Ásta Svavarsdóttir

Ritmálssafn Orðabókar Háskólans

Samtíð og saga

1 Inngangur

Orðabók Háskólans var komið á laggirnar á fimmta áratug 20. aldar með það lokamarkmið að gefa út stóra sögulega orðabók um íslensku. Áratugina á undan höfðu orðið talsverðar umræður um nauðsyn þess að semja ítarlega orðabók um íslensku með íslenskum skýringum og gerðar höfðu verið atrennur að efnissöfnun og gerð slíkrar bókar með tilstyrk Alþingis (sjá einkum Jóhannes L.L. Jóhannsson o. fl. 1920; Guðmund Davíðsson [og] Jóhannes L.L. Jóhannsson 1926–1927; Jóhannes L.L. Jóhannsson 1928). Sú vinna leiddi ekki til útgáfu nema að afar takmörkuðu leyti (Jón Ólafsson 1912–1915) en stærstur hluti þess efniviðar til orðabókagerðar sem var tekinn saman á fyrri hluta aldarinnar komst í vörslu Orðabókar Háskólans eftir að hún var stofnuð. Á umræðu- og undirbúningstíma orðabókarvinnunnar greindi menn nokkuð á um það yfir hversu langt tímabil ný vísindaleg orðabók ætti að ná, aðallega hvort hún ætti að taka til allrar sögu málsins frá upphafi ritaldar eða hvort sleppa ætti fornmálinu sem þegar höfðu verið gerð allgóð skil í orðabókum á 19. öld (sjá blaðaumræður: Alexander Jóhannesson 1919a, 1919b; Finnur Jónsson 1919a, 1919b; Jóhannes L.L. Jóhannsson 1919a, 1919b, 1919c). Þegar Háskóli Íslands tók að sér ábyrgð á orðabókarverkefninu eftir 1940 (sjá Halldór Halldórsson 1969:29–31; Jón Aðalstein Jónsson 1988) var hins vegar þegar hafinn undirbúningur að gerð orðabókar

Orð og tunga 26 (2024), 185–210, https://doi.org/10.33112/ordogtunga.26.11
© höfundar CC BY-NC-SA 4.0.

um fornmálið á vegum Árnanefndar í Kaupmannahöfn (*Ordbog over det norrøne prosasprog* (ONP)). Þess vegna var ákveðið að íslenska orðabókin skyldi taka við þar sem henni sleppti og ná yfir tímabilið frá siðaskiptum og upphafi prentaldar um miðja 16. öld til samtímans (sjá Jakob Benediktsson 1955:7–8, 1971:12–13). Saman myndu þessi tvö verk því spanna alla sögu íslensks orðaforða allt frá elstu handritum til nútímamáls. Fyrirmyndir verksins voru stórar orðabækur yfir þjóðtungur sem höfðu verið gefnar út eða voru í smíðum í öðrum löndum, einkum orðabók Grimm-bræðra í Þýskalandi (*Deutsches Wörterbuch* (DWB) í 33 bindum 1854–1971), Oxford-orðabókin í Englandi (*A New English Dictionary on Historical Principles* (OED) í 12 bindum 1884–1928), sögulega danska orðabókin (*Ordbog over det danske Sprog* (ODS) í 28 bindum 1918–1956) og orðabók sænsku Akademíunnar (*Svenska Akademiens Ordbok* (SAOB) í 39 bindum 1893–2023).

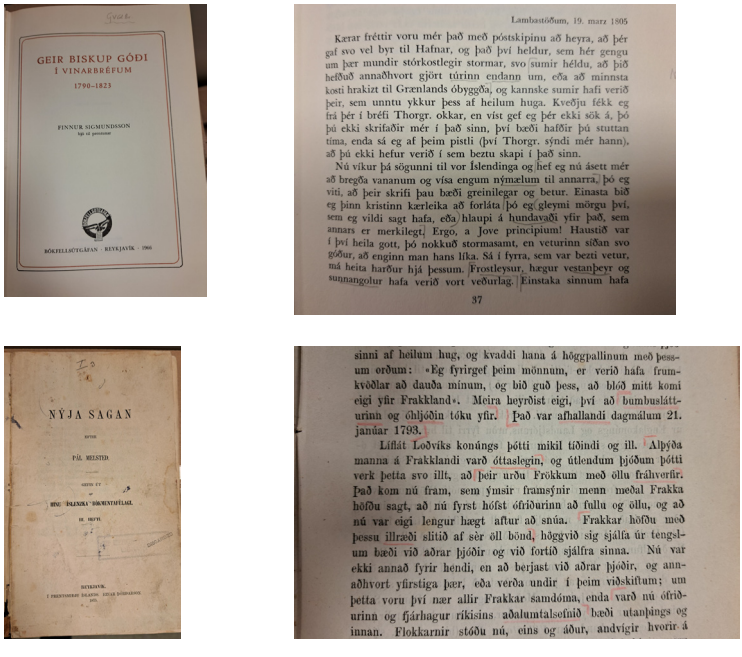
Hafist var handa við efnisöflun með því að safna dæmum um orð og orðanotkun í rituðum textum á íslensku, einkum prentuðum heimildum en einnig handritum frá síðari öldum. Þannig varð til vaxandi safn skráðra notkunardæma, hið svonefnda Ritmálssafn Orðabókar Háskólans, sem fjallað verður um í þessari grein. Í fyrri hluta hennar verður sjónum beint að tilurð og einkennum safnsins. Gerð verður grein fyrir þeim aðferðum sem beitt var við efnisöflunina í 2. kafla og fjallað um seðlasafnið sem spratt af henni í 3. kafla. Seinni hluti greinarinnar fjallar um þróun safnsins í kjölfar þeirra byltingarkenndu breytinga sem urðu á aðferðum við gerð og framsetningu orðabóka í lok 20. aldar. Í 4. kafla verður sagt frá vinnu við tölvuskráningu efnisins sem hófst á níunda áratugnum og gerð rafrænna gagnagrunna til að halda utan um efniviðinn og vinna úr honum. Þá fjallar 5. kafli um rafræna birtingu safnsins frá og með aldamótum og þar er gerð grein fyrir því hvernig efninu er nú fyrir komið þannig að almennir notendur geta nálgast það og nýtt viðmótið til þess að leita í efninu frá ýmsum hliðum. Greininni lýkur með stuttum niðurlagskafla.

2 Efniviður og aðferðir

Þegar orðabókarverkefninu var hleypt af stokkunum var stuðst við þær aðferðir sem þá þóttu bestar við gerð fræðilegra orðabóka – orðabóka sem byggðust á umfangsmikilli efnisöflun og ítarlegum rannsóknum á orðaforða og orðanotkun á grunni þeirra gagna. Efnis-

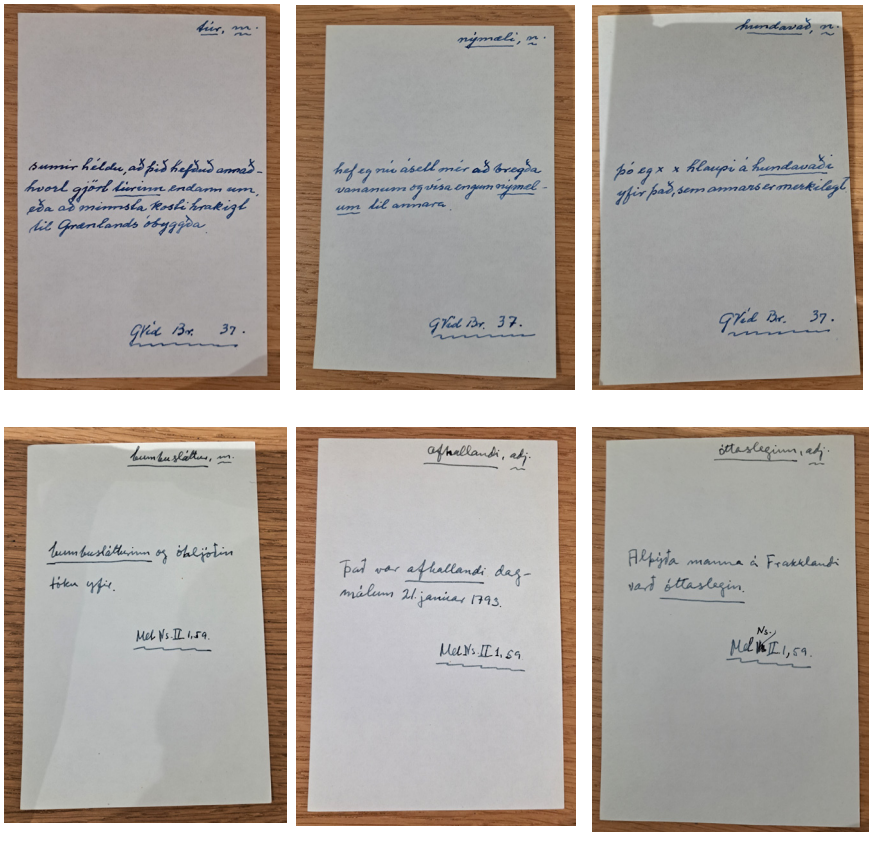
söfnunin fólst fyrst og fremst í því að lesa og orðtaka sem allra mest af prentmáli á íslensku frá 16. til 20. aldar. Framan af öldum voru prentaðar bækur svo fáar að hægt var að lesa þær nánast allar. Þær voru einsleitar, að verulegu leyti trúarlegs eðlis þar sem prentverk var tengt kirkjunni, en síðari tíma útgáfur af ritum frá sama tímabili, ekki síst veraldlegum textum, voru einnig orðteknar og jafnframt var leitað fanga í handritum frá síðari öldum til þess að auka fjölbreytni efniviðarins (sjá Jakob Benediktsson 1955). Eftir að útgáfa bóka og blaða jókst á 19. öld varð að velja úr textunum. Eftir sem áður var leitast við að fara yfir svo mikið og fjölbreytilegt efni frá 19. og 20. öld að það gæfi góða mynd af málinu á þeim tíma. Í heimildaskrá ritmálssafnsins eru alls yfir 2000 titlar en að baki sumum þeirra liggja margra binda verk, bókaflokkar eða blöð og tímarit sem sum hver voru gefin út árum eða áratugum saman.

