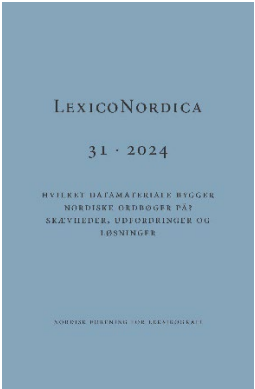


# LexicoNordica

Titel:	Ordbog over moderne islandsk – udvikling og tilføjelser	
Forfatter:	Ellert Þór Jóhannsson & Þórdís Úlfarsdóttir	
Kilde:	LexicoNordica 31, 2024, s. 127-149	
URL:	<a href="https://tidsskrift.dk/lexn/issue/archive">https://tidsskrift.dk/lexn/issue/archive</a>	

© 2024 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.

# Ordbog over moderne islandsk – udvikling og tilføjelser

*Ellert Þór Jóhannsson & Þórdís Úlfarsdóttir*

This article examines the lexicographic material found in *Dictionary of Contemporary Icelandic* (OMI). We focus on four processes for acquiring new material: 1) corpus data, 2) selective excerption, 3) user feedback, and 4) editorial additions. The result shows that from the time of inception of OMI in 2013 about 7,600 headwords have been added, of which 6,400 originate in corpus data. Other additions include related usage examples and collocations. The paper highlights the strategies needed to keep the vocabulary of OMI up to date.

## 1. Indledning

*Íslensk nútímamálsorðabók* (i dansk oversættelse *Ordbog over moderne islandsk*, herefter OMI) er et leksikografisk projekt ved Árni Magnússon-instituttet for islandske studier, redigeret af Halldóra Jónsdóttir og Þórdís Úlfarsdóttir, der har til formål at beskrive ordforrådet i det nutidige islandske sprog. OMI blev etableret i 2013 og er kun publiceret online. Ordbogen er den eneste nuværende leksikografiske ressource med udgangspunkt i moderne islandsk som bliver regelmæssigt opdateret (se Jónsdóttir & Úlfarsdóttir 2019).

OMI blev etableret ud fra arbejdet med flersprogede ordbøger på Árni Magnússon-instituttet for islandske studier. Ideen var at den nye ordbog skulle genspejle islandsk nutidssprog, især skriftsproget, ganske nøje og inkludere relevante opdateringer, nye ord og andre informationer, dvs. kollokationer og eksempler. Den skulle være brugervenlig med en nemt overskuelig grænseflade og være tilgængelig gratis online.

I forvejen fandtes der en anden islandsk ensproget ordbog, *Íslensk orðabók*, der udkom første gang på tryk i 1963 og i revideret udgave i 1983, 2002 og 2007. Femte og sidste udgave blev publiceret i 2010 og er tilgængelig mod betaling online, men bliver ikke længere opdateret. I dag er de fleste ordbogsbrugere blevet vant til webordbøger, som de også helst vil have gratis adgang til. Udviklingen af OMI var en reaktion på en efterspørgsel i samfundet efter en levende ordbog der beskriver den aktuelle sprogbrug. Samtidig skulle den være nemt tilgængelig og enkel i brug.

I denne artikel vil vi redegøre for materialet som ordbogen bygger på med fokus på dette materiales udvikling, og vi vil diskutere forskellige tilføjelser til ordbogen og motivationen bag dem. Vi vil forsøge at svare på følgende spørgsmål: Hvordan har man besluttet hvilke ord der skal medtages i ordbogen? Hvilke metoder sikrer at ordbogen bedst afspejler det aktuelle ordforråd? Hvordan bliver nye ord en del af ordbogen, og hvilke processer er blevet brugt ved supplering af lemmaer? Artiklen er struktureret på følgende måde: I afsnit 2 giver vi et overblik over baggrund og historik bag OMI hvor vi redegør for OMI's forgænger, ISLEX-projektet, og dets udvikling i perioden frem til 2011. I afsnit 3 og 4 fokuserer vi på forløbet fra OMI's etablering i 2013 og senere tilføjelser til ordbogen. Vi præsenterer her fire processer som vi har defineret som grundlag for tilføjelse af lemmaer. I afsnit 5 er der opsummerende bemærkninger.

## 2. Baggrund og historik

OMI's oprindelse går tilbage til ISLEX, et større leksikografisk projekt der går ud på at lave en flersproget ordbog med islandsk som kildesprog og andre sprog i Norden som målsprog (jf. Jónsdóttir & Úlfarsdóttir 2011, Úlfarsdóttir 2013, 2014). I sin nuværende form har ISLEX ækvivalenter på dansk, svensk, bokmål, nynorsk, fær-

øsk og finsk. ISLEX blev etableret i 2004, og første version blev publiceret online i 2011.

### 2.1. Nordisk ordbogsbase og starten af ISLEX

For at kunne redegøre for ordforrådet i OMI er det nødvendigt at se nærmere på ISLEX og det materiale som dette projekt har som fundament. ISLEX bygger på en lemmaliste der kaldes *Norrænn orðabókarstofn* (Nordisk ordbogsbase, dvs. basisgrundlag for nordiske ordbøger) som blev udarbejdet lidt før år 2000 og er baseret på forskellige tekstkilder. Arbejdet startede som et islandsk-norsk initiativ om at lave en liste over ord som kunne danne grundlag for en ny ordbog der skulle dække det almensproglige ordforråd (jf. Bjarnadóttir 1998). Man brugte materiale fra et udgivet ordbogsværk (Jónsson 1994) samt supplerende materiale efter behov fra bl.a. det store citatarkiv *ROH* og andre tekstsamlinger (jf. Bjarnadóttir 1998). Arbejdet resulterede i en bruttoliste på 200.000 lemmer, kategoriseret efter formelle og semantiske kriterier, og rangeret efter lemmapotentialer. Efter manuel gennemgang og kritisk sortering blev ca. 46.000 ord valgt som materiale til ISLEX.

