each new version of the corpus.

## 4. Availability and use

The main object of the corpus is for use in LT projects. For other uses, such as linguistics research, teaching, lexicography or other studies the data will be searchable in a web-based concordance tool. The Swedish platform KORP<sup>8</sup> (Borin et al. 2012) which in turn uses the IMS Corpus Workbench<sup>9</sup> (Evert & Hardie 2011) as a search engine is being adapted to be used for the corpus. Users of the search interface can take advantage of the annotation of the texts when specifying search criteria. Texts will be added continually to the searchable corpus.

The corpus texts will be made available for download in the TEI-conformant XML format (TEI Consortium 2017). As mentioned in Section 1 some of the corpus texts are not protected by copyright, some can be distributed with relatively open CC licenses and some texts will be made downloadable with the special license developed for the MIM corpus. This situation will be reflected in the download procedures. The corpus can be downloaded through the Icelandic LT resources website *Málföng*.<sup>10</sup>

<sup>8</sup> <<https://spraakbanken.gu.se/korp/>>.

<sup>9</sup> <<http://www.ims.uni-stuttgart.de/forschung/projekte/CorpusWorkbench.html>>.

<sup>10</sup> <<http://www.malfong.is/>>.

The corpus texts will also be searchable through an n-gram viewer based on NB N-gram viewer (Birkenes et al. 2015).

To aid developers of LT tools the corpus website will allow download of the n-grams (n up to 5) used for the n-gram viewer.

## 5. Conclusion and further work

The new Icelandic Gigaword Corpus will be a valuable resource for builders of LT tools for Icelandic. It will also be useful for researchers, lexicographers, teachers, journalists and others working with or researching the Icelandic language.

The compilation of the corpus will be an ongoing process although closed versions, which will not be changed, will be published yearly. Official texts will be added continually as well as texts protected by copyright, as long as permission for their use has been secured. The tools in the corpus pipeline will also be upgraded following the development of better tools or versions and the corpus texts reannotated to reflect improved precision and reliability of the tools.

## References

- Anna Björk Nikulásdóttir, Jón Guðnason & Steinþór Steingrímsson (2017): *Mál-tækni fyrir íslensku 2018–2022: verkáætlun*. Reykjavík: Mennta- og menning-armálaráðuneytið.
- Birkenes, Magnús B., Lars G. Johnsen, Arne M. Lindstad & Johanne Ostad (2015): From digital library to n-grams: NB N-gram. In: *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA-2015)*, NEALT Proceedings Series Vol. 23. Vilnius, Lithuania, 293–295.
- Borin, Lars, Markus Forsberg & Johan Roxendal (2012): Korp – the corpus infrastructure of Språkbanken. In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, 474–478. <[http://www.lrec-conf.org/proceedings/lrec2012/pdf /248\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf /248_Paper.pdf)> (Retrieved September 10, 2017).
- Eiríkur Rögnvaldsson, Anton K. Ingason, Einar F. Sigurðsson & Joel Wallenberg (2011): Creating a Dual-Purpose Treebank. In: *Journal for Language Technology and Computational Linguistics*, 26(2):141–152.
- Evert, Stefan & Andrew Hardie (2011): Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In: *Proceedings of the Corpus Linguistics 2011 conference*. Birmingham, UK: University of Birmingham. <<https://www.birmingham.ac.uk /documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf>> (Retrieved September 10, 2017).



- Hrafn Loftsson & Robert Östling (2013): Tagging a Morphologically Complex Language Using an Averaged Perceptron Tagger: The Case of Icelandic. In: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA-2013)*. NEALT Proceedings Series 16. Oslo. <<http://www.ep.liu.se/ecp/085/013/ecp1385013.pdf>> (Retrieved September 10, 2017).
- Hrafn Loftsson, Jökull H. Yngvason, Sigrún Helgadóttir & Eiríkur Rögnvaldsson (2010): Developing a PoS-tagged corpus using existing tools. In: *Proceedings of "Creation and use of basic lexical resources for less-resourced languages", workshop at the 7th International Conference on Language Resources and Evaluation (LREC 2010)*. Valetta. <<https://www.ru.is/~hrafn/papers/corpusTagging.final.pdf>> (Retrieved September 10, 2017).
- Jörgen Pind, Friðrik Magnússon & Stefán Briem (1991): *Íslensk orðiðinibók* [The Icelandic Frequency Dictionary]. Reykjavík: Orðabók Háskólans.
- Nelson, M. (2010): Building a written corpus. In: A. O’Keeffe & M. McCarthy (Eds.): *The Routledge Handbook of Corpus Linguistics*. New York: Routledge, 53–65.
- Sigrún Helgadóttir, Ásta Svavarsdóttir, Eiríkur Rögnvaldsson, Kristín Bjarnadóttir & Hrafn Loftsson (2012): The Tagged Icelandic Corpus (MIM). In: *Proceedings of the workshop "Language Technology for Normalization of Less-Resourced Languages" – SaLTMiL 8 – AfLaT2012 at the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul, 67–72. <<http://aflat.org/files/saltmil8-aflat2012.pdf>> (Retrieved September 10, 2017).
- Steinþór Steingrímsson, Sigrún Helgadóttir & Eiríkur Rögnvaldsson (2015): Analysing Inconsistencies and Errors in PoS Tagging in two Icelandic Gold Standards. In: *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA-2015)*. NEALT Proceedings Series Vol. 23. Vilnius, Lithuania, 287–291. <<https://aclanthology.info/papers/W15-1838/w15-1838>> (Retrieved September 10, 2017).
- TEI Consortium, eds. (2017): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 3.2.0. Last updated on 10th July 2017. TEI Consortium. <<http://www.tei-c.org/Guidelines/P5/>> (Retrieved September 10, 2017).
- Östling, Robert (2013): Staggar: An Open-Source Part of Speech Tagger for Swedish. In: *Northern European Journal of Language Technology*, 2013, Vol. 3, 1–18. Linköping: Linköping University Electronic Press. <<http://www.nejlt.ep.liu.se/2013/v3/a01/nejlt13v3a1.pdf>> (Retrieved September 10, 2017).

