

# Íslenskt orðanet: Tekstbasert kartlegging og presentasjon av leksikalske relasjonar

*Jón Hilmar Jónsson*

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

## 1. Innledning

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

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

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

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

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

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

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

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

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

## 2. Tre beslektede ordbøker

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

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

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

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

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

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

### 3. Overføring til digital form

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

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

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

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

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

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

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

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

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

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

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

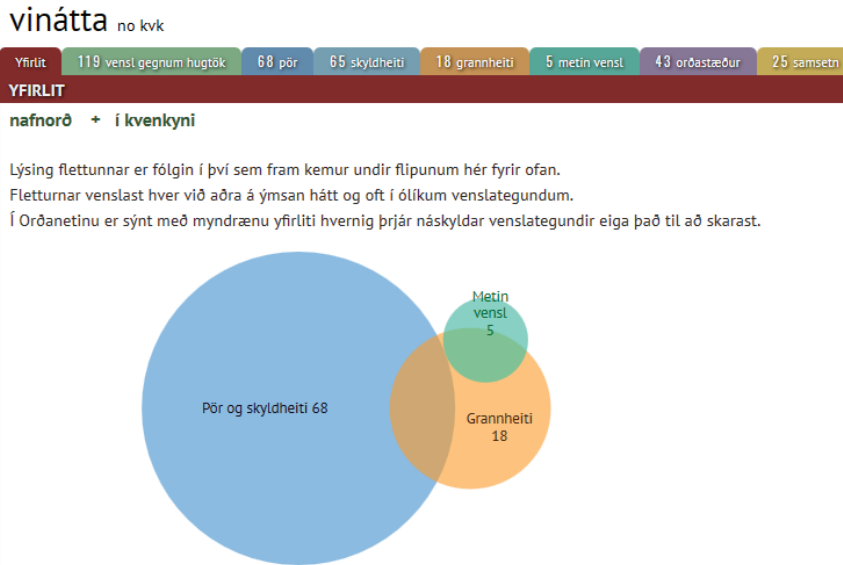
På bakgrunn av disse overveielserne ble det en sentral oppgave i den første fasen å utvide begrepsklassifiseringen, både i form av en bredere begrepsdekning og en bredere inkludering av lemmaene, også med hensyn til enkeltordene. Et viktig ledd i denne oppgaven besto i å forsyne flerordslemmaene med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

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

#### 4. Lemmaenes relasjoner som adskilte informasjonstyper

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

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



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

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

#### 4.1. Tekstbaserte relasjonar

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

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

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

#### 4.1.1. Parataktiske forbindelser

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

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

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

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



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

#### 4.1.2. Leksikalsk samhörighet

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

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

#### Skyldheiti 156

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

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

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

#### 4.1.3. Kollokasjoner

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

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

#### 4.2. Skjønnsbaserte relasjoner

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

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

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

### 4.3. Tekst- og skjønnsbaserte relasjoner

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

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

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

## Grannheiti 25

RÖÐUN: - Fjöldi merkingarvensta +

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

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

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

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

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

#### 4.4. Morfologiske relasjoner i sammensetninger

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

#### 4.5. Syntagmatiske relasjoner i flerordslemmaer

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

#### 4.6. Begrepsmessige relasjoner og begrepsklassifisering

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

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

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

**VINÁTTA (119)**

**Flettur undir hugtakinu**

- Óháð orðflokkum 119 +

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

**Venslaðar flettur í gegnum pör**

- Óháð orðflokkum 275 +

aóðáandi  
aóðáun  
aóstandandi1  
aóstandandi2  
afvopnun  
alúð  
andstæðingur  
áhangandi  
áhugamaður1

**Tengd HUGTÖK**

- Raðaó eftir vægi 137 +

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

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

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

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

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

## 5. Sammendrag

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

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

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

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

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

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

## Litteratur

### Ordbøker og digitale ressurser

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

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

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

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

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

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

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

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

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



## Annen litteratur

- Fjeld, Ruth Vatvedt & Lars S. Vikør (2008): *Ord og ordbøker*. Kristiansand: Høyskoleforlaget.
- Jón Hilmar Jónsson (2014): Oppbyggingen av en relasjonsbasert islandsk ordboksbase. I: *LexicoNordica* 21, 39–59.
- Svensén, Bo (2004): *Handbok i lexikografi*. Stockholm: Nordtedts Akademiska Förlag.
- Trap-Jensen, Lars (2014): Leksikografisk tradition og fornyelse: tre revolusjoner på 100 år? I: Ruth Vatvedt Fjeld & Marit Hovdenak (red.), *Nordiske studier i leksikografi* 12. Rapport fra Konferanse om leksikografi i Norden, Oslo 13.–16. august 2013, 42–68.
- Wörterbücher (1989–1991) = *Wörterbücher/Dictionaries/Dictionnaires*. Edited by Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta. HSK 5. Berlin/New York: Walter de Gruyter.

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