

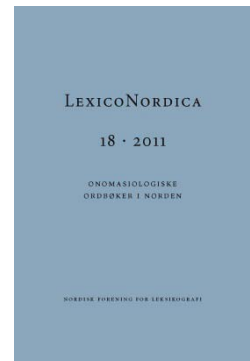
LexicoNordica

Titel: Íslenskt orðanet: Et skritt mot en allmennspråklig onomasiologisk ordbok

Forfatter: Jón Hilmar Jónsson & Þórdís Úlfarsdóttir

Kilde: LexicoNordica 18, 2011, s. 87-110

URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

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.

Íslenskt orðanet: Et skritt mot en allmennspråklig onomasiologisk ordbok

Jón Hilmar Jónsson & Þórdís Úlfarsdóttir

Íslenskt orðanet (Icelandic wordnet) is an ongoing project which aims to analyse and describe the Icelandic vocabulary and its internal relations from the onomasiological point of view. The source material is mainly derived from a comprehensive lexicographical description of Icelandic collocations and phrasemes. The semantic relations between words and other independent lexical units are traced and identified by revealing their syntactic relations, thus building up a lexicographical database with approx. 250,000 single-word and multi-word lemmas. The wordnet combines the roles of a synonym dictionary and a thesaurus, placing special emphasis on verbal constructions and other multi-word lexical units. Within the groups of semantically closely related words, the characteristic lexical and syntactic patterns of the individual words are compared, which gives clues to their degree of semantic relatedness.

1. Innledning

I vår daglige omgang med ordbøker vil vi av og til oppleve at den ordbokstradisjon og ordbokskultur som omgir oss og som vi har mange gode grunner til å respektere, har satt normer og regler vi ikke lenger synes er i takt med tiden, eller som i hvert fall fortjener å bli vurdert på nytt. I mange tilfeller henger dette sammen med at vi blir kjent med nye muligheter til formidling av leksikografisk informasjon, og innser at tidligere begrensninger for ordboksarbeid og barrierer i leksikografens kommunikasjon med brukerne ikke lenger er til stede.

Dette forholdet gjelder ikke minst for vår oppfatning av de ty-

pologiske forskjeller mellom ordbøker og hvilke funksjoner som tildeles de enkelte ordbokstyper. Her tilsier tradisjonen at det er problematisk å tildele en ordbok mange og ulike funksjoner, og at de enkelte funksjoner (i hvert fall mange av dem) best kan oppfylles gjennom spesialordbøker av ulike slag.

Denne innstillingen er preget av lang erfaring med utarbeidelse av ordbøker i trykt form. Med framveksten av elektronisk databehandling og den elektroniske ordboken er dette bildet helt endret, og leksikografien står overfor den utfordring å vurdere verdien av inngrodde teorier og å lansere og teste nye ideer og initiativer.

I denne artikkelen blir denne problematikken drøftet med spesielt hensyn til den mest sentrale og prototypiske ordbokstypen, den allmennspråklige ordboken. Det grunnleggende spørsmålet er hvordan den best kan tilpasses de forutsetninger som ligger i den elektroniske ordboken, slik at den på en dynamisk måte kan formidle mest mulig allsidig informasjon.

Fra en brukersynsvinkel er det ønskelig å kunne forene ulike bruksfunksjoner i ett og samme ordboksverk, slik at brukerne ikke behøver å ha vesentlig forhåndskunnskap om de faktorer som utgjør forskjellen mellom de ulike ordbokstyper. Dette blir enda mer aktuelt når ordboken ikke lenger eksisterer i fysisk form og dens styrke og karakter primært gjenspeiles i hvordan den gir svar på de ulike spørsmål brukerne kommer på.

2. Den semasiologiske ordboken som prototype

Forskjellen mellom den semasiologiske og den onomasiologiske ordboken har lenge stått sentralt innenfor ordbokstypologien. Vi kan trygt si at her har den semasiologiske ordboken og det semasiologiske prinsippet hatt overtaket, og de mest allmenne og sentrale ordbøkene er av den typen. De fleste håndbøker om leksiko-

grafi bærer også preg av dette, f.eks. NLO (1997), Svensén (2004) og Fjeld & Vikør (2008), for å nevne tre sentrale verk innenfor nordisk leksikografi. I alle disse verkene får det onomasiologiske aspektet en forholdsvis liten vekt.

Den allmennspråklige ordboken er tradisjonelt den vanligste semasiologiske ordbokstypen. Her gir den alfabetiske makrostruktur brukerne klar og umiddelbar tilgang til de enkelte artikler, som hver for seg innledes av et lemma i form av et konkret ord. Artiklene inneholder en strukturert beskrivelse av de ulike egenskaper som karakteriserer det aktuelle lemmaet. Hver artikkel er dermed en avgrenset enhet, og de relasjoner den har til andre artikler, er i stor grad usynlige.