Steinþór Steingrímsson  
Language Technologist  
steinthor.steingrimsson@arnastofnun.is

Sigrún Helgadóttir  
Language Technologist  
sigrun.helgadottir@arnastofnun.is

The Árni Magnússon Institute for Icelandic Studies  
Laugavegi 13  
101 Reykjavík, Iceland

Eiríkur Rögnvaldsson  
Professor  
The University of Iceland  
Faculty of Icelandic and Comparative  
Cultural Studies  
Árnagarði við Suðurgötu  
101 Reykjavík, Iceland  
eirikur@hi.is

# ***Málið.is*: A Web Portal for Information on the Icelandic Language**

*Steinþór Steingrímsson, Ari Páll Kristinsson & Halldóra Jónsdóttir*

*Málið.is* is a new web portal on the Icelandic language and language use. Initially it includes six different resources, providing reliable information on Icelandic. Work is under way to add more resources to the portal, providing more diverse information. *Málið.is* is expected to serve the Icelandic speaking population in general, although students and writing professionals could be expected to be the principal target groups. The aim of the project is to strengthen the Icelandic language in the digital era by making it easy to access information, helping people becoming more proficient and confident language users. Usage statistics shows growing user base, suggesting the project is a success.

## **1. Introduction**

The web portal *málið.is* was opened in November 2016. It is created and run by The Árni Magnússon Institute for Icelandic Studies (AMI) and gives access to a variety of databases on the Icelandic language created and/or maintained by the Institute. The portal provides abundant reliable and authoritative information, guidance and advice on Icelandic, all in one place. This concerns e.g. orthography, grammatical issues such as inflections, the lexicon, the history and etymologies of particular lexical entities, semantics, synonymity, terminologies and translations of technical vocabulary, and many other matters of language and its usage. Previously, these resources were accessible via a variety of different formats, user interfaces, web addresses, search methods and functions, which caused problems for many users as they were typically not aware of all options available.

The name (and web address) of the web portal, *málið.is*, translates to ‘the language.is’, which is descriptive and easy to remember. Access to *málið.is* is free of charge, and the portal is completely clutter free, showing no ads, notifications or other disturbances, only the results from the resources connected to the portal.

The portal is in many ways similar to the Danish web portal *sproget.dk*. And indeed, the Danish portal, initiated and operated by the Danish Language Council and Society for Danish Language and Literature, served as a model and an inspiration as we were planning this Icelandic web portal.

Students and professional language users, including writers, journalists, translators and others who work professionally with language could be considered the principal target groups for *málið.is*, as the portal gives access to abundant terminological data and directions for how to use the language “correctly”. But we also

expect it to serve the Icelandic speaking public in general and thus the web portal strives for plain and non-technical exposition and conciseness, whenever possible.

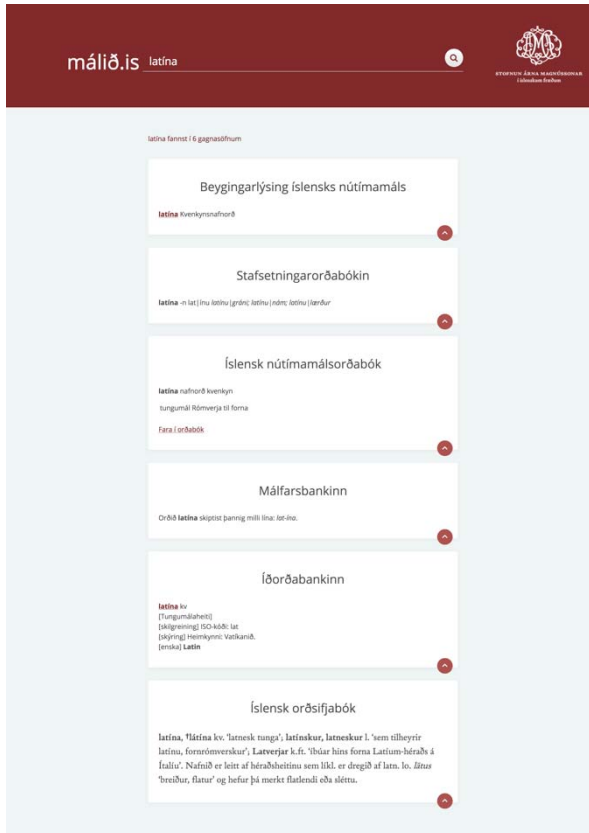


Figure 1: Example of search results on *málið.is*. The Icelandic word for 'latin', *latína*, is found in all of the six available resources.

Some of the data available on *málið.is* are explicitly of a prescriptive nature, while other data have primarily descriptive functions. We do not expect users to be immediately familiar with this fundamental distinction, but by exposing users to multiple types of information we facilitate the comparison between them and enable users to better acknowledge the distinction. Indeed, one theoretical contribution of *málið.is* is that it highlights the difference between descriptive and prescriptive language resources, for the benefit of students and researchers.

## 2. Rationale for setting up the portal

The Árni Magnússon Institute for Icelandic Studies has the role to publish prescriptive dictionaries as well as descriptive ones, both contemporary and historical.

It publishes terminologies and word-lists on language for technical and specialized purposes, it offers language consultation and advice to the general public and moreover the institute compiles text corpora and other language resources. Thus, the institute has over the years been compiling various linguistic resources and diverse language materials. For the most part, these resources have been digitized by now. Six of these resources form the contents of *málið.is* in its initial phase.

Furthermore, the institute has the duty to provide information on the Icelandic language and assistance on language use. This takes a substantial time of some of the institute's employees. With the advent of *málið.is* we hope to get fewer trivial questions, as our clients can easily get their answers on our web portal. But although this could be reason enough to set up the portal, rather than to try easing our workload, the principal aim of *málið.is* is to make it easier for the writing and speaking public to access the information they need to be more proficient and confident language users.

In Iceland, there is widespread interest among the general public in matters of the Icelandic language, in particular to etymological speculations, the coining of purist neologisms and matters of standard vs. non-standard grammar (Hilmarsson-Dunn and Kristinsson, 2013). A great portion of the queries the institute receives show this, so we tried to fulfil these needs as well as possible from the start.

There is an urgent need to make as much language advice and information available online as possible. In recent years there have been growing demands in Iceland for dictionaries and other language resources to be made available free of charge to the general public. This project aims to satisfy these demands.

### 3. The resources

Initially six resources were made accessible through the portal. They were chosen based on two main criteria: (1) How appropriate they were to the portal — we wanted to start with information most relevant to everyday use of Icelandic. (2) The state of the resources' database — whether we could connect it to the portal with relatively small effort. Based on this, the following six resources were adapted in time for the portal's launch:

- 1) The Database of Modern Icelandic Inflection (Kristín Bjarnadóttir 2012). The DMII contains 278,000 paradigms from Modern Icelandic, with over six million inflectional forms. It was created as a multipurpose resource, for use in language technology, lexicography, and as an online resource for the general public.
- 2) Spelling Dictionary (Jóhannes B. Sigtryggsson (Ed.) 2016). The 2<sup>nd</sup> edition of the Icelandic spelling dictionary, revised in accordance to updates to official Icelandic orthography, published by the Ministry for Education and Culture in 2016 (Íslensk málnefnd 2016).

