

Omdannelsen af en flersproget til en monolingval ordbog

Halldóra Jónsdóttir & Þórdís Úlfarsdóttir

This article describes the making of *Íslensk nútímamálsorðabók* (*A Dictionary of Contemporary Icelandic*), a dictionary based on the Icelandic part of ISLEX. The multilingual Scandinavian dictionary ISLEX has for several years been accessible on the web, free of charge (islex.is, islex.dk, islex.no, etc.). The dictionary is medium-sized, having Icelandic as a source language and six Scandinavian target languages. After the completion of ISLEX it became clear that its lexicographic material was well suited for other bilingual dictionaries. Furthermore, it seemed practical to develop an Icelandic monolingual dictionary on the basis of ISLEX. The project was created and named *Íslensk nútímamálsorðabók*.

When ISLEX' database was copied in preparation for the new monolingual dictionary, a large part of the contents had already been completed. What was mainly missing were good quality definitions, as ISLEX only contained rudimentary Icelandic definitions intended as an aid to the editors of the target languages. For the last few years the editorial work of the new dictionary has mainly consisted of revising and rewriting the Icelandic definitions.

1. Indledning

I denne artikel bliver der gjort rede for tilblivelsen af *Íslensk nútímamálsorðabók* (*Ordbog over moderne islandsk*), en ordbog der er baseret på den islandske del af den skandinaviske webordbog ISLEX.

Ordbogen er den første islandske monolingvale ordbog i over halvtreds år der ikke er baseret på en tidligere udgave af en eksisterende trykt ordbog. Den er desuden den anden monolingvale islandske ordbog der er udkommet overhovedet. Hovedvægten i det leksikografiske arbejde lægges på at beskrive islandsk som det tales og skrives omkring og efter året 2000.

Da arbejdet med *Íslensk nútímamálsorðabók* påbegyndtes tidligt i 2014, forelå der omfattende materiale fra ISLEX: en lemmaliste på 50 tusinde ord, kollokationer, faste ordforbindelser, sprogbrugseksempler m.m. Der var oplysninger om verbers og præpositioners kasusstyring samt interne links inde i selve ordbogen. Der fandtes endvidere en del illustrationer og opslagsordenes udtale var til stede i form af lydfile. Desuden var opslagsordene blevet delt op efter betydning. Den vigtigste og mest tidskrævende del var dog ikke afsluttet, men

det var at skrive islandske ordforklaringer, én eller flere for hvert lemma. De definitioner som forelå i ISLEX var oprindelig tænkt som en hjælpetekst til de skandinaviske redaktører som arbejdede med målsprogene, og teksten var ikke altid af god kvalitet da den ikke var tænkt til offentliggørelse.

Ordforklaringerne skulle revideres eller skrives om hvor det var nødvendigt. Da denne proces var kommet godt i gang blev ordbogen lanceret i 2016 som ”work in progress”, men arbejdet fortsatte indtil værket fik sin endelige form. Denne arbejdsmetode blev valgt gå grund af at webben har ændret folks indstilling til informationsformidling, og på webben finder man ofte materiale som ikke er hundrede procent færdigt. Ikke desto mindre ligger der nyttige oplysninger som er oplagt at gøre tilgængelige til trods for eventuelle mangler.

I de tilfælde hvor en bruger finder en fejl i ordbogen, så har han/hun en mulighed for at indsende en melding, og her kan man tale om at der sker en slags ”crowdsourcing” efter offentliggørelsen. *Íslensk nútímamálsorðabók* er gratis tilgængelig på webben (www.islenskordabok.is).