Svokölluð orðtaka fólst í því að við lesturinn var merkt við orðmyndir í textanum sem þóttu áhugaverðar af einhverjum sökum og samhengi þeirra afmarkað til þess að sýna hvernig orðið væri notað og til að varpa ljósi á merkingu þess (sjá *Mynd 1*).



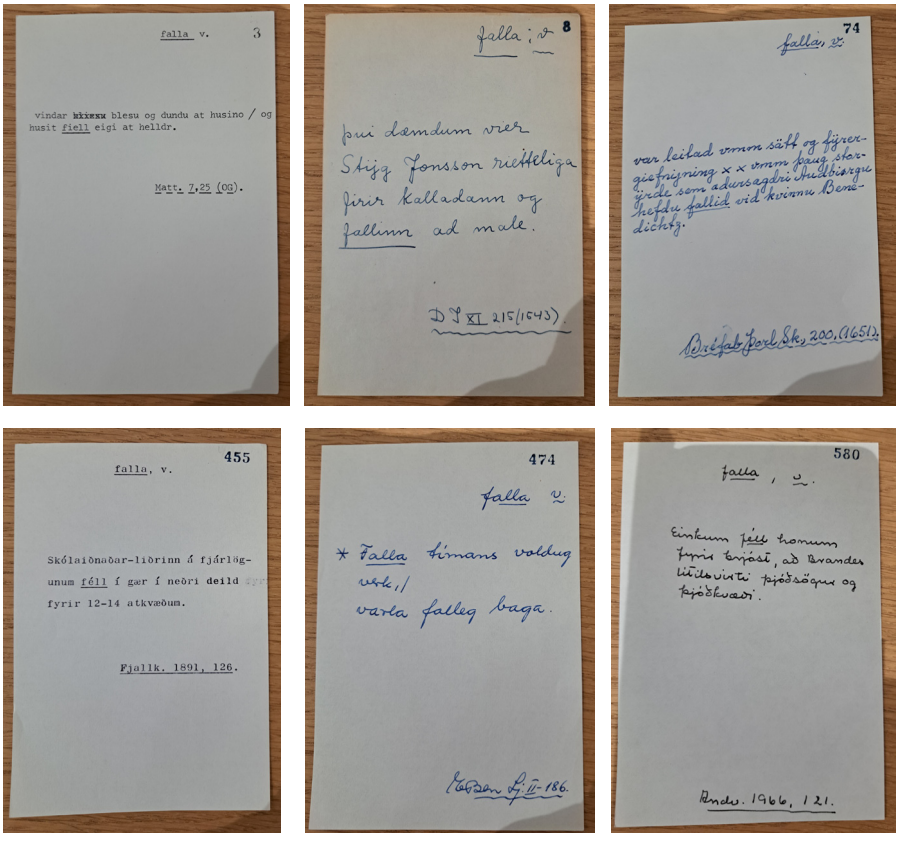
Mynd 1. Dæmi um orðtöku, annars vegar úr bréfum Geirs Vídalíns (bréf frá 1805 á síðu 37) og hins vegar úr bók Páls Melsteð frá 1875 (hluti af bls. 59), með undirstrikunum og afmörkun notkunardæma.

Síðan voru öll notkunardæmi úr viðkomandi heimild skráð stafrétt á sérstaka seðla, eitt dæmi á hvern seðil, ásamt uppflettimynd orðsins sem um ræddi og tilvísun til heimildarinnar með skammstöfun og eftir atvikum til bindis eða árgangs og blaðsíðutals. Þetta sést glögglega á *Mynd 2*. Efst í hægra horni seðils er uppflettiorðið ásamt orðflokki (m = karlkynsnafnorð, n = hvorugkynsnafnorð, adj = lýsingarorð). Á miðjum seðli er notkunardæmið eins og það birtist í orðtekna textanum (þetta er hluti dæmanna sem merkt var við í orðtökutextunum á *Mynd 1*). Loks er neðst heimildaskammstöfun (GVídBr = Geir biskup góði í vinarbréfum; MelNs = Páll Melsteð: Nýja sagan), bindi ef við á (II, 1) og blaðsíðutal.



Mynd 2. Dæmi um skráningu notkunardæma á seðla.

Ef útgáfutími heimildarritsins dugði ekki til að aldurssetja dæmið, t.d. þegar um var að ræða safnrit með textum frá ólíkum tímum eða dæmi tekið úr tilvitnun í eldri heimild, var ártal skráð í sviga aftan við heimildartilvísunina og ræður það á aldursgreiningunni (sjá *Mynd 3* sem líka sýnir ólíka seðla með sama uppflettiorði).



Mynd 3. Sýnishorn af seðlum með notkunardæmi um sögnina *falla* frá ýmsum tímum og úr mismunandi heimildum. Í sumum tilvikum fylgir ártal á eftir heimildarskammstöfuninni og það ræður þá aldursgreiningunni. Seðlarnir eru ýmist handskrifaðir eða vélritaðir og á þeim eru ólíkar rithendur.

3 Seðlasafnið

Að skráningu lokinni var seðlunum raðað í stafrófsröð eftir uppflettiorðinu í sérsniðna tréstokka eða kassa (sjá *Mynd 4*).



Mynd 4. Seðlakassi. Notkunardæmi um tiltekið uppflettiorð standa saman í aldursröð og uppflettiorðum er skipað í stafrófsröð.

Þannig stóðu öll dæmi um tiltekið orð saman og væru dæmin mörg var seðlunum raðað innbyrðis í aldursröð þannig að elstu dæmin um orðið væru fremst og þau yngstu aftast. Fremst í hvern seðlakassa var sett spjald sem sýndi hvaða bil í stafrófinu dæmin í honum spanna og kössunum var svo komið fyrir í stafrófsröð í hilluskáp (sjá *Mynd 5*).





Mynd 5. Seðlakassar merktir fremsta og aftasta orðinu í hverjum kassa. Kössunum er raðað í geymsluskápa.

Smám saman jókst safnið að vöxtum og skipuleg skráning og röðun efnisins gerði notendum kleift að ganga að dæmum um tiltekið orð á vísum stað í safninu. Við orðabókagerð eða aðrar rannsóknir

á notkun og merkingu orðs mátti síðan endurraða seðlunum eftir ólíkum viðmiðum meðan unnið var með þá, t.d. eftir formlegum, merkingarlegum eða setningarlegum einkennum. Seðlasafnið hefur ævinlega verið aðgengilegt í húsakynnum Orðabókar Háskólans og síðar Stofnunar Árna Magnússonar í íslenskum fræðum og þar hefur starfsfólk, orðabókaritstjórar, fræðimenn, háskólastúdentar og aðrir getað nýtt sér það í námi, starfi og rannsóknum. Í safninu eru liðlega tvær milljónir seðla í ríflega 700 seðlakössum. Því hefur nú verið komið fyrir í nýju húsi Árnastofnunar, Eddu, og þar verður það áfram opið fræðimönnum í handritalestrarsal ásamt öðrum söfnum Orðabókar Háskólans.

4 Rafrænt gagnasafn

Á síðustu áratugum 20. aldar urðu byltingarkenndar breytingar á vinnuaðferðum við orðabókagerð með tilkomu tölvutækninnar. Orðabók Háskólans var frumkvöðull hér á landi í nýtingu hennar í hugvísindum og laust eftir 1980 var farið að huga að tölvuvæðingu ritmálssafnsins. Fyrsti áfanginn í því starfi miðaði að því að fá betri yfirsýn yfir efniviðinn í safninu, hvaða orð þar væri að finna, hversu miklar heimildir væru um þau og frá hvaða tíma. Í kynningarbæklingi frá þessum tíma segir (OH 1984:21):

> þótt fróðleikurinn sem geymdur er í söfnum Orðabókarinnar sé geipilegur að vöxtum þá takmarkar umfangið oft þau not sem af honum er að hafa. Þetta stafar af því að seðlum er raðað í stafrófsröð […]. Þess vegna er auðvelt að afla sér upplýsinga um einstök orð í safninu en önnur vitneskja er torfundin. Þannig hefur verið nánast ógerlegt að finna bakstöðuliði (t.d. mann*legur* eða rit*höfundur*) í viðskeyttum og samsettum orðum. Sömu sögu er að segja vilji menn leita að orðafari einstakra tímabila, höfunda eða landshluta. Upplýsingarnar eru í söfnunum en hingað til hefur verið torvelt að ná til þeirra.