- 3) Dictionary of Modern Icelandic (*Íslensk nútímamálsorðabók* 2017). Dictionary of Modern Icelandic is a new dictionary only available online. It is compiled at the department of lexicography at The Árni Magnússon Institute for Icelandic studies. It contains approximately 50,000 words. The work on the dictionary commenced in 2013 and is ongoing.
- 4) Language Usage Database (*Málfarsbankinn* 2017). A compilation of short articles giving advice on language use. It contains more than 7000 articles on subjects from grammar and syntax to well-crafted and elaborate language.
- 5) The Icelandic Term Bank (<http://www.ordabanki.hi.is>). The Icelandic Term Bank (ITB) contains around 60 bilingual or multilingual glossaries, and one monolingual glossary, containing terminologies in various fields. The glossaries contain terms in Icelandic, usually with corresponding terms in other languages, most commonly English. Often the terms are accompanied by definitions or explanations. In total the ITB contains more than 180,000 terms and it is constantly growing.
- 6) Icelandic Etymological Dictionary (Ásgeir Blöndal Magnússon 1989). Icelandic Etymological Dictionary (*Íslensk orðsifjabók*) was compiled by Ásgeir Blöndal Magnússon and originally published in 1989. It is the first and only Icelandic etymological dictionary and contains approximately 25,000 entries.

In fall 2017 *Íslenskt orðanet* (Icelandic wordnet) will be added. This is a resource that was considered for the initial launch, but was deferred due to technical complications. *Íslenskt orðanet* is a semantic database with elements that allow it to be used much like an onomasiological dictionary or a thesaurus (Jón Hilmar Jónsson 2018).

Other resources to be added are the Written Language Archive (*Ritmálssafn*), a database of language use with usage examples from 1540, when the New Testament was first published in Icelandic, to recent decades; also: four historical dictionaries, compiled in the 17th, 18th and 19th centuries; and possibly ISLEX, an Icelandic-Scandinavian dictionary with six language pairs. Adding ISLEX to *málið.is* could make the portal more useful to Icelandic students of Scandinavian languages and to Scandinavian users, who might benefit from some of the other resources too.

#### 4. Usage

Usage of the portal has grown somewhat from its initial launch. The AMI does not have any advertising budget for the project and therefore has to make do with word-of-mouth. The project received valuable media coverage upon launch and in

the first six months we received on average 700 different users on working days. Most weeks we had around 2500 different users. The usage went down to 1500 weekly users during the summer 2017, but in August it picked up and the last two weeks of August and first week of September 2017, the last weeks in our data, show record numbers, 3200-3500 weekly users. This suggests that the word is spreading and the user base is growing.

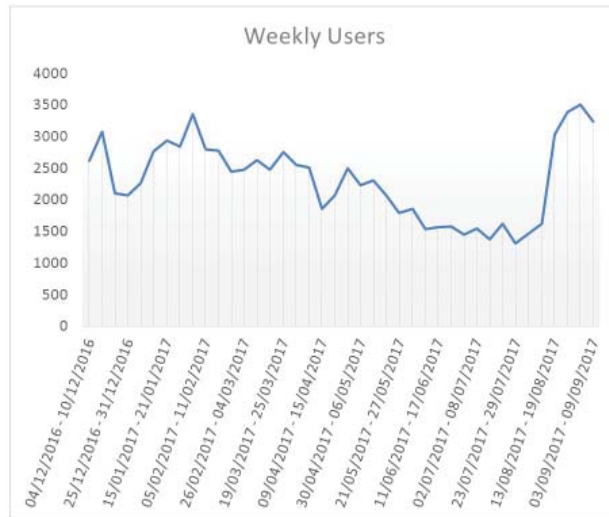


Figure 2: Weekly users.

In total, more than 50,000 users visited the site between November 2016 and September 2017. This is according to Google Analytics, which counts each browser in each device as a different user.<sup>1</sup> The actual number of users is therefore somewhat smaller. For reference, there are no more than 350,000 speakers of Icelandic in the world. Most of the portal's users (92%) are in Iceland.

The users look for all sorts of things but there seem to be some emerging patterns. Upon checking the most common words searched for on individual days, we can see that uncommon words used in news stories get many hits for a short period of time. People also search for words that are commonly misspelled, probably to make sure they spell them correctly. The distribution of search strings has a very long tail, as expected. Fewer than 300 words get more than 50 hits and 164,000 search strings get less than 10 hits.

<sup>1</sup> <<https://support.google.com/analytics/answer/2992042?hl=en>>.

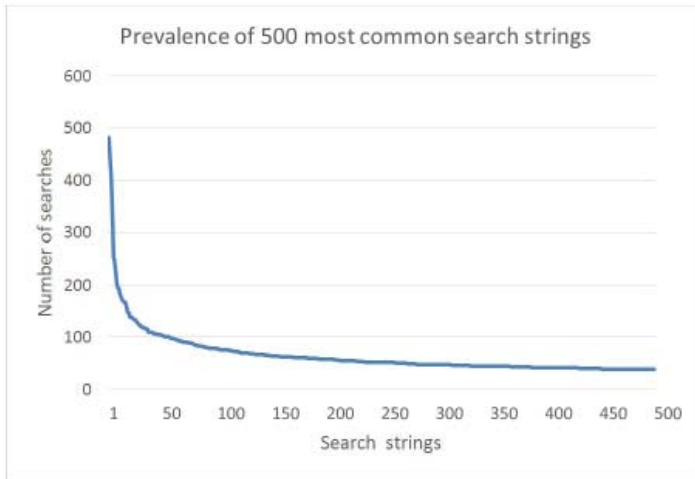


Figure 3: A few strings get looked up often, but as can be seen from the graph there is a long tail of search strings that have few lookups. The graph only shows the 500 most common, but the tail is a lot longer as over 170,000 search strings have been entered.

## 5. What next?

Most of the resources accessible through *málið.is* are constantly being updated. They will continue to be so while new resources will be added to the portal, as outlined in section 3. The data on users, shown in section 4, indicates there may be room for improvement in promoting the portal. Steps have been taken to inform teachers of the portal, but currently no plans for further promotion have been made.