2. ISLEX-projektet

Den nye monolingvale ordbog henter således sit grundmateriale fra ISLEX, en webordbog med islandsk som kildeprog og seks nordiske målsprog: dansk, bokmål, nynorsk, svensk, færøsk og finsk. ISLEX-projektet blev grundlagt i 2005 som et samarbejdsprojekt mellem seks institutioner:

- Árni Magnússon-instituttet for islandske studier, Reykjavík
- Det Danske Sprog- og Litteraturselskab, København
- Institutt for lingvistiske, litterære og estetiske studier, Universitetet i Bergen
- Institutionen för svenska språket, Göteborgs universitet
- Fróðskaparsetur Føroya, Tórshavn
- Helsingfors universitet

Et stort antal islandske og nordiske medarbejdere bidrog til ISLEX og var med til at forme ordbogens indhold. ISLEX var det første islandske ordbogsværk der udelukkende var designet til digital udgivelse, og var samtidig den eneste eksisterende islandske ordbog som forbandt mange sprog. ISLEX blev åbnet i slutningen af 2011 med målsprogene dansk, bokmål, nynorsk og svensk; senere blev færøsk (2015) og endelig finsk (2018) tilføjet. Figur 1 viser et eksempel fra ordbogen.



Figur 1: Et eksempel fra ISLEX-ordbogen, her med svensk som målsprog.

ISLEX' infrastruktur tager udgangspunkt i en beskrivelse af islandsk som kilde-sprog, og under ordbogsarbejdet blev der indsamlet et righoldigt materiale om islandsk. Det blev hurtigt klart at denne udførlige beskrivelse af moderne islandsk var oplagt at anvende til flere projekter, og ordbogen var struktureret således at det var nemt at omforme materialet til en etsproget, tosproget eller flersproget ordbog.

3. Udformning af den nye ordbog

Det behøver ikke at være en teknisk indviklet proces at omdanne en flersproget ordbog til en etsproget med udgangspunkt i kildesproget, men det forudsætter selvfølgelig en del programmeringsarbejde. Processen sker i nogle trin. Først må databasen kopieres for at danne en ny database for den nye ordbog. I dette tilfælde måtte man smide målsprogene ud samt tilknyttet materiale, så det eneste der var tilbage var det islandske kildesprogsmateriale. Figur 2 illustrerer en ordbogsartikel med det islandske materiale.

Ordbogsredaktionen kunne nu begynde sit arbejde. De midlertidige ordfor-klaringer fra ISLEX skulle revideres og/eller skiftes ud for nye, og nogle gange måtte der tilføjes flere betydninger. Der blev desuden tilføjet nyt materiale, f.eks. flere sprogbrugseksempler og ikke mindst flere lemmaer.

grein subst f

🗣️ **uttale**
→ **BØYING**

1
(trjágrein)
langur, trékenndur útvöxtur á tré, trjágrein

2
(ritsmíð)
ritsmíð, lítill ritgerð, einkum í blaði eða tímariti, sem fjallar um tiltekið efni
grein um bókmenntir

3
(námsgrein)
námsgrein, námsgrein og fræði um tiltekið efni, námsgrein, fag
hún var hæst í þremur greinum á lokaprófinu

4
(lagagrein o.þ.h.)
eitt atriði í lista með samfelldum texta, t.d. í reglugerðum og lögum (lagagrein)
20. grein fjallar um tollamál

gera grein fyrir <málinu>
útskýra <málið>

gera sér grein fyrir <ástandinu>
skilja og þekkja ástandið

Figur 2: En ordbogsartikkel hvor målsprogene er forsvundet og kun det islandske indhold vises.

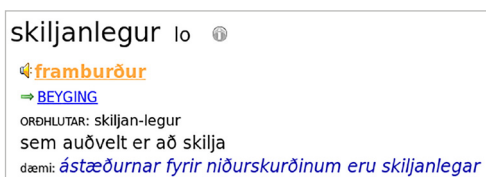
Under redigeringsprocessen er der undervejs blevet afsløret lakuner og mangler i det oprindelige værk, ISLEX-ordbogen, som der måtte rettes op på. Redigeringen har på den måde også haft en tilbagevirkende kraft på ISLEX, som er blevet opdateret samtidig med arbejdet med den monolingvale ordbog.

