

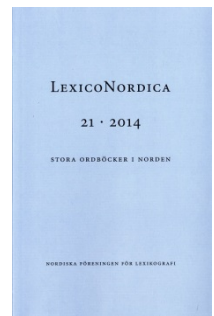
LexicoNordica

Titel: Oppbyggingen av en relasjonsbasert islandsk ordboksbase

Forfatter: Jón Hilmar Jónsson

Kilde: LexicoNordica 21, 2014, s. 39-60

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.

Oppbyggingen av en relasjonsbasert islandsk ordboksbase

Jón Hilmar Jónsson

The article describes the structuring of a relational dictionary database of Icelandic which partly derives from the lexicographic analysis of material in the Orðabók Háskólans archives and additionally builds upon the material of three related phraseological dictionaries of Icelandic from the years around 2000. The syntactic relations of words and phrases in this material reveal semantic relations building up a big collection of synonyms and other kinds of semantically related lexical items. An open and extendable lemma list (including multi-word lemmas) and a systematic presentation of the lemmas together with a comprehensive analysis of textual occurrences bridges the gap between a corpus and a ready-made dictionary.

1. Bakgrunn og utgangspunkt

Islandsk ordbokstradisjon skiller seg fra tradisjonen i de skandinaviske land der oppmerksomheten tidlig var konsentrert om utgivelsen av store nasjonale ordboksverk. På Island arbeidet man riktignok lenge mot det målet å utgi en historisk ordbok over islandsk fra reformasjonen og boktrykkingens inntog til det samtidige språket, alt fra opprettelsen av Orðabók Háskólans i året 1944. Da hadde Den Arnamagnæanske Kommission i København allerede tatt initiativ til å utarbeide *Ordbog over det norrøne prosasprog* (ONP) som ville dekke det eldre språket (jf. Battista & Johannsson 2014). Denne arbeidsfordelingen innebar at språkhistorien i sin helhet ikke ville få en sammenhengende dekning, og beskrivelsen av det yngre språket ville bli foretatt uten større hensyn til det rike kildematerialet til det eldre språket.

Innsamlingen av materiale til den ordboken *Orðabók Háskólans* skulle utarbeide, ble utført på tradisjonell vis ved en omfattende ekserpering av tekstbelegg, samtidig som det ble lagt vekt på innhenting av opplysninger om ord og ordbruk rundt omkring i landet, især gjennom et regelmessig radioprogram i mange år (se brosjyren *Orðabók Háskólans. Leksikografisk institutt ved Islands universitet* fra 1995).

Da dataalderen meldte seg tidlig i 1980-årene, var materialeinnsamlingen fortsatt i gang og det forelå ingen konkrete planer om utarbeidelse og utgivelse av et større nasjonalt ordboksverk. I stedet for å fastlegge en plan i den retning var det nå blitt mer aktuelt å utnytte den nye teknologien til å få en helhetlig oversikt over det innsamlede materialet og skaffe seg innsikt i materialets sammensetning og ut fra dette ta stilling til en videre planlegging. I samsvar med det ble det laget et oversiktsregister over oppslagsordene i instituttets hovedarkiv med noen grunnleggende markeringer som senere ble etterfulgt med inntasting av størsteparten av belegg materialet.

Det ble etter hvert klart at islandsk ikke ville få en leksikografisk beskrivelse i form av en stor ordbok som kunne måle seg med de store skandinaviske ordboksverk. Det islandske ordforrådet hadde allerede fått en utførlig beskrivelse i *Íslensk-dönsk orðabók* tidlig i 1920-årene, og den lå for en stor del til grunn for den første allmenne enspråklige islandske ordboken, *Íslensk orðabók*, som først utkom i 1963. Denne ordboken er senere blitt utvidet og nylig også utgitt i digital form.

På bakgrunn av dette har man i det akademiske ordboksarbeidet kunnet utvikle prioriteringer som tar mer hensyn til de faktiske forholdene og blir sterkere motivert av utviklingen innenfor leksikografisk databehandling og de muligheter den digitale ordboken byr på. Dette har bl.a. resultert i sammenhengende ordboksprosjekter med en sterk fokusering på leksikografiske relasjoner som foruten å forenes i et konkret ordboksverk gradvis har bygget opp en flerfunksjonell islandsk ordboksbase.

Denne artikkelen skal gi en bred oversikt over utviklingen av dette arbeidet og sammensetningen av ordboksbasen, hvordan de enkelte komponenter er relatert til hverandre og hvordan de spiller sammen i utformingen og fremstillingen av en webbasert digital ordbok.

Innledningsvis er det verdt å fremheve to viktige forhold som har vært toneangivende i hele dette arbeidet. Kartleggingen av Orðabók Háskólans beleggmateriale ble i sin tid foretatt uten begrensninger med hensyn til lemmaseleksjon, og den dekket dermed en kolossal mengde „likestilte“ oppslagsord. Disse oppslagsordene og de belegg de bygget på, hadde ennå ikke gjennomgått betydningsdifferensiering og betydningsbeskrivelse. Dermed var det semantiske aspektet ikke umiddelbart i fokus og det ble aktuelt og hensiktsmessig å anlegge andre synsvinkler.