Oprindelig indlæsning af lemmer i 2004 talte 46.440. I 2004-2005 havde man forberedt materialet til videre behandling. Under det tidlige redaktionelle arbejde blev lemmalisten revideret, og mange lemmer og opslagsformer blev slettet. Samtidig startede man gradvis med at tilføje ord som man mente manglede.

ISLEX blev gennemført i årene 2006-2011 hvor redaktørerne for alle målsprog arbejdede med den samme islandske lemmaliste. I første omgang inkluderede ISLEX målsprogene dansk, nynorsk, bokmål og svensk. Ordbogen blev publiceret online i 2011. Efterfølgende er lemmalisten blevet udvidet med flere ord, og målsprogene færøsk og finsk er blevet tilføjet.

## 2.2. Etablering af ISLEX og senere tilføjelser

Nordisk ordbogsbase var for det meste kun en liste over lemmaer, men den indeholdt dog også en grov og ufyldstgørende semantisk klassificering (191 kategorier for omkring 15 % af ordene). For at lave en ordbog ud af basen var det nødvendigt at udvikle en klar mikrostruktur og tilføje supplerende oplysninger. Den islandske redaktion sørgede for at lemmaerne blev delt op i betydninger hvor der var grund til det, og forklaringer på islandsk blev tilføjet ved hvert eneste opslag. Det skulle være til nytte ved det efterfølgende redaktionelle arbejde med målsprogene (se yderligere i Jónsdóttir & Úlfarsdóttir 2011). Disse forklaringer dannede også senere grundlag for orddefinitioner i OMI.

Yderligere leksikografiske oplysninger blev tilføjet under det redaktionelle arbejde, og ordbogsartiklerne blev udvidet med brugseksempler og forskellige kollokationer og fraser. Til det sidstnævnte har man især brugt Jón Hilmar Jónssons ordbogsværker fra 1994 og 2002.

Man opdagede tidligt at der stadigvæk manglede ord i den oprindelige indlæsning fra den nordiske ordbogsbase, især inden for nogle specifikke semantiske felter, fx samfundsrelaterede ord. Det resulterede i målrettede tilføjelser af ord eller typer af ord. I det følgende gives nogle eksempler på ord der blev tilføjet under det redaktionelle arbejde i årene 2006-2007: *hælisleitandi* 'asylsøger', *móðurmálskennsla* 'modersmålsundervisning', *ráðningarsamningur* 'ansættelseskontrakt', *upprunaland* 'oprindelsesland' og *vegabréfaskoðun* 'paskontrol'. Kilderne for disse ord inkluderede websider fra multikulturelle foreninger og samfundsorganisationer rundt omkring i Island.

Der var også mangel på lande- og indbyggernavne. For at bøde på det brugte man en liste der blev kompileret af det islandske sprogcentrum *Íslensk málstöð*, som senere blev en del af Árni Magnússon-instituttet. I tabel 1 vises eksempler på den slags ord.

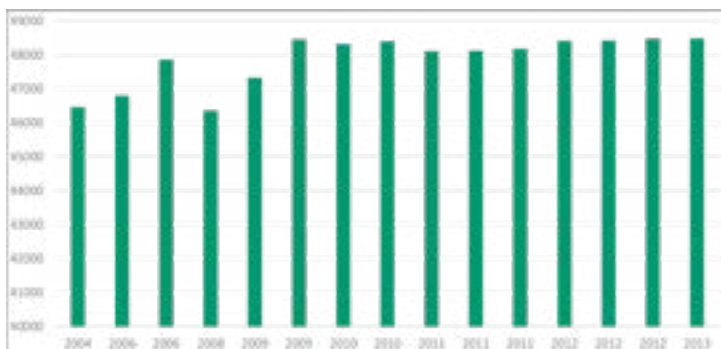
Land	Adjektiv	Indbygger
Alsír	alsírskur	Alsíringur
Argentína	argentínskur	Argentínumaður
Ástralía	ástralskur	Ástrali
Belgía	belgískur	Belgi
Brasilía	brasilískur	Brasilíumaður
Danmörk	danskur	Dani

Tabel 1: Udsnit af liste over lande med tilhørende adjektiver og indbyggernavne i ISLEX.

En liste som den man kan se i tabel 1, kan være et meget nyttigt værktøj idet det giver mulighed for at tilføje manglende oplysninger inden for et begrænset domæne.

Da man begyndte at arbejde med islandske sprogkorporer, fik man et nyt værktøj for at finde frem til manglende ord, og det var klart at en del ganske almindelige ord ikke var inkluderet i den nordiske ordbogsbase og derfor også manglede i en fyldestgørende leksikografisk beskrivelse af sproget. Et eksempel er adverbier med endelsen *-lega* (svarer til da. *-lig*). Dette er en meget aktiv orddannelsesmetode i islandsk. Ordtypen optræder i mange tekstgenrer og forekommer derfor også tit i korpusdata. En række af denne type ord blev ved en målrettet indsats tilføjet til ISLEX, fx *afbrigðilega* ‘unormalt’ el. ‘usædvanlig(t)’, *andstygðilega* ‘frygtelig(t)’, ‘forfærdelig(t)’, *ársfórðungslega* ‘kvartalsvis(t)’ og *feimnislega* ‘genert’.

Alle disse tilføjelser som er diskuteret her, dvs. ordforråd som vedrører samfundet, indbyggernavne og adverbier, blev i lighed med mange andre tilføjet til ISLEX og blev derved en del af det oprindelige ordstof i OMI nogle år senere. Udviklingen af ordforrådet i ISLEX vises i figur 1:



Figur 1: Grafik der viser udviklingen af antal lemmer i ISLEX. Der er generelt 1-3 datapunkter for hvert år.

