

Ari Páll Kristinsson, Halldóra Jónsdóttir
og Steinþór Steingrímsson

Ný útgáfa Málsins á fimm ára afmæli
vefgáttarinnar

1 Markmið, aðdragandi og fyrri útgáfa

Stofnun Árna Magnússonar í íslenskum fræðum og samstarfsaðilar hennar hafa yfir að ráða viðamiklum og verðmætum rafrænum gögnum um íslenska tungu. Heilmikið efni er tiltækt. Aðgangur, vef-föng, viðmót og framsetning einstakra gagnasafna hefur hins vegar verið með ýmsu móti í gegnum tíðina. Því vaknaði hugmyndin um eina samræmda vefgátt til að auðvelda fólki aðgang að hinu geysilega magni upplýsinga og fróðleiks sem gagnasöfnin geyma, endurgjaldslaust. Vefgáttin fékk heitið *málið.is*¹ og var opnuð 16. nóvember 2016. Vefsíðan var hönnuð með það í huga að leit væri einföld og að almennum netnotendum yrði gert það eins auðvelt og verða má að finna þær upplýsingar sem þeir leita að.

Markmið *Málsins* er að auðvelda notendum að finna upplýsingar

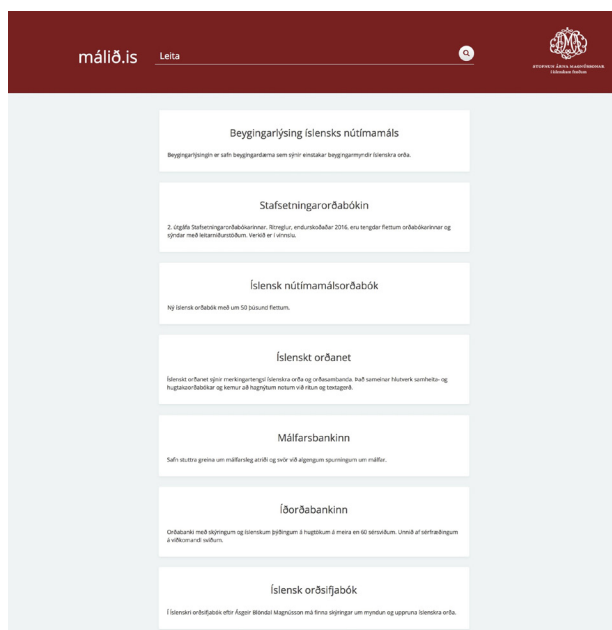
1 Til glöggvunar: Þegar heiti vefgáttarinnar *málið.is* er ritað fullum fetum, þ.e. með endingunni *.is*, er hafður lítill upphafsstafur nema í upphafi málsgreinar. Þegar notað er stytta heitið *Málið* er haft stórt M.

um mál og málfar með því að beina þeim inn í viðeigandi gagnasöfn um íslensku, hvort sem þeir sitja við tölvu, nota snjallsíma eða spjaldtölvu enda þarf notandinn aðeins einn einfaldan aðgang til að finna gögn úr þeim öllum í einu. Vefgáttin dregur fram allar tiltækar upplýsingar um það orð sem leitað er að hverju sinni, úr hinum mismunandi gagnasöfnum. Nánari upplýsingar úr hverju gagnasafni fást síðan ef smellt er á niðurstöður þaðan.

Leit í einstökum söfnum, einu í einu, er ekki aðeins tímafrekari en í gegnum *Málið* heldur getur – og hefur – auðveldlega leitt til þess að fólk finnur ekki eða veit ekki að til eru upplýsingar sem hefðu getað gagnast því. Ástæðan er m.a. sú að gögnin eru birt á ólíkan hátt sem aftur skýrist af því að gengið hefur verið frá gögnunum á mismunandi vegu og að sumu leyti í ólíkum tilgangi.