2. Syntagmatiske relasjoner i forgrunnen

Analysen av beleggmateriale i Orðabók Háskólans hovedarkiv (med belegg fra trykte språkkilder) ble i den første fasen for en stor grad rettet mot verbene, med særlig vekt på deres syntagmatiske egenskaper og mangfold i konstruksjoner (jf. *Sýnihefti sagn-orðabókar* 1993). En systematisk kartlegging av verbene og deres omfattende beleggsamling synliggjorde hvor viktig det er å belyse deres opptreden i syntagmer, ordforbindelser og setninger, også med hensyn til deres semantiske egenskaper og betydningsbeskrivelse. Denne erfaringen bidro til konstrueringen av en særskilt konstruksjons- og kombinasjonsordbok under tittelen *Orðastaður*, som først utkom i 1994 og i 2. utgave i 2001, der verb og verbkonstruksjoner inntar en forholdsvis stor plass i lemmalisten.

Arbeidet med ordforbindelser, kollokasjoner og konstruksjoner som ble innledet med *Orðastaður* ble utvidet og komplettert i 1990-årene. I disse årene ble det laget en omfattende database med ordforbindelser fra Orðabók Háskólans beleggsamling (her-

etter *Ordfb*), i alt omkring 135.000 ordforbindelser knyttet til fast 60.000 lemmaer. I 2002 ble *Orðastaður* komplettert og utvidet i form av en ny ordbok, *Orðaheimur*, der oppmerksomheten for en stor del gjaldt idiomene (se Jónsson 2005). Her var lemmalisten todelt, en hovedlemmaliste med alfabetisk ordnete begrepsbetegnelser og en tilleggsliste (register) med samtlige innholdsord i ordforbindelsene der disse var innordnet etter en fast sorteringsregel. Det siste skrittet i dette arbeidet besto i en sammenslåing av *Orðastaður* og *Orðaheimur* med en kombinert lemmaliste og et felles ordforbindelsesregister under overskriften *Stóra orðabókin um íslenska málnotkun* (heretter *Stob*) i 2005 (se Jónsson 2008).

Selv om det syntagmatiske aspektet står sentralt i alle disse prosjektene der hovedvekten legges på kollokasjoner, idiomer og konstruksjoner, er også andre leksikografiske relasjoner inne i bildet.

Orðastaður inneholder en morfologisk komponent som består i en semantisk grovklassifisert liste med sammensetninger til de enkelte lemmaer og som i alt dekker omtrent 100.000 ord. Foruten å belyse og utdype lemmaenes karakter gjenspeiler disse sammensetningene gjerne semantisk nærhet og likheter gjennom sine paradigmatisk relasjoner.

I *Orðastaður* og *Orðaheimur* (og dermed også i *Stob*) blir ledd i ordforbindelser med varierende leksikalsk innhold markert og avgrenset (med hakeparentes):

- (1) *ávarpa* <*samkomuna*> ‘adressere forsamlingen’
- (2) <*hnífurinn*> *bítur* <*vel*> ‘kniven er skarp’

Her blir det gjerne angitt to eller flere typiske representanter for det ledd det dreier seg om, som dermed står i en innbyrdes paradigmatisk relasjon og som regel vil være nært semantisk beslektet:

- (3) *setja* <*fundinn, ráðstefnuna*> ‘åpne møtet/konferansen’
- (4) <*reiðin, heiftin*> *ólgar í* <*honum, henni*> ‘sinnet koker i ham/henne’

En standardisert fremstilling av ordforbindelsene (med hensyn til faktorer som argumentstruktur, valens, kasusstyring og fakultative ledd) åpner veien for et samlet alfabetisert register over samtlige ordforbindelser der innholdsordene fungerer som lemmaer og i mange tilfeller vil dekke en stor mengde ordforbindelser av ulike slag. Dette kommer klarest til syne i *Stob* der materialet i *Orðastaður* og *Orðaheimur* er blitt slått sammen til et felles register. Her kommer de paradigmatiske relasjoner ganske sterkt til uttrykk innenfor de ulike ordklasser, blant adjektiver som kollokatorer til enkelte substantiver (*þrotlaust/hvíldarlaust/sleitulaust erfiði* ‘uavbrudt/... slit’), blant de substantiver adjektivene står sammen med (*þrotlaust erfiði/strit/þrældómur* ‘uavbrudt slit/strev/slaveri’), blant objekter til enkelte verb (*heyra hávaða/skothvell/fótatak* ‘høre bråk/smell/fottrinn’) osv.

I *Orðaheimur* tar beskrivelsen utgangspunkt i ordforbindelsenes semantiske egenskaper i og med at den i hovedsak gjelder ordforbindelser av metaforisk karakter der de enkelte ordene ikke utgjør en logisk tilgang til vedkommende forbindelse. Isteden blir ordforbindelsene dekket av en alfabetisk lemmaliste med drøyt 800 begrepsbetegnelser hvor det innenfor de enkelte artikler blir foretatt en videre gruppering med hensyn til semantiske og kommunikative faktorer. Naturlig nok dreier det seg her i hovedsak om abstrakte begreper, som f.eks. egenskaper, følelser og menneskelig atferd (med overskrifter som *Ákveðni* ‘Besluttsomhet’, *Heiðarleiki/ráðvendni* ‘Hederlighet’, *Ótti* ‘Frykt’, *Reiði* ‘Sinne’, *Ofbeldi* ‘Vold’, *Mistök* ‘Feil’), mens de konkrete begreper, som i høyere grad er knyttet til enkeltord, får mindre oppmerksomhet.