Den allmennspråklige ordbokens karakteristiske egenskaper består bl.a. i følgende faktorer:

- Lemmaseleksjonen bestemmes av ulike og ofte uklare kriterier
- Antall lemmaer blir sterkt begrenset og «lemmaverdien» varierer
- Lemmalisten består av ettordslemmaer
- Samtlige (entydige og flertydige) lemmaer får betydningsbeskrivelse (og betydningsdifferensiering)
- Grammatisk informasjon er for det meste begrenset til ordklasse- og bøyningsangivelser til lemmaene
- Fraseologiske enheter er underordnet lemmaene, og deres andel og omfang styres (delvis) av forholdene i de enkelte artikler
- Skillet mellom ordforbindelser (kollokasjoner) og bruks-eksempler er gjerne uklart
- Fast og ofte komplisert artikkelstruktur gjør det vanskelig å foreta revidering og oppdatering av ordboksteksten

Dette betyr at ordboksteksten har et statisk preg, og brukeren vil i mange tilfeller måtte ha god innsikt i strukturen for å komme fram til den informasjon han/hun leter etter. Hvis teksten foreligger i elektronisk form, kan brukeren ved fritekstsøking finne fram til indre tekstenheter og elementer, men det vil ikke gi noe helhetlig bilde av de sammenhenger og relasjoner som ligger skjult i de leksikografiske data teksten inneholder.

3. Tilpasning til det elektroniske mediet

Det elektroniske mediet byr på en mer fleksibel og dynamisk presentasjon og framstilling av leksikografiske data enn det er mulig å få til innenfor rammen av en trykt ordbok. Mange av de ovennevnte faktorer gjelder ikke elektroniske ordbøker, der leksikografen lettere kan definere og bygge opp en helhetlig ordbok som fungerer som en bred og overgripende beskrivelse av ordforrådet til et gitt språk, de leksikalske enheter det er sammensatt av og deres ulike relasjoner slik de gjenspeiles i den faktiske språkbruken. Mange fundamentale sider ved ordbøker og ordboksarbeid blir derved satt i et nytt lys. Det gjelder for lemmaseleksjon og lemmatisering, strukturer og strukturelle sammenhenger og for informasjonstyper og deres innbyrdes relasjoner. Men det gjelder ikke minst forholdet mellom lemmaenes formbaserte og innholdsbaserte egenskaper (dvs. forholdet mellom det semasiologiske og det onomasiologiske aspektet) og hvordan dette kommer til syne for brukerne og i ordboksteksten (se her også Jónsson 2009a).

På bakgrunn av dette kan en allmennspråklig ordbok i elektronisk form tildeles en rekke funksjoner som en tilsvarende ordbok i trykt form ikke kan oppfylle, og som da må knyttes sammen på en mest mulig hensiktsmessig og brukervennlig måte. Det kan gjelde funksjoner som primært har tilhørt andre og mer spesielle ordbokstyper, men også funksjoner som en allmenn ordbok i kraft av

sin datamengde og en fleksibel data- og informasjonsstruktur best er i stand til å oppfylle.

4. Fraseologien som grunnlag

Tanken om en allmennspråklig elektronisk ordbok av den typen som blir tematisert her, har vokst fram under et mangeårig arbeid med islandsk leksikografi. Dette arbeidet var i første omgang konsentrert om fraseologien, og resulterte bl.a. i tre nært beslektede ordbøker i perioden 1994 til 2005, *Orðastaður* (1994, 2. utvidede utgave 2001), *Orðaheimur* (2002) og *Stóra orðabókin um íslenska málnotkun* (2005). Den første har en alfabetisk makrostruktur og kan defineres som en kombinatorisk ordbok (med en fraseologisk og morfologisk komponent) med hovedvekt på kollokasjoner og konstruksjoner. I *Orðaheimur* er målet å dokumentere faste ordforbindelser og idiomer. Hovedlemmalisten består av begrepslemmaer (i alfabetisk orden) med krysshenvisninger til beslektede lemnaer, men i tillegg kommer et alfabetisk register over innholdsord i samtlige fraser der de enkelte frasene kan bli lokalisert ut fra en fast sorteringsregel. Den sistnevnte ordboken er resultatet av en sammenslåing av de to første med en kombinasjon av ord- og begrepslemmaer der kollokasjoner og konstruksjoner fra *Orðastaður* er inkludert i fraseregisteret. Denne ordboken utkom både som papirordbok og elektronisk på cd, og den fins nå også som nettordbok (<http://snara.is/>).

5. Utformingen av Íslenskt orðanet

Gjennom arbeidet med de fraseologiske ordbøkene ble de semantiske relasjonene mellom fraser og ordforbindelser i stigende grad trukket inn, noe som også påvirket og endret ordboksstrukturen

på viktige punkter. Erfaringen med dette ordboksprosjektet banet veien for et pågående prosjekt som forfatterne av denne artikkelen nå arbeider med i fellesskap og som har fått overskriften *Íslenskt orðanet*.¹ Her spiller dataene og datatypene til de fraseologiske ordbøkene fortsatt en viktig rolle, men det semantiske og onomasiologiske aspektet er skjerpet ytterligere ved omfattende strukturendringer og ved å utvide den semantiske markeringen og klassifiseringen til også å omfatte enkeltord av alle sentrale ordklasser.