## References

- Ásgeir Blöndal Magnússon (1989): *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Hilmarsson-Dunn, Amanda & Ari Páll Kristinsson (2013): The language situation in Iceland. In: Robert B. Kaplan, Richard B. Baldauf, Jr. & Nkonko M. Kamwangamalu (Eds.): *Language Planning in Europe: Cyprus, Iceland and Luxembourg*. London / New York: Routledge, 100–169.
- Íslensk málnefnd. (2016): *Ritreglur. Auglýsing mennta- og menningarmálaráðuneytis nr. 695/2016 með leiðréttingum*. <<http://islenskan.is/images/ritreglur-IM-2016.pdf>> (Retrieved September 10, 2017).
- Íslensk nútímamálsorðabók* (2017): Halldóra Jónsdóttir & Þórdís Úlfarsdóttir (Eds.). <<http://islenskordabok.arnastofnun.is>> (Retrieved September 10, 2017).



- Jóhannes B. Sigtryggsson (Ed.) (2016): *Stafsetningarorðabókin*. 2. Ed. <[www.malid.is](http://www.malid.is)> (Retrieved September 10, 2017).
- Jón Hilmar Jónsson (2018): Íslenskt orðanet: Tekstbasert kartlegging og presentasjon av leksikalske relasjonar. In: *Nordiske Studier i Leksikografi* 14 (this rapport). Reykjavík: NFL & Stofnun Árna Magnússonar í íslenskum fræðum, 1–17.
- Kristín Bjarnadóttir (2012): The Database of Modern Icelandic Inflection. In: *Language Technology for Normalisation of Less-Resourced Languages SALT-MIL 8 – AfLaT 2012*, 13–18 <<http://aflat.org/files/saltmil8-aflat2012.pdf>> (Retrieved September 10, 2017).
- Málfarsbankinn* (2017): Jóhannes B. Sigtryggsson (Ed.). <<http://malfar.arnastofnun.is>> (Retrieved September 10, 2017).

Steinþór Steingrímsson  
IT-ansvarlig  
[steinthor.steingrimsson@arnastofnun.is](mailto:steinthor.steingrimsson@arnastofnun.is)

Ari Páll Kristinsson  
Forskningsprofessor  
[ari.pall.kristinsson@arnastofnun.is](mailto:ari.pall.kristinsson@arnastofnun.is)

Halldóra Jónsdóttir  
Ordbogsredaktør  
[halldora.jonsdottir@arnastofnun.is](mailto:halldora.jonsdottir@arnastofnun.is)

Árni Magnússon-instituttet for islandske studier  
Laugavegi 13  
101 Reykjavík, Island

# Den islandske termbanks forskellige resurser

*Ágústa Þorbergsdóttir*

The Icelandic Term Bank, accessible at [ordabanki.hi.is](http://ordabanki.hi.is), was officially launched on the Internet in 1997 by the Icelandic Language Institute. The Árni Magnússon Institute for Icelandic Studies is now in charge of the Term Bank. The Icelandic Term Bank contains different terminological glossaries from various sources and domains. The aim of this study was to identify to what extent the data comprise an integrated whole. The examination showed that they are partly divergent and revealed that many important domains are missing. However, it is clear that The Icelandic Term Bank is of benefit to many users, who are very grateful for its existence. The study led to the conclusion that the Icelandic Term Bank still needs to be strengthened and that is necessary to encourage terminology work within more domains. Essential prerequisites are willingness to do this work, knowledge how to execute it, and sufficient financing because terminology work requires solid funding.

## 1. Indledning

Islandsk sprogsekretariats ordbank (*Orðabanki Íslenskrar málstöðvar*) blev åbnet på webben i 1997. Islandsk sprognævn og Islandsk sprogsekretariat havde lagt stor vægt på oprettelsen af en ordbank med termer og allerede i 1979 blev en elektronisk termbank planlagt af Islandsk sprognævn.

Der var store forventninger til termbanken. Den skulle gøre adgangen til islandske termer nemmere og den skulle gøre det muligt at slå op og se hvilke oversættelser man havde fundet eller foreslået. Termbanken skulle også give adgang til definitioner af termer og være til hjælp ved koordinering af sprogbrug indenfor forskellige fag (Stefán Briem 1998:75).

Termbanken består af to dele: den ene for almene sprogbrugere som søger efter termer eller definitioner af begreber i fagsproget og den anden for redaktører af termsamlinger som får adgang til en specielt tilpasset database for at redigere deres terminologisamlinger. Fra 1997–2006 administreredes termbanken af Islandsk sprogsekretariat som var et center for terminologi i Island (jf. Sigrún Helgadóttir & Kristján Árnason 1991:17–19). Fra 2006 overtog Árni Magnússon-instituttet for islandske studier Islandsk sprogsekretariats opgaver og pligter, bl.a. termbanken samt rådgivning og vejledning om terminologiske arbejdsmetoder. Det er almindeligt i Island, at forskellige fagforeninger og institutioner opretter udvalg som skal tage sig af terminologien indenfor de respektive fagforeninger. Termbankens redaktionsdel skulle være hjælp til disse termudvalg og andre som udarbejder termsamlinger.

En af de største udfordringer i oprettelsen af en termbank var at finde godt og velegnet materiale. I min artikel undersøger jeg opbygningen og den indre struktur af termbanken, og mit hovedspørgsmål er hvordan de forskellige resurser danner en helhed.

Til besvarelse af spørgsmålet vil jeg analysere og diskutere følgende under-spørgsmål:

- Hvilke metoder blev anvendt for udvælgelsen af materiale i termbanken?
- I hvilken grad kan de forskellige termsamlinger betragtes som en helhed?
- Hvor helstøbt er termbanken efter 20 år?

Til denne undersøgelse anvender jeg hovedsagelig primære data som jeg har indsamlet om termbanken samt en brugerundersøgelse som jeg foretog i maj 2017.

## 2. Valg af materiale i termbanken

I § 7 i bekendtgørelse om Islandsk sprognævn og Islandsk sprogsekretariat nr. 159 fra 1987 siger: „Sprogsekretariatet skal følge med udviklingen af termbanker i andre lande, forberede en tilsvarende termbank her og varetage dens drift“<sup>1</sup> (Reglugerð 1987). Følgelig stiledes der mod en termbank og forberedelserne bestod bl.a. i at etablere en standard for lagring af data og deres håndtering (jf. Halldór Halldórsson & Baldur Jónsson 1993:132). Islandsk sprognævn besluttede at databehandle sine termsamlinger som senere kunne blive kimen til udviklingen af en termbank.

