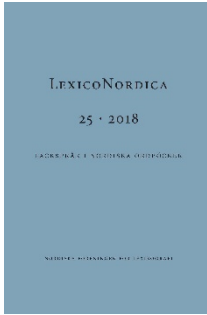


LexicoNordica

Forfatter:	Aldís Sigurðardóttir & Ellert Þór Jóhannsson [Sprogportalen málið.is]	
Anmeldt værk:	Sprogportalen málið.is <málið.is> eller <malid.is>. Stofnun Árna Magnússonar í íslenskum fræðum.	
Kilde:	LexicoNordica 25, 2018, s. 329-342	
URL:	https://tidsskrift.dk/index.php/lexn/issue/archive	

© 2018 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.

Sprogportalen málið.is

Aldís Sigurðardóttir & Ellert Þór Jóhannsson

Sprogportalen málið.is <málið.is> eller <malid.is>. Stofnun Árna Magnússonar í íslenskum fræðum.

1. Indledning

Málið.is er en frit tilgængelig islandsk sprogportal der blev lanceret på det islandske sprogs dag den 16. november 2016. Portalen er udarbejdet ved Stofnun Árna Magnússonar í íslenskum fræðum, dvs. Árni Magnússon-instituttet for islandske studier. Den indeholder forskellige typer ordbøger og andre sproglige ressourcer over moderne islandsk og er en velkommen samlet kilde til udforskning af dokumenteret materiale vedrørende det islandske sprog.

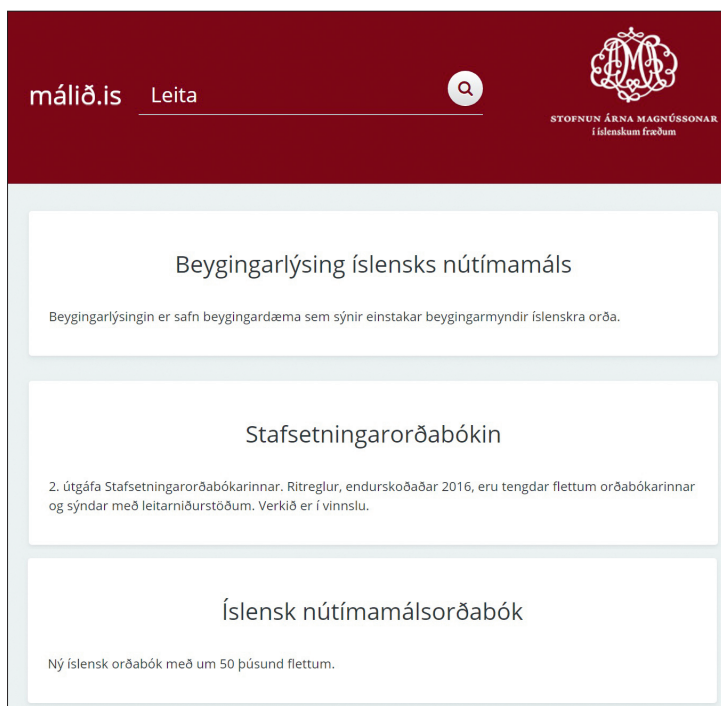
Oplysninger om tilblivelse og principper for udarbejdelsen af portalen er ikke tilgængelige på selve portalen, men i en kort artikel kan man finde oplysninger om projektet (Eva María Jónsdóttir & Steinþór Steingrímsson 2016). Formålet med at samle sproglige ressourcer på en sprogportal er ifølge artiklens forfattere at give offentligheden let adgang til det store materiale der er blevet samlet gennem tiderne. Det fremhæves også at portalen skal være med til at forhindre det islandske sprog i at lide sprogdøden. Portalen indeholder både traditionelle tidligere trykte værker og værker der specifikt er udarbejdet til elektronisk format. Med tiden er det planen at udvide portalen med flere ordbogsdata og korpora. Portalens målgruppe er alle der har brug for ordbøger, og her nævnes oversættere, journalister, forfattere, elever på alle trin og alle der beskæftiger sig med tekster og det islandske sprog.