I denne fasen av ordboksarbeidet, frem til utarbeidelsen og utgivelsen av *Stob* i 2005, kommer dermed samspillet mellom de leksikografiske relasjoner til syne i ulike sammenhenger uten at den kan sies å få en klar og selvstendig fokusering. Behovet for å få til en hensiktsmessig tilgangsstruktur ledet frem til etableringen av et altomfattende alfabetisk register. Men overfor fremveksten av den digitale ordboken med en umiddelbar brukertilgang og sine fleksi-

ble søkemuligheter er en slik tilgangsstruktur ikke lenger tilfredsstillende fra et allment brukerperspektiv. Det krever tilpasning til nye forutsetninger som samtidig innebærer et klart behov for en videre utvikling av en sammensatt ordboksmodell som forener de ulike leksikografiske relasjoner.

3. Omstrukturering, utvidet lemmabegrep og entydiggjøring

Den store utfordringen for den ordboksmodellen som her er under utvikling består i å gi de semantiske relasjonene en mer fremtredende plass og å utforme en semantisk komponent i et aktivt samspill med den syntagmatiske komponenten. En slik målsetning er ikke realiserbar innenfor en tradisjonell semasiologisk ordbok med flertydige ettordslemmaer. Hvis også ordforbindelsene skal kunne spille en aktiv rolle som betydningsbærende enheter må de få direkte tilgang til selve lemmalisten som likeverdige enheter sammen med ettordslemmaene. Innenfor en digital ordbok er dette ikke bare mulig, det er faktisk helt naturlig med hensyn til en fleksibel datastrukturering og brukertilgang.

3.1. Utvidet database, gjenspeiling av semantiske relasjoner og semantisk klassifisering

De leksikografiske data *Stob* og *Ordfb* inneholder kan sies å utgjøre et optimalt grunnlag for en semantisk orientert beskrivelse, ut fra den forutsetning at de syntagmatiske og paradigmatiske relasjoner gjenspeiler relasjoner av semantisk karakter. En fylldig samling med kollokasjoner er f.eks. av stor verdi i denne sammenheng. Med tanke på det og med hensyn til klare strukturelle likheter og en parallell fremstilling ble samtlige data fra *Stob* og *Ordfb* slått sammen i en ny og utvidet database. Det innledet en ny fase i opp-

byggingen av den tilsktete ordboksbasen som ble markert ved å gi det nye prosjektet overskriften *Íslenskt orðanet* (heretter forkortet til *Íon*, se Jónsson & Úlfarsdóttir 2011).

De semantiske relasjonene gjelder dels det gjensidige forholdet mellom enkelte ord, i form av synonymer og antonymer. Slike relasjoner er sjelden av absolutt karakter og markeringen av synonymer og antonymer vil i svært mange tilfeller være en vurderingssak. Ved en datastøttet analyse kan de semantiske relasjonene derimot vurderes på en bredere skala som antyder ulik grad av semantisk nærhet. Når det gjelder flerordsenheter må de strukturelle egenskapene også trekkes inn i bildet slik at relasjonene gjelder strukturelt likeverdige enheter (som da vil være gjensidig utskiftbare i tekstsammenheng).

Ved siden av denne markeringen og analysen består kartleggingen av de semantiske relasjonene i en overordnet begrepsklassifisering. En slik klassifisering kan bli foretatt på ulike måter og være mer eller mindre overgripende. I *ÍOn* blir den utført på samme måte som i *Orðaheimur*, slik at lemmaer av samme klasse får en felles overskrift i form av et belysende substantiv som tilhører vedkommende klasse. Disse klassene (begrepsfeltene) er av ulik omfang og overlapper i betydelig grad. I stedet for å foreta en streng hierarkisk systematisering blir begrepsklassifiseringen fremstilt i form av en fri og bevegelig konstellasjon av semantiske grupperinger, noe som i mange tilfeller gjenspeiler faktiske forhold i språkbruken, ikke minst når det gjelder flerordsenheter og ordtilfang som er knyttet til abstrakte begreper.

Det er ikke realistisk å tenke seg at begrepsklassifiseringen vil gi en helhetlig dekning av lemmabestanden til *ÍOn*. Markeringen av gjensidige relasjoner som synonymi og antonymi vil også bare dekke en del av lemmaene. Men et aktivt samspill mellom disse to operasjoner skal kunne gi en verdifull innsikt i de relasjoner som på ulike plan knytter sammen semantisk beslektede ord og andre leksikalske enheter.

3.2. Lemmatisering av betydningsbærende flerordsenheter, entydige lemmaer

Det mest overgripende og tidkrevende arbeidet ved utformingen av *ÍOn* består i å innføre flerordslemmaer og dermed å trekke betydningsbærende flerordsenheter inn i lemmalisten på samme nivå som ettordslemmaene. Dette henger sammen med kravet om semantisk entydige lemmaer, som utgjør en forutsetning for en problemfri analyse og markering av innbyrdes semantiske relasjoner. Til sammen betyr dette en vesentlig endring av ordklassenes karakter og deres innbyrdes størrelsesforhold, i første omgang særlig med hensyn til verbene som øker drastisk i antall.

