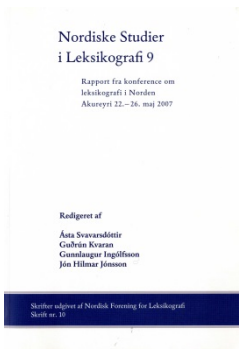


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Lemmatisering og tilgangsstruktur - Refleksjoner på bakgrunn av Íslenskt orðanet	
Forfatter:	Jón Hilmar Jónsson	
Kilde:	Nordiska Studier i Lexikografi 9, 2008, s. 269-281 Rapport fra Konference om leksikografi i Norden, Akureyri 22.-26. maj 2007	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for leksikografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

JÓN HILMAR JÓNSSON

Lemmatisering og tilgangsstruktur Refleksjoner på bakgrunn av *Íslenskt orðanet*

This article discusses how an electronic presentation of a dictionary description provides the means of a wider, but simultaneously more specific, lemmatisation than previously practised in printed dictionaries. This especially applies to multiword lemmas. At the same time users are offered a more varied access to the lexicographic information contained within the dictionary. Here, these possibilities are being evaluated with respect to an ongoing Icelandic lexicographic project named *Íslenskt orðanet* (Icelandic Wordnet). The material of the wordnet is mainly drawn from the contents of dictionaries of Icelandic phrases, and the data is used to identify semantic relations within the Icelandic vocabulary, for single words as well as for phrases. In the lemmatisation of the material the primary consideration is semantic disambiguation. Furthermore, multiword units, particularly various verbal phrases, are given a lemma status equal to that of single words. That way, the dictionary description is made accessible from many different directions, both towards particular semantic fields and related lemmas within those fields, and towards formally related lemmas.

1. Innledning

Vi som har deltatt i det nordiske leksikografiske samarbeidet helt fra starten i 1991 da Nordisk forening for leksikografi ble grunnlagt i Oslo på vår første leksikografiske konferanse, har opplevd ganske store endringer i vårt daglige ordboksarbeid. De mest iøynefallende og gjennomgripende endringene knytter seg i større eller mindre grad til tekniske fremskritt og nyheter, som har gjort det lettere å foreta innsamling og utarbeidelse av leksikografiske data, muliggjort en mer variert og målrettet søking etter leksikografisk informasjon og skapt forutsetninger for å få til en klarere og mer brukervennlig ordbokstekst. Men selv om teknologien har vært en avgjørende faktor i denne utviklingen, finnes det også andre drivkrefter bak disse endringene; de vitner nemlig også om leksikografenes interesse, vilje og evne til å definere og ta fatt på sine oppgaver på en ny måte, sprengte de rammer som lenge har begrenset formatet og dimensjonen av leksikografiske produkter, og komme fram til en strukturering og presentasjon som på best mulig måte tar hensyn til det materialet ordboken bygger på, og den leksikografiske beskrivelse den sikter mot. Mye av det som nå er blitt realistisk og aktuelt å utføre, har sikkert lenge streift leksikografenes tanker i en eller annen form, men kunne ikke bli til virkelighet mens de ytre forutsetninger sto i veien for gjennomføringen.

Når jeg ser tilbake i denne sammenheng, vil jeg gjerne framheve at for ti år siden, våren 1997, holdt NFL sin fjerde konferanse i Helsingfors. Disse vakre maidagene i Finland feiret vi bl.a. utgivelsen av Nordisk leksikografisk ordbok. NLO er altså bare ti år gammel og beholder forhåpentlig fortsatt sin verdi som hjelpemiddel og oppslagsverk for nordiske leksikografer. Uansett er det klart at visse grunnleggende forutsetninger for definisjonen og behandlingen av viktige leksikografiske begreper nå må ses i et nytt lys. Jeg vil i det følgende ta opp noen av de spørsmål det her dreier seg om, og da først og fremst på bakgrunn av et ordboksprosjekt jeg har vært opptatt av de siste årene.

2. De trykte ordbøkernes tradisjon

2.1 En fastlåst lemmaliste

Tradisjonelle ordbøker i trykt form er forbundet med mange og viktige begrensninger som i stor grad også har preget forutsetningene for definisjonen og oppdelingen av mange leksikografiske begreper. For det første så har en trykt ordbok uunnngåelig bare plass for et begrenset antall oppslagsord eller lemmaer, og denne lemmabestanden kan ikke så lett revideres etter at den er blitt konkretisert i form av en trykt ordbokstekst. Jo større og innholdsrikere ordboken er, desto vanskeligere blir det å foreta revidering og oppdatering av ordboksteksten. Og det er ikke lett å komme fram til klare kriterier for avgrensingen av det ordforrådet som får plass som oppslagsord i ordboken overfor det vokabularet som ikke blir tatt med. Dette problemet, lemmaseleksjonen og den konkrete lemmatiseringen, får sjelden noen omtale når ordboken blir presentert overfor brukerne, og de leksikografer som har sittet med dette problemet, er tilsynelatende ikke noe særlig interessert i å redegjøre for hvordan de har arbeidet med det. I ordboksreklamen blir oppmerksomheten også gjerne ledet i en annen retning ved at man framhever hvor mange lemmaer ordboken inneholder eller man utpeker enkelte lemmaer som skal være karakteristisk for lemmavalget.