Helsti hvatamaður að gerð vefgáttarinnar var Halldóra Jónsdóttir, verkefnisstjóri á Stofnun Árna Magnússonar í íslenskum fræðum, en með henni í ritstjórn hafa frá upphafi verið tveir aðrir starfsmenn stofnunarinnar, þeir Ari Páll Kristinsson rannsóknarprófessor og Steinpór Steingrímsson, verkefnisstjóri í upplýsingatækni. Auk aðstandenda gagnasafnanna sem *Málið* sækir efni sitt til hafa aðrir starfsmenn lagt ritstjórninni lið við hönnun og undirbúning, m.a. Eva María Jónsdóttir þáverandi kynningarstjóri og Trausti Dagsson, verkefnisstjóri í upplýsingatækni. Styrktarsjóður Áslaugar Hafliðadóttur veitti styrk til vefforritunar og viðmótshönnunar árið 2016 og þá naut *Málið* einnig stuðnings úr Málræktarsjóði. Steinpór Steingrímsson hefur einkum séð um tæknilega vinnslu gagnasafnanna til að gera þau sameiginlega leitarbær en Trausti Dagsson hefur hannað útlit og vefviðmót. Veffangið *málið.is* fékk stofnunin afhent til eignar af Margréti Pálsdóttur.

Við opnun Málsins 16. nóvember 2016 voru sex gagnasöfn leitarbær í vefgáttinni: *Beygingarlýsing íslensks nútímamáls*, *Íslensk stafsetningarorðabók*, *Íslensk nútímamálsorðabók*, *Málfarsbankinn*, *Íðorðabankinn* og *Íslensk orðsifjabók*. Ári síðar bættist *Íslenskt orðanet* við.



Mynd 1. Skjáskot af leitarsíðu Málsins fyrstu fimm árin.

Gagnasöfnin eru innbyrðis ólík að eðli og innihaldi enda er þeim ætlað að gegna mismunandi hlutverkum. Á vissan hátt bæta hin ólíku gagnasöfn hvert annað upp. Með hinum sameiginlega aðgangi á *Málinu* fá notendur því gleggri heildarmynd af því atriði sem leit að er að hverju sinni heldur en fæst ef leit takmarkast við aðeins eitt gagnasafn. Hin sameiginlega vefgátt dregur fram mismunandi þætti úr gagnasöfnunum sem tengjast viðkomandi leitarorði, t.d. eftir sögulegu eða samtímalegu sjónarhorni eða eftir því hvort nálgunin er í anda lýsandi eða vísandi málfræði.

Beygingarlýsing íslensks nútímamáls geymir orð og beygingar sem tíðkast í íslensku og hún er fyrst og fremst unnin í anda lýsandi málfræði – enda þótt einnig sé að finna þar athugasemdir og ábendingar um notkun. *Íslensk stafsetningarorðabók* og *Málfrásbankinn* hafa á hinn bóginn beinlínis það meginhlutverk að leiðbeina um réttan rithátt og málnotkun sem talin er samræmast hefðbundnum viðhorfum um vandað íslenskt mál. Í slíkum gagnasöfnum er áherslan því einkum á sviði vísandi málfræði.

Gagnasöfnin fyrstu fimm árin

Beygingarlýsing íslensks nútímamáls er safn beygingardæma sem sýnir allar beygingarmyndir hvers orðs, þ.m.t. afbrigði. Upprunalegt markmið var m.a. að koma upp gagnasafni til nota í máltækni. Í safninu eru nú ríflega 300 þúsund beygingardæmi. Ritstjóri *Beygingarlýsingar íslensks nútímamáls* er Kristín Bjarnadóttir.

Íðorðabankinn veitir aðgang að íslenskum þýðingum á erlendum íðorðum og hugtakaskilgreiningum íðorða á íslensku og fleiri tungumálum. Pennan sérhæfða orðaforða er ekki endilega að finna í orðasöfnum um almennt mál. Hafa ber í huga að aðgangur gegnum *málið.is* birtir aðeins niðurstöður við íslensk leitarorð og sýnir því ekki allt það gagn sem hafa má af *Íðorðabankanum* og tengist erlendum fræðiheimum. Skál í því sambandi bent á eigin vef *Íðorðabankans*. *Íðorðabankinn* er settur saman úr mörgum mismunandi íðorðasöfnum sem hvert og eitt lýtur eigin ritstjórn. Íðorðasöfnin, sem spanna meira en 60 sérsvið, eru misstór, sum geyma nokkra tugi eða hundruð hugtaka en önnur eru gríðarstór og geyma jafnvel tugi þúsunda hugtaka. Ritstjóri *Íðorðabankans* er Ágústa Þorbergsdóttir.

Íslensk nútímamálsorðabók geymir merkingarskýringar og aðrar upplýsingar um u.þ.b. 55 þúsund uppflettiorð. Reglulega bætast við ný orð og orðskýringar enda er orðabókin enn í vinnslu og ber að hafa í huga að skýringar við sum orð eru ekki fullmótaðar. Ritstjórar *Íslenskrar nútímamálsorðabókar* eru Halldóra Jónsdóttir og Þórdís Úlfarsdóttir.

