



Íslenskur námsorðaforði

Auður Pálsdóttir og Sigríður Ólafsdóttir

► Abstract ► Um höfunda ► About the authors ► Heimildir

Viðfangsefni þessarar rannsóknar er íslenskur námsorðaforði sem byggist á lagskiptingu orðaforða tungumálsins. Markmiðið var að móta lista yfir íslenskan námsorðaforða, orða í lagi 2 (LÍNO-2). Slíkur listi er mikilvægur því hann gefur upplýsingar um hvaða íslensk orð nemendur á mismunandi aldri þurfa að þekkja til að geta skilið, rætt og skrifað um það sem þau fást við í námi hverju sinni. Hér var því rannsakaður orðaforði tungumálsins en ekki orðaforði einstaklinga. Orð í lagi 1 eru hátíðniorð, algengustu orðin sem lærast í daglegum samskiptum. Í lagi 2 eru orð sem eru notuð þvert á námsgreinar og koma iðulega fyrir í fræðilegu ritmáli en síður í daglegu tali. Í lagi 3 eru sértæk og tæknileg iðorð sem hafa ákveðnar skilgreiningar innan fræðasviða og faggreina. Það eru einkum orð í lagi 2 sem kennurum hættir til að líta fram hjá í kennslu. Lítil þekking nemenda á þessum orðum er ein af meginástæðum þess að þeim gengur illa að skilja lesinn texta, sem síðan dregur úr námsárangri. Þessi rannsókn er hluti af stærra rannsóknarverkefni þar sem íslenskur námsorðaforði er meginviðfangsefnið. Sett var saman ný málheild (Málheild fyrir íslenskan námsorðaforða, MÍNO) með völdum samtímatextum *Íslensku Risamálheildarinnar* (RMH) og *Markaðrar íslenskrar málheildar* (MÍM), en þar á meðal er námsefni Menntamálastofnunar. Samtals telur MÍNO 31.680.235 lesmálorð og eru allir textar frá þessari öld. Orðum sem koma fyrir 100 sinnum eða oftar í málheildinni er raðað eftir tíðni og telur heildarorðtíðnilistinn 10.314 orð (lemmur) sem tilheyrja lagi 1, 2 og 3. Hinn nýi listi yfir íslenskan námsorðaforða (LÍNO-2) telur 2.294orð sem tilheyrja lagi 2. Þessi listi yfir íslenskan námsorðaforða mun nýtast kennurum og nemendum hér á landi við að efla orðaforða nemenda, lesskilning og tjáningarfærni auk þess að vera framlag til áframhaldandi kennslu og rannsókna sem miða að því að halda lífi í íslenskum orðum með komandi kynslóðum.

Efnisorð: Orðaforði, námsorðaforði, málheild, orðtíðni, læsi, lesskilningur, ritunarfærni, tjáningarfærni, íslenska

Inngangur

Erlendar og innlendar rannsóknir hafa sýnt sterk tengsl á milli orðaforða og lesskilnings (Freyja Birgisdóttir, 2016; Hu og Nation, 2000; Schmitt o.fl., 2011), en góður lesskilningur er forsenda farsællar námsframvindu (Auður Pálsdóttir, 2017, 2019; Sigríður Ólafsdóttir og Baldur Sigurðsson, 2017, 2019). Þá hefur komið í ljós að íslenskur orðaforði er drifkraftur framfara í lesskilningi hjá íslenskum grunnskólanemendum (Sigríður Ólafsdóttir o.fl., 2016).

Stærð orðaforða tengist þekkingu og reynslu hvers og eins (Snow, 2016). Þekking á orðum þróast frá því að kannast við orðið, þegar það heyrst eða sést, yfir í að geta beitt því á viðeigandi hátt í tali eða riti (Laufer, 2020; Nation, 2001; Roessingh o.fl., 2015; Sigríður Ólafsdóttir o.fl., 2016).

Orðaforða tungumálsins má skipta í þrjú lög, en í þessari rannsókn er sjónum beint að orðum í lagi 2. Lag 1 nær yfir hátíðniord sem eru algengustu orðin og lærast í daglegum samskiptum, t.d. hundur og lesa. Lag 3 nær yfir sértækan orðaforða, hugtök sem hafa ákveðnar skilgreiningar innan fræðasviða og ákveðinna námsgreina, t.d. *kolefni* og *margföldun*. Þessi orð eru gjarnan kölluð iðorð á íslensku. Orð í lagi 3 gegna veigamiklu hlutverki í kennslu, þekking á þeim felur í sér skilning á ákveðnum fræðasviðum viðkomandi námsgreina og eru orðin yfirleitt kennd sérstaklega (Beck o.fl., 2013; Roessingh, 2016; Snow, 2016).

Orð sem tilheyra lagi 2 eru notuð þvert á ólík fræðasvið eða námsgreinar og eru algengari í fræðilegu ritmáli en talmáli (Beck o.fl., 2013; DiCerbo o.fl., 2014; Nagy og Townsend, 2012; Roessingh, 2016; Roessingh og Douglas, 2013; Roessingh og Kover, 2002; Roessingh o.fl., 2005). Orðin gegna ekki lykilhutverki í fræðilegum textum, heldur styðja þau umfjöllun um það viðfangsefni sem fræðilegir textar fjalla um og eru ekki miðpunktur umfjöllunarinnar (Coxhead, 2000, bls. 214). Auk þess eru orð í lagi 2 reglulega notuð í samfélagslegri umræðu, þegar málefni eru rædd á ítarlegan hátt í rökræðum og af sérfræðingum. Í lagi 2 eru annars vegar hlutlæg orð, t.d. samheiti algengra orða eins og *piltur* og *stúlka* og safnorð eins og *leikföng* og *farartæki*. Hins vegar eru huglæg orð með flóknari merkingu, t.d. *vettvangur* og *skynsemi* (Beck o.fl., 2013; Sigríður Ólafsdóttir o.fl., 2017). Rannsóknir hafa leitt í ljós að nemendur eiga auðveldara með að læra hlutlæg orð en huglæg orð, bæði fyrir skilning og tjáningu (Peters, 2020). Námsorðaforði í lagi 2 byrjar að koma fram í markvissu leikskólastarfi, og gegnir lykilhutverki í öllu grunn-, framhalds- og í háskólanámi, þvert á námsgreinar og fræðasvið.

Þar sem orðin í lagi 2 eru notuð til að útskýra og fjalla um orð í lagi 3 (Beck o.fl., 2013; Coxhead, 2020; Dang, 2020) er þekking á orðum í lagi 2 mikilvæg til að ná að skilja og ræða um viðfangsefni námsgreina (Coxhead, 2020; Nation, 2016). Með því að vinna markvisst með orð í lagi 2 verður námsframvinda nemenda í ýmsum námsgreinum árangursríkari (Hyland, 2016). Af þessu má sjá að orð sem tilheyra lagi 2 og lagi 3 gegna mikilvægu hlutverki í námi nemenda og tilheyra því námsorðaforða tungumálsins (Beck o.fl., 2013; Roessingh, 2016). Þessi námsorðaforði hefur verið kallaður herra stigs orðaforði (Sigríður Ólafsdóttir o.fl., 2016) og akademískur orðaforði (Sigríður Ólafsdóttir o.fl., 2017) í skrifum íslenskra fræðimanna.