2.2 Uttrykket fremfor innholdet – den allmenne ordbokens karakter

Ordet med henvisning til sine uttrykksmessige egenskaper har lenge vært en grunnleggende enhet i allmenn leksikografisk beskrivelse, der det framtrer som lemmatisk enhet overfor brukerne samtidig som det utgjør en overskrift for en avgrenset ordboksartikkel som inneholder en selvstendig beskrivelse av vedkommende lemma. Som regel blir alle tilgjengelige opplysninger om lemmaet plassert her uansett om eller i hvilken grad ordet forekommer i beskrivelsen av andre lemmaer eller ikke. Innenfor artikkelen blir hovedvekten lagt på betydningsdifferensiering ved at det skilles mellom delbetydninger i form av nummererte ledd osv. Denne oppbyggingen er ikke minst karakteristisk for den ordbokstype som skal være mest dekkende og minst spesialisert, dvs. allmenne definisjonsordbøker.

Man anser også gjerne at denne typen for lengst har etablert seg som en autonom og grunnleggende ordbokstype som ikke har mye å hente fra de mer spesialiserte ordbøkene. Men når man ser nærmere på saken, er forholdet ikke så enkelt. Allmenne definisjonsordbøker i sin endelige form, uansett hvor omfattende og innholdsrike de er, utgjør f.eks. et ganske begrenset og ufullstendig grunnlag for de ordbøker som har til oppgave å redegjøre for ordenes semantiske relasjoner, og det gjelder ikke i mindre grad for de ordbøker som skal beskrive ordenes syntagmatiske relasjoner og fraseologi.

2.3 Problemer med struktureringen av ordboksteksten

Når det gjelder allmenne ordbøker, er det naturlig at leksikografen retter oppmerksomheten mot helhetsbildet av de enkelte lemmaer, ser etter at beskrivelsen blir strukturert på en regelrett måte, at hvert element i ordboksteksten er på sin plass osv. Det ligger i sakens natur at brukerne ikke i samme grad vil være opptatt av dette, fra deres synsvinkel er et slikt helhetsbilde sjelden av særlig interesse utover det at det viser hvor i artikkelen den aktuelle informasjon er å finne. Deres oppmerksomhet og interesse gjelder først og fremst en bestemt angivelse lemmaet gir adgang til, men som kan være adressert til et underordnet tekstelement. Her vil det selvfølgelig være en fordel for brukeren å ha en klar innsikt i struktureringen av ordboksteksten for å kunne lokalisere den informasjon han eller hun er ute etter. Det er i dette berøringsfeltet mellom leksikografens oppgave å vise et helhetlig bilde av lemmaenes egenskaper og brukerens behov for en problemfri tilgang til en bestemt informasjon ordbokens mikrostruktur får sin utforming, med sin ofte utstrakte og kompliserte oppdeling og hierarki som brukeren kan ha problemer med å orientere seg i og beherske.

Hvor mikrostrukturen er innfløkt og komplisert er makrostrukturen som regel tilsvarende enkel, slik at den først og fremst fungerer som en slags portal til mikrostrukturen som brukeren må kunne åpne, men som i seg selv ikke byr på noen vesentlig informasjon. Men dessuten viser den på sin ufullkomne måte det eneste sammenhengende bildet av ordbokens lemmabestand brukeren kommer i kontakt med, en bestand der alfabetet er den styrende faktor som også gir et glimt av den morfologiske sammenhengen. Ut fra lemmalisten kan brukeren til en viss grad danne seg et bilde av hva slags ordforråd ordboken dekker, og hvordan det blir avgrenset. Men sammenhengende deler av ordforrådet når det gjelder semantiske egenskaper og begrepsfelt, står utenfor brukerens synsfelt, og det blir sjelden nevnt om og i tilfelle i hvilken grad hensynet til slike relasjoner har påvirket lemmaseleksjonen og leksikografens avgrensning av lemmabestanden.

2.4 De semantiske og syntagmatiske relasjonene står i skyggen

De semantiske relasjonene innenfor ordforrådet står tradisjonelt utenfor den allmenne definisjonsordbokens område. Dette aspektet hører til spesialiserte ordbøker av ulike slag, først og fremst synonymordbøker og begrepsordbøker. Det

ordforrådet som disse ordbøkene inneholder, blir dermed vanligvis framstilt uten noen tilknytning til den behandling tilsvarende ord eller lemmer får i en allmenn ordboksbeskrivelse, f.eks. med hensyn til oppdeling i delbetydninger og homonymseparering.

I allmenne ordbøker blir redegjørelsen for ordenes syntagmatiske relasjoner, kollokasjoner, konstruksjoner og faste ordforbindelser i hovedsak underordnet den semantiske betraktningssmåten. En selvstendig redegjørelse for slike relasjoner som finnes i konstruksjonsordbøker, kollokasjonsordbøker og andre spesialiserte ordbøker, er heller ikke knyttet til den behandling tilsvarende ord eller lemmer får i en allmenn ordboksbeskrivelse.