4. Materialet i den nye ordbog

I dette afsnit vil vi kort gennemgå den nye ordbogs vigtigste komponenter, i makrostrukturen såvel som i mikrostrukturen.

4.1. Lemmaerne

Ordbogens lemmaliste består af 50.000 ord og afspejler sammenhængen i det islandske sprog. Der er således en del norrøne ord, ord fra ældre sprog og poetisk sprog inkluderet, og der findes endvidere fagsprog til en vis grad, f.eks. juridisk sprog, geologi og kemi. Figur 3 viser et tilfældigt lemma fra ordbogen.



Figur 3: Et eksempel fra *Íslensk nútímamálsorðabók*.

Da redigeringsarbejdet indledtes blev det hurtigt klart at der måtte tilføjes nyt ordforråd. De nye ord er især kommet ind ad følgende veje: 1) fra omgivelserne og redaktørernes løbende diskussioner, 2) ved excerpering af tekster i tilfælde af kendte lakuner i ordforrådet, f.eks. i fag som geologi og medicin, og 3) sidst men ikke mindst arbejdes der med frekvenslister fra korpuser. Med alle disse faktorer er lemmalisten vokset med 4.500 nye ord. I de tilfælde hvor der har manglet fraser så er de også blevet tilføjet løbende.

4.2. Øvrige indhold

I de følgende afsnit diskuteres nærmere ordbogens indhold og dens vigtigste bestanddele. Ligesom i de fleste digitale ordbøger består mikrostrukturen af *Íslensk nútímamálsorðabók* af følgende komponenter:

- Islandske definitioner
- Kollokationer og ordforbindelser
- Sprogbrugseksempler
- Orddeling i sammensatte ord
- Bøjning
- Oplysninger om kasusstyring af verber og præpositioner
- Illustrationer
- Udtale som lydfile
- Semantisk inddeling

4.2.1. Definitioner

Definitioner er en grundlæggende komponent i en monolingval ordbog og de skal være af god kvalitet. *Íslensk nútímamálsorðabók* indeholder 62.400 definitioner. Opslagsordene er 50.000, men forskellen beror på at mange ord har flere end én betydning, foruden at de faste udtryk også bliver forklaret. De ord

som har flest betydninger, og dermed de fleste forklaringer, er (efter frekvens) verberne *taka* ('tage'), *koma* ('komme') og *fara* ('gå'), men de har hver for sig dusinvis af betydninger. De navneord som har flest betydninger er *brot* ('brud'), *efni* ('emne'), *hlaup* ('løb'), *mál* ('mål') og *mark* ('mark').

Definitionerne har sin oprindelse i den islandske del af ISLEX-projektet, hvor otte forskellige redaktører havde arbejdet på indholdet og af den grund var de stilmæssigt ret forskellige. Eftersom de først og fremmest blev tænkt som en hjælpetekst for oversætterne var de af meget forskellig kvalitet da de ikke skulle være synlige for brugerne. Der fandtes således definitioner til alle opslagsordene, men det viste sig at mange af disse definitioner var ubrugbare. Opslagsordene måtte gennemgås kritisk, revideres, koordineres og omformuleres inden de blev offentliggjort. Endelig fik alle faste udtryk en forklaring da betydningen tit er ugenomsigtig og ikke fremgår af ordlyden.

I *Íslensk nútímamálsorðabók* bliver der gået ud fra at brugeren er fortrolig med det grundlæggende ordforråd i islandsk eftersom ordene forklares på islandsk. Ordbogens redaktører har haft følgende målgrupper i tankerne som f.eks. islandske skoleelever, udlændinge som lærer islandsk som fremmedsprog, såvel som almene brugere.