I det følgende vil vi beskrive portalens design, dens overordne-

de struktur, de værker der indgår i den, og endelig vil vi teste dens anvendelse og funktion. Testningen er foretaget på en stationær pc, men málið.is er også tilpasset mobile enheder.

2. Portalens design og overordnede struktur

Portalen er holdt i et stilrent design. Øverst til højre er instituttets (klikbare) logo placeret, til venstre portalens navn og til højre for dette søgefeltet. Farveskalaen er enkel med baggrundsfarverne bordeaux og grå med hvide bogstaver. Under søgefeltet er der syv hvide felter med titlerne på portalens værker og en kortfattet beskrivelse af dem (figur 1).



Figur 1: Udsnit af portalens grænseflade.

Portalen ser ud som ovenfor beskrevet hvis man går direkte ind på málið.is. Man kan også tilgå portalen fra instituttets site <arnastofnun.is> hvor man finder portalens søgefelt på en central plads lige under hjemmesidens topmenu. For begge fremgangsmåder gælder det at man ikke kan orientere sig i værkerne før man har skrevet et søgeord i søgefeltet. Det er efter vores mening en svaghed, for som bruger er man nysgerrig efter at vide hvad det er for værker man har foran sig før man går i gang.

Der er ingen orientering til brugeren om hvem der står bag portalen (ud over instituttets logo) eller nærmere oplysninger om de enkelte værker, og ej heller om hvordan portalen fungerer. Med hensyn til de enkelte værkers forfattere og værkernes tilblivelse må man søge oplysninger andre steder, enten på det pågældende værks hjemmeside eller i den trykte udgave. Med hensyn til hvordan portalen fungerer, må man prøve sig frem. Det ville være gavnligt med en introduktion med oplysninger om og vejledning i brug af portalen og oplysninger om de enkelte værker, især de værker der ikke er tilgængelige elektronisk andre steder.

3. Sprogressourcer på portalen

Portalen indeholder indtil videre syv værker som optræder på málið.is i samme rækkefølge som nedenfor.

- (1) *Beygingarlýsing íslensks nútímamáls* (BÍN, red. Kristín Bjarnadóttir) er et særdeles nyttigt værktøj som indeholder fuldstændige bøjningsparadigmer over ca. 278.000 islandske ord. I islandsk kan substantiver have 16 bøjningsformer, verber 48 og adjektiver 120. Det enorme materiale er her gjort tilgængeligt i en overskuelig form. Efter at BÍN blev offentliggjort i 2007, er dens popularitet steget støt, og den bruges hyppigt af en bred vifte af sprogbrugere. BÍN er en primær kompo-

nent på málið.is fordi den danner grundlag for søgningen i de øvrige værker. Ved hvert søgeord der indtastes i søgefeltet på málið.is, genereres en opslagsform eller opslagsformer på grundlag af bøjningsparadigmerne. Resultaterne fra BÍN, som vises på málið.is, er kun opslagsformerne med hyperlink, og man må klikke på det relevante link for at komme ind på BÍN's website.