Her undersøges indsamling og anskaffelse af materiale i termbanken som er en af de vigtigste faktorer ved dens forberedelse. Da termbanken blev åbnet på webben den 15. november 1997 indeholdt den 13 termsamlinger og deraf fem som tidligere var blevet udgivet i bogform af Islandsk Sprognævn: Disse termsamlinger var *Tölvuorðasafn* (EBD), *Orðaskrá úr uppeldis- og sálarfræði* (Psykologi), *Flugorðasafn* (Flyvning) og *Orðasafn úr tölfræði* (Statistik). Det havde mange fordele at have disse ordsamlinger parate i begyndelsen, fx angående ophavsret og kvalitetskontrol.

Termbankens formål var også at publicere termsamlinger fra termudvalg indenfor forskellige fagområder og andre interessegrupper. Dóra Hafsteinsdóttir, forhenværende termbankdirektør, oplyste i et interview 2001, at ”utrolig mange” tilbød deres terminologiske materiale uden anmodning (300 þúsund uppflettiórð 2001). Termsamlinger fra termudvalgene fandtes i forskellige formater. Nogle havde anvendt programmer så som Word og Excel ved udarbejdelse af deres term-

<sup>1</sup> ”Málstöðin skal fylgjast með þróun í orðabanka í öðrum löndum, undirbúa slíkan banka hér og sjá um rekstur hans.”

samlinger og de blev importeret i termbanken, men redaktører af nye termsamlinger forventedes at bruge termbankens database for redigeringen.

De første idéer om indsamling af materiale til termbanken bestod i at sprogsekretariatet skulle skaffe sig eneretten til termsamlingerne for siden at bearbejde deres indhold, revidere dem og ændre, og udelade det materiale som ikke levede op til de kvalitetskrav som der blev stillet til publicering i sprogsekretariatets navn. Ved nærmere eftersyn viste det sig at det var svært at udføre kvalitetssikringen i praksis og også meget dyrt. Derfor blev idéen opgivet og det blev besluttet, i samråd med en advokat, at indgå en samarbejdsaftale mellem Islandsk sprogsekretariat og termsamlingernes redaktører. Ifølge aftalen ejer hver redaktør sin termsamling og har ophavsretten til den. Samarbejdsaftalen med redaktørerne sikrede dem gratis adgang til termbankens brugergrænseflade og til gengæld fik de lov til at publicere deres materiale som en del af termbanken (jf. Dóra Hafsteinsdóttir 1997:6).

### **3. I hvilken grad kan de forskellige termsamlinger betragtes som helhed?**

Eftersom resurserne i termbanken er af adskillig oprindelse er termsamlingerne indbyrdes forskellige. Ved importen i termbanken formindskedes forskellen automatisk og termsamlingerne fik et helhedsindtryk pga. termbankens artikelstruktur. For at svare spørgsmålet i hvilken grad de forskellige termsamlinger betragtes som en helhed undersøger jeg disse punkter:

- Forskellige informationer i de enkelte samlinger
- Forskellig synsvinkel
- Forskellige søgeresultater
- Brugernes mening om termbanken

#### **3.1. Forskellige informationer i de enkelte samlinger**

Der kan være stor variation i hvilke informationer tages med i de enkelte samlinger. Der er kun én termsamling som er monolingval (kun islandsk), godt halvdelen af termsamlingerne er tosprogede (islandsk/engelsk), mens de andre er flersprogede (fra tre til elleve sprog). Grammatiske oplysninger, fx genus og ordklasse, angives i et mindretal af termsamlingerne. Definitioner findes i omtrent halvdelen af termsamlingerne. I nogle termsamlinger er der definitioner på andre sprog end islandsk hvilket beror på samlingens oprindelse. Redaktørerne har i de tilfælde ikke oversat definitioner til islandsk (jf. eksempel i figur 1).

Úr orðasafninu <i>Liforðasafn 2</i> ☒
Flokkun: erfðafræði
[íslenska] <b>samleitin þróun</b>
[enska] <b>convergent evolution</b> [ <b>skilgr.</b> ] Evolution in two different lineages or populations that independently produces a similar outcome in both.

Figur 1: Et eksempel på en definition på engelsk.

### 3.2. Forskellig synsvinkel

Termbankens database er baseret på en terminologisk synsvinkel hvor fokussen er på begrebet. Dette betyder at hvis der findes nogle termer for ét begreb skal redaktørerne tage stilling til hvilken term de vælger som hovedterm og registrere andre eventuelle termer som synonyme. På den måde kan man sige at termbanken skal være preskriptiv og give anbefalinger af bestemte termer.

Her kan man mærke en variation inden for termbanken på grund af hvordan synonyme registreres. Som eksempel kan nævnes termsamlingen *Tölfræðiorðasafn*, udgivet af Islandske sprognævn 1996. I termsamlingen har redaktøren valgt *hundraðshluti* som hovedterm og registreret *prósent* som synonym. Hvis man slår op *prósent* så henvises der til termen *hundraðshluti*.

Úr orðasafninu <i>Tölfræði</i> ☒	prósent -> <u>hundraðshluti</u>
[íslenska] <b>hundraðshluti</b>	[sh.] prósent [enska] percent [Tölfræði ☒]
[sh.] prósent	
[s.e.] <u>hundraðstala</u>	
[enska] <b>percent</b>	

Figur 2: Et eksempel på henvisning fra et synonym til hovedterm.

Selvom termbanken er tænkt som begrebsorienteret er det ikke alle som anvender samme metode. Her kan nævnes *Hagfræðiorðasafn* fra 2000 hvor synonyme har samme status. Termen *elasticity* registreres to gange, både med islandsk *teygni* som hovedterm og *næmi* som synonym, og omvendt.

Úr orðasafninu <i>Hagfræði</i> ☒	Úr orðasafninu <i>Hagfræði</i> ☒
[íslenska] <b>teygni</b> kv. [sh.] næmi	[íslenska] <b>næmi</b> hk. [sh.] teygni
[enska] <b>elasticity</b>	[enska] <b>elasticity</b>

Figur 3: Et eksempel fra en termsamling hvor synonymer har samme status.

Man kan kun mærke den forskel hvis termerne er suppleret med et synonym og forskellen er sandsynligvis ikke iøjnefaldende for brugerne som kun ser ét begreb ad gangen.