Som man kan se af figur 1, blev der tilføjet mange ord, men der blev også foretaget en del sletninger. Kurven bevæger sig ikke lineært opad, men går tilbage de gange redaktørerne har ryddet op i lemmalisten.

### 2.3. Sletninger fra ISLEX

Mellem 2006 og 2008 blev der slettet en del ord fra ISLEX. Da man var gået i gang med det redaktionelle arbejde, konstaterede man nemlig at den oprindelige ordbogsbase indeholdt en del ord som var forældede og/eller meget sjældne. Lemmalisten blev gennemlæst og revideret. Ordlistor over mulige sletningskandidater blev genereret og efterfølgende diskuteret til redaktionsmøder. Mange af de slettede ord var sammensatte ord som man syntes var unødvendige at medtage som selvstændige lemmer i en mellemstor ordbog. De inkluderede ord med forstærkende forled: *aðalhvatamaður* 'hovedinitiativtager', *hávísendalegur* 'højvidenskabelig', *fjallbrattur* 'stejl som et bjerg', men også andre sammensætninger, fx *kennslusjónvarp* 'undervisningsfjernsyn' og *prjónapeysa* 'strikkessweater', som redaktionen vurderede ikke hørte til

i en ordbog af denne type, da de enten er gennemskuelige eller sjældne. Tilsvarende nogle eksempler på verber der blev slettet: *þurrsalta* 'tørsalte', *þvermóast* 'nægte at gøre', *væma* 'få kvalme', *útdjöfla* 'udskælde' og *utanbókarlæra* 'udenadslære'. Der var kun få belæg på disse ord, og de er heller ikke en del af det almensproglige ordforråd.


### 3. OMI: En selvstændig ensproget ordbog

I afsnit 2 har vi diskuteret ISLEX og hvordan dette projekts materiale er blevet bearbejdet igennem årene. I forbindelse med OMI er ISLEX relevant indtil året 2013 hvor OMI blev etableret som en ny selvstændig ordbog. Første version var blot en kopi af ISLEX hvor man havde fjernet alle målsprogene (jf. Jónsdóttir & Úlfars-

**köttur** no kk

tramburður  
beyging

Eitt rándyr af kattarætt, algengt gæluðyr



www.faroea.fo

ORÐSAMBOÐ:

vera einn og grár köttur <þar>  
vera þar mikið og oft  
<þeir> eru einn og hundur og köttur  
þeim kemur mjög illa saman  
allt fer í hund og kött  
allt fer í uppnám

Skyldar færslur: heimisköttur no kk

kör no kk  
kórufótt no kk  
kórufótinn no kk  
kórufótinnasett no kk  
kórufótt no kk  
kórufót no kk  
kórufóttaflekur no kk  
kórufótt no kk  
kös no kk  
köstar no kk  
**köttur no kk**  
L skot  
labb no kk  
labba so  
labbakittur no kk  
labbitur no kk  
labba so  
labst so  
laf no kk

Figur 2: Et typisk opslag i OMI, *köttur* 'kat'.



dóttir 2020). Dengang i 2013 var ISLEX kommet op på 48.500 ord, så det var det ordforråd man brugte i den første version af OMI. Efter den tid har disse to ordbøger kørt på hvert sit spor selvom de stadigvæk har tætte forbindelser, med fælles redaktører og tit identiske opdateringer. OMI er nu den eneste islandske ensprogede ordbog der bliver regelmæssigt opdateret, og den betragtes nu som standard referencekilde over nutidens islandske sprog.

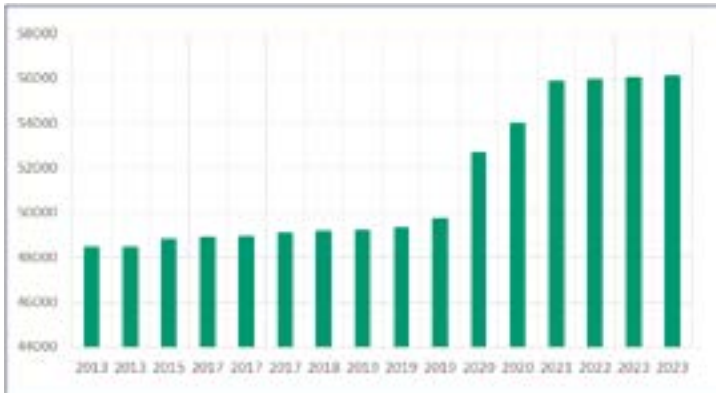
OMI er ”født” med lemmastrukturen fra ISLEX. I figur 2 vises der et typisk opslag fra OMI. Figuren illustrerer hvordan de leksikografiske oplysninger er gengivet i ordbogen. Til venstre har man opslagsordet sammen med ordklasse. Dernæst har man en klikbar knap ”Framburður” (‘udtale’) hvor man kan høre en lyd-gengivelse af det relevante ord. Man har også et link til en anden webresource, BÍN, hvor man kan se alle bøjningsformer (16 for substantiver). Nogle artikler har et billede ligesom den man kan se i figur 2. Man har også en del kollokationer og fraser med tilhørende forklaringer. Nederst er der en automatisk genereret liste over beslægtede ord, her bl.a. *heimilisköttur* ‘huskat’. Til højre i artiklen har man et ordhjul med lemmaer i alfabetisk rækkefølge.