Hyland (2016) hefur rökstutt að gagnlegt sé að horfa á orð í lagi 2 og 3 sem hópa orða er hafa ekki skörp skil sín á milli. Þó mörk á milli orða í lagi 2 og 3 séu frekar óljós er samt greinanlegur stígandi þar á milli. Sem gagnlegt viðmið, þegar flokka á orð í lag 2 og 3, benda fræðimenn á að líta beri til þess hve algeng þau eru í ýmsum tegundum fræðilegs texta (e. *frequency*), hversu mikil dreifing orða er í textategundum (e. *dispersion*) og í hve mörgum textategundum málheildarinnar orð birtast (e. *range*) (Baumann og Graves, 2010; Coxhead, 2000; Dang, 2020; Gardner og Davies, 2014). Til þess að orð falli í lag 2 er annars vegar horft til þess að orðið komi fyrir í að minnsta kosti helmingi textategundanna (Gardner og Davies, 2014). Hins vegar ef orð kemur þrisvar sinnum eða oftár fyrir í einhverri einni tegund texta hefur það ekki verið flokkað í lag 2 heldur í lag 3 (Baumann og Graves, 2010; Gardner og Davies, 2014).

Rannsóknir á orðaforða nemenda í skyldunámi benda í vaxandi mæli til þess að skilningi á orðum í lagi 2 sé víða ábótavant og hamli því að nemendur nái að skilja texta námsbóka (August og Shanahan, 2006; Collier og Thomas, 2009; Cummins, 2000; Erla Lind Þórisdóttir, 2017; Gottlieb og Ernst-Slavit, 2013; Roessingh og Elgie, 2009). Þá benda rannsóknir til að orð í lagi 2 fái sjaldan sérstaka athygli í kennslu (Bailey o.fl., 2007; Beck o.fl., 2013; Calderón o.fl., 2005; Roessingh, 2016; Sigríður Ólafsdóttir o.fl., 2020). Í ljósi þess að orð sem tilheyra lagi 2 eru svo mörg og mikilvæg í námi er áriðandi að vita hvaða orð og hve mörg orð í lagi 2 þarf að beina sjónum að með hverjum aldurshópi nemenda (Lawrence o.fl., 2016; Nation, 2016; Roessingh, 2016; Sigríður Ólafsdóttir o.fl., 2020).

Til að vel sé að verki staðið í skipulagi kennslu, þýðingu og þróun námsefnis er mikilvægt að hafa upplýsingar um hvaða íslensk orð nemendur á ólíkum aldri þurfa að þekkja. Þar skiptir mestu námsorðaforðinn, þ.e. annars vegar orðin í lagi 3 sem gjarnan eru kennd í samhengi námsgreina og hins vegar orðin í lagi 2 sem notuð eru til að skilja og túlka viðfangsefni námsins og málefni samfélagsins.

Þessi rannsókn (sem heitir *Listi yfir íslenskan námsorðaförða*) er hluti af stærra verkefni (sem heitir *Íslenskur námsorðaförði*). Stærra verkefnið hefur fengið styrki frá Rannsóknasjóði Háskóla Íslands, Markáætlun um tungu og tækni og Íslenskusjóðnum. Markmið rannsóknarinnar sem hér er fjallað um (*Listi yfir íslenskan námsorðaförða*) var að móta lista yfir íslenskan námsorðaförða í lagi 2 (LÍNO-2). Slíkur listi er mikilvægur því hann gefur upplýsingar um hvaða íslensk orð nemendur á ólíkum aldri þurfa að þekkja til að geta skilið, rætt og skrifað um það sem þau læra hverju sinni. Hér var því verið að rannsaka orðaförða tungumálsins en ekki orðaförða einstaklinga.

Fræðilegur bakgrunnur

Þegar orð eru talin, annaðhvort sem fjöldi orða í tungumáli (orðaförði tungumáls) eða safn orða sem einstaklingar kunna (orðaförði einstaklinga) eru ýmsir möguleikar á hvernig beri að skilgreina orð. Hér er gerð grein fyrir lykilhugtökum sem tengjast orðtíðnilistum, fjallað um málheildir og orðtíðnilista og sjónum beint að íslenskum málheildum og orðtíðnilistum.

Skilgreiningar á lykilhugtökum

Hugtakið lesmálsorð (e. *running words*) merkir hvert einstakt orð í texta. Fjöldi lesmálsorða í texta er því fjöldi orða í textanum. Þannig getur sama orðið komið oft fyrir og telst eitt orð í viðbót í hvert skipti (Friðrik Magnússon o.fl., 1991).

Orðafjölskylda (e. *word family*) er hugtak sem nær yfir orð sem mynduð eru af einni og sömu rót eða orðstofni, þ.e. grunnorðið, beygingarmyndir þess ásamt merkingarlega afleiddum orðum (Nation, 2001; Schmitt o.fl., 2011; Þórhildur Oddsdóttir, 2012). Sem dæmi má nefna sagnorðið fljúga og afleiddu orðin *fleygur* (lo.) og *flug* (no.). Nemendur þurfa að búa yfir þekkingu og skilningi á uppbyggingu tungumálsins (sem er hluti málvitundar) svo þeir skilji að ólík orð geti tilheyrt sömu orðafjölskyldu (Dang, 2020; Gardner, 2007; Gardner og Davies, 2014; Nagy, 2007; Nippold og Sun, 2008).

Lemma (e. *lemma*) er hugtak notað yfir uppflettmynd orða sem notuð er í orðabókum. Af nafnorðum er notað nefnifall eintala, af sagnorðum nafnháttur og af lýsingarorðum karlkyn eintala nefnifall frumstig (Dang, 2020; Kristín Bjarnadóttir, e.d.). *Lemmun* (e. *lemmatize*) er síðan aðgerð sem framkvæmd er á lesmálsorðunum til að fækka þeim í þann fjölda orða sem kemur fyrir í ýmsum myndum, en án þess að tvítelja sama orðið. Í orðaförðarannsóknum hefur reynst hentugra að telja lemmur frekar en orðafjölskyldur, þar sem skilgreiningin er einfaldari og skýrari og einnig er tölfræðivinnslan auðveldari (Vilkaitè-Lozdienè og Schmitt, 2020).

Málheild (e. *corpus*) er safn texta sem er ætlað að endurspegla ákveðið notkunarsvið tungumálsins og eru textarnir geymdir á stöðluðu sniði á rafrænu formi (Stofnun Árna Magnússonar í íslenskum fræðum, e.d.-a). Öllum textum innan málheilda fylgja upplýsingar um þau verk sem textinn kemur úr. Textarnir eru greindir á margvíslegan hátt til að þeir verði sem gagnlegastir. Til dæmis fylgir hverri orðmynd *greiningarstrengur*, þ.e. mark, sem fyrir fallorð er kyn, tala og fall og fyrir sagnorð persóna, tala og tíð. Ef bæði mark og lemma fylgja hverri orðmynd málheilda þá kallast það *mörkuð málheild* (Anton K. Ingason o.fl., 2008; Dang, 2020; Stofnun Árna Magnússonar í íslenskum fræðum, e.d.-a).