Fremstillingen av ordforbindelsene i *Stob* gjelder i hovedsak som mønster for fremstillingen av flerordsverb i *ÍOn*, ved markering av valens og fakultative ledd. Lemmatiseringen betyr bl.a. at en stor mengde idiomere fremstår som selvstendige verblemmaer og kan dermed på naturlig måte inngå i innbyrdes relasjoner. Det samme gjelder for verbkonstruksjoner med en mer varierende leksikalsk representasjon. Her kommer forskjellen gjerne direkte til uttrykk i markeringen av fakultative ledd:

- (5) a *brjóta ísinn* 'bryte isen'
 b *brjóta* <*glerið, bollann, spýtuna*>
 'knuse glasset/koppen/planken'

Innførelsen av flerordslemmaer er også av stor betydning for adverbialer som i stor utstrekning får en klar lemmastatus på linje med ettordsadverb:

- (6) a *líklega, að öllum líkindum* ('sannsynligvis')
 b *undanfarið, upp á síðkastið* ('i det siste')

Blant flerordslemmaene foregår entydiggjøringen stort sett på en

enkel måte og kommer til uttrykk gjennom selve formen, der f.eks. ulik semantisk valens gjenspeiler betydningsmessige skilnader:

- (7) a *rjúfa* <*vegginn, þakið*> ‘slå et hull i veggen/taket’
 b *rjúfa* <*sættina; heitið*> ‘bryte fredsavtalen/løftet’

Ellers er flertydighet med homonyme lemmaformer forholdsvis sjelden. I de tilfeller skilles det mellom lemmaene gjennom en homonymindeks sammen med en identifiserende kommentar til betydningen:

- (8) 1 *bera höfuðið hátt reigja höfuðið* (‘bøye hodet bakover’)
 2 *bera höfuðið hátt vera stoltur* (‘være stolt’)

Blant ettordslemmaene er flertydighet og dermed homonymi mye mer utbredt. Især blant substantivene blir lemmaene adskilt gjennom homonymindekser:

- (9) 1 *hóf hófsemi* (‘måtehold’)
 2 *hóf samkvæmi, veisla* (‘selskap, fest’)

I den faktiske språkbruken er adjektivenes betydning ganske flytende og tar i stor utstrekning farge av de substantiver de står sammen med og karakteriserer. For å gjenspeile dette og dermed få til et fungerende samspill med semantisk beslektede lemmaer får adjektivlemmaene gjerne et tilleggselement i form av et differensierende substantiv:

- (10) a *mildur [dómur]* ‘mild dom’
 b *mildur [hljómur, tónn]* ‘mild klang/tone’
 c *mildur [regn]* ‘mild regn’
 d *mildur [veður, tíðarfar, loftslag]*
 ‘mild vær/værforhold/klima’
 e *mildur [vetur]* ‘mild vinter’

Disse lemmaformene er tydeligvis relatert til adjektivene i deres attributive funksjon. Henvisning til personer blir her ikke markert, og predikativer blir fremstilt som flerordslemmaer av typen *vera* + *Adjektiv*:

- (11) *vera mildur* ‘være mild’

Tilsvarende lemmaformer gjelder også for predikative substantiver:

- (12) *vera ljúfmenni* ‘være et elskverdig menneske’
 (13) *vera eyðslukló* ‘være en sløset person’

Dermed får predikativene en selvstendig lemmastatus som trekker frem deres innbyrdes relasjoner og fremhever deres verbale karakter, især når de inngår i synonymrelasjoner med andre typer verblemmaer:

- (14) *vera vellauðugur* ‘være stenrik’
 (15) *vera margfaldur milljónamæringur* ‘være multimiljonær’
 (16) *vita ekki aura sinna tal* ‘ha flust med penger’

Denne omfattende omstruktureringen av den opprinnelige lemmalisten i *ÍOn* åpner ikke bare veien for en bred kartlegging av lemmaenes semantiske relasjoner, den endrer faktisk hele ordboksbasens karakter. Hele den leksikografiske analyse og klassifisering foregår nå på lemmaplanet og dekker en betydelig større mengde lemmaer enn tidligere.

Tilgangen til store islandske tekstsamlinger og korpusmateriale er blitt vesentlig forbedret de siste årene. For oppbyggingen av ordboksbasen har dette vært av stor betydning. Det digitale biblioteket *Tímarit.is* dekker mesteparten av islandske aviser og tidsskrifter helt fra slutten av 1800-tallet til dags dato. Denne store tekstsam-

lingen har vært til stor nytte i ordboksarbeidet. Det gjelder også for den første grammatisk taggede korpusen over islandsk nåtids-språk, *Mörkuð islensk málheild*, som ble ferdiggjort i 2011. Den leksikografiske analysen har i økende grad hentet tilleggsmateriale fra disse kildene for dermed å få til en bredere dekning av det aktuelle analyseobjektet.

3.3. Grammatisk tagging av flerordslemmaer

Det tilhører den leksikografiske tradisjon å angi ordklassebetegnelse ved de enkelte lemnaer. I en statisk omgivelse som i en trykt ordbok spiller denne angivelsen ikke noen aktiv rolle, men den utgjør et klassifiseringselement som kan ha en praktisk betydning i andre sammenhenger, f.eks. som søkefaktor i en digital ordbok. Man kan godt tenke seg at ordklasseskilnaden får en klarere funksjon i leksikografisk sammenheng, men uansett dreier det seg om en forholdsvis grov grammatisk klassifisering.