#### 4. Processer ved tilføjelser i OMI

En vigtig del af det redaktionelle arbejde i en ordbog over samtidens sprog er at den konstant skal opdateres med nye ord og andre leksikografiske oplysninger. Motivationen for at tilføje nye lemmaer til OMI er at opretholde ordbogen som afbalanceret og repræsentativ for islandsk nutidssprog.

I figur 3 kan man se hvordan antallet af lemmaer har ændret sig. Datapunkterne er lidt tilfældige fordi ordbogen bliver opdateret et antal gange om året, men ikke på bestemte datoer. Når redaktørerne skønner at de har forberedt en del lemmaer til publicering, bliver ordbogen opdateret med de nye ord. På den måde

er ordbogen vokset gradvis med ca. 7.600 nye ord siden den blev etableret.



Figur 3: Overblik over udviklingen af antal lemmaer i OMI.

Figur 3 viser udviklingen fra 2013. Man kan se at der sker en gradvis stigning i antallet af ord frem til 2019. Mellem 2019 og 2021 sker der en markant stigning, og efter 2021 bliver stigningen igen mere dæmpet.

Vi var interesseret i at se nærmere på disse nye lemmaer, og hvordan man har fundet frem til dem. Hvis man ser på ordbogsdataene i historisk blik og undersøger de tilføjelser der er kommet til, kan man inddele de nye lemmaer i nogle overordnede kategorier efter hvordan man har fundet frem til dem. Efter at vi havde studeret materialet, kunne vi identificere fire processer som bidrager til tilføjelser af nye ord i OMI: 1) nye korpusdata, 2) målrettet excerpering, 3) brugerbidrag, 4) redaktionelle forbedringer. I de følgende afsnit vil vi redegøre for disse fire processer og den påvirkning de har haft på indholdet i OMI.

#### 4.1. Nye korpusdata

Integreringen af nye korpusdata er afgørende for at holde OMI op-

dateret og for at bevare ordbogens aktualitet i forhold til hvordan sproget bliver brugt. Indførelsen af friske sprogdata hjælper med til at sikre ordbogens relevans.

OMI bliver opdateret med nye ord nogle gange om året, fra forskellige kilder. I 2019-2021 blev der foretaget et særligt initiativ for at udvide OMI's lemmaliste ud fra korpusdata. Ord blev hentet fra *Risamálheild* [Gigaword Corpus] (Rmh), et nyere tekstkorpus (se Steingrímsson et al. 2018 og Barkarson et al. 2022). Dette korpus er udviklet hos Árni Magnússon-instituttet for islandske studier, ligesom OMI og ISLEX. En frekvensbaseret liste blev brugt, hvor de tre hovedordklasser, substantiver, adjektiver og verber, blev behandlet hver for sig. Listerne blev sammenlignet med OMI's lemmaliste, hvor de ord der allerede var i ordbogen, blev sorteret fra.

Dataene fra Rmh er dog ikke uden problemer. Der er en vis ubalance i korpusdataene, idet visse genrer fylder meget i tekst-kilderne (jf. diskussion hos Sigurðsson & Steingrímsson 2024). En forholdsvis stor andel administrative tekster og reklameagtigt materiale kan give skævhed i frekvenslisterne som man skal være opmærksom på når korpusmateriale skal indgå i ordbogsarbejde. Ud over det er ordene i nogle tilfælde i en anden grundform end i OMI og bliver derfor ikke fanget af sammenligningsprocessen, fx Rmh *perluvinur* 'perleven', OMI *perluvinir* 'perlevenner'. Almindelige tastefejl eller stavfejl i Rmh kan også forekomme, fx *auðævi* i stedet for *auðæfi* 'rigdom'. I andre tilfælde kan der opstå variantformer fordi retskrivningsregler er blevet revideret, fx *verzlun* ved siden af *verslun* 'handel'. Ud over sådanne tilfælde er der også en del "støj" i korpusdata som skal sorteres fra, fx på grund af fejl i lemmatisering. Derfor er det klart at korpusdata skal bruges med kritisk omtanke i leksikografisk sammenhæng.

Der var imidlertid ikke en åbenlys og etableret procedure for hvordan ord fra Rmh bedst kunne integreres i OMI. Resultatet blev en excerperingsprocedure som inkluderede en blanding af maskinel og menneskelig indsats. I første omgang blev der genere-

ret frekvenslister. Disse lister blev udskrevet og siden gennemgået manuelt af ordbogsredaktørerne. Ved gennemlæsning blev ordene vurderet ud fra subjektive kriterier, med udgangspunkt i redaktørernes sprogforfølelse og om ordet hører til i en almensproglig ordbog. Her kigger redaktørerne ikke kun på frekvens, men også på andre faktorer, dvs. hvor bekendt ordet er, og i hvilken slags tekster det forekommer. Godkendte ord blev markeret med rødt x (jf. figur 4).

gæzluvarðhald	,nafnorð,426,247
göngudagur	,nafnorð,176,170
göngufélagi	,nafnorð,254,240
göngufjarlægð	,nafnorð,275,263
gönguhraði	,nafnorð,256,243
gönguklúbbur	,nafnorð,329,317
göngukona	,nafnorð,288,258
gönguljós	,nafnorð,198,186
göngumynstur	,nafnorð,394,364
gönguseiði	,nafnorð,464,345

Figur 4: Et udvalg fra en udskrevet liste med potentielle lemmer genereret fra korpusdata.

Disse ord blev derefter tilføjet til OMI's lemmalister og behandlet i stil med andre lemmer i selve ordbogen.

Det er også interessant at se nærmere på tilføjelserne til OMI's lemmalister i lyset af semantiske felter. OMI bruger et semantisk klassificeringssystem hvor alle ord er markeret som hørende til én eller flere af over 1.500 betydningskategorier. Dette system var oprindeligt udviklet til ISLEX, men blev også en del af OMI.