Orðtíðnilistar sem byggjast á málheildum með samtímatextum segja okkur hvaða orð eru notuð, í hve miklum mæli og í hvers konar samhengi (Beck o.fl., 2013; Staples o.fl., 2018; Zipf, 1935). Því algengari sem orð eru (og ofar á orðtíðnilistum) þeim mun líklegri er að nemendur rekist á þau í námi sínu (Dang, 2020). Orðtíðnilistar veita því kennurum nákvæmar upplýsingar um hvaða orð nemendur þurfa að þekkja til að ná góðum námsárangri (Lawrence o.fl., 2016; Nation, 2016; Roessingh, 2016).

Þróun málheilda og orðtíðnilista

Í rannsóknum síðustu áratuga á orðaforða nemenda og tengslum hans við námsárangur hefur komið í ljós að mikilvægt er að skipta orðum (tungumálsins eða ákveðinnar málheildar) í flokka eftir tíðni orðanna. Fræðafólk hefur rökrætt mörk milli orðtíðniflokka en er sammála um gagnsemi þeirra bæði í rannsóknum og kennslu (Dang, 2020).

Nation (2001), sem er einn af frumkvöðlum í rannsóknum á enskum orðaforða, hefur rökstutt að gagnlegt sé að í hverjum tíðniflokki séu 1.000 orð. Í hæsta tíðniflokknum (fyrstu 1.000 orðin) eru algengustu orðin sem dreifast jafnt á margs konar textategundir (orð í lagi 1). Orð í næsthæsta tíðniflokknum (orð númer 1.001–2.000) hafa minni dreifingu á textategundir en orðin í fyrsta tíðniflokknum. Því lengra sem tíðniflokkarnir eru frá þeim algengasta þeim mun sjaldgæfari eru orðin og dreifing þeirra að jafnaði minni. Á þessum grundvelli skipti Nation (2001) orðum í fjóra flokka; hátíðniorð, lágtíðniorð, almennan námsorðaforða, og svo tæknileg orð. Hátíðniorð kallast orðin í hæsta tíðniflokknum og þau tilheyra lagi 1. Lágtíðniorð eru orðin sem eru í tíðniflokknum fyrir neðan þann hæsta, bæði almennur námsorðaforði, sem tilheyrir lagi 2, og einnig tæknileg orð, sem á íslensku kallast íðorð, og tilheyra lagi 3. Þessi flokkun orða eftir bæði tíðni og lögum er viðurkennd nálgun og hefur haft veruleg áhrif á orðaforðarannsóknir (Dang, 2020).

Eftir því sem rannsóknum á orðaforða hefur undið fram, hefur staðsetning skila á milli há- og lágtíðniorða verið rökrædd. Í vaxandi mæli hefur ekki aðeins verið litið til þess hversu algeng orð eru heldur líka til þess hversu líklegt er að þau lærist án beinnar kennslu (Dang, 2020). Í nýrri orðtíðnilistum hafa skilin milli há- og lágtíðniorða verið færð niður, þannig að algengustu 2.000 orðin hafa verið talin til hátíðniorða (Brezina og Gablasova, 2015, 2017; Dang, 2020; Nation, 2006, 2012). Schmitt og Schmitt (2014) kynntu nýtt hugtak þegar þau rökstuddu að betra væri að færa skilin enn neðar, og telja algengustu 3.000 orðin til hátíðniorða en kalla skyldi orð nr. 3001 til 9.000 orð af millitíðni (e. *mid-frequency level*) (Vilkaitè-Lozdiènè og Schmitt, 2020). Samkvæmt þeirri skilgreiningu teljast þá orðin sem eru neðar en algengustu 9.000 orð á listanum sem lágtíðniorð.

Í samantekt á nýjustu orðaforðarannsóknum benda Vilkaitè-Lozdiènè og Schmitt (2020) á að mörkin milli há- og lágtíðniorða eigi að flytja aftur upp tíðnilistann. Rökin eru þau hversu mikilvægt er að nemendur læri öll algengustu orð tungumálsins. Horfa eigi til rökstudnings Nation (2001) um að miða eigi við að 1.000 algengustu orðin séu hátíðniorð sem flestir nemendur læra fljótt og auðveldlega (og tilheyra lagi 1). Orð sem eru neðar á listanum eru ekki nógu algeng til að tryggja að þau lærist án beinnar kennslu (Vilkaitè-Lozdiènè og Schmitt, 2020).

Við mótun tíðnilista orðaforðans hefur reynst gagnlegt að setja fram aðskilda orðtíðnilista fyrir orð í hverju lagi. Þá er settur fram sérstakur listi yfir hátíðniorð í lagi 1, aðrir listar yfir tæknileg íðorð í lagi 3 (*íðorðalistar*), og svo er mótaður sérstakur listi yfir orð sem tilheyra lagi 2 (Dang, 2020). Til að setja saman orðtíðnilista með annars vegar 1.000 algengustu orðunum (lag 1) og hins vegar orðum sem tilheyra lagi 2 og lagi 3 er nauðsynlegt að byggt sé á ólíkum málheildum. Málheild, sem listi yfir algengustu orð tungumálsins byggist á, þarf að innihalda mun almennari textategundir en listi yfir fræðilegan námsorðaforða. Orðtíðnilistar með orðum í lagi 2 eða 3 þurfa að byggjast á málheildum með fræðilegum textasöfnum (Browne o.fl., 2013a, 2013b; Coxhead, 2020; Dang, 2020; Gardner og Davies, 2014).

Fræðafólk víða um heim sem hefur rannsakað orðaforða, mótun málheilda og orðtíðnilista, hefur horft til rannsókna Browne og féлага (2013a, 2013b) á enskum málheildum. Árið 2013 gáfu Browne og félagar (2013b) út orðtíðnilistann *New General Service List* (NGSL) sem tók við af eldri lista *General Service List* (West, 1953). NGSL-listinn byggist á 273 milljón lesmálsorða málheild úr *Cambridge English Corpus* þar sem valdir voru textar úr bókum, tímaritum, námsbókum skóla og háskóla, og rituðu og töluðu máli í fréttamiðlum og öðrum vefmiðlum. Á NGSL-listanum eru samtals 2.800 lemmur sem eru algengustu orð málheildarinnar sem Browne og félagar skilgreindu sem hátíðniorð í lagi 1 (Browne o.fl., 2013b).

Á sama hátt gáfu Browne og félagar (2013a) út *New Academic Word List* (NAWL) sem tók við af *Academic World list* (AWL) frá 1998 (Coxhead, 1998). NAWL-listinn byggist á 288 milljón lesmálsorða málheild með fræðilegum textum sem sóttir eru í fræðileg tímarit, fræðirit, ritgerðir nemenda, fræðilega umræðu, akademískt mál og fagbækur af toppsölulistum. Á NAWL-listanum eru 960 lemmur sem eru orð sem tilheyra lagi 2 (Browne o.fl., 2013b).

Við gerð ensku orðtíðnilistanna NGSL og NAWL voru höfð tvö meginmarkmið að leiðarljósi. Annars vegar að málheildirnar yrðu stærri en þær fyrri. Hins vegar að telja lemmur en ekki orðafjölskyldur eins og gert var í fyrri listum (Gardner og Davies, 2014; Schmitt og Zimmerman, 2002).