### 3.3. Forskellige søgeresultater

Termbanken kan give forskellige søgeresultater og det kan der være mange grunde til. Forældede termer kan være en grund, fx findes der i nogle termsamlinger termer som er gået af brug og anvendes ikke længere. Et eksempel herpå er adjektivet *kynhvarfur* for ‘homoseksuel’ i termsamlingen *Orðaskrá í uppeldis- og sálarfræði* fra 1986, som ikke er blevet opdateret, mens termen *samkynhneigður* anvendes i termsamlinger i biologi og medicin. Et andet eksempel i den samme termsamling er *sjálfhverfur* for ‘autistisk’, men nu er *einhverfur* den brugte term. Til gengæld er *sjálfhverfur* den brugte islandske term for ‘egocentrisk’. En anden grund kan være at ikke alle er enige i opfattelsen af hvad er en god term. I de adskillige termsamlinger i termbanken forekommer fx ni islandske termer for ‘dysleksi’: *lesblinda*, *leshömlun*, *dyslexía*, *lestregða*, *lesröskun*, *torlæsi*, *lestrar-glöp*, *lestrarörðugleikar* og *lesvandkvæði*. Det er usædvanlig mange termer til ét og samme begreb, men i dette tilfælde er grunden at *lesblinda*, som er den mest almindelige term, opfattes af mange som stødende og de andre er forslag til nye termer som skulle erstatte *lesblinda*.

Úr orðasafninu <i>Læknisfræði</i> ☒	Úr orðasafninu <i>Ónæmisfræði</i> ☒	Úr orðasafninu <i>Liforðasafn</i> ☒
[íslenska] <b>hnýttill</b> [sh.] meðvaki	[íslenska] <b>hnýttill</b> [sh.] smávaki	[íslenska] <b>meðvaki</b> [sh.] hnýttill [sh.] hjálparvaki
[enska] <b>haptent</b>	[enska] <b>haptent</b>	[enska] <b>haptent</b>

Figur 4: Et eksempel på forskellige søgeresultater fra tre forskellige samlinger i termbanken.

I figur 4 kan man se et eksempel om terminologisk forvirring i tre beslægtede fag-

sprog (medicin, immunologi og biologi) angående en islandsk term for det engelske begreb *haptan*. Det er ikke noget særsyn og efter min mening er det uheldigt med hensyn til termbankens kvalitet.

Enslydende termer giver sjældent anledning til misforståelser, når de optræder indenfor forskellige fag. Bygningsingeniørens *brú* forveksles ikke med sømandens som heller ikke forveksles med *brú* i tandlægevidenskab.

### 3.4. Brugernes mening

Den første og eneste brugerundersøgelse om termbanken blev foretaget i maj 2017 og den havde først og fremmest det formål at opnå mere viden om brugerbehov, benyttelse og tilfredsheden med termbanken. I undersøgelsen var der ikke stillet et direkte spørgsmål om brugernes oplevelse af termbanken som helhed. Det viste sig dog at nogle brugere kom med kommentarer om at der hersker en vis uoverensstemmelse mellem termsamlinger i beslægtede fag, fx mellem medicin og immunologi, og af nogle opfattedes det som en ulempe. De vil kunne stole på termbanken og de ønsker mere harmoni. Brugerne ønskede definitioner i flere termsamlinger og flere målsprog, som man kan muligvis tolke som et ønske om større harmoni eller måske kun om mere oplysninger og øget kvalitet.

På den anden side er det tydeligt at ikke alle brugere er klare over at termbanken er sammensat af termsamlinger af forskellig oprindelse og med forskellige redaktører. Som et enkelt eksempel kan man nævne e-mails fra brugere til en uvedkommende redaktør. Det kan tyde på at ikke alle brugere har en opfattelse af hvordan termbanken er opbygget, dvs. faktisk som en portal til mange forskellige termsamlinger. Ganske vist kan man læse om termbankens opbyggelse på dens hjemmeside, men tydeligvis er det ikke alle som læser disse oplysninger.

## 4. Hvor helstøbt er termbanken efter 20 år?

Termbankens indhold er vokset meget fra dens oprettelse og den er blevet den største terminologisamling i Island. Da termbanken blev åbnet var der kun 13 publicerede termsamlinger og 30.000 termposter i termbanken, men nu er der godt 60 termsamlinger og 180.000 termposter.

Når man iagttager termbankens indhold ser man, at den har stærke domæner, især medicin som indeholder omkring 33.000 begreber. Andre relativt stærke domæner er fx elektrotekniske termer (19.000 begreber) og plantetermer (11.500 begreber) foruden biologi (10.600 begreber). Det kan desværre ikke afvises, at der er ubalance i termbanken. Det er nemt at finde fagområder hvor der mangler terminologiarbejde, fx inden for humaniora og socialvidenskab, og desuden kan man finde termlister som overlapper hinanden og berører samme område. Når man overvejer spørgsmålet hvor helstøbt termbanken er efter 20 år kan man ikke overse denne ubalance.

Idéen eller konceptet med termbanken var at den skulle blive mere og mere righoldig (Dóra Hafsteinsdóttir 1997:6), men når drivkraften for fagfolks termarbejde almindeligvis er baseret på deres interesse bliver konsekvenserne ret tilfældigt termarbejde. Efter min mening er kvantitet en absolut succesfaktor for termbanken og det er selvfølgelig vigtigt at termsamlingerne der laves er af høj kvalitet. Det er vigtigt at der skabes tillid til termbanken. Der skal være tilstrækkeligt materiale for at brugerne kan stole på termbanken når de leder efter termer. Det er min opfattelse at hvis dækningen i termbanken har kraftig slagside og der mangler vigtige domæner kan man ikke sige at termbanken er helstøbt. Det kræver planlægning og en dedikeret og struktureret indsats at opnå succes med en termbank som indfrier brugernes behov og ønsker.

## 5. Konklusion

Mit hovedspørgsmål var at analysere hvordan og i hvilket omfang termbankens forskellige resurser danner en helhed. Jeg har stillet tre underspørgsmål som efter min mening er nødvendige for at opnå en samlet besvarelse af spørgsmålet. Det første angår indsamling og anskaffelse af materiale i termbanken som i sig selv var en vældig udfordring. En nærmere undersøgelse viste at materialet i termbanken er af forskellig oprindelse. Det er et faktum som man ikke kan overse ved behandling af forskelligheden.