Denne strenge typologiske oppdelingen har satt sterke grenser for utfoldelsen av allmenn ordboksbeskrivelse. Det gjelder ikke bare den allmenne ordbokens autonomi, de spesialiserte ordbokstypene jeg har nevnt, har tradisjonelt heller ikke vært i kontakt med hverandre, og det er ganske uvanlig å se en sammenkobling av en begrepsorientert beskrivelse på den ene side og en syntagmatisk eller fraseologisk orientert beskrivelse på den andre. Her gjelder den langvarige tradisjonen som går ut på at det semasiologiske og det onomasiologiske aspektet skal være adskilt, og at det semasiologiske skal ha prioritet som en slags grunntype.

3. Framveksten av den elektroniske ordboken

Framveksten av elektronisk databehandling og den elektroniske ordboken har endret dette bildet og skapt forutsetninger for revidering av den tradisjonelle typologien. I og med at man nå i mange tilfeller med en enkel søkeoperasjon kan gå direkte til den ønskete informasjon, blir det mindre aktuelt for brukeren å støtte seg til en fast makro- eller mikrostruktur. Avgrensingen av det ordforrådet ordboken dekker, er også usynlig, det kan gjennomgå en kontinuerlig revidering og oppdatering der brukerne til og med kan opptre som aktive rådgivere. Som følge av dette får vi nå en sterkere fokusering på ordforrådet som helhet og dets interne relasjoner istedenfor å være opptatt av en selvstendig beskrivelse av et stort antall uavhengige og isolerte lemmer. Da spørres det bl.a. om den allmenne ordboksbeskrivelsen kan få en utvidet funksjon ved å få en direkte tilknytning til en omfattende og detaljert redegjørelse for de semantiske og syntagmatiske relasjoner ordene opptre i.

4. *Íslenskt orðanet*

4.1. *Semantisk ordnett på grunnlag av syntagmatiske og morfologiske relasjoner*

Disse refleksjonene har sin konkrete bakgrunn i et ordboksprosjekt jeg har arbeidet med de siste to årene sammen med Þórdís Úlfarsdóttir. Dette prosjektet, som har fått navnet *Íslenskt orðanet*, går i hovedsak ut på å utnytte ordenes syntagmatiske og fraseologiske relasjoner med utgangspunkt i konkrete leksikografiske

produkter for å avdekke ordenes semantiske relasjoner og bygge opp et sammenhengende semantisk ordnett der også den syntagmatiske og fraseologiske sammenheng blir ivaretatt.¹

Oppbyggingen av dette ordnettet foregår i mange etapper, og mye arbeid gjenstår før prosjektet er ferdig. Et grunnleggende spørsmål i den innledende fasen besto i å vurdere og bestemme hvilke leksikalske enheter skulle fungere som grunnelementer i ordnettet og dermed inngå i de semantiske relasjoner ordnettet ville vise, og samtidig være tilgjengelige som søkeelementer for brukerne. Oppfatningen av dette har bl.a. en avgjørende betydning for hvor omfattende ordnettet vil være, om det først og fremst skal knytte sammen enkeltord, eller om det også skal kunne dekke flerordsenheter.

4.2 Ordnettets kildemateriale

Det konkrete materialet bak ordnettet er ganske omfattende og mangfoldig, men kjernen består av ordforbindelser, kollokasjoner og fraser av ulike slag som er hentet fra tre forskjellige leksikografiske kilder: a) to fraseologiske ordbøker, *Orðastaður* (Jón Hilmar Jónsson 2001) og *Orðaheimur* (Jón Hilmar Jónsson 2002), senere slått sammen i *Stóra orðabókin um íslenska málnotkun* (STOB, Jón Hilmar Jónsson 2005), og b) et register over ordforbindelser og fraser i seddelarkivet til *Orðabók Háskólans* (Íslands universitets leksikografiske institutt; jf. *Orðasambandaskrá Orðabókar Háskólans*), til sammen over 200.000 ordforbindelser av ulike slag. Dertil kommer omtrent 100.000 sammensetninger fra *Orðastaður* som i ordboken er semantisk klassifisert og morfologisk knyttet til sine orddeler.

I tillegg til dette materialet inneholder ordnettets database leksikografiske opplysninger av semantisk karakter som er hentet fra to ulike kilder. For det første dreier det seg om en samling med islandske ekvivalenter hentet fra to tospråklige ordbøker med islandsk som målspråk, *Ensk-íslenska orðabókin* og *Dönsk-íslensk orðabók*. Disse ekvivalentene er valgt slik at de i ordboksteksten står sammen med en eller flere ekvivalenter slik at man kan gå ut fra at det (i de fleste tilfeller) dreier seg om en synonymirelasjon. For det andre inneholder databasen et ordtilfang fra en eldre arkivsamling med opplysninger fra informanter om betydningen av ord og ordforbindelser, som i samlingens database får en markering som viser at de står i en synonymi- eller hyponymirelasjon til hverandre.