Með vaxandi þekkingu á orðaforða tungumáls og gagnsemi orðtíðnilista í námi og kennslu hefur orðið hröð þróun í gerð málheilda og orðtíðnilista. Mikill meirihluti rannsókna hefur snúist um enskt tungumál sem notað er í námi, bæði í enskumælandi samfélögum og annars staðar.

Orðtíðnilistar sem byggjast á málheildum með samtíðatextum hafa reynst sérstaklega árangursríkir fyrir nám- og kennslu tví- og fjöltyngdra nemenda (Browne o.fl., 2013a, 2013b; Dang, 2020). Ástæða þess að markviss vinna með orðaforða er einkum mikilvæg fyrir slíka nemendahópa er að orðaforði þeirra þróast að jafnaði á báðum eða öllum tungumálum þeirra. Það kann að leiða til þess að orðaforði þeirra í hverju máli verði minni en hjá nemendum sem aðeins nota eitt tungumál (Oller o.fl., 2007; Sigríður Ólafsdóttir og Hrafnhildur Ragnarsdóttir, 2010; Sigríður Ólafsdóttir o.fl., 2016).

Á Íslandi nota flest efekki öll börn og ungmenni bæði íslensku og ensku daglega (Birna Arnbjörnsdóttir og Hafdís Ingvarsdóttir, 2018). Birna og Hafdís draga þá ályktun, af viðamiklum rannsóknum á enskunotkun Íslendinga á ýmsum aldri, að enska geti tæplega lengur verið talin með erlendum tungumálum hér á landi, heldur sé enska komin í stöðu viðbótarmáls við íslenskuna. Þær leggja áherslu á að nú sé þörf fyrir markvissari kennslu í íslensku, að ekki sé lengur hægt að treysta því að nemendur læri mikilvæg íslensk orð af sjálfu sér. Í þessu ljósi er áriðandi að vinna stig af stigi, og með beinni kennslu, með íslenskan orðaforða sem er forsenda námsframvindu nemenda í íslenskum skólum. Til þess að það sé mögulegt þarf vita hvaða orð og hversu mörg orð þarf að taka fyrir með hverjum aldurshópi nemenda. Slíkar upplýsingar má nálgast með því að taka saman fræðilega málheild með íslenskum samtímatekstum og námsefni nemenda, raða orðum málheildarinnar eftir tíðni og draga fram orð sem eru algeng, en ekki of algeng né of sértæk, sem sagt orð sem tilheyra hvorki lagi 1 né lagi 3. Á þann hátt fæst listi yfir íslenskan námsorðaforða sem tilheyrir lagi 2, orð sem liggja til grundvallar námi á ýmsum sviðum skólans. Orð í lagi 2 leika ekki lykilhlutverk í texta en eru nauðsynleg til dæmis þegar iðorð í lagi 3 eru skilgreind og tekin til umfjöllunar. Þá ber að horfa til þess að því algengari sem orð í lagi 2 eru þeim mun mikilvægari eru þau. Þess vegna er sérstaklega gagnlegt að vinna með algengustu orðin í skólastarfi með ungum börnum og feta sig niður tíðnilistan með hækkandi aldri. Á sama hátt ber að vinna með orðin efst á tíðnilistanum í upphafi kennslu íslensku sem annars tungumáls með nemendum á ýmsum aldri sem eru nýflutt til landsins og feta sig síðan niður eftir tíðnilistanum.

Íslenskar málheildir og orðtíðnilistar

Hér á landi má rekja rannsóknir á tíðni íslenskra orða til fyrri hluta síðustu aldar.

Ársæll Sigurðsson rannsakaði orðtíðni íslenska tungumálsins árið 1940. Tilgangurinn var að gera stafsetningarkennslu árangursríkari. Hann rannsakaði stíla barna, sendibréf fullorðinna og námsbækur í náttúrugreinum, sögu og landafræði. Málheildin sem hann vann með var samtals 100.227 lesmálsorð og orðmyndirnar voru samtals 13.636. Niðurstöður Ársæls leiddu í ljós að tiltölulegar fáar orðmyndir mynduðu meginhluta lesmálsins en meirihluti einstakra orðmynda myndaði aðeins lítinn hluta af lesmálinu. Ársæll skrifar:

Þetta sannar því það, að um íslenskuna gildir hið sama og um aðrar menningartungur, að ritað mál (talmál sennilega ekki síður) er samsett af fáum orðmyndum, sem oft eru endurteknar, og mörgum orðmyndum, sem sjaldan eða aldrei eru endurteknar (Ársæll Sigurðsson, 1940, bls. 19–20; Friðrik Magnússon, 1988, bls. 5).

Árið 1991 kom *Íslenska orðtíðnibókin* út, en hún markaði þáttaskil í íslenskum orðaforðarannsóknnum (Friðrik Magnússon o.fl., 1991). Bókin byggist á 590.297 lesmálsorðum úr nærri 100 mismunandi textategundum og telur samtals 31.876 lemmur. Á þessum tíma var tölvutæknin að ryðja sér til rúms og gerði rannsókn á orðaforða markvissari en áður. Síðan hafa verið settar fram tvær rafrænar íslenskar málheildir sem eiga að endurspegla tungumálið í heild.

Önnur rafræna málheildin er *Mörkuð íslensk málheild* (MÍM) sem inniheldur 25 milljón lesmálsorð úr fjölbreyttum textategundum frá tímabilinu 2000–2010 (Sigrún Helgadóttir o.fl., 2012; Stofnun Árna Magnússonar í íslenskum fræðum, e.d.-a). Stærsti flokkur texta í MÍM eru útgefnar bækur en einnig er þar að finna texta úr dagblöðum, prentuðum tímaritum og frá opinberum aðilum. Ætlunin var að gefa eins raunsæja mynd og hægt var af íslensku máli í samtímanum (Stofnun Árna Magnússonar í íslenskum fræðum, e.d.-a). MÍM var svo gerð aðgengileg í opnum aðgangi árið 2013 (Sigrún Helgadóttir o.fl., 2012).

Hin rafræna málheildin er svokölluð *Risamálheild* (RMH) og er mörkuð eins og sú fyrri. Nýjasta útgáfan kom út árið 2019 og inniheldur 1.550 milljón lesmálsorð (Stofnun Árna Magnússonar í íslenskum fræðum, e.d.-b). Risamálheildin samanstendur af textum í sex meginflokkum; úr fréttamiðlum, stjórnýsly, fræðslumiðlum, greinum með skoðunum og greinum með sérefni, ásamt útgefnum bókum (Steinþór Steingrímsson o.fl., 2018).

Þó að tíðni íslenskra orða hafi verið viðfangsefni fræðafólks síðustu áratugi er ekki vitað um vinnu með íslenskan orðaforða sem gefur vísbindingar um hvar mörk á milli íslenskra há- og lágtíðniorða eigi að liggja, né hvaða íslensk orð beri að skilgreina í flokk orða í lagi 2. Ekki er heldur vitað um að könnuð hafi verið þekking barna og ungmenna á íslenskum orðum af mismunandi tíðni. Því hefur hingað til ekki verið mögulegt að bjóða nemendum náms- og kennsluhætti eða námsefni þar sem íslensk orð eru valin af gaumgæfni eftir tíðni þeirra og hlutverki í náms- og samfélagstextum samtímans.