Til að bæta úr slíkum annmörkum var ráðist í gerð rafrænnar yfirlitsskrár um öll uppflettiorð í safninu þar sem eftirfarandi upplýsingar voru tilgreindar:

(1) uppflettimynd orðsins (t.d. **haus**, **hausta**);

(2) orðflokkur og kyn nafnorðs (t.d. **m** = masculinum, þ.e. nafnorð í karlkyni, **v** = verbum, þ.e. sagnorð);

(3) fjöldi dæma um orðið (t.d. **3** = þrjú dæmi, **>** = fleiri en 5 dæmi);

(4) tímabilið sem dæmin spanna (miðað við aldarþriðjunga, t.d. **17f–20s** = frá fyrsta þriðjungi 17. aldar til síðasta þriðjungs 20. aldar)

(5) gerð orðsins (**g** = grunnorð, **v** = viðskeytt orð, **f** = forskeytt orð, **s** = samsett orð)

(6) elsta heimild um orðið í safninu (skammstöfun, t.d. **ÓDavÞj. II, 37** = Ólafur Davíðsson: Íslenzkar þjóðsögur, 2. bindi, bls. 37)

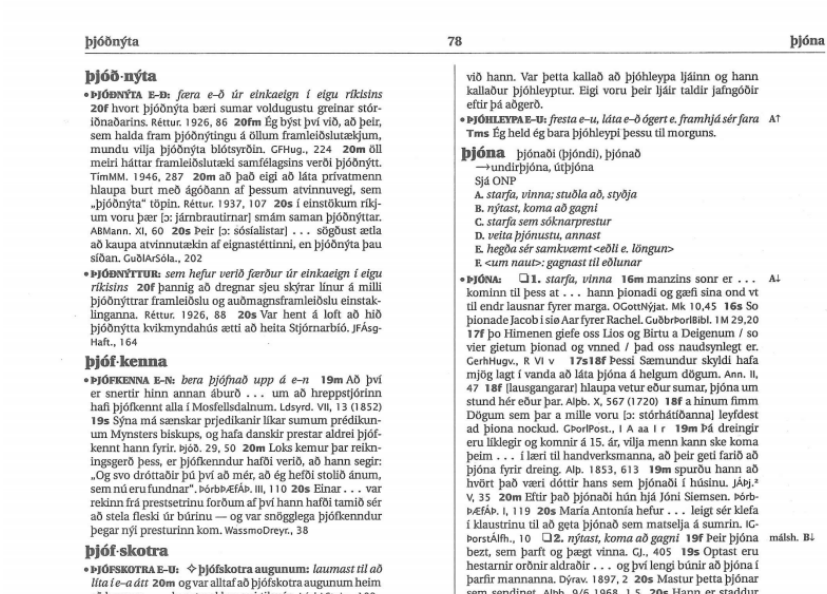
Samhliða var gerð yfirlitsskrá um þær heimildir sem höfðu verið orðteknar. Í henni var skráð skammstöfun heimildarinnar í safninu (sbr. (6) að ofan; sjá líka sýnishorn af seðlum á *Mynd 2* og *3*) ásamt hefðbundnum bókfræðilegum upplýsingum um viðkomandi rit: höfundur þar sem það á við, titill, útgáfuár, útgefandi o.s.frv. Jafnframt var hver heimild tímasett eftir ákveðnu kerfi þar sem hverri öld var skipt í þrennt, t.d. 19f = 1801–1833, 19m = 1834–1867, 19s = 1868–1900 (sbr. (4) hér að ofan). Oftast miðast aldurinn við útgáfuár heimildarritsins en ef notkunardæmi úr viðkomandi heimild eru almennt sérmerkt með ártali ræður það aldursgreiningu dæmanna. Í heimildaskránni stendur þá „ármerkt“ við ritið í stað viðeigandi aldursmerkingar m.t.t. útgáfuárs. Loks voru skráðar upplýsingar um einkenni heimildarinnar, s.s. efnisflokk og tegund ritsins.

Þessar tvær skrár veittu áður óþekkta sýn á efniviðinn og skipuleg skráning upplýsinganna í gagnagrunn gaf færi á að leita í efninu frá ólíkum hliðum. Meðal þess sem draga mátti út úr skránum voru ýmsar tölulegar upplýsingar eins og fjöldi uppflettiorða í safninu sem reyndist vera ríflega 600 þúsund.[1] Einnig kom í ljós að notkunardæmin dreifðust mjög ójafnt á uppflettiorðin. Um stærstan hluta þeirra var bara eitt dæmi í safninu (rúmlega 60% orðanna), um ríflega fjórðung orða reyndust vera 2–5 dæmi og einungis 12% uppflettiorðanna

[1] Nýrri tölur benda til þess að í rafræna gagnasafninu séu nú u.þ.b. 700 þúsund uppflettiorð.

höfðu sex eða fleiri dæmi á bak við sig. Áætlaður seðlafjöldi er nálægt 2,3 milljónum. Dreifing eftir orðflokkum reyndist líka mismunandi. Nafnorð eru langflest, yfir 80% af uppflettiorðunum, en jafnframt sá orðflokkur þar sem stakdæmi eru mest áberandi. Um næstum 65% nafnorða er bara eitt notkunardæmi og fleiri en sex dæmi um einungis 10% þeirra. Aftur á móti eru sagnorð ekki nema tæplega 3% uppflettiorðanna. Um ríflega 30% sagna eru a.m.k. sex dæmi og bara stakdæmi um 44% þeirra (sjá nánar Ásta Svavarsdóttir o.fl. 1993:9–11, 13).

Í framhaldi af gerð yfirlitsskránna var hafist handa við að skrá sjálf notkunardæmin í áföngum. Í kjölfarið voru gerðar tilraunir með ritstjórn efnisins til þess að láta á það reyna hversu vel efniviðurinn dygði sem grundvöllur orðabókarlýsingar. Fyrsta atrennan beindist að stafabilinu *roskinn–rúm* sem svaraði til einnar opnu í orðabók Sigfúsar Blöndals (1920–1924). Jón Hilmar Jónsson (1988:124–125) hefur gert grein fyrir þessu verkefni og þeim lærdómi sem draga mátti af tilrauninni. Að fenginni reynslu beindist næsta atrenna að sögnum. Notkunardæmi um fjölmargar algengar sagnir voru tölvuskráð og í kjölfarið var unnið að greiningu þeirra til undirbúnings frekari ritstjórnartilrauna (sjá Jón Hilmar Jónsson 1988 og sýnishorn þar af sögnunum *berja* og *geyma*). Sú vinna leiddi til útgáfu *Sýniheftis sagnorðabókar* (Ásta Svavarsdóttir o.fl. 1993) þar sem fjallað var um efniviðinn í söfnum Orðabókar Háskólans og gerð rækileg grein fyrir ritstjórnarvinnunni í inngangi og síðan sýnd dæmi um orðabókartexta sem náði til allra sagna á bilinu *þagga–þjösna* ásamt forskeyttum og samsettum sögnum sem myndaðar eru á grundvelli þeirra (t.d *afþakka*, *hampþétta*, *helþegja*, *vanþakka* o.fl.), alls 113 orðabókargreinar eða flettur (sjá *Mynd 6*).



þjóðnýta 78 þjóna

þjóð·nýta

• **ÞJÓÐNÝTA E–D:** *færa e–ð úr einkaeign í eigu ríkisins* **20f** hvort þjóðnýta bæri sumar voldugustu greinar stóriðnaðarins. Réttur. 1926, 86 **20fm** Ég býst því við, að þeir, sem halda fram þjóðnýtingu á öllum framleiðslutækjum, mundu vilja þjóðnýta blótsyrðin. GFHug., 224 **20m** öll meiri háttar framleiðslutæki samfélagsins verði þjóðnýtt. TímMM. 1946, 287 **20m** að það eigi að láta privatmenn hlaupa burt með ágóðann af þessum atvinnuvegi, sem „þjóðnýta" töpin. Réttur. 1937, 107 **20s** í einstökum ríkjum voru þær [þ: járnbrautirnar] smám saman þjóðnýttar. ABMann. XI, 60 **20s** Þeir [þ: sósíalistar] . . . sögðust ætla að kaupa atvinnutækin af eignastéttinni, en þjóðnýta þau síðan. GuðlArSóla., 202

• **ÞJÓÐNÝTTUR:** *sem hefur verið færður úr einkaeign í eigu ríkisins* **20f** þannig að dregnar sjeu skýrar línur á milli þjóðnýttrar framleiðslu og auðmagnsframleiðslu einstaklinganna. Réttur. 1926, 88 **20s** Var hent á loft að hið þjóðnýtta kvikmyndahús ætti að heita Stjórnarbíó. JFÁsg-Haft., 164

þjóf·kenna

• **ÞJÓFKENNA E–N:** *bera þjófnað upp á e–n* **19m** Að því er snertir hinn annan áburð . . . um að hreppstjórinn hafi þjófkennt alla í Mosfellsdalnum. Ldsyrd. VII, 13 (1852) **19s** Sýna má sænskar þjedikanir líkar sumum prédikunum Mynsters biskups, og hafa danskir prestar aldrei þjófkennt hann fyrir. Þjóð. 29, 50 **20m** Loks kemur þar reikningsgerð þess, er þjófkenndur hafði verið, að hann segir: „Og svo dróttaðir þú því að mér, að ég hefði stolið ánum, sem nú eru fundnar". ÞórbÞÆfÁÞ. III, 110 **20s** Einar . . . var rekinn frá prestsetrinu forðum af því hann hafði tamið sér að stela fleski úr búrinu — og var snögglega þjófkenndur þegar nýi presturinn kom. WassmoDreyr., 38

þjóf·skotra

• **ÞJÓFSKOTRA E–U:** ✧**þjófskotra augunum:** *laumast til að líta í e–a átt* **20m** og var alltaf að þjófskotra augunum heim að bænum . . . hvort nokkur sæi til mín. ÞórbÞStein., 189

þjóf·starta

• **ÞJÓFSTARTA:** *fara of snemma af stað, byrja of snemma á e–u* **20m** Í úrslitunum þjófstörtuðu þeir Árni og Guttormur tvisvar. Árbíþr. 1945, 20 **20s** Í herbúðum Bandaríkjamanna lak fýlan af nánustu ráðgjöfum forsetans sem kvörtuðu yfir því að Rússarnir væru að þjófstarta. IHJHrafn., 94

þjóf·stela

Sjá ONP

• **ÞJÓFSTELA E–U:** *taka e–ð ófrjálsri hendi, stela e–u* **20m** Það vantaði nú ekki annað en að maður færi að þjófstela fyrir hann kaffi, þennan jómfrúgulu-gelding. GHagalKon., 46

• **ÞJÓFSTOLINN:** *sem hefur verið tekinn ófrjálsri hendi e. stolið* **16s** að . . . Þurijdur hefdi jetid þiofstolid hiáá fódur sijnum. Alþb. II, 336 (1592) **17f** að . . . Gissur heffdi borgat vt þioffstoled fie vr sijnum husum. Alþb. III, 240 (1602) **18f** Fóli. . . . Það merkir þjófstolið fé; en það hefir sitt nafn af því, að þjófar vilja gjarnan hafa fólgið við hann. Var þetta kallað að þjóhleypa ljáinn og hann kallaður þjóhleyptur. Eigi voru þeir ljáir taldir jafngóðir eftir þá aðgerð.