Før vi går inn på sammensetningen og oppbyggingen av ordnettet, vil vi framheve noen grunnleggende trekk i den ordboksmodellen det skal representere:

- Betydningsbærende flerordsenheter (fraser) blir omformet til flerordslemmaer og får samme status som ettordslemmaer
- Frasematerialet og dermed flerordslemmaene har en primær funksjon i oppbyggingen av databasen, noe som preger ordbokens karakter
- Enhetlig behandling av lemmaene krever (etter mønster fra flerordslemmaene) semantisk entydiggjøring av samtlige ettordslemmaer (innholdsord i ordforbindelser)
- De enkelte relasjonstyper og deres innbyrdes sammenhenger står i fokus
- Gjennom analysen av dataene bygges det opp en semantisk komponent som dekker beskrivelsen av begrepsammenhenger og semantiske relasjoner innenfor ordforrådet
- Grammatisk tagging av flerordslemmaene muliggjør grammatisk sortering i vekselvirkning med den semantiske klassifiseringen
- Tilleggsdata i form av teksteksempler, ordforbindelser og nye lemmaer sikrer kontinuerlig oppdatering av ordboksbasen i takt med fortsatt analyse av dataene

¹ Prosjektet er blitt støttet av det islandske forskningsrådet RANNIS.

5.1. Materiale og omfang

Materialet til ordnettet blir lagret i to adskilte databaser. Bearbeidelsen og analysen foregår i en redigeringsbase, men ved siden av den fins det en åpen nettbasert database der produktet av analysen blir tilgjengeliggjort i form av en leksikografisk beskrivelse og løpende blir supplert ettersom analysen skrider fram. Materialet er i hovedsak hentet fra følgende kilder:

- *Stóra orðabókin um íslenska málnotkun*
- En omfattende samling fraser basert på sitater i sitatarkivet til Orðabók Háskólans (www.arnastofnun.is/page/arnastofnun_gagnasafn_ordasambond)
- Korpus med tekster fra islandske aviser og tidsskrifter (<http://timarit.is>)
- Sideordnede islandske ekvivalenter til fremmedspråklige kildeord fra tre tospråklige ordbøker

Alle innholdsord som forekommer i frasene er inkludert i lemmalisten, sammen med alle sammensetninger (ca. 100 000) som opprinnelig tilhørte den morfologiske komponenten i *Orðastaður*. Gjennom analysen av materialet er lemmabestanden utvidet og supplert, for en stor del ved lemmatisering av betydningsbærende fraser som flerordslemmaer. Lemmalisten inneholder nå over 250 000 lemnaer fordelt på følgende måte:

125 000 substantiver
 35 000 adjektiver
 80 000 verb (og verbale fraser)
 11 000 adverb og adverbialer

Denne store mengden lemnaer er et resultat av at ordnettet er ment å skulle dekke ordforrådet i dets mangfold og bredde og

gjengi de innbyrdes relasjonar som blir avdekket gjennom analysen av materialet. Det betyr at lemmalisten er åpen og uavgrenset, og beskrivelsen av de enkelte lemmaene blir komplettert og utvidet i takt med framdriften av analysen.

5.2. Lemmatisering og bearbeiding av materialet

Grunntanken bak ordnettet er at de syntagmatiske relasjonene innanfor kollokasjoner og andre typer fraser gjenspeiler semantisk slektskap mellom ord og andre leksikalske enheter. På samme måte vil paradigmatiske relasjonar mellom ordledd i sammensetningar i mange tilfeller avdekke semantisk nærhet.

Forutsetningen for at denne kartleggingen av lemmaenes innbyrdes relasjonar kan realiseres, er at lemmaene blir entydiggjort. Dette kommer klarest til uttrykk blant substantivene. I lemmalisten blir flertydige substantiv framstilt som adskilte enheter i form av en homonymindeks (etterfulgt av en identifiserende forklaring):

- 1 *ráðning* no kvk (hirting)
- 2 *ráðning* no kvk (skipan í stöðu)
- 3 *ráðning* no kvk (lausn)

Eksemplet ovenfor viser hvordan det flertydige ordet *ráðning* med delbetydningene 'straff', 'ansettelse' og 'løsning' blir framstilt som tre adskilte lemmaer (no kvk står for substantiv (nafn-ord) hunkjónn (kvenkyn)). Sammen med lemmatisering av betydningsbærande ordforbindelser som flerordslemmaer utgjør entydiggjøring et grunnleggjende tiltak for å få til en konsistent og standardisert lemmaliste som dekker hele datamengden og kan utvides etter behov.

Bearbeidingen av materialet består ellers i hovudsak i å gjennomgå fraser og sammensetningar og på bakgrunn av det å mar-

kere og gruppere semantisk nærstående ord. Det foregår delvis ved mekanisk prosessering og samgruppering, men for en stor del vil prosessen styres av leksikografens vurdering og innsikt i materialet. Figur 1 viser noen glimt fra artikkelen *þjóð* ‘nasjon’, der kollokasjoner ordet opptrer i, blir gruppert for å forene synonyme kollokatorer.

45:	
OSAMB: efnahagsástand þjóðarinnar	efnahagsástand no hvk
OSAMB: efnahagur þjóðarinnar	efnahagur no kk
OSAMB: fjárhagur þjóðarinnar	fjárhagur no kk
46:	
OSAMB: aldersdreifing þjóðarinnar	aldursdreifing no kvk
OSAMB: aldurssamsetning þjóðarinnar	aldurssamsetning no kvk
47:	
OSAMB: athafnalíf þjóðarinnar	athafnalíf no hvk
OSAMB: atvinnuhættir þjóðarinnar	atvinnuhættir no kk flt
OSAMB: atvinnulíf þjóðarinnar	atvinnulíf no hvk
48:	
OSAMB: bókhneigð þjóð	bókhneigður lo
OSAMB: bókmenntahungruð þjóð	bókmenntahungraður lo

Figur 1: Ordforbindelser (kollokasjoner) i artikkelen *þjóð* ‘nasjon’. De tre første gruppene inneholder genitivkonstruksjoner, den siste forener attributive adjektiver.