Lematisering av flerordsenheter byr her på en mye mer detaljert grammatisk markering og dermed en mer finkornet klassifisering av lemnaene. En slik markering vil kunne styrke flerordslemnaenes egenverdi og selvstendighet i betydelig grad. Med hensyn til dette er det blitt foretatt en grammatisk tagging av samtlige flerordslemnaer i ordboksbasen, der faktorer som ordklasse, kasus og bestemthet står i forgrunnen. Det betyr f.eks. at verblemnaene kan klassifiseres ut fra kasusstyring og at verb med objekt i akkusativ kan bli avgrenset som en klasse for seg. Det viser at taggen “so <no-ag>” (verb med substantivisk objekt i akkusativ i bestemt form, f.eks. *lesa <bókina; söguna; textann>* ‘lese boken/historien/teksten’) omfatter omtrent 2500 lemnaer, mens søkestrengen “so <no-ag>*” (der objektet kan etterfølges av et komplement i en eller annen form) matcher drøyt 4000 lemnaer. Derimot dekker tilsvarende tagger med dativobjekt, “so <no-dg>” og “so <no-dg>*”, henholdsvis ca. 400 og 1200 lemnaer. Til

sammenlikning inneholder ordboksbasen omtrent 1500 lemmaer der verbet forutsetter et dativsubjekt (lemmaer som <mér> *leiðist* ‘jeg kjeder meg’, <mér> *líður* <vel> ‘jeg har det bra’).

Et slikt leksikalsk uavhengig klassifiseringsgrunnlag er av stor verdi, både med hensyn til oppbyggingen av databasen og med tanke på praktiske brukerbehov overfor en webbasert digital ordbok. Når det gjelder ordboksbasen, er det bl.a. til direkte nytte ved den semantiske klassifiseringen av lemmaene, for de grammatiske likhetene viser seg i svært mange tilfeller å gjenspeiles i semantiske likheter.

3.4. Semantisk analyse og markering av semantisk nærhet

Markeringen av semantiske relasjoner som synonymi og antonymi vil i mange tilfeller kunne bygge på leksikografens intuisjon og innsikt, ikke minst i de tilfeller der de leksikografiske data ikke bidrar med noen klar vitnesbyrd. Ord og ordforbindelser kan til og med være så sjeldne at de ikke eller kun sporadisk forekommer i store tekstsamlinger. Men ved å vurdere ordenes semantiske egenskaper og relasjoner på bakgrunn av deres forekomst og tilknytninger i tekstsammenheng gis det et bedre grunnlag for den semantiske analyse og beskrivelsen av det innbyrdes forholdet mellom nærstående leksikalske enheter.

Når leksikografens oppmerksomhet gjelder enkeltordenes forekomst i kollokasjoner og konstruksjoner, blir deres opptreden som sideordnete enheter mest ute av fokus. Når fokuseringen overføres til likestilte leksikalske enheter og deres innbyrdes relasjoner slik det skjer ved omstruktureringen av den islandske ordboksbasen, blir forbindelser av denne typen derimot et høyaktuelt objekt i ordboksarbeidet.

I arbeidet med *ÍOn* har parataktiske forbindelser og de paradigmatiske og dermed semantiske relasjoner de avdekker, fått sta-

dig økt betydning, både for den leksikografiske analyse og som et direkte middel til å fortette dekingen av enkelte begrepsfelter og til å trekke inn nye lemmaer. Verdien ligger ikke minst i det forholdet at parataktiske forbindelser karakteriserer jevnt bruken av enkeltord som betydningsbærende flerordsenheter (*frost og kuldi* 'frost og kulde', *styrjaldir og vopnuð átök* 'krig og væpnede konflikter'; *einbeittur og viljasterkur* 'besluttsom og viljesterk', *einbeittur og fastur fyrir* 'besluttsom og bestemt'; *biðjast afsökunar og skammast sín* 'be om unnskyldning og skamme seg', <*aðgerðin*> *heppnast og ber árangur* 'operasjonen lykkes og bærer frukt', *það kólnar og hvessir* 'det blir kaldere og det blåser opp') og de er karakteristiske for alle åpne ordklasser og deres flerordsforbindelser. Ut fra det synspunktet er de av spesiell betydning for analysen av flerordslemmaene, for her er hypotaktiske forbindelser ikke inne i bildet som et alternativt analysegrunnlag på samme måte som i ettordslemmaene.

I mange begrepsfelter i *ÍÖn* er det bygget opp et tettmasket nett av lemmaer som er relatert gjennom parataktiske forbindelser i tekstsammenheng og er dessuten på forskjellig måte knyttet til lemmaer innenfor semantisk nærstående felt. Det gjenspeiler mangfoldigheten og dynamikken i ordboksbasen og knytter sammen beslektet ordforråd over begrepsfelter. Men disse forbindelsene bidrar også med data som inngår i den semantiske analysen og gir markeringen av synonymi og semantisk nærhet en ny dimensjon.