• **ÞJÓHLEYPA E–U:** *fresta e–u, láta e–ð ógert e. framhjá sér fara* **Tms** Ég held ég bara þjóhleypi þessu til morguns. A†

þjóna þjónaði (þjóndi), þjónað
→undirþjóna, útþjóna
Sjá ONP
A. *starfa, vinna; stuðla að, styðja*
B. *nýtast, koma að gagni*
C. *starfa sem söknarprestur*
D. *veita þjónustu, annast*
E. *hegða sér samkvæmt <eðli e. lögun>*
F. *<um naut>: gagnast til eðlunar*

• **ÞJÓNA:** ❑**1.** *starfa, vinna* **16m** manzins sonr er . . . kominn til þess at . . . hann þionadi og gæfi sina ond vt til endr lausnar fyrer marga. OGottNýjat. Mk 10,45 **16s** So þionade Jacob i sio Aar fyrer Rachel. GuðbÞorlBibl. 1M 29,20 **17f** þo Himenen giefe oss Lios og Birtu a Deigenum / so vier gietum þionad og vnned / þad oss naudsynlegt er. GerhHugv., R VI v **17s**/**18f** Þessi Sæmundur skyldi hafa mjög lagt í vanda að láta þjóna á helgum dögum. Ann. II, 47 **18f** [lausgangarar] hlaupa vetur eður sumar, þjóna um stund hér eður þar. Alþb. X, 567 (1720) **18f** a hinum fimm Dögum sem þar a mille voru [þ: stórhátíðanna] leyfdest ad þiona nockud. GÞorlÞost., I A aa I r **19m** Þá dreingir eru líklegir og komnir á 15. ár, vilja menn kann ske koma þeim . . . í læri til handverksmanna, að þeir geti farið að þjóna fyrir dreing. Alþ. 1853, 613 **19m** spurðu hann að hvört það væri dóttir hans sem þjónaði í húsinu. JÁþj.[2] V, 35 **20m** Eftir það þjónaði hún hjá Jóni Siemsen. ÞórbÞÆfÁÞ. I, 119 **20s** María Antonía hefur . . . leigt sér klefa í klaustrinu til að geta þjónað sem matselja á sumrin. IGÞorstÁlfh., 10 ❑**2.** *nýtast, koma að gagni* **19f** Þeir þjóna bezt, sem þarft og þægt vinna. GJ., 405 **19s** Óptast eru hestarnir orðnir aldraðir . . . og því lengi búnir að þjóna í þarfir mannanna. Dýrav. 1897, 2 **20s** Mastur þetta þjónar sem sendinet. Alþb. 9/6 1968, 1,5 **20s** Hann er staddur í allstórum sal sem greinilega á að þjóna sem setustofa. GuðlArSóla., 9 **20s** Kirkja eða bænhús í hlaði hefur . . . öðrum þræði þjónað sem stöðutákn. ÍslÞjm. V, 93 ❑**3.** *starfa sem söknarprestur við kirkju* **17s** hann villdi giarnann vid platzed blifa, so leinge sem hann giæte þionad og soknínn villdi med nægiast. ÞréfabBrSv., 232 **18s** Nú kann einn að eiga barn í einni sveit, sem hann þiónar fyrir. LFR. IV, 134 **19f** batnaði þá nokkuð, svo hann [þ: prestur] gat þjónað til veturnátta. Konbréf., 124 (1830) **19m** Hann segist hafa þjónað þar, síðan presturinn hefði horfið þaðan. JÁÞj. II, 37 ✧**þjóna fyrir altari:** *annast helgiathafnir fyrir altari við guðsþjónustu* **19f** Sialands Stiftprófastr þiónadi fyrir Altarinu. Sagnabl. II, 94 **20s** var fjölmenn altarisganga í dómkirkjunni þarsem séra Bjarni . . . þjónaði fyrir altari. SAMFugl., 230 ❑**4.** ✧**þjóna fyrir borðum:** *ganga um beina* **16m** þat er eigi hæfiligt at vier forlatum guds ord / og þionum fyrer bordum. OGottNýjat. P 6,2 **17f** Síra Jón Björnsson þjónaði fyrir borðum. JEBisk., 69 **19s** Engelar bióna fyrir borðum. Þjóð. 34, 105 ✧**þjóna** A↓ málsh. B↓ C↓ a↓ D↓ a↑

Mynd 6. Sýnishorn af orðabókartexta úr *Sýnihefti sagnorðabókar* (1993).

Þessi vinna skilaði dýrmætri reynslu innan stofnunarinnar auk þess sem sýniheftinu var dreift víða til kynningar.

Þriðja atrennan í tölvuskráningu efnisins naut stuðnings Lýðveldissjóðs, sem stofnað var til í tilefni af 50 ára afmæli íslenska lýðveldisins og skyldi m.a. styrkja verkefni á sviði íslenskrar tungu. Vinnan, sem fór fram á árunum 1995–1997, miðaði að því að koma sem stærstum hluta dæmasafnsins í rafrænt form. Þótt styrkurinn væri myndarlegur og skipti sköpum fyrir verkefnið var ljóst frá upphafi að safnið væri of stórt til að hann nægði til þess að tölvuskrá það allt. Frekar en að þurfa að hætta í miðju stafrófi var ákveðið að ná utan um þann hluta orðaforðans sem var algengur og langlífur, þ.e.a.s. orð sem verulegar

heimildir voru um og frá löngu tímabili. Í fyrstu var því afráðið að sleppa orðum með einungis eitt notkunardæmi í safninu á bakvið sig, m.a. með þeim rökum að í uppflettiskránni væru þegar talsverðar upplýsingar um þau orð (aldur og uppruni dæmisins). Fljótlega varð þó ljóst að þetta dygði ekki til og var þá farið að sleppa orðum sem örfá dæmi voru um við tölvuskráninguna og síðar einnig dæmum úr bundnu máli sem oft eru torráðin og gefa takmarkaða vitneskju um venjulega notkun og merkingu orða. Með þessu móti náðist það markmið að komast til enda í stafrófinu þótt bagalegt sé að það hafi þurft að skilja eftir eyður í dæmasafninu. Síðar hefur að einhverju leyti verið stoppað í slík göt en eigi að síður eru enn fjölmörg notkunardæmi einungis varðveitt í seðlasafninu þótt þau séu fá í hlutfalli við þau sem finna má í rafræna dæmasafninu sem nú geymir um 1,8 milljónir notkunardæma. Í gagnasafninu eru einnig notkunardæmi sem ekki eru í seðlasafninu. Þetta eru dæmi úr heimildum sem orðteknar voru eftir að tölvuvinnsla hófst á stofnuninni og voru tölvuskráð beint eftir heimildinni. Í upphafi voru þessi dæmi prentuð á seðla og raðað inn í safnið en því var síðar hætt og þau því aðeins varðveitt í rafræna safninu. Þetta eru því hlutfallslega fá dæmi en notendur verða samt að hafa í huga að þótt efniviðurinn í seðlasafninu og í rafrænni gerð ritmálssafnsins sé að langmestu leyti sá sami þá er samræmið ekki algert. Yfirlitsskráin nær eigi að síður til allra uppflettiorðanna, hvort sem dæmi um þau eru á báðum stöðum, einungis í seðlasafninu eða bara í rafræna safninu. Rafrænu gögnunum var komið fyrir í gagnagrunni þar sem ólíkir hlutar þeirra eru samtengdir þannig að öll (tölvuskráð) dæmi um tiltekið uppflettiorð eru tengd saman og við upplýsingar um orðið í yfirlitsskránni og heimildartilvísanir eru sömuleiðis tengdar við heimildaskrána og þær upplýsingar sem þar eru um viðeigandi rit.

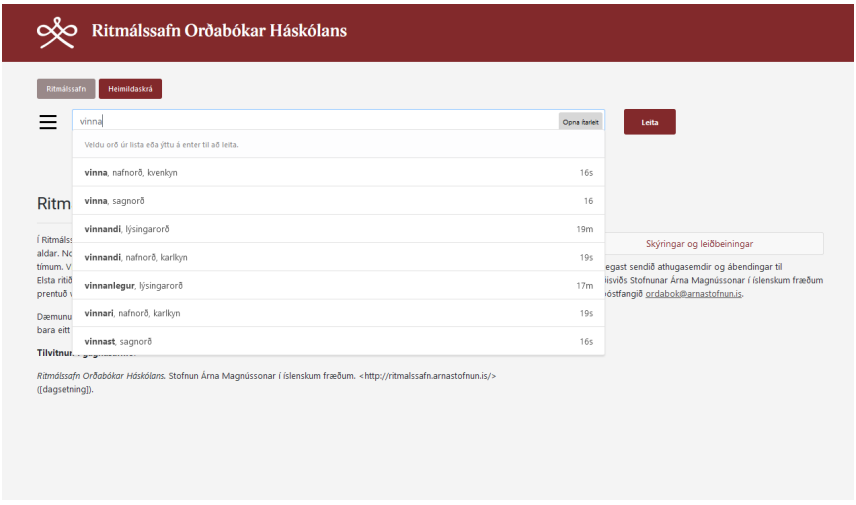
5 Vefaðgangur

Í upphafi voru rafrænu gögnin aðeins aðgengileg á tölvum Orðabókar Háskólans og nýttust því fyrst og fremst starfsfólki stofnunarinnar við vinnu sína. Þegar nettengingar urðu öflugri og almennari undir aldamótin og með tilkomu veraldarvefsins var ráðist í að koma upp vefaðgangi að gögnunum sem nýst gæti fleirum, bæði innan og utan stofnunar. Síðan hefur safnið verið öllum opið til leitar þótt hún hafi verið ýmsum takmörkunum háð. Lengi framan af einskorðaðist

leit í safninu að mestu við uppflettiorð og notendur þurftu því að vita eftir hverju þeir væru að slægjast. Þessar takmarkanir á leitarmöguleikum spruttu í upphafi af því að tölvukerfið var ekki nógu öflugt til að leyfa fjölbreytilegri möguleika og síðar urðu tafir á nauðsynlegri endurnýjun gagnagrunnsins ásamt manneklu til þess að þær endurbætur sem þurfti til að gera leitina sveigjanlegri létu á sér standa. Í árslok 2018 voru gögnin loks flutt í nýjan gagnagrunn og í kjölfarið var komið upp nýju vefviðmóti fyrir ritmálssafnið á vef Stofnunar Árna Magnússonar í íslenskum fræðum. Síðan hefur verið unnið að því að auka og efla safnið til hagsbóta fyrir notendur þess, ekki síst með því að nýta betur þær upplýsingar sem þegar var að finna í gagnasafninu til að auka sveigjanleika í leit. Úrbætur hafa líka verið gerðar á sjálfum efniviðnum, t.d. hefur dæmum í rafræna safninu fjölgað talsvert, ýmiss konar samræming hefur átt sér stað og margvíslegar villur hafa verið leiðréttar. Í þessum kafla er leitast við að gera grein fyrir helstu möguleikum til að nálgast efniviðinn í safninu eins og það birtist almennum notendum á vef stofnunarinnar.