Synonym- og antonymrelasjoner står i fokus for dette arbeidet, slik at det bygges opp en synonym- og antonymkomponent innenfor ordnettet som både dekker ettords- og flerordslemmaer. Vurdering og markering av synonymrelasjoner er problematisk, bl.a. fordi det i praksis dreier seg om ulik grad av synonymi. Av hensyn til dette blir det ved analysen skilt mellom to typer, synonymer og nærsynonymer.

Ordnettet dekker hermed funksjonen til en synonymordbok som bygges opp kontinuerlig under bearbeidingen av materialet.

Men ordnettet skal også fungere som begrepsordbok ved å knytte sammen lemmaer som tilhører samme begrepsfelt. Oppbyggingen av denne komponenten foregår også etappevis uavhengig av synonymgrupperingen, men vil i mange tilfeller dra nytte av den samtidig som den fører sammen beslektede grupper synonymer.

Markeringen av antonymer bygger for det meste på leksikografens skjønn og vurdering og henger nøye sammen med markeringen av synonymer.

Mekanisk prosessering av de leksikografiske data underletter redaksjonsarbeidet i ulike sammenhenger og har dessuten en selvstendig verdi. Resultatet av slike operasjoner kan i visse tilfeller vises fram direkte til orientering for brukerne, selv om det dreier seg om grovsortering av materialet. Men det tjener først og fremst som grunnlag for å kunne definere og presisere det semantiske forholdet mellom beslektede ord og lemmaer, enten det gjelder synonymer eller mer overgripende semantiske relasjoner, og gjøre dette forholdet synlig for brukerne av ordboken.

Dette kan illustreres nærmere. En kollokasjon av typen adjektiv + substantiv kan gjennom en konstant substantivisk basis forene en rekke semantisk beslektede adjektiviske kollokatorer, f.eks. sammen med ordet *sjúkdómur* 'sykdom': *hættulegur* / *mannskæður* / *banvænn sjúkdómur* 'farlig / som koster mange livet / dødelig sykdom'. De enkelte adjektivene kan gi tilknytning til andre substantiver av samme karakter: *hættulegur faraldur* 'farlig epidemi', *mannskæð inflúensa* 'influenza som koster mange livet', *banvæn drepsótt* 'dødelig pest'. Disse substantivene kan trekke til seg flere ord gjennom andre syntagmatiske og paradigmatiske relasjoner, slik at det samles opp en lang rekke relaterte lemmaer. En god del av dette vokabularet vil høre sammen betydningsmessig. Iblant dreier det seg om ord fra andre semantiske felt, jo mer desto videre rammen er om relasjonene.

5.3. Flerordslemmaenes sentrale posisjon

Som allerede nevnt utgjør flerordslemmaene en betydelig del av lemmalisten og er spesielt karakteristiske for verbene og adverbialene (se også en nærmere beskrivelse i Jónsson 2009b:180f.). For å gi dem selvstendig lemmastatus og sorteringsgrunnlag på linje med ettordslemmaene får de grammatisk markering i form av tagger. Framstillingen er standardisert etter mønster av frasene i de ordbøkene de er hentet fra. Lemmalisten i figur 2 viser en liten del av de verblemmaer som innledes med verbet *leggja* 'legge', der den grammatiske taggen (som primært henviser til ordklasse og kasus) står bakerst i hakeparenteser.

395	<i>leggja niður vinnu sína</i>	<i>so [so ao no-a fn-a]</i>
396	<i>leggja niður vopn</i>	<i>so [so ao no-a]</i>
397	<i>leggja niður vopnin</i>	<i>so [so ao no-ag]</i>
398	<i>leggja niður völd</i>	<i>so [so ao no-a]</i>
399	<i>leggja niður <byrðina></i>	<i>so [so ao <no-ag>]</i>
400	<i>leggja niður <embeitti sitt></i>	<i>so [so ao <no-a fn-a>]</i>
401	<i>leggja niður <fornan> óvanda</i>	<i>so [so ao <lo-a> no-a]</i>
402	<i>leggja nótt með degi</i>	<i>so [so no-a fs no-d]</i>
403	<i>leggja nótt við dag</i>	<i>so [so no-a fs no-a]</i>
404	<i>leggja ofuráherslu á <þetta></i>	<i>so [so no-a fs <fn-a>]</i>

Figur 2: En liten del av de verblemmaer som innledes med verbet *leggja* 'legge'; lemmaene er grammatisk tagget. Taggen *so* står for verb (sagnorð), *no* for substantiv (nafnorð), *lo* for adjektiv (lýsingarorð), *fn* for pronomen (fornafn), *ao* for adverb (atviksorð), *fs* for preposisjon (forsetning). Sammensatte tagger viser kasusmarkering, *a* står for akkusativ, *d* for dativ, *g* for genitiv; bestemt form angis ved *g* etter kasusmarkeringen.

Denne taggingen av verbene gir interessante muligheter til sortering og søking, og når enkelte verb kommer til syne i en lang liste med flerordslemmaer, gir det en oversikt over karakteristiske mønstre og dermed grunnlag for sammenlikning mellom ulike verb. Men man kan også fokusere på enkelte konstruksjonsmønstre og deres utbredelse innenfor lemmaene og dermed knytte sammen formmessig sammenhørende lemmaer, som vist i figur 3.