Íslensk orðsifjabók er eftir Ásgeir Blöndal Magnússon (1909–1987). Hún miðlar fróðleik, skýringum og skýringartilgátum um uppruna og söguleg vensl íslenskra orða og orðmynda, jafnt úr fornu máli sem úr nútímaíslensku. Orðsifjabókin kom út á prenti 1989 á vegum Orðabókar Háskólans. Bókin geymir um 25 þúsund uppflettiorð. Hafa verður í huga að orðsifjabókin er höfundarverk Ásgeirs og efnið hefur ekki verið uppfært eða endurbætt í ljósi nýrra viðhorfa eftir að bókin kom út 1989.

Íslensk stafsetningarorðabók (áður *Stafsetningarorðabókin*) er opinber réttitunarorðabók um íslensku. Hlutverk hennar er að leiðbeina um rithátt og beygingar í samræmi við hefðbundin viðhorf um vand- að ritmál. Byggt er á ritreglum Íslenskrar málnefndar með nánari

útfærslu. Um 73 þúsund flettiorð voru í fyrstu útgáfu 2006. Önnur útgáfa, *Íslensk stafsetningarorðabók* (2016–) er einungis aðgengileg á rafrænu formi. Orðabókin er uppfærð eftir þörfum og útlit og virkni er í frekari þróun. Ritstjóri er Jóhannes B. Sigtryggsson.

Íslenskt orðanet er umfangsmikið yfirlit um íslenskan orðaforða, orðasambönd og orðanotkun. Það hefur að geyma óviðjafnanlega lýsingu á innra samhengi orðaforðans þar sem byggt er á greiningu á merkingarvenslum orða og orðasambanda. Þá sameinar *Íslenskt orðanet* í raun hlutverk samheitaorðabókar og hugtakaorðabókar og nýtist því vel við hvers konar ritun og textagerð. Höfundur *Íslensks orðanets* er Jón Hilmar Jónsson.

Málfarsbankinn geymir safn stuttra greina um málfarsleg efni. Þar er leitast við að benda á málnotkun sem almennt hefur verið talin best við hæfi í málsniðum vandaðs og hefðbundins íslensks ritmáls og í vönduðu töluðu máli. Efni *Málfarsbankans* á meðal annars rætur í svörum sem tekin hafa verið saman við málfarsráðgjöf á Íslenskri málstöð og síðar Stofnun Árna Magnússonar í íslenskum fræðum. Enn fremur er að finna í *Málfarsbankanum* fjöldamarga pistla um málfar eftir Jón G. Friðjónsson, prófessor emerítus við Háskóla Íslands. Ritstjóri *Málfarsbankans* er Jóhannes B. Sigtryggsson.

Notendur og notkun

Vefgáttinni var strax vel tekið við opnun hennar 2016. Vikulegar heimsóknir voru að jafnaði rúmlega 4.900 fyrsta árið, og uppflettingar um 14 þúsund í hverri viku. Notkunin hefur vaxið á hverju ári um 20–65% og á fimmta árinu voru vikulegar heimsóknir orðnar á 17. þúsund og vikulegar flettingar að nálgast 50 þúsund. Notkunin er mest á morgnana klukkan 8 til 12. Þá fara flettingar stundum yfir 1000 á klukkustund sem þýðir að á 3–4 sekúndna fresti sé einhvers staðar flett upp í vefgáttinni á þeim tíma dags. Notendum vefgáttarinnar hefur fjölgað stöðugt frá upphafi, eins og sjá má í *Töflu 1*.

ár	heimsóknir	flettingar	heimsóknir á viku	flettingar á viku
1	256.441	726.699	4.932	13.975
2	341.572	1.010.887	6.569	19.440
3	432.744	1.305.166	8.322	25.099
4	716.553	2.172.236	13.780	41.774
5	862.980	2.465.020	16.596	47.404

Tafla 1. Heimsóknir og flettingar á Málið fyrstu fimm árin.

Mynd 2 sýnir sérstaklega þróun heimsókna á viku fyrstu fimm árin.



Mynd 2. Fjöldi heimsókna á viku árin 2016–2021.

Vefgáttin Málið stuðlar ótvírætt að kynningu og útbreiðslu einstakra gagnasafna sem þar er að finna. Sem dæmi má nefna að Íslenskt orðanet fór inn á málið.is í desember 2017 og eftir það fjölgaði margfalt vikulegum heimsóknum inn á það gagnasafn. Mynd 3 sýnir glögglega stórauðna notkun orðanetsins eftir að það bættist við Málið í desember 2017.