- (2) *Stafsetningarorðabókin* (SOB) er en ny og revideret udgave af papirudgaven (Dóra Hafsteinsdóttir 2006). Den trykte udgave indeholdt 73.000 lemmaer, og denne første udgave er tilgængelig gennem ordbogsportalen Snara. 2. udgave er, som det fremgår af den korte præsentation på málið.is, stadig under revision. Søgeresultaterne vises som et statisk billede på portalen, dvs. at man kun får præsenteret resultatet af éns søgning uden mulighed for at kunne klikke rundt i ordbogens elektroniske version.
- (3) *Íslensk nútímamálsorðabók* (ÍN, red. Halldóra Jónsdóttir & Þórdís Úlfarsdóttir) er en forholdsvis ny netordbog over moderne islandsk med ca. 50.000 lemmaer. Den bygger på det flersprogede netordbogsprojekt ISLEX som er et samarbejdsprojekt mellem Island, Danmark, Finland, Færøerne, Norge og Sverige (tilgængelig på <islex.is>, hvor der er adgang til alle de involverede sprog). På málið.is vises opslagsordets første (primære) betydning med link til ÍN's website, hvor man bl.a. kan finde flere betydninger, idiomer, kollokationer og en alfabetisk lemmaliste. Ved hvert lemma oplyses at værket er under udarbejdelse.
- (4) *Íslenskt orðanet* (ÍNÉT, red. Jón Hilmar Jónsson) er skabt på basis af Jón Hilmar Jónssons værker *Orðastaður* (1994), *Orðaheimur* (2002) og *Stóra orðabókin um íslenska málnotkun* (2005). ÍNÉT er født som en elektronisk sprogressource og indeholder en meget stor samling sprogdata, suppleret med materiale fra tekstkorpora. Materialet er ordnet i en lemmali-

ste med over 250.000 etords- og flerordsenheder. ÍNET giver eksempler på fraser og kollokationer og viser syntagmatiske og semantiske relationer. Søgeresultaterne på málið.is består af en liste over opslagsord der hver især giver adgang til de relevante opslag på ÍNETs eget website.

- (5) *Málfarsbankinn* (MB, udarbejdet på instituttets sprogrøgtsafdeling, tidligere Íslensk málstöð) består af kortfattede artikler som hver især belyser et bestemt problem i forbindelse med “korrekt” sprogbrug. Samlingen er baseret på henvendelser fra offentligheden og en række institutioner og ministerier med spørgsmål af forskellig art (om udtale, retskrivning, låneord, navne osv.) ledsaget af svar på spørgsmålene. MB blev lanceret som webresource i 2002 med ca. 7000 lemmaer, og vi formoder at antallet er væsentligt forøget i de mellemliggende år, for det oplyses på hjemmesiden at sprogrøgtsafdelingen behandler ca. 2000 henvendelser pr. år. Resultatet af søgning på málið.is er artikler hvor opslagsordet enten optræder som lemma eller indgår i en artikel. Udviklerne af málið.is har valgt at gengive søgeresultaterne statistisk, dvs. at der ikke er link til MBs egen hjemmeside via portalen.
- (6) *Íðorðabankinn* (ÍOB, red. Ágústa Þorbergsdóttir et al.) er en fagordssamling som har været åben for adgang på internettet siden 1997. Den bygger på en lang tradition i Island med udvikling af islandske fagord inden for så godt som alle fagområder. Samlingerne findes i en søgbar database, hvor man får oplysninger om tilsvarende fagord på engelsk og andre sprog. Søgeresultatet i dette værk på málið.is varierer alt efter om ordet findes inden for et eller flere fagområder. I tilfælde af det sidste dukker der en liste op med links til forskellige artikler. Søgeresultatet ledsages af en forkortet artikel.
- (7) *Íslensk orðsifjabók* (ÍOS, red. Ásgeir Blöndal Magnússon) er en etymologisk ordbog som indtil lanceringen på málið.is i 2016 kun har været tilgængelig på tryk. ÍOS er det eneste værk

af sin art over moderne islandsk. Det indeholder ca. 25.000 opslagsord med oplysninger om endnu flere ord som ikke optræder som selvstændige lemmaer. På málið.is fremtræder søgeresultaterne som statiske billeder, og brugeren har ikke mulighed for at bevæge sig rundt mellem artiklerne. Brugeren har heller ikke adgang til det trykte værks grundige indledning, hvor der bl.a. er en gennemgang af principperne for værket samt lister over forkortelser.