4.3 Oppbyggingen av ordnettet






















4.3.1 En liste med nøkkelord dekker samtlige ordforbindelser og sammensetninger
Hele fraseologien i ordnettets kildemateriale blir knyttet sammen i et register over samtlige nøkkelord (fortrinnsvis innholdsord) i ordforbindelsene. Den ene delen av dette registeret ble til under utarbeidelsen av STOB, den andre er hentet

1 For nærmere informasjon om prosjektet se: <http://www.ordanet.is/>

fra det store fraserregisteret til seddelarkivet. Dette nøkkelordregisteret, som også dekker alle sammensetninger fra *Orðastaður* og de orddeler de består av, utgjør kjernen i en lemmaliste som i samsvar med de regler som gjelder for lemmatiseringen, etter hvert blir bygget opp. Lemmatiseringen er dermed ikke gitt på forhånd, men inngår som en del av bearbeidelsen av det leksikografiske materialet. Foruten å dekke de konkrete ordforbindelser og sammensetninger lemmaet forekommer i (i alfabetisk rekkefølge), viser de enkelte artikler også en særskilt liste med de lemmaer (nøkkelord) disse ordforbindelsene og sammensetningene knytter seg til, sortert etter ordklasse og alfabetisk innenfor hver ordklasse for seg (jf. figur 1), med umiddelbar tilgang til disse lemmaene. Dermed får lemmaene en utvidet funksjon som gjenspeiler hvordan de knytter sammen formelt og betydningsmessig sammenhengende vokabular.

4.3.2 Typemarkering av ordforbindelsene

Den videre bearbeidelsen av grunnmaterialet består bl.a. i å markere skilnaden mellom de enkelte typer ordforbindelser det dreier seg om. Det blir gjort for det første ved å markere idiommer som en særskilt type som da skiller seg fra de mer allmenne ordforbindelsene. For det andre blir ordpar (med sideordnete ledd) definert som en egen type. Denne typemarkeringen blir synliggjort blant nøkkelordene i de enkelte artikler i form av ikoner (jf. figur 1), som da viser hvilket forhold de enkelte nøkkelordene har til det aktuelle lemmaet. Som det framgår av figur 1 dekker denne ikoniseringen også de nøkkelord som er relatert til lemmaet gjennom en allmenn ordforbindelse eller en sammensetning.

5 dýrlegur lo 	<input type="checkbox"/> PAR: hof og hallir	hof no hvk
6 glæsilegur lo  	<input type="checkbox"/> PAR: skrautbyggingar og hallir	skrautbygging no kvk
7 glæstur lo  	<input type="checkbox"/> PAR: stórhýsi og hallir	stórhýsi no hvk
8 gulllegur lo 	<input type="checkbox"/> OSAMB: byggja höll	byggja <hús; verksmiðju>
9 háreistur lo 	<input type="checkbox"/> OSAMB: dýrlegar hallir	dýrlegur lo
10 hátimbraður lo 	<input type="checkbox"/> OSAMB: endurbyggja höllina	endurbyggja <húsið> so
11 stór lo 	<input type="checkbox"/> OSAMB: glæsileg höll	glæsilegur lo
12 veglegur lo 	<input type="checkbox"/> OSAMB: glæstar hallir	glæstur lo
13 aðsetur no hvk 	<input type="checkbox"/> OSAMB: gullleg höll	gulllegur lo
14 aðseturshöll no kvk 	<input type="checkbox"/> OSAMB: gæta hallarinnar	gæta <hallarinnar> so
15 álfahöll no kvk 	<input type="checkbox"/> OSAMB: háreist höll	háreistur lo
16 álfur no kk 	<input type="checkbox"/> OSAMB: hátimbruð höll	hátimbraður lo
17 bíó no hvk 	<input type="checkbox"/> OSAMB: höll konungs	konungur no kk
18 bíóhöll no kvk 	<input type="checkbox"/> OSAMB: höllin rís <uppi á hæðinni, í miðjum garðinum>	<húsið> rís so3bn
126 stórbygging no kvk 	<input type="checkbox"/> OSAMB: höllin stendur <uppi á hæðinni, í miðjum garðinum>	<húsið> stendur <við götuna> so3bn
127 stórhýsi no hvk 	<input type="checkbox"/> OSAMB: loftsvallir hallarinnar	loftsvallir no kvk flt
128 sumar no hvk 	<input type="checkbox"/> OSAMB: reisa höll	reisa <hús; skóla> so
129 sumarhöll no kvk 	<input type="checkbox"/> OSAMB: vegleg höll	veglegur lo
130 sund no hvk 		

Figur 1: Íslenskt orðanet. Artikkelen **höll** i forkortet form. Den venstre spalten inneholder nøkkelordene til lemmaet, sortert etter ordklasse. Spalten i midten inneholder de enkelte ordforbindelser, sortert etter ordforbindelsestype og alfabetisk innenfor hver type.

4.3.3 Markering av semantiske relasjoner blant lemmaene

Gjennomgåingen av fraseologien til de enkelte lemmaer sikter mot det formål å knytte sammen semantisk sammenhørende ord og andre leksikalske enheter som i de enkelte artikler opptrer som nøkkelord. Her synliggjør adjektivartiklene sammenhørende substantiver og omvendt, substantivartiklene gir dessuten adgang til sammenhørende verb, verbartikler viser sammenhørende substantiver og adverb, og adverbiale artikler knytter sammen beslektede verb. I tillegg kommer de sideordnete kollokasjoner der oppslagsordet er knyttet til semantisk nærstående ord, typisk et nærsynonym eller antonym.