OMI		
Semantisk felt	Tilføjelser	Antal ord i alt
mad	171	1217
medicin	113	1168
rejse	124	531

skole/uddannelse	154	584
samfund	68	467
IT	75	428
musik	177	538
administration	138	279

Tabel 2: Antal lemmer i OMI inden for otte udvalgte semantiske felter.

I tabel 2 vises kun et udvalg af tilføjelser til nogle af de større semantiske felter. Når ordene blev tilføjet, var det dog ikke ud fra kriterier baseret på semantiske oplysninger, men samtidig med at ordene bliver en del af OMI, bliver de markeret som hørende til en eller flere semantiske felter. Man kan se at nogle af kategorierne voksede med op til 25 % ved tilføjes af korpusdata. Nogle ord falder ind under mere end én kategori, fx hører *heimsbikarmót* ‘VM-turnering’ til både <sport> og <samling>, og *sammemandi* ‘medstuderende’ hører til både <skole> og <sociale relationer>. I alt blev ca. 6.400 ord tilføjet fra korpusdata.

#### 4.2. Målrettet excerpering

Selvom de mest markante tilføjelser til OMI stammer fra Rmh, er der en del ord der er blevet tilføjet efter andre metoder, både før og efter korpusindsatsen 2019-2021. På grund af overvægt af visse genrer i korpusmaterialet fanger den mere automatiske proces ikke alle de ord man skønner hører til i et ordbogsværk som OMI. I afsnit 2.2. har vi vist nogle eksempler på målrettet excerpering i forbindelse med supplerung af materiale til ISLEX. Man har også udført lignende tilføjelser i forbindelse med OMI for at udfylde huller i lemmalisten. I figur 5 ses et eksempel på hvordan ord er blevet excerperet fra en brochure fra det islandske rigshospital vedrørende hjerteproblemer og stødbehandling (*Rafvending* 2021).



Figur 5: En excerperet side fra en brochure om hjerte problemer.

Som man kan se på figur 5, har man ved en manuel gennemgang af materialet markeret ord som kunne være kandidater til OMI. Ordene er derefter tjekket mod lemmalisten i OMI og tilføjet hvis de manglede. Ved gennemgangen af denne bestemte udgivelse er der tilføjet en række ord i OMI, fx *rafvending* 'stødbehandling', *gáttatíf* 'hjertereflimmer', *gáttasflökt* 'hjertereflagren', *hjartsláttaróregla* 'hjerterytmí' og *blóðþyningarmeðferð* 'blodfortyndingsbehandling'. Ordene overstreget med lilla, *þjóstverkur* 'brystsmerte' og *blóðþyningarlýf* 'blodfortyndende medicin', viste sig allerede at være i ordbogen.

En lignende excerpering fra en brochure om vulkansk luftforurening resulterede også i en del nye ord. Efter at en række jordskælv og vulkanudbrud begyndte på Reykjanes-halvøen i marts 2021, dukkede ordforråd relateret til dette op i medier og samfundsdiskussioner. For det meste var ordene allerede i ord-

bogen, men der var grund til at tilføje flere ord. Derfor blev der samlet et tekstmateriale hvor mange af de relevante ord forekom. Et eksempel er en brochure med titlen *Hætta á heilsutjóni vegna loftmengunar frá eldgosum* 'Fare for helbredet på grund af luftforurening fra vulkanudbrud' der blev udgivet i fællesskab af seks institutioner, bl.a. det islandske meteorologiske institut og sundhedsministeriet (*Hætta ...* 2022). Brochuren indeholder fx ord som *brennisteinsdíoxíð* 'svovldioxid', *loftgæði* 'luftkvalitet', *bráðarugl* 'delirium' og *skyndidauði* 'akut død'.

Målrettet excerpering afspejler en bevidsthed om behovet for at finde frem til forskellige typer ord som måske ikke er repræsenteret i de tilgængelige korpusdata. Eventuelle skævheder i korpusdata har betydning for spørgsmålet om lemmasammensætningen, og det er nødvendigt at gøre en supplerende indsats for at sikre at terminologi og ordforråd inden for specifikke, vigtige emner er fyldestgørende repræsenteret i ordbogen. Her er det dog nødvendigt at vælge omhyggeligt hvilke termer der er relevante for almindelige brugere og ikke medtage ord der kan være for specialiserede, og som kun bruges i meget begrænset omfang.

### 4.3. Brugerbidrag

Ordbogsbrugere påpeger af og til at visse ord mangler i en ordbog. I nogle tilfælde resulterer dette i at der tilføjes ord. For eksempel har oversættere mellem islandsk og dansk fremsat mange forslag til lemmaer som de ønskede blev tilføjet til ISLEX, og alle disse tilføjelser blev også inkluderet i OMI. Eksempler på disse ord er *gagnaver* 'datacenter', *ffarfundur* 'videokonference' og *geðrof* 'psykose'. Ikke mange dinosaurer har fået et islandsk navn, men arterne *snareðla* 'velociraptor' og *þórseðla* 'brontosaurus' er undtagelser og er blevet optaget i OMI efter brugerhenvendelser. Der har også været andre forslag til lemmaer, af meget forskellig art, som fx *hipsúmhaps* 'hip som hap', *heilsuójöfnuður* 'helbredsulighed' og *menningarnám* 'kulturel appropriation'.

Brugerkommentarer retter sig tit mod betydninger eller definitioner, frem for mod tilføjelser af ord. Et eksempel er en brugerhenvendelse vedrørende lemmaet *trilljón*: “der mangler betydningen 1.000.000.000.000.000.000”. Dette har resulteret i at man har tilføjet betydningen ‘tallet ét med 18 nuller’ til den oprindelige ordbogsartikel som kun havde betydningen ‘meget stor mængde’.