4. Test af sprogportalen

I det følgende vil vi fokusere på hvordan málið.is fungerer og på samspillet mellem portalen og værkerne. På denne baggrund vil vi kommentere enkelte værker på portalen. Noget af det vi skriver i det følgende, kan meget vel være ugyldigt i morgen eller om en uge, for som det er med internetbaserede ressourcer, er også denne underlagt så hyppige ændringer at følgende gennemgang må blive et øjebliksbillede for perioden februar-marts 2018.

4.1. Generelt om søgning

Al søgning på málið.is fremgår ved at man indtaster et ord eller en ordform i søgefeltet. Som tidligere nævnt danner BÍN grundlaget for søgningen så man ikke nødvendigvis behøver at indtaste den gængse lemmaform (subst.: nom. sg., adj.: nom. sg. mask, vb.: infinitiv etc.), men kan indtaste en hvilken som helst bøjningsform. Dette kan være nyttigt, men medfører ofte at man får resultater fra mere end ét opslagsord, hvilket kan virke forvirrende. Her ville det være en fordel at kunne begrænse søgningen til et bestemt lemma og en bestemt ordklasse, men som portalen er sat op nu, er det ikke muligt.

Selv om det ikke er nødvendigt at indtaste grundformen af et

ord, er det nødvendigt at indtaste alle islandske særtegn. Eksempelvis kan man ikke indtaste “tina” eller “tyna” for verberne *tína* ‘samle, plukke’ og *týna* ‘tabe, miste’. Accenter volder ofte problemer, især for udlændinge. Søgningen på ordene uden accent giver ikke noget svar, og vi kunne derfor tænke os en mere fleksibel løsning som fx “mente du?” med forslag som kunne hjælpe én videre, eller en alternativ søgemulighed hvor der ikke blev skelnet mellem særlandske og andre tegn. En anden mulighed ville være at indsætte et felt med de islandske tegn ved siden af søgefeltet, sådan som man kender det fra fx ISLEX. Endelig har vi prøvet at søge med wildcards, men det er ikke en mulighed på málið.is. Det var værd at overveje om det ikke burde tilbydes brugeren.

4.2. Test af forskellige ord

Med ovenstående in mente besluttede vi at den bedste måde at finde ud af hvordan portalen fungerer, ville være at teste den gennem at prøve forskellige ord og ordformer. Vi begyndte med et almindeligt ord og håbede at der var hits i alle værkerne. Vores første forsøg var søgestrengen “hafa”, som er opslagsform af verbet *hafa* ‘have’. Resultatet fra de første tre værker ses i figur 2 på næste side.

Øverst på siden står der “hafa fannst í 6 gagnasöfnum” (hafa blev fundet i seks ressourcer). Det oplyses ikke hvilket af værkerne der ikke er repræsenteret, men vi kunne hurtigt konstatere at det var MB. Det er måske ikke så overraskende fordi denne ressource primært er orienteret mod vanskelige og problematiske ord og vendinger, og man forventer måske ikke at finde et så almindeligt ord som *hafa* her, selv om det i princippet godt kunne tænkes at vendinger hvori *hafa* indgår, kunne volde problemer.

Når man kigger nærmere på resultaterne, viser det sig at det ikke kun er verbet *hafa*, men også substantivet *haf* ‘hav’ der gives som svar på søgningen. Det har sin forklaring i at formen *hafa* også kan være en bøjningsform (genitiv pluralis) af substantivet

hafa fannst í 6 gagnasöfnum

Beygingarlýsing íslensks nútímamáls

haf Hvorugkynsnafnorð
hafa Sagnorð, þátið *hafði*

Stafsetningarorðabókin

haf -ið hafs; höf *haf|kul; haf|skrímsli*
hafa hafði, haft *ég hef(i) lágt; hann hefur/hefir lagt sig*

Íslensk nútímamálsorðabók

haf nafnorð hvorugkyn
sjór, einkum úthaf
oplið haf
<bátinn rekur> á haf út
<vera staddur> á hafi úti
<vera staddur> úti á hafi
<bærinn stendur> fyrir opnu hafi
<halda> til hafs
<sigla> <vestur> um haf