TEMATISKE BIDRAG

249	brunatryggja <húsið, innbúið> so [so <no-ag>]
250	brúa <lækinn, ána> so [so <no-ag>] cc
251	brúka <hestinn> so [so <no-ag>]
252	brúleggja <ána: stéttina> so [so <no-ag>] cc
253	brúna <steikina: kartöflurnar> so [so <no-ag>]
254	brúnelda <steikina> so [so <no-ag>]
255	brydda <skóna> so [so <no-ag>] cc
256	brytja <kiótið, mörinn> so [so <no-ag>] cc
257	brýna <
258	bræða <
	56 festa <plankanum> so [so <no-dg>]
	57 fjarstýra <tækinu> so [so <no-dg>]
	58 fletta <blaðinu> so [so <no-dg>] cc
	59 fletta <trénu: borðunum> so [so <no-dg>]
	60 fleygja <ruslinu> so [so <no-dg>] cc
	61 flota <bátnum> so [so <no-dg>] cc
	62 flugríða <hestinum> so [so <no-dg>]
	63 flýta <suðunni> so [so <no-dg>]

Figur 3: Den grammatiske taggingen forener verb som viser samme konstruksjonsmønster. I spalten til venstre består konstruksjonen av et verb, etterfulgt av et varierende objekt i akkusativ (bestemt form). Spalten til høyre viser en tilsvarende dativkonstruksjon.

Flerordslemmaene muliggjør også i mange tilfeller leksikalsk betingede søk, f.eks. på bestemte substantiver som opptrer som nominaler i de sammensatte lemmaene. Substantivet *eldur* ‘ild, brann’ knytter f.eks. sammen omtrent 50 lemmaer der det står i ulike funksjoner, som subjekt (<*eldurinn*> *gýs upp* ‘brannen bryter ut’, <*eldurinn*> *logar* ‘ilden brenner’, <*eldurinn*> *breiðist út* ‘brannen sprer seg’) eller objekt (*glæða* <*eldinn*> ‘friske på ilden’, *slökkva* <*eldinn*> ‘slukke ilden’). Funksjonelle enheter innenfor lemmaene kan også være av interesse, som kombinasjonen *eins og* ‘som’ i sammenlikningsledd, som forekommer bortimot 900 ganger i ulike lemmatyper, bl.a. i lemmaene *drekka eins og svín* ‘drikke som et svin’, *lifa eins og kóngur* ‘leve som en konge’, *reykja eins og strompur* ‘røyke som en skorstein’.

Den grammatiske taggingen av flerordslemmaene er ikke minst til stor nytte ved den semantiske klassifiseringen av lemmaene, enten det gjelder synonym- eller begrepsrelasjoner, for formell likhet henger i ganske stor grad sammen med semantisk slektskap.

Denne beskrivelsen av flerordslemmaene har vært konsentrert om verbene, men flerordsforbindelser blir også lemmatisert innenfor andre ordklasser: substantiver (nominaler), adjektiver og ikke minst adverbialer. Blant adjektivene blir entydiggjøringen synliggjort i sammensatte lemmaformer av typen *beittur* [*hnífur*] ‘skarp kniv’, der nominalet i hakeparenteser har en identifiserende og differensierende funksjon overfor andre betydninger av adjektivet og samtidig utgjør et forenende element for adjektiver som karakteriserer vedkommende nominalbegrep: *bitur* [*hnífur*] ‘skarp kniv’, *bitgóður* [*hnífur*] ‘skarp kniv’, *egghvass* [*hnífur*] ‘kniv med skarp egg’ osv.

For adverbialenes vedkommende gir den grammatiske markeringen et verdifullt grunnlag for semantisk klassifisering av leksikalske enheter som tradisjonelt ikke har kommet til syne som selvstendige lemmaer, men som her inngår i semantiske og syntagmatiske relasjoner på linje med andre lemmaer. Figur 4 gir innsikt i omfanget av og variasjonen i de adverbiale lemmaene. Den første listen har et felles innledende element (*þegar* ‘når’) og den andre listen forener lemmaer med samme grammatiske tagg (konjunksjon + substantiv i nominativ + verb).

76	<i>þegar til lengdar lætur</i> <small>ao [st fs no-g so]</small>
77	<i>þegar til reyndanna kemur</i> <small>ao [st fs no-gg so]</small>
78	<i>þegar til stykkisins kemur</i> <small>ao [st fs no-gg so]</small>
79	<i>þegar tímar líða</i> <small>ao [st no-n so]</small>
80	<i>þegar tímar líða fram</i> <small>ao [st no-n so ao]</small>
81	<i>þegar um hægist</i> <small>ao [st fs so]</small>
82	<i>þegar upp er staðið</i> <small>ao [st ao so so]</small>
83	<i>þegar út á lífi</i> <small>ao [st no-n so]</small>
84	<i>þegar úthalle</i> <small>ao [st no-n so]</small>
6	<i>ef þörf krefur</i> <small>ao [st no-n so]</small>
7	<i>meðan aldir renna</i> <small>ao [st no-n so]</small>
8	<i>meðan kraftar endast</i> <small>ao [st no-n so]</small>
9	<i>meðan land byggist</i> <small>ao [st no-n so]</small>
10	<i>meðan öld lifir</i> <small>ao [st no-n so]</small>
11	<i>sem fætur toga</i> <small>ao [st no-n so]</small>
12	<i>sem lög tilstanda</i> <small>ao [st no-n so]</small>
13	<i>þegar ferð fellur</i> <small>ao [st no-n so]</small>
14	<i>þegar færi gefst</i> <small>ao [st no-n so]</small>

Figur 4:
Adverbiale
flerordslemmaer.