4.2.2. Kollokationer og ordforbindelser

En af ordbogens stærkeste sider er det store antal kollokationer og faste udtryk (14.500), herunder 5.000 fraser med verber. Kollokationernes betydning er oftest transparent: *rød i hovedet*, *en udslukt vulkan*, *gå i kloster*.

Betydningen af andre typer faste ordforbindelser kan være helt utransparente da kernen ofte har lidt eller ingenting at gøre med selve ordlyden: *vaxa úr grasi* ('vokse op'), *láta greiðar sópa* ('rage til sig'). Der kan være tale om en metafor og i nogle tilfælde idiomer.

Naveord som indeholder de fleste ordforbindelser er (efter frekvens): *hönd* ('hånd'), *fötur* ('fod'), *mál* ('sag, sprog'), *orð* ('ord'), *auga* ('øje'), *maður* ('mand'), *vegur* ('vej'), *dagur* ('dag'), *líf* ('liv') og *ráð* ('råd'). Sammenlagt findes der i ordbogen omkring 10.000 forskellige slags ordforbindelser.

4.2.3. Sprogbrugseksempler

Ordbogen indeholder mange sprogbrugseksempler og deres rolle er at vise opslagsordet i en typisk sætningssammenhæng og desuden at understrege ordets forskellige betydninger, i de tilfælde hvor ordet har flere end en betydning. Der er ikke tale om autentiske citater med kildehenvisninger, som man kender fra mange andre ordbøger, sprogbrugseksemplerne er derimod specielt konstruerede for ordbogen. En mere udførlig redegørelse for sprogbrugseksemplerne finder man hos Thórdís Úlfarsdóttir (2014).

4.2.4. Orddeling i sammensatte ord

Ligesom de andre nordiske sprog af germanisk oprindelse er islandsk rigt på sammensatte ord. Når opslagsordene er sammensatte vises ordet igen med mindre typer med en bindestreg mellem leddene. F.eks. *samnings-tímabil* ('kontrakt-periode'), *ffallgöngu-maður* ('bjerg-vandrer'), *bak-herbergi* ('bag-lokale'). Dette bliver gjort for at illustrere ordenes led, foruden at det kan fungere som hjælp til rigtig orddeling mellem tekstlinjer.

4.2.5. Bøjning

Opslagsordenes bøjning fremstår med links til bøjningsparadigmerne i BÍN (*Beygingarlýsing íslensks nútímamáls*), en database med bøjninger af flere hundrede tusinde islandske ord. Når man trykker på linken *bøjning* åbnes ordets bøjningsparadigme i sin helhed. Alle ord som bøjes har links: navneord, tillægsord, udsagnsord og stedord. For en del ords vedkommende vises desuden bøjninger i forkortet form i selve ordbogsartiklen.

4.2.6. Kasusstyring

Da islandsk er et bøjningssprog er det i visse tilfælde nødvendigt at give oplysninger om kasusstyring (valens). Dette gælder selvfølgelig for verber og præpositioner, som kan tage med sig forskellige kasus alt efter funktion. Som et eksempel kan nævnes verbet *svara* ('svare') som tager to kasus i dativ: *svara henni* (dat.) *engu* (dat.) ('svare hende ingenting', 'give hende ikke noget svar'). Præpositioner kan tage tre forskellige kasus, akkusativ, dativ eller genitiv, f.eks. *um* + akk. ('om'), *frá* + dat. ('fra') og *til* + gen. ('til'). Disse oplysninger er af stor hjælp for sprogbrugerne, ikke mindst de som lærer islandsk sem fremmedsprog.

4.2.7. Illustrationer

Illustrationer har længe eksisteret i ordbøger, især for unge læsere, og *Ordbog over moderne islandsk* indeholder 3.200 illustrationer. I nogle tilfælde kan et billede forklare et ords indhold bedre end man kan gøre med ord (f.eks. *egern*, *saks*, *tandhjúl*, *trádrulle* og *tøjklemme*). Nogle gange kan det også være praktisk at have billeder for at afdække et ords mange betydninger, jfr. ordet *keila* som bl.a. betyder fisken 'brosme' og 'trafikkegle'. Billederne er af forskellig oprindelse, men ca. 700 stammer fra illustratoren Jón Baldur Hlíðberg, hovedsagelig af naturfænomer: planter, fugle, fisk og flere dyr. Desuden findes der i ordbogen en del fotografier og tegninger.