Af framangreindu má sjá að mörg verkefni á sviði íslensks orðaforða eru áriðandi. Markmið þessarar rannsóknar var að móta lista yfir íslenskan námsorðaforða, orða í lagi 2 (LÍNO-2). Slíkur listi er mikilvægur því hann gefur upplýsingar um hvaða íslensk orð nemendur á mismunandi aldri þurfa að þekkja til að geta skilið, rætt og skrifað um það sem þau læra hverju sinni.

Aðferð

Í þessari rannsókn er spurt hvaða íslensk orð tilheyra námsorðaforða í lagi 2. Fyrst er gerð grein fyrir mótun málheildar fyrir íslenskan námsorðaforða (MÍNO) sem inniheldur orð í lagi 1, 2 og 3. Þá er útskýrt hvaða orð málheildarinnar eigi að skilgreina í lagi 2 og hvar mörk á milli orða í lagi 1 og 2 eigi að liggja. Að lokum er skilgreint hvar mörkin á milli orða í lagi 3 og 2 eigi að liggja. Eftir standa þá orð sem fara á lista yfir íslenskan námsorðaforða í lagi 2 (LÍNO-2).

Málheild fyrir íslenskan námsorðaforða (MÍNO)

Þegar velja á orð sem mikilvæg eru fyrir nám nemenda í íslenskum skólum þarf að byggja á málheild með samtímatextum, þ.e. textum sem eru í notkun í skólum, eins og vefsíðum og skólabókum, en líka textum úr þjóðlífinu. Slík málheild þarf eins og kostur er að endurspegla tungumálið sem nemendur þurfa að skilja og geta beitt. Sett var saman málheild með íslenskum fræðilegum textum líkt og gert var við mótun málheildar fyrir NAWL-listann (Browne o.fl., 2013b). Til upprifjunar inniheldur málheildin NAWL fræðileg tímarit, fræðirit, ritgerðir nemenda, fræðilega umræðu, akademískt mál og fagbækur af toppsölulistum. Í nýju málheildinni fyrir íslenskan námsorðaforða (MÍNO) var reynt eftir fremsta megni að hafa eins sambærilegar textategundir og eru í málheild NAWL.

Þó ber að hafa í huga að vegna þess að enskt málsamfélag er margfalt stærra en það íslenska er mikill munur á magni útgefinna enskra texta og íslenskra. Einnig ber að nefna að fræðilegir textar eftir

Íslenska sérfræðinga eru að verulegu leyti skrifaðir á ensku og einnig er misjafnt eftir fræðasviðum í hve miklum mæli það er.

Í MÍNO voru valdar textategundir úr bæði MÍM og RMH, margs konar textar skrifaðir á 21. öld og eru tengdir námi og samfélagi. Þetta eru tíu megintextategundir, fimm úr MÍM og fimm úr RMH. Textategundirnar eru tilgreindar í *Töflu 1*, ásamt fjölda lesmálsorða innan hverrar tegundar og hvorri málheildinni hver tegund tilheyrir.

Tafla 1. Textategundir sem valdar voru fyrir nýja málheild íslensks námsorðaorða (MÍNO).

Textategund	Fjöldi lesmálsorða	Málheild
Skýrslur og greinargerðir af vefsetrum ráðuneyta	489.798	MÍM
Lokaritgerðir háskólastúdenta	486.677	MÍM
Fréttir útvarps og sjónvarps (RÚV)	262.219	MÍM
Efni til upplestrar	432.287	MÍM
Stúdentsprófsritgerðir í íslensku	179.365	MÍM
Frumvörp (2020–2021)	2.158.576	RMH
Alþingistillögur (2020–2021)	464.560	RMH
Alþingisræður (2020–2021)	4.878.917	RMH
Fræðileg tímarit	6.501.519	RMH
Vísindavefurinn og tímarit um náttúruvísindi	10.328.918	RMH
Bækur (ekki skáldsögur)	4.406.633	RMH
Námsefni frá Menntamálastofnun	1.090.766	RMH
Heildarfjöldi lesmálsorða	31.680.235	

Eins og sjá má í *Töflu 1* er langstærsta textategundin efni af Vísindavefnum og tímarit um náttúruvísindi, samtals 10.328.918 lesmálsorð. Þá telja textar sem sóttir eru á vefsetur ráðuneyta ásamt frumvörpum, alþingistillögum og -ræðum samtals 7.991.851 lesmálsorð. Þá ber að geta að námsefni frá Menntamálastofnun inniheldur allt útgefið námsefni fyrir grunnskólanemendur frá aldamótum sem aðgengilegt er á stafrænu formi og telur rúmlega eina milljón lesmálsorða. MÍNO er þegar birt á CLARIN-IS (Sigríður Ólafsdóttir o.fl., 2023a).

Málheildin MÍNO var lemmuð. Þær lemmur sem koma fyrir 100 sinnum eða oftar í málheildinni fóru á orðtíðnilista málheildar MÍNO, samtals 10.313 lemmur. Á MÍNO eru orð sem tilheyra lagi 1, 2 og 3, það eru fallorð, sagnorð og smáorð. Orðtíðnilisti MÍNO er þegar birtur á CLARIN-IS (Sigríður Ólafsdóttir o.fl., 2023b).

Mótun lista yfir íslenskan námsorðaforða í lagi 2 (LÍNO-2)

Næsta skref fól í sér að ákveða hvaða orð af orðtíðnilista MÍNO ættu að fara á LÍNO-2 og því mikilvægt að taka fram hvar mörkin á milli há- og lágtíðniorda ættu að liggja, það er að segja hvar setja ætti mörk á milli orða í lagi 1 og 2 og einnig hvar mörk á milli orða í lagi 2 og 3 ættu að liggja (Nation, 2001; Vilkaitè-Lozdienè og Schmitt, 2020). Beitt var bæði hlutlægu og huglægu mati. Fyrst verður gerð grein fyrir hlutlægu mati á hvort orð tilheyra lagi 1 eða 3 og fara því ekki á LÍNO-2 og síðan hvernig listinn, sem þá var tilbúinn, var metinn huglægt. Að lokum var farið yfir orð sem höfðu dottið út við hlutlægu matið og metið huglægt hvort þeim bæri að eiga sæti á LÍNO-2.

Þegar verið var að ákveða hlutlægt hvar mörkin milli íslenskra há- og lágtíðniorda ættu að liggja var rökstuðningur Nation (2001) hafður til hliðsjónar. Því var ákveðið að mörkin milli hátíðniorda í lagi 1 og lágtíðniorda í lagi 2 og 3 væru við 1.000 algengustu íslensku orðin. Þar sem Risamálheildin

inniheldur samtímatexta af ýmsu tagi sem endurspeglar íslenska tungumálið í heild sinni (en ekki eingöngu fræðilega texta eins og MÍNO) voru 1.000 algengustu orðin í RMH fundin og tekin út og því ekki valin á LÍNO-2 (Browne, 2013a, 2013b; Dang, 2020).