5.1 Leit í ritmálssafni

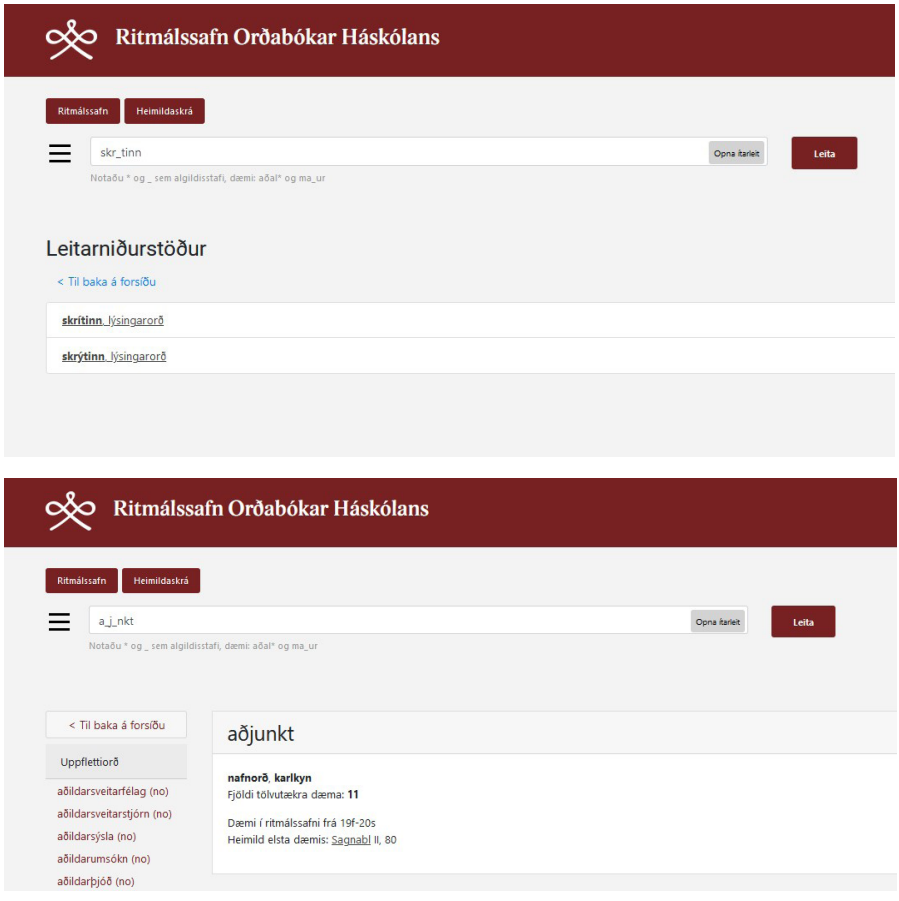
Leit í safninu byggist á grunnmynd uppflettiorða. Þar er átt við þá mynd sem almennt er höfð sem fulltrúi orðsins í orðabókum og öðrum uppflettiritum – nefnifall eintölu af nafnorðum (t.d. *bolli*, *lærdómur*, *kápa*, *blað*), nefnifall eintölu karlkyn af lýsingarorðum (t.d. *gulur*, *skemmtilegur*) og nafnháttur af sagnorðum (t.d. *detta*, *ferðast*). Einfaldasti leitarmöguleikinn er að leita að einu tilteknu uppflettiorði með því að slá grunnmynd þess inn í leitargluggann (sjá *Mynd 7*). Þá birtist fellilisti með tillögum um möguleg uppflettiorð ásamt orðflokki og kyni og velja má viðeigandi orð af honum. Þetta er ekki síst gagnlegt þegar tvö eða fleiri orð hafa sömu uppflettimynd, t.d. *vinna* (nafnorð eða sagnorð) eða *sykur* (karlkyn eða hvorugkyn).



Mynd 7. Upphafsskjár Ritmálssafns Orðabókar Háskólans á slóðinni <https://ritmalssafn.arnastofnun.is/>.

Á *Mynd 7* sést hvernig stafastrengurinn „vinna“ hefur verið sleginn inn í leitargluggann og þá birtist fellilisti með orðum sem til greina koma: efst nafnorðið *vinna*, þá samhljóma sögn og síðan önnur uppflettiorð sem byrja á strengnum *vinna-*. Notandi getur valið orð af fellilistanum eða smellt á „Leita“ og þá birtast einungis þau orð sem svara nákvæmlega til leitarstrengsins, í þessu tilviki nafnorðið og sögnin *vinna*.

Viðmótið býður auk þess upp á möguleika til að takmarka eða útvíkka leitina á ýmsan hátt. Algildistáknin _ og * eru gagnleg, m.a. til að finna fleiri en eitt uppflettiorð í einu. Það fyrra, undirstrikið, stendur fyrir einn bókstaf sem getur verið hvaða stafur sem er og nýtist einkum ef vafi leikur á um ritháttt leitarorðsins, annaðhvort vegna þess að notandinn er ekki viss um hver hann er, t.d. hvort uppflettimyndin er *skrítinn* eða *skrýtinn*, eða ef fleiri en einn möguleiki kemur til greina, einkum í gömlum sjaldgæfum orðum eða tökuorðum sem hafa ekki fastmótaðan ritháttt, t.d. er ýmist skrifað *adjunkt*, *adjúnkt*, *aðjunkt* eða *aðjúnkt* í notkunardæmunum (sjá *Mynd 8*).



Mynd 8. Dæmi um leit með algildistákninu _, sem svarar til hvaða bókstafs sem er.

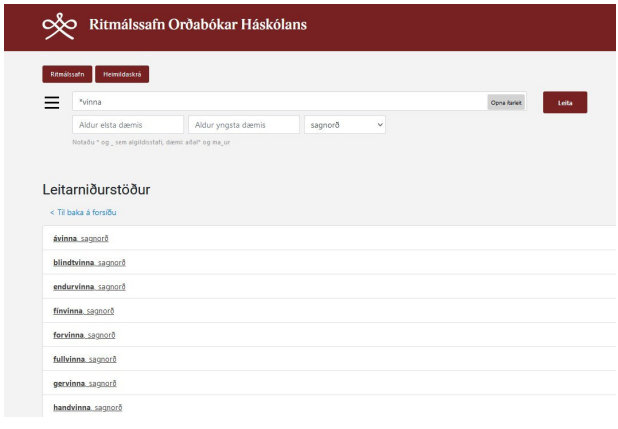
Á *Mynd 8* sést þegar annars vegar var leitað að strengnum „skr_tinn“ og hins vegar „a_j_nkt“. Í fyrra tilvikinu reyndust uppflettimyndirnar *skrítinn* og *skrýtinn* báðar vera í safninu en í því síðara einungis uppflettimyndin *aðjunkt*.

Almennt eru uppflettiorðin rituð samkvæmt reglum um nútímastafsetningu og unnið hefur verið að samræmingu samkvæmt því. Eigi að síður má víða finna ýmiss konar misræmi í rithætti innan safnsins og dæmi eru um að sama orðið hafi fleiri en eina uppflettimynd þótt slíkum tilvikum fari fækkandi. Auk þess getur verið álitamál hvaða afbrigði eigi að fella undir eitt og sama uppflettiorð og hver ekki. Notendur þurfa því að vera á varðbergi gagnvart mögulegum tilbrigðum, annaðhvort með því að nota algildistákn eða prófa aðra

möguleika ef þeir finna ekki það sem þeir leituðu fyrst að því að orðið getur leynst undir öðrum rithætti eða jafnvel undir fleiri uppflettimyndum en þeirri sem leitað var að í upphafi.[2]

Síðara algildistáknið, stjarna (*), stendur fyrir ótilgreindan fjölda bókstafa, einn eða fleiri. Það má nota í sama tilgangi og strikið en það er þó einkum gagnlegt til að finna hópa orða með sameiginleg einkenni, t.d. orð sem byrja eða enda á sama hátt. Þannig má finna öll orð í safninu sem enda á viðskeytinu *-legur*, öll samsett orð með *-tjörn* að síðari eða síðasta lið eða öll orð með forliðnum *kven-*.

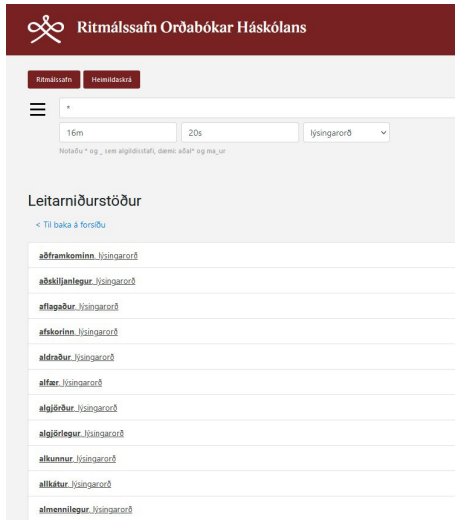
Með því að nota ítarleit má afmarka leitina frekar, með tilliti til aldurs dæmanna og/eða orðflokks (sjá *Mynd 9*). Með leitarstrengnum „*vinna“ finnast öll uppflettiorð með *-vinna* sem síðasta lið. Leitin skilar 884 orðum sem ýmist eru nafnorð eða sagnorð. Hafi notandinn einkum áhuga á sögnunum er einfaldara og skýrara að velja „sagnorð“ í ítarleitinni en að vinsa þau úr handvirkt og þá stendur eftir listi með 31 orði.