Mynd 3. Vikulegar heimsóknir inn á Íslenskt orðanet 2017–2018. Orðanetið fór inn á Málið í desember 2017.

Tvær kannanir hafa verið gerðar á notkun á vefgáttinni. Í notendakönnun sem framkvæmd var á vefgáttinni sjálfri árið 2018 kom í ljós að um 40% notenda væru undir 35 ára, hátt í 60% voru konur og um þrír af hverjum fjórum höfðu háskólapróf. Snemma árs 2021 var spurt um málið.is í könnun Félagsvísindastofnunar. Þar kom fram að um 33% aðspurðra höfðu einhvern tíma farið inn á málið.is og 48% fólks á aldrinum 18–29 ára. Ekki kom fram mikill munur á konum og

körlum en fylgni var með aukinni notkun vefgáttarinnar og aukinni menntun. Um 7% aðspurðra sögðust nota vefinn í hverjum mánuði, 13% fólks á aldrinum 18–29 ára. Um 39% fólks á aldrinum 18–29 ára voru skráð í framhaldsskóla eða háskóla árið 2020. Um þriðjungur þess fjölda notar *málið.is* reglulega.

Ný útgáfa *Málsins* 2021

Ný útgáfa vefgáttarinnar *málið.is* var kynnt á degi íslenskrar tungu, 16. nóvember 2021, á fimm ára afmæli *Málsins*. Bætt hefur verið við aðgerðum eins og að fela niðurstöður úr gagnasöfnum sem notandinn hefur ekki áhuga á og að breyta röð gagnasafna í niðurstöðum. Val notandans vistast sjálfkrafa á því tæki og í þeim vafra sem notandinn gerir breytingarnar í. Þá hafa grunnupplýsingar um hvert gagnasafn verið gerðar aðgengilegar með einföldum hætti við allar leitarniðurstöður, upplýsingavalmynd hefur verið endurbætt, notendum til hægðarauka og dæmi um verkefni fyrir framhaldsskólanema hafa verið gerð aðgengileg. Skólaverkefni unnu þær Hugufrún Hólmgeirsdóttir og Halldóra Björt Ewen, íslenskukennarar í Menntaskólanum við Hamrahlíð. Ef smellt er á línurnar þrjár vinstra megin efst á vefsíðunni má finna almennar upplýsingar um vefgáttina og lýsingar á öllum gagnasöfnum, á íslensku og ensku. Enn fremur er hægt að koma á framfæri ábendingum og athugasemdum. Í vinstra horni neðst á leitarsíðunni er hnappur með tenglum í alla vefi Stofnunar Árna Magnússonar í íslenskum fræðum og samstarfsaðila. Þar má m.a. finna fleiri vefi um mál og málnotkun, t.a.m. *Risamálheildina*.

Mikilvægasta breytingin á fimm ára afmælinu í nóvember 2021 fólst í því að við fyrstu sjö gagnasöfnin bættust nú tólf ný. Nú sækir *Málið* því efnivið til samtals nítján gagnasafna.

Gagnasöfnin tólf sem bættust við 2021 eru talin hér á eftir. Athuga ber að *ISLEX-orðabókin* inniheldur sex íslensk-norrænar orðabækur og *LEXÍA* geymir í raun tvær tvímála orðabækur. *LEXÍA* er fyrst og fremst íslensk-frönsk orðabók en hún geymir einnig margar þýskar þýðingar og jafnheiti þar sem íslensk-þýskur orðabókarhluti er í vinnslu.

ISLEX-orðabókin er margmála orðabók á vefnum. Grunnmálið er íslenska og markmálin eru danska, sænska, norskt bókmál, nýnorska, færeyska og finnska. *ISLEX* er samstarfsverkefni sex norrænna fræðastofnana á Íslandi, í Danmörku, Noregi, Svíþjóð, Færeyjum og Finn-

landi. Þær eru Stofnun Árna Magnússonar í íslenskum fræðum í Reykjavík, Det Danske Sprog- og Litteraturselskab í Kaupmannahöfn, Institutt for lingvистiske, litterære og estetiske studier við Háskólann í Bergen, Institutionen för svenska språket við Háskólann í Gautaborg, Fróðskaparsetur Føroya í Þórshöfn og Háskólinn í Helsinki. Íslenska ritstjórnin ber ábyrgð á íslenska hluta orðabókarinnar ásamt mótun og þróun gagnagrunnsins, en vinna við markmálin er í umsjón ritstjórna í hverju landi fyrir sig. Aðalritstjóri *ISLEX-orðabókarinnar* er Þórdís Úlfarsdóttir.