Hvis vi tar utgangspunkt i et innholdsrikt lemma som substantivet *saga* 'historie', kan vi forvente at materialet der vil avdekke relevante semantiske relasjoner mellom de adjektiver ordet står sammen med, og at det fører sammen semantisk beslektede verb og verbale konstruksjoner. Det får vi også et konkret bevis på:

eftirminnileg ('minneverdig') saga
minnisstæð ('uforglemmelig') saga

þýða ('oversette') söguna
snara ('oversette') sögunni

hlusta á ('lytte til') söguna
hlýða á ('lytte til') söguna

Men substantiviske konstruksjoner viser seg også å ha denne funksjon:

meginefni ('hovedinnhold') sögunnar
efniskjarni ('kjerneinnhold') sögunnar
meginþema ('hovedtema') sögunnar

Og blant sammensetningene finnes det gjerne nært beslektede ord:

kjaftasaga 'sladderhistorie'
slúðursaga 'sladderhistorie'
sögupersóna 'romanperson'
söguhetja 'romanhelt'

Det er et aktuelt spørsmål i denne sammenheng hvor gjennomgripende og presis denne semantiske sammenføyingen kan bli. Svaret vil først og fremst være avhengig av hvordan de lemmatiske enhetene blir utformet og presentert, om man kan nøyes med ettordsenheter i samsvar med den leksikografiske tradisjonen, eller om det er behov for flerordslemmaer. Et annet viktig spørsmål gjelder homonymi og polysemi: Kan man operere med flertydige lemmaer, eller stilles det krav om entydiggjøring av lemmaene?

4.3.4 Lemmatisering av flerordsforbindelser

Ved utarbeidelsen av ordnettet har det vært av avgjørende betydning at det fraseologiske grunnmaterialet allerede er framstilt på en standardisert måte i de leksikografiske kilder det er hentet fra. Det kommer sterkest til uttrykk når det dreier seg om verb, verbale konstruksjoner og verbforbindelser av ulike slag, både kolokasjoner og mer faste forbindelser som idiomer. Det er nettopp disse ordbokselementene som lenge har vært mest påtrengende og plasskrevende i ordbøkens mikrostruktur, noe som har betydd at de ikke i tilstrekkelig grad er synlige i tekstbildet, og at de ikke kan ses i en aktiv sammenheng med betydningsmessig eller formelt beslektede enheter. I *Orðabæimur* og senere i STOB blir det tatt et skritt mot å gjøre slike forbindelser til selvstendige leksikografiske enheter ved å vise den semantiske sammenhengen under begrepsoverskrifter og tilføye en særskilt registerkomponent der samtlige ordforbindelser blir angitt i alfabetisk rekkefølge under vedkommende verb (jf. figur 2). I ordnettet blir skrittet tatt fullt ut, og verbforbindelsene blir gjort til regelrette lemmaer. Dermed kan de fritt inngå i semantiske relasjoner på linje med ettordslemmaene, samtidig som de byr på sortering og klassifisering med hensyn til deres syntagmatiske egenskaper.

láta ekki sitt eftir liggja → DUGNADUR, FRAMLAG,
FRAMTAKSSEMI, liggja, OSÉRHLÍFNI/FORNFÝSI
láta ekki sjá sig → FJARVERA, sjá
láta ekki svo lítið að <heilsa gestunum>
→ DRAMB/HROKI
láta ekki teyma sig á asnaeyrunum → STADFESTA
láta ekki teyma sig á eyrunum → STADFESTA
láta ekki troða á sér → DUGUR, VIDNÁM
láta ekki vaða ofan í sig → DUGUR, ²vaða
láta ekki vaða ofan í sig á skitugum skónum
→ DUGUR
láta ekki þar við lenda → ÁFRAMHALD
láta ekki þar við sitja → ÁFRAMHALD
láta ekki <tækifærið> ganga sér úr greipum
→ FRAMTAKSSEMI
láta ekki <tækifærið> hjá líða → FRAMTAKSSEMI
láta engan bilbug á sér finna → ÁKVEDNI, bilbug-
ur, DUGUR, KIARKUR
láta fallast ofan í stól → stóll
láta fallast með <honum> → ÁSTARSAMBAND
láta fangann lausan → fangi
láta fara fram athugun á <málinu> → athugun
láta fara illa með sig → DUGLEYSI
láta fara sem fara vill → KÆRULEYSI
láta fara vel um sig → fara, MAKINDI, VELLÍÐAN
láta fá sig til að <taka á sig sökina> → VEIKLYNDI
láta fé af hendi rakna → fé
láta félagsmál til sín taka → félagsmál
láta fjálmár skarast → fjöl
láta fjúka í kveðlingum → kveðlingur
láta fjúka í kviðlingum → kviðlingur
láta fjölinna fljóta → KÆRULEYSI
láta flakka lausu → KÆRULEYSI
láta fótinn fæða sig → ÁGENGNI, FRAMFÆRSLA
láta frá landi → BROTTFOR

Figur 2: *Stóra orðabókin um íslenska málnotkun*. Et avsnitt av artikkelen *láta* i registerdelen, der de enkelte ordforbindelser er angitt i alfabetisk rekkefølge med henvisninger til vedkommende lemma(er) i ordbokens hoveddel.