Mynd 9. Leit með algildistákninu *, sem stendur fyrir ótilgreindan stafastreng („*vinna“ = öll orð sem enda á *-vinna*), ásamt orðflokknum „sagnorð“ í ítarleit. Á myndinni er einungis hluti sagnanna.

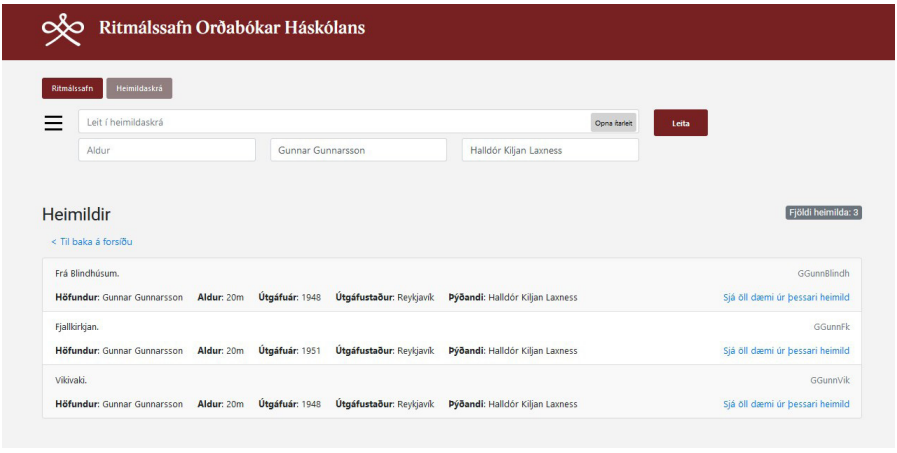
[2] Hér er einnig vert að nefna að í safninu leynast ýmiss konar villur sem eru leiðréttar eftir því sem þær koma í ljós en aldrei er hægt að komast algjörlega fram hjá þeim í svo stóru safni. Þetta eru einkum innsláttarvillur eða ritvillur sem hafa sprottið af mislestri á seðlunum, sem margir eru handskrifaðir, og þótt þær séu hvimleiðar eru þær flestar skaðlitlar. Ætli notendur að taka upp dæmi úr safninu má nýta heimildartilvísunina til að bera þau saman við frumtextann ef ástæða þykir til. Í stöku tilvikum geta slíkar villur verið í sjálfum uppflettiorðunum eða að villur í innsláttarskrám hafa valdið því að dæmi hafa lent undir röngu uppflettiorði. Ábendingar frá notendum um hvers kyns villur sem þeir rekast á í safninu eru afar vel þegnar.

Glöggur lesandi tekur fljótt eftir að sumar sagnanna sem enda á *-vinna* hafa í rauninni *-tvinna* sem síðari samsetningarlið. Þar sem leitin byggist á bókstafastrengjum er ekki hægt að koma í veg fyrir að hún skili oft líka óviðkomandi niðurstöðum. Einnig má nota ítarleitina án leitarstrengs til þess að finna orð af tilteknum orðflokki í safninu, t.d. allar upphrópanir eða allar forsetningar, eða til þess að finna öll orð eða orð af tilteknum flokki sem dæmi eru um á ákveðnu tímaskeiði, t.d. lýsingarorð sem koma fyrir á öllu tímabilinu frá miðri 16. öld til seinasta þriðjungs 20. aldar (sjá *Mynd 10*; þau reyndust vera 575 af alls rúmlega 92 þúsund lýsingarorðum).



Mynd 10. Leit að öllum lýsingarorðum í safninu (* = hvaða stafastrengur sem er) þar sem elsta dæmið er frá miðri 16. öld og það yngsta frá síðasta hluta 20. aldar. Á myndinni sjást einungis fyrstu orðin í stafrófinu sem samræmast leitarskilyrðunum.

Notendur geta líka leitað í safninu gegnum heimildaskrána með því að smella á viðeigandi flipa á forsíðu vefsins. Í leitargluggann má færa heimildarskammstöfun og þá birtast upplýsingar um viðkomandi heimild. Þarna má líka nota algildistákn, t.d. HKL* til að kalla fram öll orðtekin rit Halldórs Kiljan Laxness og með því að nýta ítarleit má skilgreina leitina nánar, t.d. er hægt að kalla fram öll orðtekin rit frá ákveðnu tímabili, rit eftir tiltekinn höfund og rit sem voru íslenskuð af ákveðnum þýðanda (sjá *Mynd 11*).

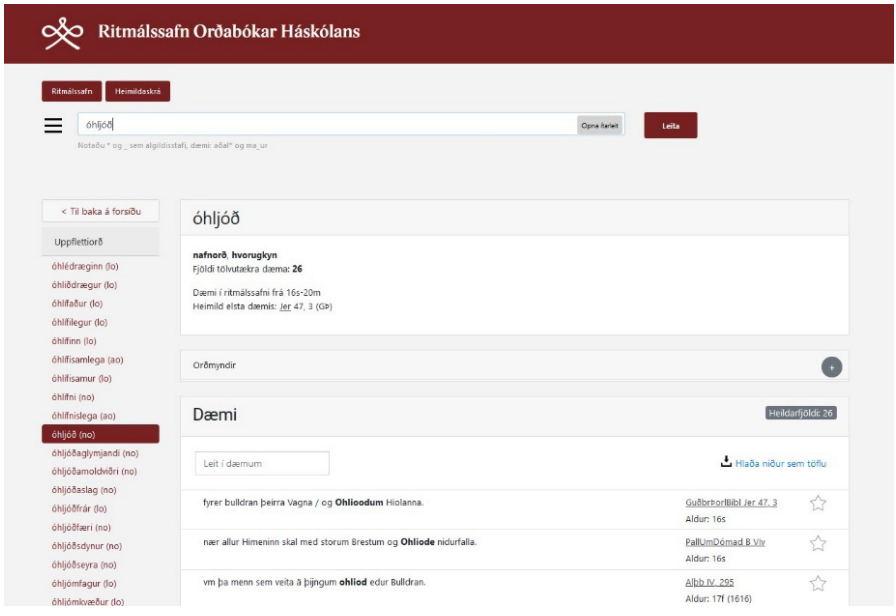


Mynd 11. Ítarleit í heimildaskrá ritmálssafnsins að orðteknum ritum Gunnars Gunnarssonar í þýðingu Halldórs Kiljan Laxness.

Auk upplýsinga um hverja einstaka heimild er mögulegt að kalla fram öll dæmi úr henni sem tiltæk eru í rafræna gagnasafninu, óháð uppflettiorði, með því að velja „Sjá öll dæmi úr þessari heimild“ hægra megin við heimildina (sjá *Mynd 11)*. Í dæmalistanum sem þá birtist er svo tengill í hverju dæmi sem leiðir notendur áfram að viðkomandi uppflettiorði í ritmálssafninu.

5.2 Niðurstöður leitar

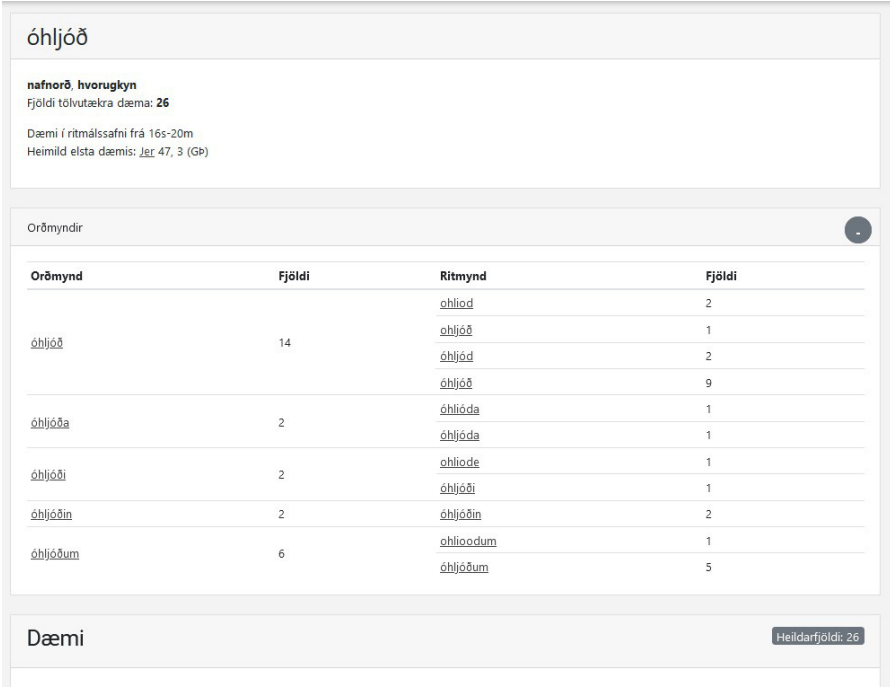
Leit að tilteknu uppflettiorði skilar upplýsingum og dæmum um orðið auk lista til vinstri yfir orð sem standa næst því í stafrófsröðinni (sjá *Mynd 12*). Þau eru oft skyld því, ekki síst samsetningar sem hafa leitarorðið sem fyrri lið, og gagnlegt getur verið að skoða orðin í samhengi hvert við annað. Hægt er að smella á hvert þessara orða og kalla þannig fram upplýsingar og eftir atvikum dæmi um orðið og notkun þess.



Mynd 12. Dæmi um niðurstöðu leitar að uppflettiorðinu *óhljóð*. Í efsta rammanum eru upplýsingar um orðið, hversu miklar heimildir um það eru í safninu og frá hvaða tímabili.