[Fara í orðabók](#)

hafa sagnorð
FALLSTJÓRN: ÞOLFALL
vera með (e-ð)
ég hef mitt eigið skrifborð í vinnunni
hann hafði dagblað í hendinni
við höfum stóran garð

[Sjá 34 merkingar í orðabók](#)

Figur 2: Udsnit af málið.is. Søgeresultat af *hafa* fra BÍN, SOB og ÍN.

haf, og fordi søgningen er baseret på bøjningsformer og ikke opslagsord, bliver dette resultatet. Søgning på *haf* giver et næsten identisk resultat eftersom *haf* også er imperativ sg. af verbet *hafa*. Det er ikke muligt at begrænse søgningen til kun at omfatte verbet eller kun substantivet, og søgeresultatet bliver derfor lidt mudret.

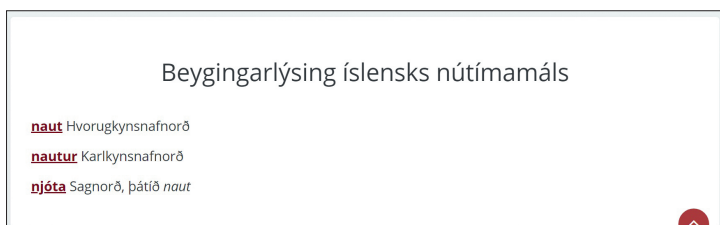
I tilfældet *hafa* eller *haf* handler det kun om to mulige ord, mens der ved andre homonymer kan være tale om et helt uoverskueligt antal hvor det ikke klart fremgår hvilke ord det drejer sig om. Det er tydeligt at systemet har svært ved at håndtere homonymer generelt, fx viser søgning af *sá* (demonstrativt pronomener) ‘den’ også resultat for *sár* (subst. neut.) ‘sår’ selvom *sá* ikke kan være bøjningsform af dette ord. Det skyldes formentlig at *sá* kan være bøjningsform af *sár* (subst. mask.) ‘kar’, som er homonym til *sár* (subst. neut.), og derfor er søgningen udvidet til også at inkludere dette ord. I sådanne tilfælde virker resultaterne meget forvirrende.

Det næste ord vi testede, var *slíkur* ‘sådan, slig’, et ord der i sig selv byder på en del vanskeligheder. Søgeresultatet viser fire værker på málið.is: BÍN, SOB, ÍN og ÍOS. Ordets ordklassetilørsforhold er ikke så enkelt, hvilket afspejles i uenigheden mellem værkerne. BÍN og ÍOS klassificerer ordet som demonstrativt pronomener, hvorimod ÍN klassificerer det som adjektiv. SOB angiver ikke ordklasseforhold.

ÍNÉT overraskede ved ikke at vise noget søgeresultat. Går man uden om málið.is og direkte ind på ÍNET, får man dog et andet resultat. Ordet er der skam, og det klassificeres som pronomener. I dette værk ser det ud til at søgeresultatet i forbindelse med pronomener er én til én, forstået på den måde at kun den opslåede ordform vises i et meget stort antal fraser. For at se eksempler på de øvrige bøjningsformer af ordet må man indtaste dem hver især. Vi undersøgte i denne forbindelse flere pronomener (*þessi*, *hann* og *hún*), og det viste sig at ingen af disse kommer op som resultat

í ÍNET gennem málið.is, men at de alle findes på ÍNET når man går direkte ind på sitet, og vi må konkludere at der er mangler i koblingen mellem ressourcerne.

Det næste ord vi testede, var substantivet *naut* ‘okse’. Grundformen (neutr. nom.) af dette ord falder sammen med akkusativ og dativ af substantivet *nautur* (mask. nom.) ‘kammerat; (gave) giver’ og med præteritum singularis af verbet *njóta* ‘nyde’, nemlig formen *naut*. Slår man substantivet *naut* op på málið.is viser resultatet derfor både verbet og to substantiver, se figur 3, repræsenteret ved BÍN.