4.2.8. Udtale

Opslagsordenes udtale på lydfiler er en af de komponenter der blev taget i arv fra ISLEX. Det er muligt at lytte til udtalen ved at klikke på et ikon neden for selve lemmaet. Da læseren indlæste ordene i alfabetisk orden viste der sig en tendens til at oplæsningen blev monoton, noget som ikke var ønskværdigt når det drejede sig om udtalen af enkelte ord. Af den grund blev ordene læst i en tilfældig orden og dermed forsvandt monotonien. Foruden de 50.000 opslagsord blev der indlæst godt 700 faste udtryk som forekommer i ordbogsartiklerne.

4.2.9. Semantisk inddeling

Ordbogens samtlige ordforråd er blevet inddelt semantisk og redaktørerne har arbejdet efter den inddeling fremfor den traditionelle alfabetiske orden. Store semantiske grupper er f.eks. *beskæftigelse, indbo, dyr, tøj, vind, vej, geologi* (hovedsagelig navneord); *alder, tilfredshed, tilstand, uduelighed, sundhed, farve, form* (hovedsagelig tillægsord); og *ændring, død, ophold, bevægelse, tale* (hovedsagelig udsagnsord). De fleste lemmaer tilhører flere end et semantisk felt. Den semantiske inddeling er tænkt som et redaktionelt værktøj og er af den grund ikke synlig for brugeren.

Grupperne er af forskellig størrelse, hvor de største indeholder over 1.000 ord, men de mindste kun nogle dusin. Denne arbejdsmetode var en nyhed i islandsk ordbogsarbejde, og har vist sig at være af stort gavn for redaktørerne, bl.a. for at sikre bedre overensstemmelse i behandlingen af semantisk beslægtede ord, hvor der målrettet kan gøres rede for sproglige nuancer.

5. Brugen af korpuser

En af sprogteknologiens vigtige opgaver er dannelsen af tekstkorpuser, en udvikling som ordbogsfolk har høstet frugterne af. Nu er det blevet normen at ordforrådet m.m. i nye ordbøger er baseret på korpuser, og islandske ordbøger er her ingen undtagelse. Man kan endda sige at det er utænkeligt at indlede et ordbogsprojekt uden at anvende korpuser.

Oplysninger som man finder i korpuser er, bl.a.:

- Ordnes frekvens
- Ordnes alder, eller første registrering
- Betydningsbestemmelse, én eller flere

- Sprogligt register (fx slang, formelt)
- Fraser og forskellige faste udtryk

I de seneste to årtier er der sket en hastig udvikling af islandske tekstkorporer, hvoraf de vigtigste er: *Íslenskt textasafn*, *tímarit.is* og *Risamálheildin* (2018). Ovennævnte tre korporer er indbyrdes forskellige. Den førstnævnte indeholder mangeartede teksttyper, mens de to sidstnævnte hovedsagelig består af store mængder nyheds- og dagbladsstof. Alle disse korporer er til gavn for leksikograferne for at udvide deres forståelse for leksikalske enheder.¹

6. Modtagelser og opdateringer

Arbejdet med Ordbog over moderne islandsk påbegyndtes i januar 2014 og der blev åbnet for adgang til ordbogen sent i 2016. På dette tidspunkt var dog visse elementer stadig under bearbejdning, og ordbogens første udgave blev ikke klar før end i marts 2019.

Íslensk nútímamálsorðabók har sin egen hjemmeside (www.islenskordabok.is), og er samtidig tilgængelig på portalen www.málið.is (jfr. Steinþór Steingrímsson et al. 2018a) sammen med andre opslagsværker om det islandske sprog.