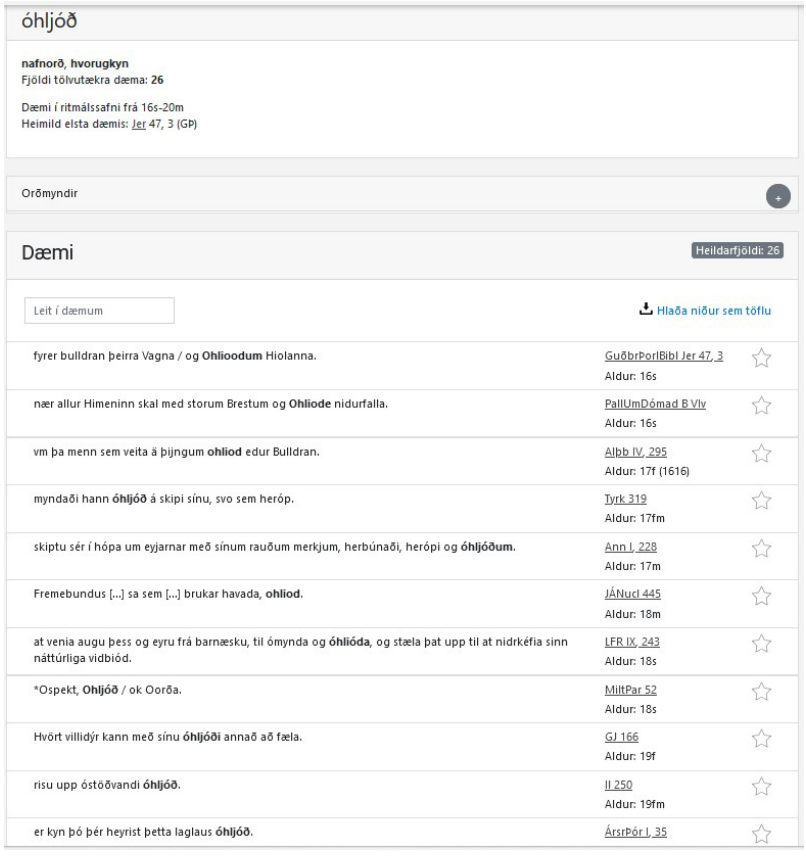
Í efsta rammanum til hægri birtast helstu upplýsingar um orðið: orðflokkur auk kyns nafnorða, fjöldi tiltækra dæma (í rafræna gagnasafninu), tímabilið sem dæmin spanna (þ.e. aldur elsta og yngsta dæmisins í safninu) ásamt tilvísun til elstu heimildar um orðið (með því að smella á skammstöfunina sjást nánari upplýsingar um ritið). Næsti rammi, sem þarf að opna með því að smella á plúsmerkið hægra megin (sjá *Mynd 13*), sýnir töflu með þeim orðmyndum sem koma fyrir í dæmunum; í hægri dálkinum birtast stafréttu myndirnar úr dæmunum sjálfum en í þeim vinstri er samsvarandi stöðluð nútímamálsmynd og búið er að fella tilbrigði undir hana, t.d. „ohlioodum“ og „óhljóðum“ undir orðmyndina *óhljóðum*.[3]

[3] Stöðlun orðmynda er ekki að fullu lokið og þess vegna er stafrétta myndin í sumum tilvikum líka í vinstri dálki þótt hún sé ólík nútímamálsmyndinni.



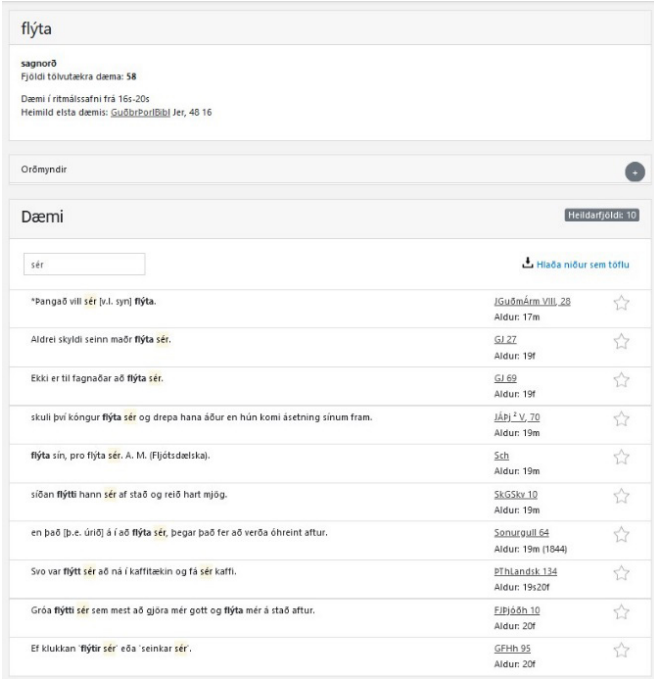
Mynd 13. Yfirlitstafla um þær orðmyndir sem koma fyrir í dæmum um orðið *óhljóð* ásamt fjölda dæma um hverja mynd. Í hægri dálki (Ritmynd) eru stafréttar myndir eins og þær koma fram í dæmunum en í þeim vinstri (Orðmynd) eru þær stöðluðu nútímamálsmyndir sem stafréttu myndirnar svara til. Orðmyndirnar birtast í stafrófsröð í töflunni.

Í þriðja rammanum birtast öll dæmi um orðið í rafræna gagnasafninu í aldursröð (sjá *Mynd 14*). Til vinstri er notkunardæmið sjálft með upp-flettiorðinu feitletruðu og hægra megin við það er heimildartilvísun og aldursmerking dæmisins.



Mynd 14. Hluti dæma um uppflettiorðið *óhljóð* í safninu (11 elstu dæmin af 26) ásamt heimildartilvísun og aldri hvers dæmis.

Séu dæmin ekki mjög mörg er tiltölulega auðvelt að fá yfirsýn yfir þau en annars eru tvær leiðir til þess að takmarka fjöldann. Annars vegar má smella á einhverja myndanna í orðmyndatöflunni og þá birtast einungis dæmi þar sem sú mynd kemur fyrir í dæmalistanum. Hins vegar má nota leitargluggann ofan við dæmin til þess að leita í þeim. Með strengnum „kald“ í dæmum um nafnorðið *hlátur* finnast t.d. tvö dæmi um orðastæðuna *kaldur hlátur* og leit að myndinni „sér“ í dæmum um sögnina *flýta* skilar einungis dæmum þar sem hún kemur fyrir (sjá *Mynd 15*).



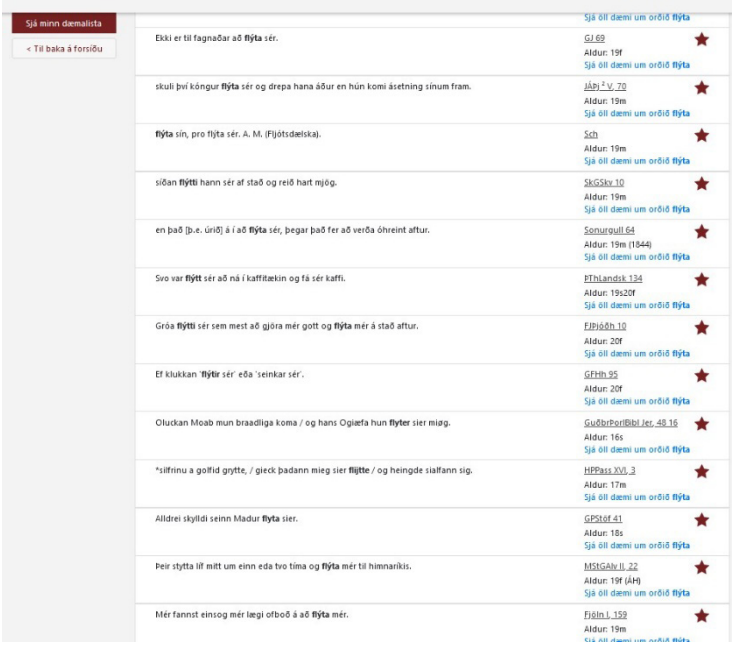
Mynd 15. Niðurstöður leitar að strengnum „sér“ í dæmum um sögnina *flýta* sem skilaði 10 dæmum um sambandið *flýta sér*.

Þessi leitarkostur getur verið gagnlegur ef leitað er að ákveðnum orðasamböndum en þar sem dæmin eru frá ólíkum tímum og tekin upp stafrétt gefur hann misgóða raun, fornafnið *sér* er t.d. stundum ritað „sier“ í gömlum dæmum og ef notandinn áttar sig ekki á því getur hann misst af áhugaverðum dæmum. Það getur því borgað sig að fara í gegnum allt dæmasafnið og leita handvirkt að dæmum um það sem er til athugunar (sjá um aðferðir til þess í 5.3).

5.3 Vinna með niðurstöður

Með nýju vefviðmóti var leit í safninu ekki einungis gerð sveigjanlegri heldur er notendum líka gert auðveldara að vinna með niðurstöðurnar. Aftan við hvert dæmi í leitarniðurstöðunum er óútfyllt stjarna (sjá *Mynd 12*, *14* og *15* í 5.2). Með því að smella á hana er viðkomandi dæmi merkt (þá fyllist stjarnan) og um leið verður til tímabundinn dæmalisti þar sem safna má saman dæmum sem snerta athugunarefnið hverju sinni. Þessi valkostur er ekki síst gagnlegur ef

notendur vilja bera saman notkun mismunandi orða sem eru formlega eða merkingarlega tengd, því að hægt er að safna á listann dæmum sem tilheyra fleiri en einu uppflettiorði. Listinn birtist þegar smellt er á „Sjá minn dæmalista“ vinstra megin á síðunni (sjá *Mynd 16*).



Mynd 16. Sýnishorn af hluta dæmalista sem hefur verið búinn til um orðasambandið *flýta sér*, þar á meðal í 1. og 2. persónu (*flýta mér*/*þér*) og með eldri rithætti (*sier*, *sjer*), með því að nota stjörnumerkingu.