Næst var að ákveða hversu langt niður eftir orðtíðnilista MÍNO ætti að fara þegar orð voru flokkuð í lag 2 eða 3. Ákveðið var að fara ekki neðar en að orði númer 5.000 því að hér er verið að móta lista sem viðmið fyrir nám og kennslu á grunnskólastigi (NAWL er ætlaður háskólanemendum og nær niður að lemmu nr. 9.000).

Eftir að algengustu 1.000 orð RMH höfðu verið tekin út var afgangur af orðunum (samtals 4.027 orð) greindur eftir því hvort þau tilheyra lagi 2 eða lagi 3. Þegar orð hafa verið valin á lista yfir enskan námsorðaförða, orð sem tilheyra lagi 2, er litið til þess hversu oft orð birtist í heildarmálheildinni (e. *frequency*) sem listinn byggist á, í hve mörgum textategundum (e. *range*) og hversu jafnt orðið birtist í ólíkum textategundum (e. *dispersion*) (Dang, 2020). Því oftast sem orðið birtist og því dreifðara og því jafnari sem dreifingin er, þeim mun betur fellur orðið undir skilgreiningu á orðum í lagi 2. Slík orð eru, eins og áður segir, ekki hluti af sérstökum námsgreinum, eða námsgreinasviðum, heldur eru notuð á öllum sviðum náms.

Við mat á hvar mörk ættu að liggja á milli orða í lagi 3 og lagi 2 var horft til enskra viðmiða (Dang, 2020). Þó ber að geta að sökum þess að íslenska málheildin MÍNO er miklu minni en enskar málheildir er samsetning textategunda ekki sú sama. Færri lesmálsorð eru í hverri textategund og dreifing orða (lemma) er því ekki alveg sambærileg. Í þessari rannsókn var ákveðið að til að orð fengi sæti á LÍNO-2 yrði það að koma fyrir í öllum textategundum (e. *range*) MÍNO og í hlutfallslega svipuðum mæli (e. *dispersion*).

Næsta skref í mótun LÍNO-2 fólst því í niðurskurði eftir tíðnihlutfalli orða í textategundum. Til að ákveða hvaða orð tilheyra lagi 3 (og fara því ekki á LÍNO-2) voru fundin orð sem birtast hlutfallslega í meiri mæli í einni textategund en öðrum. Fundið var hlutfall p (e. *proportion*) tíðnihlutfalls f_i (e. *relative frequency*) einstakrar textategundar af öðrum textategundum samanlagt eftir formúlunni:

$$p = \frac{f_i}{f_{i1} + f_{i2} \dots f_{in}}$$

Hlutfallið gefur upplýsingar um hversu þungt tiltekið orð vegur í þeirri textategund sem um ræðir hverju sinni. Prófað var fyrst að gera þessa útreikninga fyrir hverja einustu textategund og þannig taka út öll orð sem birtust oftast en þrisvar sinnum (3x) eða oftast en tvisvar sinnum (2x), sama hvaða textategund var um að ræða. Fljótlega kom í ljós að orð sem áttu greinilega að vera áfram inni á LÍNO-2, samkvæmt fræðilegri skilgreiningu á orðum í lagi 2, dattu út. Því var ákveðið að prófa að taka einungis út ákveðnar tegundir sérhæfðra orða. Margar textategundirnar eru nokkuð almennar að undanskildum annars vegar stjórnáttatengdum textategundum, þ.e. alþingisræðum, alþingistillögum, frumvörpum og vefsíðum ráðuneyta, og hins vegar náttúruvísindatengdum textategundum, þ.e. efni af Vísindavefnum og tímaritum um náttúruvísindi.

Reiknað var út sameiginlegt tíðnihlutfall stjórnáttatengdu textategundanna af tíðnihlutföllum hinna textategundanna fyrir hvert orð. Með þessu fengu orð sem voru algeng í öllum fyrrgreindum fjórum textategundum meira vægi og því var auðveldara að ákvarða hvort þau ættu að tilheyra lagi 3 eða ekki. Þá var gerð sams konar greining fyrir textategundir um náttúruvísindi.

Eftir að hlutfall stjórnáttatengdra orða var reiknað út ásamt hlutfalli náttúruvísindatengdra orða urðu til tveir listar, sem sýndu annars vegar orð sem birtust tvisvar sinnum eða oftast og hins vegar orð sem birtust þrisvar sinnum eða oftast í þessum textategundum en hinum. Þegar listi var keyrður út sem byggðist á því að ef orð birtist tvisvar sinnum eða oftast í stjórnáttatengdum textum eða náttúrufræðitengdum textum ætti það ekki heima á LÍNO-2, sátu eftir á listanum samtals 3.404 orð. Þá kom í ljós að fjölmörg orð dattu út þó þau gætu talist mjög dæmigerð fyrir orð í lagi 2, samkvæmt ofangreindri skilgreiningu á orðum af því tagi. Dæmi um slík orð eru *adgengilegur*, *búseta*, *óvissa*, *samþykki*, *starfsemi*, *ungmenni*, *viðeigandi*, *ungmenni*.

Þegar horft var á lista sem byggðist á að ef orð birtist þrisvar sinnum eða oftar í stjórnmálatengdum textum eða náttúrufræðitengdum textum ætti það ekki heima á LÍNO-2, sátu eftir 3.583 orð. Orðin sem duttu þá út voru að verulegu leyti sértæk iðorð á þessum sviðum. Dæmi um slík orð eru *erfðaefni*, *prótín*, *sólkerfi*, *rikisábyrgð*, *háþrýstingur*, *þingsályktunartillaga* og *fjárlaga-frumvarp*.

Samt sem áður voru þarna orð inni á milli og í raun mest efst á orðtíðnilistanum sem ættu að fara á LÍNO-2 þar sem þau falla nokkuð skýrt undir skilgreiningu orða í lagi 2. Dæmi um slík orð eru *álykta*, *foreldri*, *forgangsröðun*, *sambúð*, *starfshópur*, *upplýsingagjöf* og *ökutæki*. Því var strax ljóst að huglægt mat yrði að gera síðar.

Byggt á þessum útreikningum var ákveðið að miða við efri mörkin, þ.e. að orð sem birtust þrisvar sinnum eða oftar í fyrrgreindum tveimur textategundum færu ekki á LÍNO-2. Þetta er líka í samræmi við mörkin sem sett hafa verið á milli orða í lagi 2 og 3 í enskum námsorðaforðalistum (Baumann og Garves, 2010; Browne o.fl., 2013b; Coxhead, 2000; Dang, 2020; Garner og Davies, 2014).

Næsta skref við mótun LÍNO-2 var huglægt mat. Líkt og Dang (2020) mælir með voru þrír hópar sérfræðinga fengnir til að vinna huglæga matið. Ásamt greinarhöfundum voru það læsis- og prófsérfræðingar hjá Menntamálastofnun og sérfræðingar iðorða og málheilda hjá Stofnun Árna Magnússonar í íslenskum fræðum.