Íslensk-dönsk orðabók eftir Sigfús Blöndal kom út á árunum 1920–1924. Orðabókin er lykilverk í íslenskri orðabókarsögu og er hún ein stærsta íslenska orðabókin sem hefur verið gerð til þessa. Allar orðskýringar í bókinni eru á dönsku sem kann að vekja furðu nú um stundir, en á þeim árum sem verkið var í vinnslu (1903–1920) var Ísland enn hluti af danska konungsveldinu. Styrkir til starfsins og útgáfunnar komu frá dönskum sjóðum, svo og dönskum og íslenskum stjórnvöldum. Sigfús vann að orðabókinni í fjölda ára ásamt eiginkonu sinni, Björgu Þorláksdóttur Blöndal, og síðar bættust við Jón Ófeigsson og Holger Wiehe. Stafræn gerð orðabókarinnar var opnuð í febrúar 2021. Verkefnisstjórn skipa Halldóra Jónsdóttir (formaður), Steinþór Steingrímsson og Þórdís Úlfarsdóttir.

Íslensk-rússnesk orðabók var samin af Valeríj P. Bérkov með aðstoð Árna Böðvarssonar. Bókin kom út árið 1962 og er fyrir löngu uppseld. Vefútgáfa bókarinnar er unnin í samstarfi Stofnunar Árna Magnússonar í íslenskum fræðum og Stofnunar Vigdísar Finnbogadóttur í erlendum tungumálum.

LEXÍA-orðabókin er veforðabók milli íslensku og frönsku, unnin í samstarfi Árnastofnunar og Stofnunar Vigdísar Finnbogadóttur í erlendum tungumálum. Íslenski hluti verksins er unninn hjá Stofnun Árna Magnússonar í íslenskum fræðum sem og gagnagrunnur og önnur tæknivinna við orðabókina. Stofnun Vigdísar Finnbogadóttur í erlendum tungumálum hefur umsjón með franska markmálinu. Aðalritstjóri er Þórdís Úlfarsdóttir. Ritstjóri frönsku er Rósa Elín Davíðsdóttir.

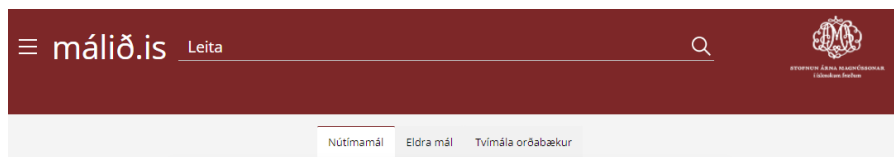
Nýyrðavefur Stofnunar Árna Magnússonar í íslenskum fræðum birtir sýnishorn söfnunar undanfarinna ára en einnig er hægt að senda inn nýyrði. Innsend nýyrði eru tekin til skoðunar af ritstjóra

sem ákveður hvort og hvenær þau birtast á nýyrðavefnum en öllum nýyrðatillögum er haldið til haga. *Nýyrðavefurinn* byggist upp eftir því sem fleiri nýyrði verða skráð í grunninn. *Nýyrðavefurinn* er í umsjón Stofnunar Árna Magnússonar í íslenskum fræðum. Ritstjóri er Ágústa Þorbergsdóttir.

Í *Ritmálssafni Orðabókar Háskólans* eru dæmi um notkun orða úr íslenskum ritum. Þau spanna tímabilið frá miðri 16. öld til loka 20. aldar. Notkunardæmin, sem sýna orðin í samhengi, eru flest úr prentuðum bókum eða blöðum en einnig úr handritum frá síðari tímum. Við efnissöfnunina var nær allt prentað mál frá upphafi fram á 19. öld lesið og orðtekið auk fjölmargra rita frá 19. og 20. öld. Ritstjóri *Ritmálssafns* er Ásta Svavarsdóttir.

Þrískipting gagnasafna á leitarsíðu

Til að auðvelda notendum að finna hratt og örugglega þau orðasöfn sem þeir þurfa að nota hefur verið sett upp aðgreining efnisins í þrjá flipa: *Nútímamál*, *Eldra mál* og *Tvímála orðabækur*. Á *Mynd 3* má sjá flipana þrjá sem efnið flokkast í.