5.4. Henvisninger og indikasjoner på semantisk slektskapsgrad

En sentral målsetting for arbeidet med ordnettet er å få innsikt i slektskapsforholdet mellom semantisk beslektede ord og lemnaer. Prosessering av leksikografiske data byr på interessante muligheter i denne sammenheng, og det har vært utviklet ulike metoder for å vurdere og måle slektskapsgraden (se bl.a. Nikulásdóttir og Whelpton 2010). I ordnettet er arbeidet med dette fortsatt under utvikling, men hittil består kartleggingen først og fremst i to observasjoner: a) sammenlikning av konkrete lemnaer ut fra felles tilknytning i relasjoner til andre lemnaer og b) sammenlikning ut fra antall felles partnere i parataktiske forbindelser (ordpar).

5.4.1. Felles lemmatilknytning

Artiklene i ordnettet blir strukturert slik at samtlige lemnaer som opptrer som partnere i ordforbindelser eller sammensetninger, blir oppført som lenker. Figur 5 viser to bruddstykker fra artikkelen *sigur* 'seier', som i alt dekker omtrent 70 ordforbindelser og fraser og ca. 100 sammensetninger, og som har lenker til bortimot 300 ettords- og flerordslemnaer. Spalten i midten inneholder de fraser og sammensetninger lemnaet opptrer i (sortert primært etter type og sekundært alfabetisk), spalten til høyre viser hvilke lemnaer som også er lenket til den aktuelle frasen eller sammensetningen, og spalten til venstre inneholder samtlige lenker (nøkkelord), sortert primært etter ordklasse og sekundært alfabetisk. I det første utsnittet er adjektivene i fokus, men i det andre får vi se (sammensatte) verblemnaer i den venstre spalten og sammensetninger i midten.




Figur 5: To bruddstykkir fra artikkelen *sigur* ‘seier’.

Det framgår av dette at lemmatilknytningen til de enkelte lemmaene gir et godt grunnlag for sammenlikning mellom lemmaer, ut fra det resonnement at jo større likheter det dreier seg om, desto mer vil lemmaene være i slekt med hverandre. Figur 6 viser et avsnitt av en slik sammenlikning mellom lemmaene *krá* ‘kro’ og *knæpa* ‘kneipe’. Spalten i midten inneholder de lemmaer som er lenket til begge to (47), mens spaltene til venstre og høyre inneholder de lemmaer som utelukkende er knyttet til det ene av de to. Antallet felles nøkkelord er her ganske høyt og det vil, ut fra sammenlikning med flere lemmaer i vedkommende felt, gi et verdifullt vitnesbyrd om den semantiske nærheten mellom disse to ordene.



Figur 6: Sammenlikning av lemmaene *krá* ‘kro’ og *knæpa* ‘kneipe’ med hensyn til (felles) lemmatilknytning. Ikonene foran de relaterte lemmaene står for de ulike typer relasjoner, O står for ordforbindelse (kollokator), P for ordpar, H for sammensetningsledd, R for sidestilt ekvivalent.

5.4.2. Antall felles partnere i parataktiske forbindelser

I sammenlikningskartet i figur 6, som i selve artiklene, får de enkelte data- og relasjonstyper blant nøkkelordene en særskilt markering i form av et ikon. Blant disse har ikonet , som står for parataktisk relasjon (ordpar, som *diskótek og krár* ‘diskotek og kroer’, *kaffistofur og knæpur* ‘kaféer og kneiper’), høy frekvens. Dette gjenspeiler den erfaringen at slike relasjoner i svært mange tilfeller belyser det semantiske forholdet mellom nærstående ord, og en nærmere kartlegging av dette forholdet kan gi et godt bilde av hvor nært et gitt lemma er beslektet med andre lemmaer.

I ordnettet blir dette kartlagt på den måten at det fra databasen blir hentet tall om i hvor mange tilfeller et gitt lemma er relatert til andre lemmaer gjennom en felles partner i parataktiske relasjoner. Troverdigheten av denne sammenlikningen vil være avhengig av datamengden (antall ordpar) til de enkelte lemmaene. En sammenlikning av lemmaene *krá* og *knæpa* vitner klart om et nært semantisk slektskap mellom de to ordene, som det framgår av figur 7. Her blir det for hvert nøkkelord angitt i den nest siste kolonnen hvor mange ordpar vedkommende artikkel inneholder. Kolonnen foran angir antall ordpar med lemmaet i overskriften, og den siste kolonnen viser antall felles partnere. For eksempel forekommer *krá* i 90 ordpar med ulike partnere, mens *veitingahús* ‘restaurant’ er knyttet til 138 partnere i slike forbindelser, og 49 av disse partnerene er felles for begge to.