En større samling parataktiske forbindelser kan gi et interessant grunnlag for sammenlikning av lemmaer som gjennom felles partnere er knyttet til hverandre. Her vil antallet felles partnere gi en sterk antydning om semantisk nærhet. Det betyr at sett fra et bestemt lemma kan denne nærheten vises frem i form av en gradert liste over nærstående lemmaer med antatt forskjellig nærhetsgrad. Her må det naturligvis tas hensyn til hvorvidt de aktuelle lemmaer har fått likeverdig deking slik at deres innbyrdes fre-

kvens i parataktiske forbindelser også blir trukket inn i bildet. Men uansett hvordan denne nærheten blir tolket og fremstilt, dreier det seg om en viktig informasjon som belyser hvor nært enkelte lemmaer med lik opptreden er relatert til hverandre.

Som et eksempel kan vi se på to svært nærstående substantiver, *andvaka* og *svefnleysi* 'søvnløshet'. I databasen forekommer *svefnleysi* i 167 ordpar, *andvaka* i 67. Blant disse dreier det seg om 27 felles partnere: *áhyggjur* ('bekymring') og *svefnleysi*, *áhyggjur* og *andvökur*; *kviði* ('angst') og *svefnleysi*, *kviði* og *andvökur*; *svefnleysi* og *ofþreyta* ('utmattelse'), *ofþreyta* og *andvökur*. Mange av disse felles ordene viser seg også å stå høyt oppe på lista over de ord som har de fleste felles partnere med *andvaka* og *svefnleysi*: *vökur* 'våking' (24 med *andvaka*, 27 med *svefnleysi*), *sálarstríð* 'indre strid' (17 med *andvaka*, 16 med *svefnleysi*), *kviði* 'angst' (11 med *andvaka*, 20 med *svefnleysi*), *þunglyndi* 'depresjon' (9 med *andvaka*, 16 med *svefnleysi*). Et slikt vitnesbyrd understreker en sterk semantisk nærhet mellom vedkommende ord. En analyse av ordenes ulike tilknytninger i forbindelser av denne typen vil dessuten i mange tilfeller kunne trekke fram differensierende faktorer og dermed være til nytte ved presisering av ordenes semantiske egenskaper og karakterer.

Foruten å trekke frem semantisk nærhet på denne måten gir direkte søking på parataktiske forbindelser til de enkelte lemmaer et viktig bidrag til oppbyggingen av ordboksbasen i to henseender. For det første vil det oppdages nye aktuelle lemmakandidater som på en enkel måte kan inkluderes i basen og vil berike vedkommende begrepsfelt eller begrepsfelter. For det andre vil utbyttet i svært mange tilfeller bestå i en lang og innholdsrik liste forbindelser med utstrakt semantisk likhet blant partnerne som i sin tur byr på klassifisering og vil være et viktig input i en nærmere semantisk analyse.

4. Webbasert ordbok, semantiske relasjoner i forgrunnen

Hensikten med etableringen av *ÍOn* var dels å utvikle og levere et konkret leksikografisk produkt i form av en webbasert digital ordbok. Denne ordboken er tilgjengelig på websiden <www.ordanet.is> og er fortsatt under utvikling. Med hensyn til det er det ikke grunn til å gå nøye inn på avgrensingen og utformingen av denne ordboken her. Hittil har den vært konsentrert om bestemte datatyper der de semantiske relasjonene står i forgrunnen. Tilgangen til de enkelte begrepsfelter går gjennom søk på hvilket som helst lemma som tilhører feltet, og det samme gjelder for synonymer og antonymer. Ordklassebetegnelse og den grammatiske markeringen av flerordslemmaene spiller også en aktiv rolle ved at brukeren kan velge hvordan lemmaene innenfor enkelte begrepsfelter skal sorteres, enten alfabetisk eller etter den grammatiske taggen. Her ville det være interessant å by på en mer direkte tilgang til markeringen slik at brukeren kan søke på bestemte tagger eller kombinasjoner av disse (se Jónsson & Úlfarsdóttir 2011:97–99).

De parataktiske forbindelsene setter et sterkt preg på mange lemmaer i ordboken slik den ser ut nå der de enkelte partnerne utgjør interaktive lenker. Men det gjenstår å bygge inn informasjon om hva denne datatypen antyder om semantisk (mest) nærstående lemmaer.

Uansett hvordan den webbaserte ordboken kommer til å utvikles vil den fortsatt bære preg av å bygge på en større ordboksbase som kan kompletteres og utvides i ulike retninger og ikke har noen skarpe grenser. Det vil bl.a. gjenspeiles i betydelig ubalanse mellom de enkelte lemmaer der noen viser mange og mange slags tilknytninger til andre lemmaer, mens andre kan bygge sin tilhørighet på en enkel og ensidig relasjon.

Ordboken inneholder nå drøyt 150.000 lemmaer, mens selve ordboksbasen omfatter ca. 260.000 lemmaer. Denne forskjellen

kan sies å gjenspeile ulik lemmaverdi. De lemmaer som hverken inngår i en semantisk relasjon (synonymi, antonymi, begrepsfelt) eller opptrer som ledd i en parataktisk forbindelse, vil til så lenge stå utenfor ordboken slik den blir definert nå. En del av disse lemmaene kan allikevel forekomme i syntagmatiske relasjoner (kollokasjoner og idiomer) i ordboksbasen, mens andre (og da først og fremst sammensetninger) utelukkende forekommer i morfologiske relasjoner. Ordboksbasen inneholder dermed en hel del latente lemmaer som i første omgang har en liten aktiv funksjon, men som vil aktiveres når de inngår i relevante relasjonstyper.