Fyrsta skrefið var að fara yfir orðin sem sátu á LÍNO-2 og athuga hvort þar væru orð sem örugglega tilheyrðu lagi 2, en ekki lagi 1 eða lagi 3 (Dang, 2020). Meginskilgreiningin sem höfð var til hliðsjónar við huglæga matið á hvort orð eigi að tilheyra lagi 2 eða lagi 3, var sú að ef orð felur í sér þekkingu eða hugsun sem þarf að taka fyrir í skólastarfi telst það til lags 3 en ekki til lags 2. Í ljós kom að fjölmörg orð duttu þá út sem eining var um að ættu ekki heima á LÍNO-2. Dæmi um orð sem duttu út: *adildarríki*, *atvinnuleysi*, *eldgos*, *fjárhagslegur*, *fóstur*, *hagkerfi*, *heilbrigðiskerfi*, *hjúkrunarheimili*, *kynlíf*, *landlæknir*, *lífvera*, *lunga*, *lýðræði*, *samkynhneigður*, *trúarbrögð*, *umbodsmadur* og *þjóðkirkja*.

Næsta skref var að fara yfir 1.000 algengustu orð í RMH sem höfðu verið tekin út, til að athuga hvort þar væru orð sem ættu samt heima á LÍNO-2. Ástæðan er sú að hlutlæg ákvörðun um hvar mörkin milli há- og lágtíðni orða í íslensku eigi að liggja byggðist eingöngu á rannsóknum á ensku tungumáli (Coxhead, 2020). Við þessa yfirferð bættust við þónokkur orð, til dæmis *almennur*, *ásamt*, *einnig*, *erlendur*, *hljóta*, *jafnframt*, *raun*, *samkvæmt*, *tillaga* og *undanfariinn*. Hér ber að hafa í huga að þekking nemenda á ýmsum aldri á íslenskum orðum eftir tíðni þeirra hefur aldrei verið könnuð. Því var í huglæga matinu ákveðið að hafa frekar fleiri orð en færri á LÍNO-2, en mörg orðanna sem bættust við í huglæga matinu eiga samheiti í RMH sem eru mjög algeng í talmáli.

Að lokum var farið yfir orðin sem höfðu fallið brott þegar mörkin á milli orða í lagi 2 og 3 voru sett við það að orð sem birtast þrisvar sinnum eða oftar í náttúruvísindatextum eða stjórnmálatengdum textum færu ekki á LÍNO-2 því þau tilheyra lagi 3. Þessi orð voru metin huglægt á sama hátt og áður, þ.e. ef fjalla þyrfti um merkingu þeirra í skólastarfi ættu þau ekki heima á LÍNO-2 heldur væru hluti af orðum í lagi 3. Við þá yfirferð bættust við orð eins og *andsvar*, *álykta*, *árétta* og *foreldri*.

Eftir hlutlægt og huglægt mat á orðum MÍNO, sátu eftir á LÍNO-2 samtals 2.294 orð (lemmur).

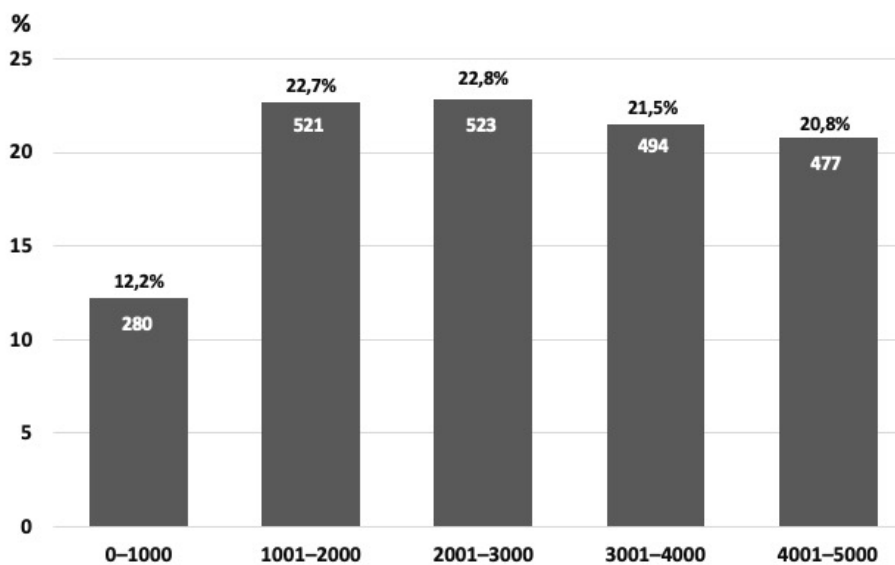
Hér ber að hafa í huga að rannsóknirnar sem hér er stuðst við hafa allar verið gerðar fyrir ensku. Ensk orð eiga gjarnan uppruna sinn í latínu og grísku en íslensk orð eru iðulega samsett úr algengum orðum eða orðhlutum og eru því gagnsærri og auðskiljanlegri en þau ensku. Til dæmis er íslenska orðið *yfirgrípsmikill* (sett saman úr þremur orðum) *extensive* á ensku (af latneskum uppruna). Því gildir ekki um íslensku hið sama og um ensku að lengri orð séu erfiðari en styttri orð (Auður Pálsdóttir og Sigríður Ólafsdóttir, 2019; OECD, 2016, bls. 11). Á ensku eru auk þess miklu fleiri orð en í íslensku, og fallorð í íslensku eru beygð í kyni, tölu og falli auk þess sem sagnorðin í íslensku beygjast í meiri mæli en ensk sagnorð.

Niðurstöður

Markmið þessarar rannsóknar var að móta lista yfir íslenskan námsorðaförða, orða í lagi 2 (LÍNO-2). Slíkur listi er mikilvægur því hann gefur upplýsingar um hvaða íslensk orð nemendur á mismunandi aldri þurfa að þekkja til að geta skilið, rætt og skrifað um það sem þau fást við í námi hverju sinni. Hér var því rannsakaður orðaförði tungumálsins en ekki orðaförði einstaklinga.

Líkt og hjá Nation (2001) var LÍNO-2 skipt í fimm tíðniflokka orða úr MÍNO. Í fyrsta tíðniflokki eru orð meðal 1.000 algengustu orða málheildarinnar MÍNO. Í öðrum tíðniflokki eru orð sem eru á tíðnibilinu 1.001–2.000 í málheildinni MÍNO. Á *Mynd 1* má sjá hlutfallslega dreifingu og fjölda orða í lagi 2 (á LÍNO-2) í fimm orðtíðniflokkum MÍNO, sem fóru á lista yfir íslenskan námsorðaförða (LÍNO-2).

Mynd 1. Hlutfallsleg dreifing og fjöldi orða á LÍNO-2 í fimm tíðniflokkum MÍNO, samtals 2.294 orð.



Líkt og sést á *Mynd 1* eru hlutfallslega færst orð LÍNO-2 í algengasta tíðniflokki MÍNO. Ástæður fyrir þessu lága hlutfalli (12,7%) eru að í þessum tíðniflokki eru algengustu orð málheildarinnar og flest þeirra tilheyra lagi 1. Þá voru 1.000 algengustu orð RMH tekin út sem eru einmitt orð í þessum algengasta tíðniflokki. Þau sem eftir sitja voru metin huglægt að eiga heima á LÍNO-2, eins og áður hefur komið fram. Hlutfallslega flest orð í LÍNO-2 eru í öðrum og þriðja tíðniflokkunum, þ.e. 1.001–2.000 og 2.001–3.000 í MÍNO. Í fjórða og fimmta tíðniflokki MÍNO lækkar hlutfall orða í LÍNO-2, því hlutfall sérhæfðra orða eykst með lækkandi tíðni, enda höfðu sérhæfðu orðin í lagi 3 ekki farið með inn á LÍNO-2.