Figuren viser at 11 av de 12 lemmaene som har flest felles partnere med henholdsvis *krá* og *knæpa*, er de samme for begge lemmaene. En slik sammenlikning vil kunne være aktuell for ulike grader av semantisk slektskap, også for ord der det gjelder å bekreftede at et slikt forhold ikke er til stede.

krá:	knæpa:
1 veitingahús no hvk 90 138 49	1 veitingahús no hvk 61 138 35
2 veitingastaður no kk 90 115 44	2 krá no kvk 61 90 33
3 kaffihús no hvk 90 144 43	3 veitingastaður no kk 61 115 27
4 sjoppa no kvk 90 106 34	4 kaffihús no hvk 61 144 26
5 knæpa no kvk 90 61 33	5 bjórstofa no kvk 61 41 19
6 bjórstofa no kvk 90 41 24	6 sjoppa no kvk 61 106 19
7 öldurhús no hvk 90 33 20	7 öldurhús no hvk 61 33 16
8 vínstofa no kvk 90 32 16	8 gildaskáli no kk 61 30 16
9 gistihaus no hvk 90 37 16	9 vínstofa no kvk 61 32 14
10 gildaskáli no kk 90 30 14	10 ölkrá no kvk 61 22 13
11 skemmtistaður no kk 90 14 10	11 gistihaus no hvk 61 37 13
12 kvikmyndahús no hvk 90 12 9	12 skemmtistaður no kk 61 14 11

Figur 7: Antall felles partnere i parataktiske forbindelser:
Sammenlikning av lemmaene *krá* og *knæpa*.

Lemmaenes slektskapsforhold innenfor ordnettet er ennå ikke blitt kartlagt på denne måten i større målestokk. Men erfaringen hittil og de eksempler vi har vist, gir forventninger om at denne metoden kan være av stor verdi for en vesentlig del av lemmaene, spesielt substantivene og adjektivene. Ordpar som datatype er dessuten svært nyttig med hensyn til behovet for å utvide og komplettere ordforrådet på de enkelte semantiske felt.

6. Elektronisk ordbok på nettsiden www.ordanet.is

Ordnettet er hittil beskrevet ut fra dets oppbygning i den sentrale ordboksbasen. Den fungerer som redaksjonsbase, og den vil ikke være åpen for vanlige brukere. Men datastrukturen byr på ulike muligheter til å bygge opp et sammenhengende leksikografisk produkt i form av en elektronisk ordbok. En særskilt database er blitt utformet for dette formålet, og oppbyggingen av en slik ordbok har foregått i takt med framdriften i redaksjonsarbeidet. Produktet blir synliggjort på nettsiden *ordanet.is*, der det blir oppdatert regelmessig.

Her blir det særlig fokusert på semantiske og begrepsmessige

relasjoner, slik at ordboken kan sies å ha en kombinert funksjon som synonym- og begrepsordbok. Men samtidig bevarer den sterke elementer fra de kollokasjons- og konstruksjonsordbøker den opprinnelig bygger på.

Nettordboken omfatter per juni 2011 omtrent 150 000 lemmer, der hvert enkelt lemma står i en relasjon (én eller flere) til andre lemmer, som da fungerer som lenke(r) innenfor vedkommende artikkel. Disse lemmaene har svært ulik status. Noen er knyttet til et større begrepsfelt som dekker en hel rekke lemmer, andre er (foreløpig) bare relatert til ett bestemt lemma. Dette bildet endrer seg stadig ettersom redaksjonsarbeidet og analysen skrider fram, tidligere artikler blir utvidet, nye lemmer kommer inn og analysen blir oppdatert på ulike punkter.

En ordboksmodell av denne typen er svært ulik ordbøker der strukturen er innviklet og fastlåst og der brukerne står overfor en ferdig og urokkelig ordbokstekst. Her gjelder det å gjøre brukerne bevisste på at ordboken er under kontinuerlig komplettering og oppdatering. Men den største forskjellen består i at de enkelte lemmaene ikke framstår som avgrensede og isolerte enheter, men står i et aktivt samspill med andre lemmer som kan utvide brukerens synsfelt i ulike retninger.

Som det framgår av figur 8, kan søk i ordboken rettes mot to adskilte lemmalister. Hovedlisten (FLETTA) er en formbasert lemmaliste som dekker samtlige lemmer i basen, og som man kan søke i via søkefeltet til høyre. Denne listen står i konsentrert form i den øverste rammen til venstre, der man kan klikke på en bestemt alfabetisk del. Nedenfor fins det en annen begrepsbasert alfabetisert lemmaliste (HUGTAKAFYRIRSAGNIR) med overskrifter til begrepsklassene i form av et sentralt og belysende lemma i vedkommende klasse.



Figur 8: Elektronisk ordbok på nettsiden *ordanet.is*. En formbasert lemmanliste gir adgang til enkelte lemnaer og de begrepsfelt de tilhører. En særskilt liste med begrepsoverskrifter gir direkte tilgang til de enkelte feltene.

Brukeren kan fritt definere sitt søkeobjekt i søkefeltet ved hjelp av jokertegn og dermed søke på ulike elementer i lemnaene. Med hensyn til standardiseringen av lemnaformene betyr det at strukturelle elementer kan trekkes fram, på lik linje med enkeltord og andre leksikalske enheter.