5. Flerfunksjonell ordboksbase

I det forangående har beskrivelsen av den islandske ordboksbasen vært konsentrert om hvordan den er bygget opp omkring leksikografiske relasjoner med de syntagmatiske relasjonene som grunnlag for en omfattende leksikografisk analyse der de semantiske relasjonene står i sentrum. Termen *ordboksbase* i denne sammenheng har en flertydig funksjon. Den kan simpelthen henvise til det samlede datamateriale som ligger til grunn for utarbeidelsen av en konkret ordbok, altså *ÍOns* webordbok i dette tilfellet. Men termen har også en bredere henvisning til en mer allmenn funksjon som utgangsmateriale og grunndatabase for leksikografisk arbeid, både med hensyn til forskning og praktiske ordboksprosjekter (se også Jónsson 2013). Denne allmenne funksjonen har etter hvert fått større vekt i utformingen av arbeidet og fortjener derfor en nærmere presentasjon.

5.1. Lemmaseleksjon og lemmaverdi

Behovet for tilgang til en større leksikografisk database kan aktualiseres i ulike sammenhenger. Med hensyn til islandsk vil det i

praksis primært oppstå ved utarbeidelse av konkrete ordbøker, såvel enspråklige som to- eller flerspråklige. Når det gjelder lemaseleksjon og vurdering av lemmaverdi, kan ordboksbasen bidra med veiledende opplysninger. Her kommer bl.a. antall og variasjoner i kollokasjoner, parataktiske forbindelser og konstruksjoner inn i bildet. Ordboksbasen vil også kunne underlette en begrepsmessig oversikt over lemmabestanden og en spesiell vurdering av lemaseleksjonen i forhold til enkelte begrepsfelter. Den semantiske klassifiseringen av lemmaene i basen kan dessuten ha en kontrollerende funksjon ved betydningsbeskrivelsen, og valg og fremstilling av ekvivalenter for behandlingen av enkelte lemmaer kan kontrolleres mot relaterte lemmaer innenfor samme felt. Det betyr til og med at fremdriften av ordboksbeskrivelsen kan led-sages av begrepsklassifiseringen. Det har allerede vist seg å være av praktisk betydning i arbeidet med den islandsk-skandinaviske ordboken *Islex*, der en grov semantisk markering av de islandske lemmaene, som for en stor del er hentet fra ordboksbasen, gjorde det mulig å arbeide med semantisk sammenhengende ordtilfang i hver enkel redigeringsfase (se Úlfarsdóttir 2013).

5.2. Semantisk nærhet

Markering av semantisk nærhet og det datamaterialet den bygger på kan ha stor verdi for ordboksarbeid i ulike sammenhenger. Foruten å utdype en tradisjonell synonymbeskrivelse styrker den grunnlaget for lemaseleksjon og gir viktig veiledning for valg av ekvivalenter bygget på konkrete tall. Her er det verdt å understreke at opplysninger om leksikalsk frekvens er av spesiell betydning når de gjelder innbyrdes sammenhengende ord eller andre leksikalske enheter, men slike opplysninger har lenge vært sparsomme i leksikografisk arbeid.

En ordboksbase som prioriterer en dyptgående og omfattende semantisk klassifisering av ordforrådet, er spesielt verdifull når det

gjelder å finne motsvarigheter på målspråkssiden til kildepråklige elementer i en to- eller flerspråklig ordbok. Her må valget utelukkende styres av innholdet til de aktuelle ekvivalenter, og for å kunne treffe et veloverveid valg må leksikografen ha oversikt over et bredt leksikalsk spektrum. Ut fra en slik oversikt er det også mulig å foreta en mer generell kontrastiv analyse av forholdet mellom to språk der motsvarigheter og likheter inngår i en større kontrastiv sammenheng (jf. Hannesdóttir & Jónsson 2001, Hannesdóttir & Tingsell 2011).

5.3. Grammatisk tagging

Grammatisk tagging av betydningsbærende flerordsenheter gir et sorterings- og klassifiseringsgrunnlag som i flere henseender vil være til nytte i praktisk ordboksarbeid, og kan på ulike måter kobles til leksikalske og semantiske elementer og deres opptreden. Den kan også gi det grammatiske aspektet sterkere fokusering i utformingen av konkrete ordbøker, både i enspråklig og flerspråklig sammenheng.

5.4. Grunnlag for leksikografisk arbeid

Alle disse elementene har en viktig funksjon i den islandske ordboksbasen som her har vært skildret og diskutert. En ordboksbase av denne type skal dermed kunne utgjøre et viktig grunnlag for leksikografisk arbeid og konstrueringen av konkrete ordbøker med ulike genuine formål. Det beror bl.a. på at den gjelder et stadium i ordboksarbeidet som kan sies å stå midt i mellom et grovanalysert datamateriale, f.eks. i form av tagget korpus, og en konkret ordbok med et klart genuint formål. Utarbeidelsen av en relasjonsbasert ordbok som en integrert del av selve ordboksbasen understreker og konkretiserer den rolle ordboksbasen tillegges i denne sammenheng.