Denne lemmatiseringen fører til en voldsom økning av verblemmaene ut over det som er vanlig i leksikografisk sammenheng. I alt dreier det seg om bortimot 70.000 verbale lemmer mot omtrent 125.000 substantivlemmer.

Motivasjonen bak lemmatisering av flerordsforbindelser er ikke begrenset til verbene selv om den kan sies å ha størst virkning i den sammenheng. Adverbiale flerordsforbindelser fortjener også klart lemmastatus, slik at semantisk beslektede adverbialer kan knyttes sammen og bygge opp et større semantisk felt. Substantiviske flerordsforbindelser bør også kunne lemmatiseres, særlig når det dreier seg om klare begreper som kan være semantisk relatert til andre substantiver. Blant adjektivene er forholdene litt mer komplisert, som det framgår av redegjørelsen i neste avsnitt.

4.3.5 Entydiggjøring av lemmer

En utstrakt lemmatisering av flerordsforbindelser innebærer i grunn at mikrostrukturen blir oppløst samtidig som flerordslemmerne blir gjort til makrostrukturelle enheter på linje med ettordslemmerne. For verbene har denne operasjonen en avgjørende betydning for såvidt at hvert enkelt verblemma i hvert fall langt på vei blir entydiggjort, foruten at selve lemmerformen i mange tilfeller gjenspeiler betydningen og viser et syntagmatisk mønster som gjerne er felles for flere semantisk beslektede lemmer.

Behovet for entydiggjøring gjelder i prinsippet for hele lemmabestanden, med hensyn til at de semantiske relasjoner kollokasjonene avdekker, i alle tilfeller gjelder en bestemt delbetydning av flertydige ord. Spørsmålet er hvordan man best kan foreta en klar entydiggjøring av flertydige ettordslemmer der selve lemmerformen ikke uten videre kan avsløre den aktuelle betydningen.

Generelt kan flertydighet blant substantivene oppheves ved å tildele de enkelte delbetydningene ulike indeksnumre, på liknende måte som homonymer blir adskilt med homonymindeks i en tradisjonell lemmaliste. En slik separering vil være tilstrekkelig for å få til en feilfri sammenkobling av beslektede ord, men overfor brukeren må den kompletteres med en markering som identifiserer den aktuelle betydningen.

Blant adjektivene er skillelinjene mellom de enkelte delbetydningene mye mer uklare. Men her gjelder det også å forhindre at flertydighet fører sammen ord med ulik betydning, i dette tilfellet substantiviske kontekstpartnere. Den konkrete løsningen i ordnettet består delvis i å utvide lemmerformen med et veiledende substantiv som identifiserer den aktuelle betydningen, f.eks. av **hlýr** 'varm': **hlýr** <**vindur**> ('vind'), **hlýr** <**sumar, vetur**> ('sommer, vinter'), **hlýr** <**flík**> ('klesplagg'), **hlýr** <**bros**> ('smil'). Men dessuten blir den bruken av adjektiver som er relatert til personer, realisert i form av en predikativ konstruksjon: **vera hlýr** ('være hyggelig'), **vera hlýlegur** ('være blid'). Denne fremstillingen blir også overført på karakteriserende substantiver med en liknende funksjon: **vera dugnaðarforkur** 'være en dyktig person', **vera letingi** 'være lat'.

4.4 Ordnettets tilgangsstruktur

Her er det ikke anledning til å gi en utførlig beskrivelse av ordnettets tilgangsstruktur, den er heller ikke fastlagt i detaljer. Følgende kommentarer er for det meste begrenset til noen allmenne forutsetninger som etter min mening preger utformingen av brukerens tilgang til den leksikografiske informasjon han eller hun er ute etter.

4.4.1 Verbale konstruksjonsmønstre

Lemmatiseringen av flerordsforbindelser og kravet om mest mulig entydiggjøring av flertydige lemmaer er bestemmende faktorer for hvordan ordnettets tilgangsstruktur kan utformes. Før vi går inn på den semantiske komponenten i ordnettet, er det verdt å fremheve betydningen av å kunne foreta klassifisering og sortering av de verbale flerordslemmaene ut fra deres formelle egenskaper. Grammatisk tagging av lemmaformene, som allerede er blitt foretatt, har vist seg å være hensiktsmessig både når det gjelder oversikt over konstruksjonsmønstre for enkelte verb, og for å knytte sammen lemmaer som viser samme konstruksjonsmønster (se Þórdís Úlfarsdóttir 2006). Den medfører også at en semantisk sortering av lemmaene kan vurderes og klassifiseres i forhold til lemmaenes formelle egenskaper og omvendt.