Mynd 3. Skjáskot af hluta af leitarsíðu *Málsins* í nýju útgáfunni.

Undir flokkinn *Nútímamál* falla þau orðasöfn sem eru í virkri ritstýringu og því uppfærð reglulega og fjalla um eða lýsa íslensku máli eins og það er talað í dag. Þar eru öll þau orðasöfn sem fyrir voru á vefgáttinni, fyrir utan *Íslenska orðsifjabók* sem var færð í flokkinn *Eldra mál*. Þá hefur *Nýyrðavefnum* verið bætt við.

Í flokknum *Eldra mál* eru eldri orðabækur, sem voru prentaðar en hafa nú verið færðar yfir á stafrænt form, auk *Ritmálssafns Orðabókar Háskólans*. *Ritmálssafnið* sýnir dæmi um notkun orða úr íslenskum ritum frá miðri 16. öld og fram á síðari hluta 20. aldar. Prentuðu orðabækurnar, sem gerðar hafa verið stafrænar, eiga það sameiginlegt að útgáfa þeirra er endanleg og efni þeirra verður ekki breytt. Þær eru: *Íslensk orðsifjabók* frá 1989, sem hefur verið aðgengileg á *Málinu* frá upphafi, *Íslensk-dönsk orðabók* eftir Sigfús Blöndal, sem kom út á

árunum 1920–1924, og *Íslenzk-rússnesk orðabók* eftir Valeríj P. Bérkov sem kom út árið 1962. Þær tvær síðarnefndu eru nýjar í vefgáttinni, rétt eins og *Ritmálssafnið*.

Í flokknum *Tvímála orðabækur* eru íslensk-norrænu orðabækurnar úr *ISLEX* gerðar leitarbærar. Það eru orðabækur þar sem grunnmálið er íslenska og markmálin danska, sænska, norskt bókmál, nýnorska, færeyska og finnska. Í þessum flokki er líka leitað í tveimur orðabókum sem tengjast *LEXÍU*-verkefninu, íslensk-franskri og íslensk-þýskri orðabók, sem er í vinnslu. Tvímála orðabækurnar eru, eins og orðasöfnin í flokknum *Nútímamál*, í virkri ritstjórn. Það þýðir að við orðabækurnar getur bæst efni eða að efni þeirra sé breytt ef þurfa þykir. *ISLEX-orðabækurnar* geyma í kringum 53 þúsund flettur hver, íslensk-franska orðabókin um 40 þúsund og íslensk-þýska orðabókin ríflega 18.500 flettur.

Eins og sjá má í *Töflu 2* var hægt að kalla fram yfir 2 milljónir mismunandi leitarniðurstaðna á vefgáttinni við opnun hinnar nýju útgáfu í nóvember 2021, og fjöldinn eykst stöðugt.

Nútímamál		Eldra mál		Tvímála orðabækur úr íslensku. Markmál eru:	
Beygingarlýsingin	303.871	Blöndal, Ísl.-dönsk	148.580	Danska	53.371
Ísl. nútímamáls-orðabók	55.553	Ritmálssafnið	707.149	Sænska	51.781
Ísl. stafsetningar-orðabók	67.563	Ísl. orðsifjabók	41.575	Norskt bókmál	47.555
Málfarsbankinn	17.225	Bérkov, Ísl.-rússnesk	31.490	Nýnorska	47.532
Íðorðabankinn	176.835			Færeyska	53.362
Nýyrðavefurinn	666			Finnska	46.886
Ísl. orðanet	119.520			Franska	39.632
				Þýska	18.538

Tafla 2. Við opnun nýrrar útgáfu er að finna 2.028.684 leitarniðurstöður á *Málinu*. Í heildina voru mismunandi leitarorð 943.813.

*Ari Páll Kristinsson
Stofnun Árna Magnússonar í íslenskum fræðum
Laugavegi 13
IS-101 Reykjavík
ari.pall.kristinsson@arnastofnun.is*

*Halldóra Jónsdóttir
Stofnun Árna Magnússonar í íslenskum fræðum
Laugavegi 13
IS-101 Reykjavík
halldora.jonsdottir@arnastofnun.is*

*Steinþór Steingrímsson
Stofnun Árna Magnússonar í íslenskum fræðum
Þingholtsstræti 29
IS-101 Reykjavík
steinthor.steingrimsson@arnastofnun.is*