Ordbogens modtagelser har været over forventning og i 2018 var brugerne 130.000, som må anses for meget godt hvis man tager hensyn til det antal mennesker som taler islandsk. Brugergrænsefladen har nu fået nyt design som blev åbnet i slutningen af 2019.

I begyndelsen var der 50.000 opslagsord i *Íslensk nútímamálsorðabók*, ligesom i ISLEX, men efter at første version af ordbogen blev åbnet var redaktørerne blevet klare over at der skulle tilføjes en del nye ord. Det nye ordforråd stammer hovedsagelig fra det nyeste islandske tekstkorpor, *Risamálheildin* (2018), og i 2020 bliver der tilføjet 4.500 nye lemmaer som skal redigeres. Det er klart at den omfattende brug af ordbogen giver et fingerpeg om at der er stor efterspørgsel efter et sprogværktøj, der opdateres løbende og er gratis tilgængeligt på webben.

¹ Nærmere oplysninger om *Risamálheildin* er i Steinþór Steingrímsson et al. (2018b).

Litteratur

Ordbøger

- Ordbog over moderne islandsk (Íslensk nútímamálsorðabók)*. Halldóra Jónsdóttir & Þórdís Úlfarsdóttir (red.). Árni Magnússon-instituttet for islandske studier. <www.islenskordabok.is> (november 2019).
- ISLEX-ordbogen*. Þórdís Úlfarsdóttir (red.). Árni Magnússon-instituttet for islandske studier. <www.islex.is> (november 2019).

Anden litteratur

- BÍN = Kristín Bjarnadóttir (red.): *Beygingarlýsing íslensks nútímamáls*. Árni Magnússon-instituttet for islandske studier. <www.bin.arnastofnun.is> (november 2019).
- Risamálheildin* (2018). Árni Magnússon-instituttet for islandske studier. <malheildir.arnastofnun.is> (november 2019).
- Steinþór Steingrímsson, Sigrún Helgadóttir & Eiríkur Rögnvaldsson (2018b): An Icelandic Gigaword Corpus. I: *Nordiske Studier i Leksikografi* 14. Rapport fra 14. Konference om Leksikografi i Norden (NFL) i Reykjavík 30.5-2.6.2018: 246-254.
- Steinþór Steingrímsson, Ari Páll Kristinsson & Halldóra Jónsdóttir (2018a): Málið.is. I: *Nordiske Studier i Leksikografi* 14. Rapport fra 14. Konference om Leksikografi i Norden (NFL) i Reykjavík 30.5-2.6.2018: 255-261.
- Þórdís Úlfarsdóttir (2014): ISLEX—a Multilingual Web Dictionary. I: *Proceedings from LREC 2014, Ninth International Conference on Language Resources and Evaluation*. Reykjavík 26-31 May 2014.
- Timarit.is*. Tímaritavefur Landsbókasafns Íslands – Háskólabókasafns. <www.timarit.is> (november 2019).
- Þórdís Úlfarsdóttir (2013): ISLEX - norræn margmála orðabók. I: Orð og tunga 15. Reykjavík: Árni Magnússon-instituttet for islandske studier.
- Þórdís Úlfarsdóttir (red.): *Íslenskt textasafn*. Árni Magnússon-instituttet for islandske studier. <corpus.arnastofnun.is> (november 2019).

Halldóra Jónsdóttir
ordbogsredaktør
Árni Magnússon-instituttet
for islandske studier
Laugavegur 13
IS-101 Reykjavík
halldora.jonsdottir@arnastofnun.is

Þórdís Úlfarsdóttir
ordbogsredaktør
Árni Magnússon-instituttet
for islandske studier
Laugavegur 13
IS-101 Reykjavík
thordis.ulfarsdottir@arnastofnun.is