4.4.2 Søking i lemmalisten

Man kan gå ut fra at en direkte søking i lemmalisten vil være den mest naturlige søkeoperasjon for de fleste brukere, i hvert fall i første omgang. Lemmatisering av flerordsforbindelser åpner her for nye søkemuligheter, f.eks. slik at en bestemt del av lemmaformen blir valgt som søkeelement. Dette vil f.eks. være aktuelt overfor mange idiommer med en stor formell variasjon der et konstant ledd kan knytte sammen de enkelte varianter. Man kan også regne med at brukeren vil ha behov for en samlet oversikt over de verbale lemmaer av ulik syntagmatisk karakter som inneholder et bestemt verb, for dermed å kunne foreta en målrettet søking. En slik liste vil kunne fremstilles på ulike måter, bl.a. sortert etter konstruksjonsmønstre. I en større sammenheng kan en slik fremstilling by på en interessant sammenlikning mellom enkelte verb. Også i andre tilfeller vil det være aktuelt å fokusere på en formelt sammenhengende del av lemmalista, enten det gjelder et avgrenset ordforråd innenfor en ordklasse, et bestemt konstruksjonsmønster eller forekomsten av et bestemt ord eller en orddel.

4.4.3 Valens og tilknytning til nøkkelordindeksen

De verbale flerordslemmaene gir en eksplisitt informasjon om verbets syntaktiske og semantiske valens. Valensangivelsen er avgrenset innenfor lemmaformen med hakeparentes der de konkrete ord og ordkombinasjoner fungerer som kategorisatorer for leksikalske elementer som syntagmatisk og semantisk kan stå i den aktuelle omgivelsen: *lesa* <*bókina; söguna; textann*> 'lese <boken; historien; teksten>', <*fyrrtækið*> *skilar* <*gróða, hagnaði*> 'bedriften gir <overskudd>'.

For å få innsikt i verbenes syntaktiske valens vil det være mest hensiktsmessig å utnytte den grammatiske markeringen av lemmaformene. Når det gjelder de elementer som uttrykker verbenes semantiske valens, så er de tilgjengelige for direkte søking som i mange tilfeller vil føre sammen semantisk og/eller syntagmatisk beslektede lemmaer. Men her bør man tenke på at de konkrete kontekstpartnerne blir valgt som representanter for betydningslike ord som kan stå i den aktuelle omgivelsen. For å kunne aktivisere hele det ordfeltet og dermed gjøre søkingen uavhengig av et subjektivt valg av kontekstpartner, kreves det at søkefeltet blir utvidet fra å være begrenset til lemmaene til også å dekke nøkkelordindeksen eller i hvert fall deler av den. En slik utvidelse vil være nyttig i mange sammenhenger, men det spørs allikevel om den ikke vil komplisere tilgangsstrukturen i for høy grad.

4.4.4 Søking på semantiske relasjoner

Det er naturlig å tenke seg at de semantiske relasjonene innenfor ordforrådet står i sentrum for brukernes oppmerksomhet, og det overordnede målet for leksikografen består også i å gi brukeren tilgang til en klar og systematisk oversikt over semantiske felt og relasjonstyper, enten det gjelder synonymer og antonymer eller større ordfelt. Her spiller kollokasjonenes vitnesbyrd en hovedrolle ved å føre sammen semantisk beslektet vokabular. I den forbindelse kan man tenke seg ulike forutsetninger for samkjøring av sammenhengende vokabular som da vil gi ulike resultater når det gjelder omfanget og innholdet av et semantisk felt. Slike alternative grupperinger av semantisk beslektede ord kan være interessant i mange brukssammenhenger. Men brukeren vil også gå ut fra at det foreligger en bevisst semantisk sammenkobling og gruppering av lemmabestanden ut fra leksikografens vurdering, enten det gjelder synonymi, antonymi eller andre semantiske relasjoner. Brukerens tilgang til denne informasjon kan utføres på ulike måter. Fra leksikografens synsvinkel spørs det om det er hensiktsmessig å operere med konkrete og synlige overskrifter for de enkelte felt slik man kjenner fra trykte ordbøker (f.eks. *Orðaheimur*), eller om det er mer fleksibelt og brukervennlig å tenke seg at brukeren i alle tilfeller går ut fra de formelle egenskapene (et ord, en frase eller et konstruksjonsmønster), og at veien til de semantiske relasjonene og ordfeltene alltid starter som en utvidet informasjon til et bestemt lemma. I det tilfellet kan konkrete overskrifter være overflødige og til og med ha forstyrrende effekt på den indre dynamikken i begrepsrelasjonene. Her ville man eventuelt kunne nøye seg med veiledende elementer i form av sentrale ord eller fraser innenfor et semantisk felt for å markere det semantiske innholdet i et mindre (underordnet) felt, mens tilgangen til et utvidet felt ikke ville kreve en særskilt markering. Men uansett må leksikografen foreta en omfattende og presiserende etikettering av lemmaene for å gi brukeren tilgang til semantisk sammenhengende enheter.