Figur 9 viser et enkelt eksempel der søkeordet har vært *sofa* 'sove'. Det gir en liste med bortimot 60 lemnaer der *sofa* står øverst. Ved å klikke på *sofa* åpnes det en artikkel som består av tre felt. Det øverste feltet viser antonymet *vaka* 'være våken', i midten er det et større felt med lemnaer som tilhører den samme begrepsklassen, og det nederste feltet rommer lenker til beslektede klasser. Midtfeltet inneholder 143 lemnaer som står i alfabetisk rekkefølge. Bare den første delen er synlig, men listen kan utvides ved å klikke på *meira*. I øverste høyre hjørne er det to andre lenker. Ved å klikke på *sýna setningargerð* ('vise konstruksjonsmønstre') får vi en liste med grammatisk sortering av lemnaene (jf. bildet til

høyre i figur 9). Ikonet ved siden av fører til en side med forklaringer på de grammatiske taggene.



Figur 9: Artikkelen *sofa* 'sove' på nettsiden www.ordanet.is.

7. Sluttord

Man kan sikkert se det slik at den ordboksmodellen vi har presentert her, ikke kan måle seg med en tradisjonell allmennspråklig ordbok, ikke minst fordi betydningsbeskrivelsen tilsynelatende mangler. På andre punkter går den langt ut over rammen for en allmennspråklig ordbok i tradisjonell forstand: med hensyn til avgrensning og omfang, den vekt som legges på fraser og flerordsenheter, og ikke minst den sentrale stilling de semantiske relasjonene får i modellen. Grunnen til at betydningsbeskrivelse og definisjoner ikke blir prioritert, er at ordet (i kraft av sine formbaserte egenskaper) ikke er den sentrale leksikografiske enheten. Men betydningsbeskrivelsen har sin plass i modellen, selv om fokuset er

blitt flyttet fra det semasiologiske til det onomasiologiske aspektet. I første omgang har den en identifiserende funksjon i forbindelse med entydiggjøring av lemmaene ved å skille mellom homonymer (lemmatiserte betydningsvarianter). Men den vil også komme inn på et senere stadium for å utdype og komplettere den informasjon som formidles gjennom andre elementer i ordboksbasen. Hovedsaken er at betydningsbeskrivelse og definisjoner ikke er obligatoriske, og at lemmaenes betydningsmessige karakter også gjenspeiles i de relasjoner de inngår i og – især for flerordslemmaenes vedkommende – i selve lemmaformene.

Vi har i denne artikkelen lagt vekt på at det onomasiologiske aspektet bør få en mer framtrædende plass i allmenn leksikografisk beskrivelse. Om noen får en ny oppfatning av hva begrepet *allmennspråklig ordbok* innebærer, spiller i og for seg en mindre rolle. Motivet er å finne fram til en hensiktsmessig og adekvat formidling og presentasjon av leksikografiske data uten å være bundet av eldre tradisjoner og et fastlåst system. Det innebærer bl.a. at ordboken kan kombinere egenskaper og funksjoner som tradisjonelt tilhører ulike typer. Den store utfordringen i den forbindelse er nettopp å få til et aktivt og balansert samspill mellom det semasiologiske og det onomasiologiske aspektet.

Litteratur

Ordbøker

- NLO 1997 = Henning Bergenholtz/Ilse Cantell/Ruth Vatvedt Fjeld/Dag Gundersen/Jón Hilmar Jónsson/Bo Svensén: *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Orðastaður 2001 = Jón Hilmar Jónsson: *Orðastaður. Orðabók um íslenska málnotkun*. 2. utgave. Reykjavík: JPV-útgáfa. [1. utgave 1994.]

TEMATISKE BIDRAG

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

Stóra orðabókin um íslenska málnotkun 2005 = Jón Hilmar Jónsson: *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV-útgáfa. <http://snara.is>.

Annen litteratur

Fjeld, Ruth Vatvedt & Lars Vikør 2008: *Ord og ordbøker*. Kristiansand: Høyskoleforlaget.

Jónsson, Jón Hilmar 2009a: Lexical description. An onomasiological approach on the basis of phraseology. I: Sandro Nielsen & Sven Tarp (red.): *Lexicography in the 21st Century. In honour of Henning Bergenholtz*. Amsterdam / Philadelphia: Benjamins, 257–280.

Jónsson, Jón Hilmar 2009b: Lemmatisation of Multi-word Lexical Units: Motivation and Benefits. I: Henning Bergenholtz, Sandro Nielsen & Sven Tarp (red.): *Lexicography at a Crossroads*. Bern: Peter Lang, 165–194.

Nikulásdóttir, Anna Björk & Matthew Whelpton 2010: Lexicon Acquisition through Noun Clustering. I: *LexicoNordica 17*, 141–161.

Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.

Nettsteder

Fraseologisk register til *Orðabók Háskólans*:

www.lexis.hi.is/osamb/osamb.pl

Íslenskt orðanet: www.ordanet.is

Nettside med islandske ordbøker: <http://snara.is>

JÓNSSON & ÚLFARSDÓTTIR

Jón Hilmar Jónsson
forskningsprofessor
Árni Magnússon-instituttet for
 islandske studier
Íslands universitet
Neshagi 16
IS-107 Reykjavík
jhj@hi.is

Þórdís Úlfarsdóttir
ordboksredaktør
Árni Magnússon-instituttet for
 islandske studier
Íslands universitet
Neshagi 16
IS-107 Reykjavík
disa@hi.is