6. Sammenfatning

Flere omstendigheter bidro til at islandsk ikke på samme måte som de skandinaviske språk fikk et stort nasjonalt ordboksverk i trykt form. Allikevel ga islandsk ordboksarbeid omkring og etter midten av 1900-tallet, der hensikten var å utarbeide et stort ordboksverk over islandsk helt fra boktrykkingens inntog midt på 1500-tallet til samtiden, et rikelig utbytte, bl.a. i form av en verdifull beleggssamling. Dette materialet ble tilgjengelig på en ny måte da datamaskinell registrering og analyse av ordboksmateriale ble aktuell sent på 1900-tallet.

Denne beleggssamlingen utgjør en viktig kilde i arbeidet med noen islandske ordbøker og leksikografiske databaser som ble ferdiggjort omkring århundreskiftet 2000. Her gjelder oppmerksomheten især ordforbindelser og dermed analyse av syntagmatiske relasjoner. Under det arbeidet kom forholdet til de semantiske relasjonene stadig mer inn i bildet. I ordboksprosjektet *Íslenskt orðanet*, som for en stor del bygger på disse ordbøkene og deres materiale, får de semantiske relasjonene en sterk fokusering, med analyse av synonymi og semantisk nærhet og en omfattende og detaljert begrepsklassifisering. Her har innførelsen av flerordslemmer og deres grammatiske tagging en avgjørende betydning for en effektiv og hensiktsmessig analyse.

Gjennom dette arbeidet er det etter hvert bygget opp en flerfunksjonell ordboksbase som både skal kunne støtte utarbeidelsen av nye islandske ordbøker og ha en generell betydning for forskning omkring islandsk ordforråd og islandsk språkbruk. En webbasert digital ordbok utgjør en integrert komponent i denne ordboksbasen og er med på å fremheve den betydning som de leksikografiske relasjoner og deres innbyrdes samspill tildeles.

Litteratur

Ordbøger og databaser

ÍOn = *Íslenskt orðanet*.

Íslensk-dönsk orðabók (1920–1924): Sigfús Blöndal (red.). Reykjavík.

Íslenskt orðanet. <www.ordanet.is> (april 2014).

Íslenzk orðabók handa skólum og almenningi (1963): Árni Böðvarsson (red.). Reykjavík: Bókaútgáfa Menningarsjóðs. [2. utgave 1983, 3. utgave 2007; nettutgave på <<http://snara.is>>].

Mörkuð íslensk málheild. <mim.arnastofnun.is> [Tagget korpus med islandske tekster fra perioden 2000–2010.] (april 2014).

Ordfb = *Skrá um orðasambönd*. <arnastofnun.is/page/skra_um_ordasambond> [Fraseologisk register til Orðabók Háskólans.] (april 2014).

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 JPV-útgáfa 2001.]

Stob = *Stóra orðabókin um íslenska málnotkun*.

Stóra orðabókin um íslenska málnotkun (2005) = Jón Hilmar Jónsson: *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV-útgáfa. <<http://snara.is>> (april 2014).

Sýnihefti sagnorðabókar (1993). Ásta Svavarsdóttir, Guðrún Kvaran, Jón Hilmar Jónsson & Kristín Bjarnadóttir (red.). Reykjavík: Orðabók Háskólans.

Tímarit.is. <<http://timarit.is/>> [Digitalt bibliotek med islandske avistekster.] (april 2014).

Annen litteratur

- Battista, Simonetta & Ellert Thor Johannsson (2014): Ordbog over det norrøne prosasprog 2004–2014 – fra trykt udgave til net-version. I: *Leda-Nyt* 57, 6–23.
- Hannesdóttir, Anna Helga & Jón Hilmar Jónsson (2001): Að hafa í sig og á. Ísländsk fraseologi i ett isländskt-svenskt perspektiv. I: *LexicoNordica* 8, 67–91.
- Hannesdóttir, Anna Helga & Sofia Tingsell (2011): Svensk fraseologi för islänningar – om den isländsk-svenska ordboken i ISLEX och en onomasiologisk svensk resurs: Frاسبanken. I: *LexicoNordica* 18, 61–86.
- Jónsson, Jón Hilmar (2005): Orðaheimur – en fraseologisk begrepsordbok. I: Ruth Vatvedt Fjeld & Dagfinn Worren (red.): *Nordiska studiar i leksikografi 7. Rapport frá Konferanse om leksikografi i Norden Volda 20.–24. mai 2003*. Oslo, 228–236.
- Jónsson, Jón Hilmar (2008): Lexicographic 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 (2013): Að byggja undir íslensk-erlenda orðabók. Forgreining og orðabókarefni. I: *Orð og tunga* 15, 1–22.
- Jónsson, Jón Hilmar & Þórdís Úlfarsdóttir (2011): Íslenskt orðanet: Ett skritt mot en allmennspråklig onomasiologisk ordbok. I: *LexicoNordica* 18, 87–109.
- Orðabók Háskólans. Leksikografisk institutt ved Islands universitet* (1995): Ásta Svavarsdóttir (red.). Reykjavík: Orðabók Háskólans.
- Úlfarsdóttir, Þórdís (2013): ISLEX – norræn margmála orðabók. I: *Orð og tunga* 15, 41–71.

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