Brugerne bidrager ofte med tilføjelser og forslag til ordbogen, og inkludering af dem bringer ordbogen tættere på de personer der anvender materialet. En kritisk gennemgang og evt. ændringer på grund af brugerhenvendelser sikrer at ordbogen er inkluderende og repræsenterer det aktuelle sprog der anvendes af forskellige brugergrupper.

#### 4.4. Redaktionelle tilføjelser

Nogle gange hvor redaktørerne tilføjer et enkelt ord, muligvis efter excerpering, bliver der efterfølgende tilføjet flere ord af samme type, fx verber på *-era* (tit låneord fra dansk), fx *flambera*, *formúlera*, *gratínera*, *impónera* og *stílísera*. Andre eksempler er substantiver på *-sjón* (låneord fra dansk/engelsk), fx *tradisjón*, *sessjón*, *frústrasjón* og *próduksjón*, og IT-relaterede ord, fx *snjallúr* ‘smartur’, *skalanlegur* ‘skalerbar’, *vélþýðing* ‘maskinoversættelse’ og *veggja þátta auðkenning* ‘tofaktorgodkendelse’. Et enkelt ord eller et bestemt semantisk domæne kan derfor resultere i mange flere ord hvis det sætter gang i en frugtbar tankeproces hos redaktørerne.

Tit opstår der nye ord ved arbejdet på tosprogede ordbøger. Når redaktører på ISLEX eller andre tosprogede ordbogsprojekter (fx LEXÍA) støder på ord som de synes savner en islandsk ækvi-valent, bliver den ofte identificeret og tilføjet. Det gælder især mange neologismer som måske ikke har en lang historie i sproget og ikke er blevet en del af det tidligere ordbogsmateriale, fx *skortsala* ‘blankosalg’, *auðsöfnun* ‘kapitalakkumulation’, *hámhorf* ‘bingewatching’, *spjallmenni* ‘chatbot’, *ljóslestur* ‘OCR-behandling’, *ofmettaður* ‘overmættet’ og *kynhlutlaus* ‘kønsneutral’.



Disse eksempler på redaktionelle forbedringer illustrerer at det kan være vigtigt at leksikografer tilpasser de eksisterende data. Denne form for redaktionelt arbejde mindsker risikoen for at ordbogen muligvis overser nogle relevante ord, og arbejdet spiller en afgørende rolle for at nydannede ord hurtigt bliver optaget i ordbogen.

#### 4.5. Andre tilføjelser

Ud over tilføjelser af enkelte opslagsord er ordbogens indhold blevet forøget på andet sæt. En opslagskategori som er blevet udvidet, er flerordslemmaer, dvs. et lemma der inkluderer faste fraser, dvs. mere end ét ord som allerede findes i ordbogen. I en række tilfælde danner den slags fraser en selvstændig betydningsenhed som formelt ikke kan ændres. I tabel 3 kan vi se eksempler på adverbielle flerordslemmaer som er blevet tilføjet fra 2013.

Flerordslemma	Dansk oversættelse
<i>ef að líkum lætur</i>	‘sandsynligvis’
<i>eða þar um bil</i>	‘eller deromkring’
<i>ef á þarf að halda</i>	‘om nødvendigt’
<i>eins og stendur</i>	‘for tiden’
<i>ekki alls kostar</i>	‘ikke helt’
<i>fyrir alla muni</i>	‘for enhver pris’
<i>fyrir langa löngu</i>	‘for længe siden’

Tabel 3: Eksempler på flerordslemmaer tilføjet fra 2013.

De fleste flerordslemmaer er adverbielle fraser der har status som adverbier i sproget. De forekommer tit i tekster, men kan være svære at finde frem til i ordbøger fordi det er uklart hvilket ord man skal slå op under (se fx Jónsson 2005:25-26, 2015). I OMI behandles flerordslemmaer på samme måde som étordslemmaer

og kan søges både som tekststreng i sin helhed eller som individuelle dele. Hvis søgningen giver flere resultater, får man en liste af mulige kandidater og kan klikke på det relevante lemma. Flerordslemmaer er en væsentlig del af sproget, og det er vigtigt at man nemt kan finde frem til dem i leksikografiske ressourcer som OMI.

Derudover er der tilføjet materiale der ikke resulterer i flere lemnaer, men som figurerer i andre dele af artiklerne i ordbogen, fx fraser, kollokationer og brugseksempler. Når et nyt ord bliver etableret, tilføjes der nogle gange også fraser eller kollokationer til det nye opslag. Tit bliver også brugseksempler tilføjet til nyoprettede ordbogsartikler.

#### 4.6. Sammenfatning om ændringer i antal opslagsord

Vores undersøgelse har vist at antallet af lemnaer er vokset med 7.600 fra OMI's etablering i 2013, fra omkring 48.500 til omkring 56.100, dvs. ca. 15 %. Hvis man inkluderer ISLEX-delen, viser det sig at der i alt er blevet tilføjet omkring 17.000 ord siden den oprindelige nordiske ordbogsbase blev konstrueret i slutningen af halvfemserne. I samme periode er omkring 7.000 ord blevet slettet.

Når man ser nærmere på perioden fra 2013 og disse 7.600 nye ord, fremgår det at 6.400 ord stammer fra korpusdata, især gennem den særlige indsats der fandt sted i årene 2019-2021. Omkring 1.200 ord stammer fra andre kilder som vi har identificeret, dvs. målrettet excerpering, brugerbidrag og redaktionelle tilføjelser. Det er ikke altid nemt at skelne imellem disse underkategorier fordi ordene ikke bliver konsekvent markeret i OMI's database efter hvilken metode de er blevet tilføjet. Nogle gange er det dog klart, og vi har givet nogle eksempler på forskellige typer tilføjelser.