Figur 3: Udsnit af málið.is. Søgeresultat af *naut* fra BÍN.

ÍNET viser to ord, nemlig det neutrale substantiv *naut* og verbet *njóta* (figur 4).



Figur 4: Udsnit af málið.is. Søgeresultat af *naut* fra ÍNET.

Et hurtigt blik på søgeresultatet fra ÍNET på málið.is afslører at noget ikke stemmer, nemlig at de to sidste eksempler under “*njóta* so” (*njóta* vb.) er forkert placeret. Der er ikke tale om verbet, men om substantivet *naut*.

Hvis man klikker på *njóta*, kommer man ind på ÍNETs site hvor man til venstre finder en liste med eksempler. Eftersom det er verbet *njóta* vi har valgt, burde substantivet *naut* ikke figurere her, men det gør det (se figur 5).

Sagnorð	
njóta	
fá að njóta sín	
fá að njóta <hæfifleika sinna, verka sinna>	
glápa á <hann, hana> eins og naut á nývirki	
góna á <hann, hana> eins og naut á nývirki	
haga sér eins og naut í glervörubúð	
láta <hann, hana> njóta drauma sinna	
láta <hann, hana> njóta sannmæla	

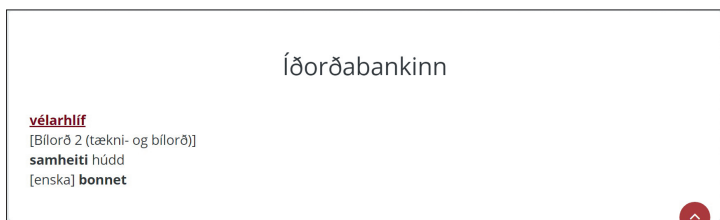
Figur 5: Udsnit af fraser med verbet *njóta* på ÍNETs site.

Her har tre eksempler med substantivet forvildet sig ind i listen over fraser med verbet *njóta*. Det fremhævede eksempel i figur 5 og de to efterfølgende eksempler viser substantivet *naut* og ikke præteritumsformen af verbet *njóta*. Den fejlagtige identificering og placering af substantivet i listen af fraser med verbet må stamme fra en teknisk fejl på ÍNET, som afspejles i resultatet på málið.is.

De sidste to eksempler vi vil omtale, er substantiverne *rostungur* ‘hvalros’, og *húdd* ‘motorhjeml’. *Rostungur* gav hits i fem af værkerne på málið.is, og alle fem værker fungerede efter hensigten. MB og ÍOS fremviste ingen resultater. Eftersom MB indeholder forholdsvis få lemmaer, var det ikke så overraskende at den ikke var repræsenteret. Til gengæld havde vi forventet at finde *rostungur* i den etymologiske ordbog (ÍOS), og vi slog derfor ordet op i den trykte udgave, hvor det da også var at finde. En hurtig test af

25 opslag i papirudgaven af ÍOS (Ásgeir Blöndal Magnússon 1989) viste dog at alle 25 lemnaer også figurerede på málið.is, så vi håber at det manglende *rostungur* er et enkeltstående tilfælde.

Med hensyn til *húdd* ‘motorhjeml’ var vi især interesseret i om ordet var belagt i ÍOB. I figur 6 kan man se at *húdd* ikke figurerer som lemma, men anføres som “samheiti”, dvs. synonym for lemnaet *vélarhlíf*, og at det svarer til *bonnet* på engelsk (figur 6).



Figur 6: Udsnit af málið.is. Søgeresultat af *húdd* fra ÍOB.

Det undrer os at kun det engelsk-engelske ord er repræsenteret, for vi er klar over at *húdd* ikke er en islandsk dannelse, men et lån fra amerikansk, og vi klikkede os derfor videre ind på ÍOBs website for at se om vi kunne finde oplysninger om det amerikansk-engelske ord.



Figur 7: Søgeresultat fra ÍOBs site.