4.5 Noen sammenliknende kommentarer

Jeg har her redegjort for hvordan ordnettets genuine formål bestemmer prinsippene for lemmatisering av de leksikalske enhetene det inneholder, og hvordan en utstrakt lemmatisering av flerordsforbindelser samt et krav om entydige lemmaer preger utformingen av brukerens tilgang til den leksikografiske beskrivelse. Hvis vi sammenlikner med den leksikografiske tradisjon vi kjenner fra den allmenne trykte definisjonsordboken, består ordnettets særpreg i følgende: a) Kjernen i lemmabestanden er i utgangspunktet nøkkelord til de ordforbindelser og sammensetninger ordnettet inneholder; b) Oppbyggingen av lemmalisten pågår kontinuerlig i takt med utarbeidelsen av det leksikografiske materialet; c) Lemmaene er entydige; d) Lemmaene kan klassifiseres og sorteres utfra ulike kriterier i samsvar med brukerens søkemotiv; e) Den leksikografiske beskrivelsen knytter sammen det semasiologiske og onomasiologiske aspektet. I tillegg kommer at bak lemmabestanden ligger en selvstendig fraseologisk beskrivelse ordnettet kan knyttes til.

5 Sluttord

Hvilken posisjon har så en ordboksbeskrivelse som er utformet på denne måte? Er den begrenset av den ramme som den her er blitt presentert innenfor, eller er det tenkelig å bygge den ut med flere komponenter? Innenfor ordnettet blir det ikke regnet med at enkelte lemmaer, enten de består av ett eller flere ord, vil få betydningsforklaringer eller definisjoner slik man kjenner det fra allmenne definisjonsordbøker. Men det er faktisk logisk at slike forklaringer og definisjoner blir tilføyet når det semantiske ordnettet er blitt ferdiggjort. Det henger ikke minst sammen med verdien av å kunne arbeide med definisjonene ut fra de opplysninger ordnettet kan formidle om forskjellige leksikografiske egenskaper til ordene, og om betydningsgrupperinger innenfor ordforrådet. Ordnettets vil også kunne fungere som veiledende verktøy ved organiseringen og framdriften av definisjonsarbeidet, med hensyn til hvilket ordtilfang skal prioriteres, hvilke ord eller lemmaer som ikke krever noen betydningsforklaringer, og kunne nøye seg med den kombinatoriske og klassifiserende informasjon som foreligger i ordnettet. Det ville gjenspeile den realistiske innstilling at lemmaene har ulikt behov for betydningsforklaringer, og at det finnes lemmaer som godt kan settes i en leksikografisk sammenheng uten å bli definert på en tradisjonell måte. En slik innstilling bidrar til å oppheve de strenge restriksjoner lemmabestanden gjerne blir utsatt for, og den medfører at oppmerksomheten i større grad får gjelde ordforrådet i sin fulle bredde og mangfoldighet. Ordnettets kan også veilede om en innbyrdes vurdering av de delbetydninger ordene opptrer i, f.eks. ut fra antall kollokasjoner og andre forbindelser vedkommende lemma er knyttet til.

Det er verdt å framheve at det ordforrådet som ligger til grunn for ordnettet, er blitt satt sammen ut fra helt andre vurderinger enn de som karakteriserer

lemmavalget i en allmenn ordbok. Ordnettet er i grunnen uavgrenset og åpent for tilføyelser og nye lemmaer, som dels vil oppstå under utarbeidelsen av materialet. Og det gir for såvidt et mer virkelighetstro bilde av ordforrådet og ordbruken at de mest sentrale og aktive ordene får størst vekt, de trekker til seg de fleste og mest varierende forbindelsene og de krever mest leksikografisk analyse. Ordnettet inneholder også viktige opplysninger som kan utnyttes ved vurdering og en eventuell markering av ordenes verdi innenfor ordforrådet i ulike sammenhenger. Her kunne f.eks. antall nøkkelord være en aktuell målestokk, og det kunne også være interessant å se hvilke ord som er mest karakteristiske for idiomer, hvilke konstruksjonstyper som kjennetegner bruken av enkelte verb osv.

Men overfor de muligheter leksikografen har til å bygge opp en helhetlig ordboksbeskrivelse med en fleksibel tilgangsstruktur og mangfoldige søke- og sorteringsrutiner for aktive og målbevisste brukere, står de ytre omstendigheter leksikografien og utgivelsen av ordbøker er avhengig av, der markedet bestemmer, og nye ordbøker må kunne appellere til de gjennomsnittlige ordboksbrukerne og oppfylle deres forventninger og behov. Og utgivelsen må forrente seg. Det er ikke gitt på forhånd, aller minst i våre små språksamfunn.

Litteratur

- Dönsk-íslensk orðabók*. 1992. Ritstj. Hrefna Arnalds, Ingibjörg Johannesen. Reykjavík: Ísafoldarprentsmiðja.
- Ensk-íslenska orðabókin*. 2006. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2001. *Orðastaður. Orðabók um íslenska málnotkun*. Önnur útgáfa, aukin og endurskoðuð. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2002. *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandaskrá*. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson. 2005. *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.
- Orðasambandaskrá Orðabókar Háskólans*: http://www.arnastofnun.is/page/arnastofnun_ord_ordasambond, 8. april 2008.
- STOB = Jón Hilmar Jónsson 2005.
- Þórdís Úlfarsdóttir. 2006. Málfræðileg mörkun orðasambanda. *Orð og tunga* 8:117–144.

Jón Hilmar Jónsson
 forskningsprofessor, f. 1947
 Stofnun Árna Magnússonar í íslenskum fræðum – Háskóli Íslands
 Neshaga 16
 IS-107 Reykjavík
 jhj@hi.is