En prominent kategori som hører under redaktionelle forbedringer, er flerordslemnaer. Antallet af dem er steget med en tred-

jedel fra ca. 600 til nuværende 900 i alt. Vi kan også se fra databasen at der er ca. 915 nye fraser og kollokationer der er kommet til i de seneste ti år, og de er nu omkring 10.000 i alt, hvilket nogenlunde er det som man forholdsmeæssigt kunne forvente hvis man sammenligner med stigningen i antal lemmer. Brugseksempler i ordbogsartiklerne som primært kommer fra korpusdata, er også steget omkring 15 %, hvilket er proportionelt i forhold til stigningen i antal lemmer, nemlig fra omkring 29.500 til 34.000 i alt.

## 5. Opsummerende bemærkninger

I denne artikel har vi diskuteret ordforrådet i OMI og vist hvordan ordbogen er blevet til, og hvordan det leksikografiske materiale er blevet udvidet og forbedret i løbet af de 10-11 år der er gået siden ordbogen blev etableret. Vi har illustreret hvordan OMI (og dens slægtning ISLEX) er blevet udviklet igennem årene med forskellige tilføjelser og ændringer. Vi kan opsummere hvordan vi har svaret spørgsmålene i indledningen, om hvordan man har besluttet hvilke ord hører til i ordbogen, hvilke metoder der sikrer at ordbogen bedst afspejler det aktuelle ordforråd, hvordan nye ord bliver en del af ordbogen, og hvilke processer der er blevet brugt ved supplering af lemmer.

Ideen med OMI var at den skulle være en mellemstor ordbog der fanger ordforrådet i samtidssproget. Størstedelen af lemmerne i OMI er arvet fra ISLEX hvor de går tilbage til Nordisk ordbogsbasis som blev kompileret i slutningen af sidste århundrede med det formål at belyse moderne islandsk. Samtidig har det været nødvendigt at ordbogen fortsat kan afspejle ord der bruges i samfundet og deres betydninger. Lemmalisten skal derfor revideres og nye ord tilføjes i takt med at sprogbrugernes behov ændres. Det kræver at ordbogens redaktion er opmærksom på de faktorer der

kan være relevant for ordforrådet og er klar til at reagere med revideringer og suppleringer af ordbogsmaterialet.

Når vi ser på de metoder der er blevet anvendt, fremgår det klart at den menneskelige indsats spiller den største rolle for at sikre ordbogens relevans, hvor redaktørerne vurderer ordene ud fra deres formelle karakteristika og skønnet frekvens. Mange beslutninger når det gælder nye ord, er baseret på subjektive indtryk og diskussioner af individuelle ord eller grupper af ord ved redaktionelle møder.

Korpusdata er vigtigt og kan bidrage med afgørende oplysninger om nye ord og betydninger. I lyset af semantiske oplysninger kan man se at mange af de nye ord hører til relativt få semantiske felter. Det kunne indikere at visse semantiske kategorier er mere produktive end andre og derfor burde få større opmærksomhed af OMI's redaktører. Forskellen i det semantiske indhold af de tilføjede ord kan også tyde på at der muligvis er grund til at gå målrettet efter suppleringsmateriale til underrepræsenterede semantiske kategorier.

Vi har identificeret fire processer der er blevet brugt ved supplerings af lemmaer og redegjort for de metoder man har brugt i forbindelse med dem for at sikre at ordbogen bedst kan afspejle det aktuelle ordforråd. Af disse processer er brug af korpusdata det som er mest afgørende. Andre processer er dog også vigtige for at modvirke skævheder i korpusdata og til at tilpasse materialet til samfundsmæssig udvikling og for at reagere på brugerrelaterede behov. Vi har forklaret motivationen bag tilføjjelserne og diskuteret relevante problemstillinger som er i spil når supplerende ordforråd vurderes og evt. nye ord bliver en del af ordbogen.

OMI stræber efter at registrere og belyse det nutidige islandske ordforråd. Ordbogen har etableret sig som et værdifuldt referenceværktøj for alle der er interesseret i det islandske sprog, og den er en meget populær ressource. Antallet af brugere har øgedes markant år fra år, og tallene fra en seks måneders periode fra

august 2023 til februar 2024 viser at der er over 234.000 unikke brugere som har besøgt ordbogen over 670.000 gange. Vedvarende arbejde med at forbedre og udvikle ordbogen så den bedst repræsenterer sproget, fortsætter, og redaktørerne udvikler processer for at holde ordbogen relevant med opdateringer og tilføjelser af nyt materiale. I den nærmeste fremtid har man planer om at bruge nyudviklede sprogteknologiske værktøjer som en hjælp til automatisk excerpering af digitalt materiale. Man er også i gang med at lave et korpus over udgivne bøger fra det 20. århundrede som repræsenterer tekstgenrer der ikke er så prominente i de nuværende korpusressourcer. Begge disse initiativer kommer sikkert til at resultere i flere tilføjelser til OMI de næste måneder eller år.

OMI er en forholdsvis ung ordbog. Den har dog på kort tid etableret sig som en vigtig islandsk sprogresurse. Med konstant revidering og udvikling af nye metoder til at fange ordforrådet i det levende sprog kan man sikre at ordbogen bliver relevant mange år fremover.

## Litteratur

### Ordbøger, korpusser og digitale resurser

BÍN = *Beygingarlýsing íslensks nútímamáls*. Kristín Bjarnadóttir (red.). Árni Magnússon-instituttet for islandske studier. <bin.arnastofnun.is> (marts 2024).

ISLEX. Þórdís Úlfarsdóttir (hovedred.). Árni Magnússon-instituttet for islandske studier. <islex.is>, <islex.dk> (marts 2024).