Det andet underspørgsmål drejede sig om i hvilken grad de forskellige termsamlinger kan betragtes som en helhed. Her er opmærksomheden rettet mod hvor man helst kan mærke forskellighed. Det viste sig at det var varierende hvilke informationer blev taget med i termsamlingerne, fx angående antal sprog. Det var undertiden muligt at mærke forskellige arbejdsmetoder hos redaktørerne, dvs. en del af termsamlingerne er lavet efter terminologiske principper hvor udgangspunktet tages i begreberne, men andre ikke. Brugerundersøgelsen viste at for mange og for varierende søgeresultater ser ud til at være blandt de ting, som brugerne bryder sig mindst om. Konsekvenserne kan være at tilliden til termbankens troværdighed som giver mange forskellige svar på en søgning kan formindskes.

Det sidste underspørgsmål drejede sig om i hvor høj grad man kan argumentere for at termbanken er helstøbt. Min undersøgelse viste at der er stor bredde i termsamlingerne og at termbanken har stærke domæner, men det er nemt at finde fagområder hvor der mangler terminologiarbejde og det ser brugerne som en ulempe. Efter min mening er ubalancen mellem domæner et af de største problemer. Det kan let mindske tilliden til termbanken hvis der mangler vigtige domæner i termbanken.

Nu er der 20 år siden termbanken blev oprettet og derfor er der god grund til at overveje termbankens status og evaluere resultaterne og se mod fremtiden. Konceptet med en termbank hvor den frivillige græsrod laver termsamlinger er interessant, men man kan ikke undgå at spørge hvor godt det har fungeret. Ter-



minologiarbejde opstår ikke af sig selv og man kan ikke stole på at termbanken vokser frem af sig selv.

Min konklusion har frembragt en række opmærksomhedspunkter angående hvilken lære man kan drage af termbankens udvikling de seneste 20 år. Min undersøgelse bekræfter at det er ønskværdigt at forny og opdatere termbanken. Visse ting har ændret sig siden termbanken blev oprettet. Jeg vil først og fremmest nævne Islandske sprogpolitik (Íslenska til alls 2008) som der kan henvises til. I sprogpolitikken siges der at terminologi er af afgørende vigtighed for den faglige diskurs på modersmålet og forudsætningen for at man kan behandle faglige emner på islandsk.

Trods termsamlingernes indbyrdes forskellighed er termbanken til stor gavn for brugerne. Forskelligheden anses ganske vist som en ulempe, men ikke et uoverkommeligt problem og brugerundersøgelsen viste at der er mange meget taknemmelige brugere. Termbanken skal gerne vokse og blive større og man skal kunne stole på den. Den islandske sprogpolitik kan være en støtte, men for at opnå disse mål er nogle forudsætninger nødvendige: *Vilje* – eksperter bør ønske at drive med terminologiarbejde og se en værdi i det. *Kendskab* – eksperter må lære hvordan de kan drive terminologiarbejde. Og sidst, men ikke mindst: *Finansiering* – terminologiarbejde koster penge og derfor er finansiering en nødvendig forudsætning for en ny og forbedret termbank.

## Litteratur

- Dóra Hafsteinsdóttir (1997): Orðabanki Íslenskrar málstöðvar. I: *Tölvumál* 22(5.–6), 5–6.
- Halldór Halldórsson & Baldur Jónsson (1993): Íslensk málnefnd 1964–1989. Afmælisrit. Rit Íslenskrar málnefndar 8. Reykjavík: Íslensk málnefnd.
- Íslenska til alls (2008). Tillögur íslenskrar málnefndar að íslenskri málstefnu. Reykjavík: Menntamálaráðuneyti.
- Orðabanki Íslenskrar málstöðvar*. <ordabanki.hi.is> (15.9.2017).
- Reglugerð (1987) = *Reglugerð um Íslenska málnefnd og starfsemi Íslenskrar málstöðvar* nr. 159/1987. <<https://www.reglugerd.is/reglugerdir/allar/nr/159-1987>> (august 2017).
- Sigrún Helgadóttir & Kristján Árnason (1991): Áætlun um tilhögun iðorðastarfsemi. Tillögur. Reykjavík: Íslensk málnefnd/ Menntamálaráðuneytið.
- Stefán Briem (1998): Islandske fagtermer og informationsteknologi. I: *Lexico-Nordica* 5, 67–76.
- 330 þúsund uppflettiorð (2001). I: *Morgunblaðið* 28. apríl 2001, 8. <[http://timarit.is/view\\_page\\_init.jsp?issId=249193&pageId=3394829&lang](http://timarit.is/view_page_init.jsp?issId=249193&pageId=3394829&lang)> (15.9.2017).

Ágústa Þorbergsdóttir  
Árni Magnússon Instituttet for islandske studier  
Laugavegi 13  
101 Reykjavík  
agusta.thorbergsdottir@arnastofnun.is

# Program



## 14. Konfernce om Leksikografi i Norden

30. maj – 2. juni 2017

Reykjavík

### 30. maj – tirsdag

18:00–20:00 Velkomstreception og registrering: Árni Magnússon-institutet for islandske studier, Laugavegi 13 (4. sal)

### 31. maj – onsdag

9:00–10:00	Registrering – kaffe fra ca. 9:30	
10:00–11:00	Åbning og plenumforelæsning – Askja 132 Ordstyrer: Steinþór Steingrímsson	
	Åbning af konferencen	
	Bolette Sandford Pedersen: Semantisk processering på basis af leksikografisk viden	
11:00–12:00	Session 1A – Askja 131 Ordstyrer: Henrik Hovmark	Session 1B – Askja 130 Ordstyrer: Sturla Berg-Olsen
	Carsten Hansen & Jakob Steensig: Ord i samtaler – findes de?	Petri Lauerma: Finska språkets lånordförråd i ljustet av Martti Rapolas ordsamling
	Emma Sköldberg & Lena Wenner: Amatörlexikografiska insatser på sajten Folkmun.se	Taru Laanti: Ord och sammansättningsled av svenskt ursprung i Vanhan kirjasuomen sanakirja
12:00–13:30	FROKOST – Háma	
13:30–14:30	Session 2A – Askja 131 Ordstyrer: Emma Sköldberg	Session 2B – Askja 130 Ordstyrer: Ágústa Þorbergsdóttir
	Jonas Jensen, Lars Trap-Jensen, Henrik Lorentzen, Sanni Nimb & Mette Marie Møller Svendsen: Thaipiger, muskelhunde og fulde svenskere: nedsættende ord, stereotyper og ligestilling i Den Danske Ordbog	Caroline Sandström: Finlandssvenska dialektordböcker i ljustet av insamlat material
	Lennart Larsson: Får flickor vara vildbasare? Om könstillhörighet och genus hos Dalin och i SAOL	Henrik Hovmark: Udflytning som formidlingsstrategi i ordbøger