Önnur nýjung er sú að nú er hægt að hlaða niður dæmum í töflu-formi (sem excel-skjali), t.d. dæmum um tiltekið uppflettiorð í heild eða að hluta, dæmum úr ákveðinni heimild (eftir leit í gegnum heimildaskrá, sbr. 5.1 og *Mynd 11*) eða dæmum sem safnað hefur verið á eigin dæmalista með því að stjörnumerkja valin dæmi eins og fyrr var lýst (sjá *Mynd 16*). Það gerir notendum kleift að vinna með dæmin á ýmsan hátt. Auðvelt er t.d. að eyða óviðkomandi dæmum og fá þannig betri yfirsýn yfir það sem athyglin beinist að, t.d. tiltekið orðasamband, setningagerð eða merkingu, og sömuleiðis að bæta við eigin greiningarþáttum sem síðan má nýta til að flokka dæmin á margvíslega vegu.

6 Niðurlag

Ritmálssafn Orðabókar Háskólans hefur orðið til á löngum tíma og þar hafa nokkrar kynslóðir starfsfólks orðabókarinnar og síðar Stofnunar Árna Magnússonar í íslenskum fræðum lagt hönd á plóg við skipulag, umsjón, orðtöku, uppskrift notkunardæma, röðun seðla, tölvuskráningu, forritun, vefhönnun og ýmiss konar greiningu og úrvinnslu. Hópurinn er stór og því eru engin nöfn nefnd hér en í honum hafa verið málfræðingar, aðstoðarfólk og fjölmargir háskólastúdentar í sumarvinnu. Þótt eðli safnsins sé í grundvallaratriðum hið sama og lagt var upp með hefur hver nýr áfangi haft mótandi áhrif á framsetningu þess og aðgang að því fyrir orðabókaritstjóra, fræðimenn, stúdenta og almenning eins og lýst hefur verið hér að framan. Óhætt er að fullyrða að safnið sé mikilvægasta heimildin sem tiltæk er um íslenskan orðaforða, þróun hans og sögu á síðustu 450–500 árum og það hefur nýst á margvíslegan hátt, ekki síst við málrannsóknir og orðabókagerð, eins og fjölmargar tilvísanir til þess í ræðu og riti bera vitni um. Tækniþróun á undangengnum áratugum hefur haft mikil áhrif á aðferðir og verklag við orðabókagerð og á útgáfu orðabóka. Tölvutæknin hefur m.a. leitt til þess að nú hafa rafræn textasöfn og málheildir með samfelldum leitarbærum textum leyst efnissöfnun með orðtöku og gerð stórra seðlasafna af hólmi sem aðalefniviðurinn við gerð orðabóka um samtímamálið og rannsóknir á því. Þess vegna hefur hefðbundinni orðtöku verið hætt og þar af leiðandi bætast ekki fleiri notkunardæmi í safnið. Það þýðir þó alls ekki að ritmálssafnið hafi misst gildi sitt sem heimild um íslensk orð og orðaforða á því tímabili sem það spannar. Í fyrsta lagi hefur farið fram talsverð úrvinnsla og greining á efniviðnum þannig að þar liggja upplýsingar sem ekki eru annars staðar og þó að ekki verði bætt við efnið er ætlunin að vinna áfram að frekari flokkun og greiningu efniviðarins og gera safnið þar með aðgengilegra og nýtilegra. Í öðru lagi ná þau rafrænu íslensku textasöfn og málheildir sem nú eru tiltæk fyrst og fremst til nýlegra texta, efnis frá þessari öld og síðustu 10–15 árum 20. aldar, og einungis að mjög litlu leyti til texta frá fyrri tímum.[4] Unnið er

[4] Í *Íslensku textasafni* á vef Stofnunar Árna Magnússonar í íslenskum fræðum er allnokkuð af eldri leitarbærum textum sem m.a. voru notaðir til að fylla í göt í efniviðnum við gerð *Sýniheftis sagnorðabókar* (Ásta Svavarsdóttir o.fl. 1993:16–17; sjá líka Jörgen Pind o.fl. 1993). Textarnir í því safni eru þó algjörlega ógreindir (orðmyndirnar eru m.ö.o. hvorki málfræðilega markaðar né tengdar viðeigandi

að því að koma upp slíkum söfnum með eldri textum og talsvert er þegar til af 19. og 20. aldar textum í rafrænni gerð og þeim fer fjölgandi. Uppbygging og framsetning málheilda með gömlum textum er þó að mörgu leyti flóknari og tímafrekari en málheilda með samtímatextum. Þangað til slíkar málheildir verða tiltækar er mikilvægi ritmálssafnsins ótvírætt og auk þess eru í safninu gögn og greining sem gætu komið að góðum notum við gerð málheilda með eldri textum.

Heimildir

Alexander Jóhannesson. 1919a. Orðabókin. *Ísafold* 1919(3):2–3.
Alexander Jóhannesson. 1919b. Enn um orðabókina. *Ísafold* 1919(5):2.
Ásta Svavarsdóttir, Guðrún Kvaran, Jón Hilmar Jónsson og Kristín Bjarnadóttir. 1993. *Sýnihefti sagnorðabókar*. Rannsóknar- og fræðslurit 3. Reykjavík: Orðabók Háskólans.
Finnur Jónsson. 1919a. „Vísindalega orðabókin“. *Lögrjetta* 1919(14):1.
Finnur Jónsson. 1919b. Orðabókin. Svar. *Lögrjetta* 1919(29):2–3.
Guðmundur Davíðsson [og] Jóhannes L.L. Jóhannsson. 1926–1927. *Orðabókarmál Íslendinga*. [Án útgefanda og útgáfustaðar.]
Halldór Halldórsson. 1969. Alexander Jóhannesson. Háskólamaðurinn. *Andvari* 1969:3–38.
Íslenskt textasafn. Stofnun Árna Magnússonar í íslenskum fræðum. <http://corpus.arnastofnun.is/>.
Jakob Benediktsson. 1955. Det islandske ordbogsarbejde ved Islands universitet. *Scripta Islandica* 6:7–12.
Jakob Benediktsson. 1971. *Íslenzk orðabókastörf*. Erindi dr. Jakobs Benediktssonar á fundi Stúdentaakademíu 6. desember 1969. Rit Stúdentaakademíu II. Reykjavík: Stúdentaakademía.
Jóhannes L.L. Jóhannsson, Jakob Jóh. Smári [og Þórbergur Þórðarson (sýnishorn)]. 1920. *Álit og tillögur um vísindalega íslenzka orðabók ásamt sýnishorni*. Reykjavík: Prentsmiðjan Gutenberg.
Jóhannes L.L. Jóhannsson. 1919a. Vísindalega orðabókin. *Lögrjetta* 1919(3):10.
Jóhannes L.L. Jóhannsson. 1919b. Orðabókin. *Lögrjetta* 1919(18):66.
Jóhannes L.L. Jóhannsson. 1919c. Orðabókin enn. *Lögrjetta* 1919(32):2–3.
Jóhannes L.L. Jóhannsson. 1928. *Fréttir af orðabókarmálinu*. Sérprent úr Verði. Reykjavík: Prentsmiðjan Gutenberg.

uppflettiorði) og leitin einskorðast því við einfalda strengjaleit. Vert er að benda á að meðal textanna er allt dæmasafnið úr Ritmálssafni Orðabókar Háskólans og þarna er því hægt að leita að orðmyndum í dæmasafninu óháð því undir hvaða uppflettiorði þau standa í ritmálssafninu.

Jón Aðalsteinn Jónsson. 1988. Alexander Jóhannesson og Orðabók Háskólans. *Orð og tunga* 1:105–122.
Jón Hilmar Jónsson. 1988. Sagnorðagreining Orðabókar Háskólans. *Orð og tunga* 1:123–174.
Jón Ólafsson. 1912–1915. *Orðabók íslenzkrar tungu að fornu og nýju* í fjórum bindum. [Einungis tvö hefti 1. bindis voru gefin út: 1: A–áætlun; 2: ávirðing–brýnn.] Reykjavík: Orðabókarfélagið.
Jörgen Pind, Kristín Bjarnadóttir, Jón Hilmar Jónsson, Guðrún Kvaran, Friðrik Magnússon og Ásta Svavarsdóttir. 1993. Using a Computer Corpus to Supplement a Citation Collection for a Historical Dictionary. *International Journal of Lexicography* 6(1):1–18. <https://doi.org/10.1093/ijl/6.1.1>.
OH 1984 = *Orðabók Háskólans*. Kynningarbæklingur. Ritstj. Jörgen Pind. Reykjavík: Orðabók Háskólans.
Ritmálssafn Orðabókar Háskólans. Stofnun Árna Magnússonar í íslenskum fræðum. <https://ritmalssafn.arnastofnun.is/>.
Sigfús Blöndal. 1920–1924. *Íslensk-dönsk orðabók*. Reykjavík.

Lykilorð

ritmálsheimildir, orðtaka, seðlasafn, rafrænt gagnasafn

Keywords

written language sources, excerpting, citation archive, digital database

Abstract

The article describes the so-called Written Language Archive (WLA) of the University of Iceland Dictionary (or the Institute of Lexicography), now preserved and maintained by the Árni Magnússon Institute for Icelandic Studies. Work on the WLA started in the early 1940s applying the method of traditional excerpting from written texts, and thus gradually building up a collection of dictionary slips with excerpted citations showing the use of Icelandic words in the period from the mid-16th to the late 20th centuries. This collection now contains over 2 million citations excerpted from more than 2000 different sources, and the number of headwords or lemmata is around 700 thousand. In the 1980s the first steps were taken in the digitalization of the collection, first by creating complete registers over the headwords and the sources, and later by digitizing most of the citations. All the digital material was organized in a database indicating relations between different parts of the collection, and around the turn of the century the digital version of the WLA was made generally accessible for search on the web. The database and the interface (on the website <https://ritmalssafn.arnastofnun.is>) were updated in 2018 and the means of searching the collection made more flexible by using wildcards and advanced search, not only full headwords.

Ásta Svavarsdóttir
Stofnun Árna Magnússonar í íslenskum fræðum
Arngrímsgötu 5
IS-107 Reykjavík
asta.svavarsdottir@arnastofnun.is