Og ganske rigtigt, her er det opført nederst i den korte artikel (figur 7). Det står ikke klart hvorfor denne del af artiklen er udeladt på málið.is. Det kan ikke skyldes pladshensyn idet det kun drejer sig om en linje på et par ords længde. Det ville hjælpe brugeren hvis oplysninger om eventuelle forkortelser af artiklerne blev anført på málið.is. En løsning kunne også være at indsætte en mulighed for at udfolde hele artiklen ved et klik med musen.

5. Konklusion

Sprogportalen málið.is giver adgang til en række forskellige islandske sproressourcer som hele offentligheden kan få glæde af. Det er en åbenlys fordel at samle disse ressourcer ét sted hvor man kan søge i mange værker samtidig. Det er imidlertid en svaghed ved portalen at man ikke kan begrænse sin søgning til ét bestemt ord eller til udvalgte værker. Det medfører at søgeresultaterne ofte indeholder mange forskellige informationer om flere ord, og det kræver i visse tilfælde lidt sorteringsarbejde at finde frem til det man søger svar på.

Under testen af portalen stødte vi på nogle tekniske brister både på selve portalen og i enkelte værker. Det er dog vores indtryk at de fleste af disse brister forholdsvis let vil kunne afhjælpes. En forbedret kobling af data og en klarere struktur i søgeresultaterne ville øge portalens brugervenlighed og effektivitet markant.

Eva María Jónsdóttir & Steinþór Steingrímsson (2016) gør opmærksom på at málið.is ikke er et færdigt produkt, men en dynamisk ressource som er under udvikling, og som efterhånden vil blive forbedret og udvidet. Vi forventer at portalen vil opnå stor udbredelse blandt islandske sprogbrugere og alle andre der er interesseret i det islandske sprog og glæder os til at følge udviklingen.

Litteratur

- Ásgeir Blöndal Magnússon (1989): *Íslensk orðsifjabók*. Reykjavík: Mál og menning.
- BÍN = *Beygingarlýsing íslensks nútímamáls*. <bin.arnastofnun.is> (marts 2018).
- Dóra Hafsteinsdóttir (2006): *Stafsetningarorðabókin*. Reykjavík: JPV útgáfa.
- Eva María Jónsdóttir & Steinþór Steingrímsson (2016): *Vefgáttin málið.is. I: Tölvumál. Tímarit Skýrslutæknifélags Íslands 1, 16-17*.
- ÍN = *Íslensk nútímamálsorðabók*. <islenskordabok.arnastofnun.is> (marts 2018).
- ÍNÉT = *Íslenskt orðanet*. <www.ordanet.is> (marts 2018).
- ÍOB = *Íðorðabankinn = Orðabanki Íslenskrar málstöðvar*. <www.ordabanki.hi.is> (marts 2018).
- ÍOS = *Íslensk orðsifjabók*. Tilgængelig på <malid.is> (marts 2018).
- ISLEX = <islex.is> (marts 2018).
- Jón Hilmar Jónsson (1994): *Orðastaður. Orðabók um íslenska málnotkun*. Reykjavík: Mál og menning. 2. udgave 2001. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson (2002): *Orðaheimur. Íslensk hugtakaorðabók*. Reykjavík: JPV útgáfa.
- Jón Hilmar Jónsson (2005): *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa.
- MB = *Málþarsbankinn*. <malfar.arnastofnun.is> (marts 2018).
- Snara = <snara.is> (marts 2018).
- SOB = *Stafsetningarorðabókin*. 2. udgave. Tilgængelig på <malid.is> (marts 2018).

Aldis Sigurðardóttir
ordbogsredaktør, mag.art.
Københavns Universitet
Njalsgade 136
DK-2300 København S
cws301@hum.ku.dk

Ellert Þór Jóhannsson
ordbogsredaktør, ph.d.
Københavns Universitet
Njalsgade 136
DK-2300 København S
ellert@hum.ku.dk