14:30–15:00	KAFFEPAUSE	
15:00–16:30	Session 3A – Askja 131 Ordstyrer: Pär Nilsson	Session 3B – Askja 130 Ordstyrer: Lars Trap-Jensen
	Jørgen Schack: Perfektum participium og ordklassen adjektiver	Sturla Berg-Olsen, Knut E. Karlsen & Dagfinn Rødningen: Norsk historisk normeringsordbok
	Carina Nilstun: Når breen kalver – leksikalisering av bøyde former	Ari Páll Kristinsson: Holdninger og formål påvirker brugen af (fremmed)sproglige resurser: behandling af statsnavne i en offentlig statsnavnefortegnelse
	Lars Brink: Adjektiv eller adverbium?	Jóhannes B. Sigtryggsson: Den islandske retskrivningsordbog og nye islandske skrivregler
17:00–19:00	RECEPTION i Nordens hus (de nordiske ambassader)	

### 1. juni – torsdag

9:15–10:00	Plenumforelæsning – Askja 131 Ordstyrer: Anna Helga Hannesdóttir	
	Jón Hilmar Jónsson: <i>Íslenskt orðanet</i> : Tekstbasert kartlegging og presentasjon av leksikalske relasjoner	
10:00–10:30	KAFFEPAUSE	
10:30–12:00	Session 4A – Askja 131 Ordstyrer: Jóhannes B. Sigtryggsson	Session 4B – Askja 130 Ordstyrer: Caroline Sandström
	Ellert Jóhannsson & Simonetta Battista: Middelaldertekster som sproglig ressource	Christian-Emil Ore & Oddrun Grønvik: Bokmål og nynorsk samindeksert - Metaordboka som verktøy for jamføring og utforskning av ordtilfang
	Bodil Rosqvist and Bo-A. Wendt: Gammelt projekt och nya resurser – SAOB:s hantering av nyttillkomna elektroniska textmängder	Henrik Køhler Simonsen & Patrick Leroyer: Sprogresourcer kontra ordbøger: hvad er bedst?
	Asgerd Gudiksen & Henrik Hovmark: Ømålsordbogen som leksikografisk og sproglig resurse	
12:00–13:30	FROKOST – Háma	
14:00	EKSKURSION (fra/til Askja)	

## 2. juni – fredag

9:15–10:00	Plenumforelæsning – Askja 131 Ordstyrer: Ásta Svavarsdóttir	
	Matthew Whelpton: The Variability of Semantic Categories. An Experiment in Extensional Semantics	
10:00–10:30	KAFFEPAUSE	
10:30–11:30	Session 5A – Askja 131 Ordstyrer: Helga Hilmisdóttir	Session 5B – Askja 130 Ordstyrer: Oddrun Grønvik
	Pär Nilsson: Metaftonymi i praktiken: vad är en bildlig betydelse i SAOB?	Ann-Kristin Hult: Användningen av en inlärningsordbok på nätet: fallet Lexins svenska lexikon
	Kristine Eide: Syntaktiske opplysninger i ettspråklige ordbøker med en gruppe intransitive verb som eksempel	Tove Bjørneset: Med Norsk Lexin gjennom 20 år
11:30–12:30	NFLs GENERALFORSAMLING – Askja 131	
12:30–13:30	FROKOST – Askja	
13:30–14:30	Session 6A – Askja 131 Ordstyrer: Ellert Jóhannsson	Session 6B – Askja 130 Ordstyrer: Ann-Kristin Hult
	Tarrin Wills: Lexicon Poeticum: a new dictionary of Old Norse skaldic poetry	Helga Hilmisdóttir & Marjakaisa Matthíasson: Naturbundna uppslagsord i den isländsk-finska nätordboken Islex
	Ágústa Þorbergsdóttir: Termbankens forskjellige resurser	Rósa Elín Davíðsdóttir & Karl Erland Gadelii: En isländsk-fransk nätordbok: utmaningar och utvägar
14:30–15:00	POSTERSESSION – Askja (foyer)	
15:00–15:30	KAFFEPAUSE	
15:30–16:30	Session 7A – Askja 131 Ordstyrer: Jakob Steensig	Session 7B – Askja 130 Ordfører: Håkan Jansson
	Sturla Berg-Olsen: Deskriptivitet og preskriptivitet i allmennordbøker	Aurelija Griskeviciene: Norskhet sett fra litauisk perspektiv
	Anna Helga Hannesdóttir & Benjamin Lyngfelt: Kontrastiv konstruktionsgrammatik och tvåspråkig lexikografi	Poul Hansen: Internettet som korpus: Hvor anvendelig er internettet til at fastsætte frekvenser af ordparsvariationer i dansk og svensk?
16:30–17:00	Afslutning af konferencen – Askja 131	
19:00	FESTMIDDAG i Iðnó	

## Postere

Ellert Jóhannsson, Simonetta Battista, Þorbjörg Helgadóttir & Aldís Sigurðardóttir (Nordisk Forskningsinstitut, København): *Redigering af adjektiver i ONP Online*.

Rune Kyrkjebø & Oddrun Grønvik (Universitetsbiblioteket i Bergen): *Leksikografisk arbeid ved Universitetet i Bergen*.

Mette Marie Møller Svendsen (Det Danske Sprog- og Litteraturselskab, København): *Ordbogsbehov i en zapperkultur — en brugerundersøgelse for Den Danske Ordbog*.

Pär Nilsson (Svenska Akademiens ordboksredaktion, Lund): *r. l. m. l. f. (ss. ig.) — om problematisk genusbeteckning i en diakronisk samtidsordbok*.

Steinþór Steingrímsson, Sigrún Helgadóttir & Eiríkur Rögnvaldsson (Árni Magnússon-instituttet for islandske studier & Islands universitet, Reykjavík): *Risamálheildin: An Icelandic Gigaword Corpus*.

Steinþór Steingrímsson, Ari Páll Kristinsson & Halldóra Jónsdóttir (Árni Magnússon-instituttet for islandske studier, Reykjavík): *MÁLID.IS: A web portal for information on the Icelandic language*.