*Íslensk orðabók* (2010). Mörður Árnason & Laufey Leifsdóttir (red.). 5. útgáfa. [1. udg. (1963), 2. udg. (1983), 3. udg. (2002), 4. udg. (2007)]. Reykjavík: Forlagið.

Jónsson, Jón Hilmar (1994): *Orðastaður*. Reykjavík: Mál og menning.

- Jónsson, Jón Hilmar (2002): *Orðaheimur*. Reykjavík: JPV-útgáfa.
- LEXÍA. Þórdís Úlfarsdóttir (hovedred.), Rósa Elín Davíðsdóttir (red. fransk). Árni Magnússon-instituttet for islandske studier. <lexia.hi.is> (marts 2024).
- OMI = *Íslensk nútímamálsorðabók* [Ordbog over moderne islandsk]. Þórdís Úlfarsdóttir & Halldóra Jónsdóttir (red.). Árni Magnússon-instituttet for islandske studier. <islenskordabok.is> (marts 2024).
- Rmh. = *Risamálheildin* [Icelandic Gigaword corpus]. Stofnun Árna Magnússonar í íslenskum fræðum. <malheildir.arnastofnun.is> (februar 2024).
- ROH = *Ritmalssafn Orðabókar Háskólans* [Det leksikografiske instituts historiske citatarkiv]. <ritmalssafn.arnastofnun.is> (maj 2024).

## Anden litteratur

- Barkarson, Starkaður, Steinþór Steingrímsson & Hildur Hafsteinsdóttir (2022): Evolving Large Text Corpora: Four Versions of the Icelandic Gigaword Corpus. I: Nicoletta Calzolari et al. (eds.): *Proceedings of the Language Resources and Evaluation Conference, Marseille, France*. Marseille: European Language Resources Association. 2371-2381.
- Bjarnadóttir, Kristín (1998): *Norræna verkefnið*. Upubliceret rapport.
- Bjarnadóttir, Kristín (2012): The Database of Modern Icelandic Inflection. I: Guy Pauw et al. (eds.): *LREC 2012 Proceedings of the workshop of Language Technology for Normalization of Less-Resourced Languages, SaLTMiL 8 – AfLaT 2012*. Istanbul: European Language Resources Association. 13-18.
- Bjarnadóttir, Kristín, Kristín Ingibjörg Hlynsdóttir & Steinþór Steingrímsson (2019): DIM: The Database of Icelandic Morphology. I: *Proceedings of the 22nd Nordic Conference on*

- Computational Linguistics*, (NoDaLiDa 2019, Turku, Finland). Linköping Electronic Conference Proceedings, No. 167, NE-ALT Proceedings Series, No. 42. Linköping: Linköping University Electronic Press. 146-154.
- Hætta á heilsutjóni vegna loftmengunar frá eldgosum. Leiðbeiningar fyrir almenning* [Fare for helbredsconsekvenser på grund af vulkanudbrud. Vejledning for almenheden] (2022). Reykjavík: Sóttvarnalæknir, Umhverfisstofnun, Veðurstofa et al. <almannavarnir.is/frettir/baeklingur-radleggingar-vegna-gasmengunar/>.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2011): ISLEX – en flersproget nordisk ordbog. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi* 11. Lund: Nordisk förening för lexikografi. 353-366.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2019): Íslensk nútímamálsorðabók. I: *Orð og tunga* 21, 1-25. doi.org/10.33112/ordogtunga.21.2.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2020): Omdannelsen af en flersproget til en monolingval ordbog. I: Caroline Sandström et al. (red.). *Nordiska studier i lexikografi* 15. *Rapport från 15 konferensen om lexikografi i Norden. Helsingfors 4-7 juni 2019*. Helsingfors: Skrifter udgivet af Nordisk Forening for Leksikografi. 175-186.
- Jónsson, Jón Hilmar (2005): Aðgangur og efnisskipan í íslensk-erlendum orðabókum – vandi og valkostir. I: *Orð og tunga* 7, 21-40.
- Jónsson, Jón Hilmar (2015): Flerordslemmaer. Form og funksjoner. I: Caroline Sandström et al. (red.): *Perspektiv på lexikografi, grammatik och språkpolitik i Norden* (Vol. 39). Helsinki: Institutet för de inhemska språken. 116-133.
- Sigurðsson, Einar & Steinþór Steingrímsson (2024): Representativeness and biases in Icelandic corpora. *LexicoNordica* 31 (dette bind).

Steingrímsson, Steinþór, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson & Jón Guðnason (2018): Risamálheild: A Very Large Icelandic Text Corpus. I: Nicoletta Calzolari et al. (eds.): *Proceedings of LREC 2018, Eleventh International Conference on Language Resources and Evaluation*. Myazaki: European Language Resources Association. 4361-4366.

Úlfarsdóttir, Þórdís (2013): ISLEX – norræn margmála orðabók. I: *Orð og tunga* 15, 41-71.

Úlfarsdóttir, Þórdís (2014): ISLEX – a Multilingual Web Dictionary. I: Nicoletta Calzolari et al. (eds.): *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation*. Reykjavík: European Language Resources Association. 2820-2825.

Ellert Þór Jóhannsson  
forskningslektor  
Árni Magnússon-instituttet for  
íslandske studier  
Edda, Arngrímsgötu 5  
IS-107 Reykjavík  
ellert.thor.johannsson@arnastofnun.is

Þórdís Úlfarsdóttir  
orðbogsredaktør  
Árni Magnússon-instituttet for  
íslandske studier  
Edda, Arngrímsgötu 5  
IS-107 Reykjavík  
thordis.ulfarsdottir@arnastofnun.is