Á *Mynd 2* má sjá dæmi um orð á LÍNO-2 í tíðniflokkunum fimm á MÍNO.

Mynd 2. Dæmi um orð á LÍNO-2 (orð í lagi 2) eftir tíðniflokkum MÍNO.

0–1.000	einnig, almennur, skilgreina, ábyrgð, umfjöllun
1.001–2.000	nýting, losa, eiginlega, upprunalegur, viðhald
2.001–3.000	ímynd, misseri, tímabundið, loksins, orsaka
3.001–4.000	hentugur, útskýring, afhjúpa, uns, hugleiða
4.001–5.000	heimkynni, villandi, giska, ótvírætt, feðgar

Eins og sjá má á dæmunum á *Mynd 2* eru orð á LÍNO-2 ýmist nafnorð, sagnorð, lýsingarorð eða smáorð. Þetta eru orð sem tilheyra lagi 2, eru notuð í margs konar textum samtímans, bæði námstextum og fræðitextum. Orð af þessu tagi leika ekki lykilhlutverk í texta heldur eru þau notuð þegar verið er að fjalla um og útskýra sértæk og tæknileg iðorð ólíkra námsgreina (orð í lagi 3). Skilningur á þessum orðum er því grundvöllur þess að nemendur skilji útskýringar á sértækum og tæknilegum orðum (í lagi 3). Jafnframt þurfa nemendur að geta notað orð í lagi 2 til að geta bæði tekið þátt í samræðum og skrifað um viðfangsefni námsins.

Í hnotskurn er hér kynntur listi yfir íslenskan námsorðaförða (LÍNO-2), samtals 2.294 orð sem öll koma úr nýrri málheild fyrir íslenskan námsorðaförða (MÍNO). Orðin á LÍNO-2 eru mikið notuð í textum samtímans, í samfélagslegri og fræðilegri umfjöllun og í námsgögnum grunnskólanema. Listann LÍNO-2 má nálgast í heild á CLARIN-IS (Sigríður Ólafsdóttir og Auður Pálsdóttir, 2023).

Umræður

Markmið þessarar rannsóknar var að móta lista yfir íslenskan námsorðaförða, orða í lagi 2 (LÍNO-2). Slíkur listi er mikilvægur því hann gefur upplýsingar um hvaða íslensk orð nemendur á ólíkum aldri þurfa að þekkja til að geta skilið, rætt og skrifað um það sem þau fást við í námi hverju sinni. Hér var því rannsakaður orðaförði tungumálsins en ekki orðaförði einstaklinga.

Listi af þessu tagi hefur fram til þessa ekki verið til fyrir íslenska tungu, svo hér er um mikilvæga nýjung að ræða sem ætlað er að vera framlag til að efla íslenskan orðaförða, lesskilning og tjáningarfærni nemenda á Íslandi.

Mótuð var ný íslensk málheild fyrir íslenskan námsorðaförða (MÍNO) og mótaður var tíðnिलisti yfir íslenskan námsorðaförða (LÍNO-2) fyrir grunnskólanemendur. Líkt og hjá Nation (2001) var LÍNO-2 skipt upp í fimm tíðniflokka. Fyrsti flokkurinn hafði orð sem tilheyra 1.000 algengustu orðum MÍNO, annar flokkurinn næstu þúsund og svo framvegis. Hlutfall orða sem fóru á LÍNO-2 í hverjum tíðniflokki var lægst í fyrsta tíðniflokknum (12,2%), hæst í öðrum og þriðja (22,7% og 22,8%), en rétt rúm 20% í fjórða og fimmta (21,5% og 20,8%).

Mikilvægt er að hafa íslenskan námsorðaförðalista sem skiptist í tíðniflokka til að geta unnið markvisst með orð í kennslu, við gerð námsefnis og matstækja til að greina stöðu nemenda og framfarir. Þannig ætti að beina sjónum að orðum í fyrsta tíðniflokknum með yngstu börnunum og bæta orðum úr næstu tíðniflokkum með hækkandi aldri og aukinni íslenskufærni. Þetta þýðir líka að nú er kominn mikilvægur orðaförðagrunnur fyrir markvissa kennslu í íslensku sem öðru tungumáli sem hefur það meginmarkmið að undirbúa nemendur sem flytja til Íslands á ólíkum aldri fyrir nám í íslenskum skólum. Með LÍNO-2 er komið íslenskt verkfæri sem reynst hefur árangursríkt til að minnka mun

á frammistöðu nemenda í námi (Lawrence o.fl., 2016; Quigley, 2018; Snow, 2014). Markviss vinna með LÍNO-2 getur því spornað gegn vaxandi mun á námsgengi nemenda af íslenskum og erlendum uppruna (Sigríður Ólafsdóttir o.fl., 2016) því góður orðaforði í tungumáli skólans er forsenda og drifkraftur framfara nemenda í lesskilningi yfir grunnskólaárin (Hu og Nation, 2000; Laufer, 1989, 1992; Laufer og Ravenhorst-Kalovski, 2010; Schmitt o.fl., 2011). Í þessu samhengi leika orð í lagi 2 lykilhlutverk (Beck o.fl., 2013; Coxhead, 2020; Peters, 2020).

Lokaorð

Niðurstöður þessarar rannsóknar kynna nýja málheild fyrir íslenskan námsorðaforða (MÍNO) og lista yfir íslenskan námsorðaforða (orð í lagi 2, LÍNO-2). Nú þegar er hafin vinna við þrjú meginverkefni sem byggja á listanum LÍNO-2. Í fyrsta lagi verður LÍNO-2 þýddur á algengustu tungumál fólks af erlendum uppruna hér á landi, fyrir kennslu í íslensku sem öðru tungumáli. Í öðru lagi verður könnuð þekking nemenda á ýmsum aldri á orðum LÍNO-2. Í þriðja lagi verða samdir textar með orðum LÍNO-2, frá þeim algengustu og svo niður eftir tíðnilistanum, og textarnir þýddir á sömu tungumál. Í fjórða lagi verða þróaðar kennsluleiðbeiningar um notkun textanna í umræðum og ritunarverkefnum. Í gegnum slíkt skólastarf gefst nemendum tækifæri til að nota orðin, og þannig verða orðin þeirra orð, ekki aðeins til skilnings heldur líka til tjáningar.

Önnur mikilvæg verkefni eru til dæmis að byggt verði á LÍNO-2 við þróun prófa, bæði lesskilnings- og ritunarprófa. Þá verður gagnlegt að skilgreina náms- og kennsluáferðir, og móta kennsluleiðbeiningar þar sem unnið er markvisst með orð LÍNO-2, bæði til skilnings og tjáningar. Eins þyrfti að kanna viðhorf og reynslu kennara af notkun kennsluleiðbeininganna. Enn fremur mætti nýta LÍNO-2 til að semja námsefni í ýmsum faggreinum, þar sem orð eru valin markvisst af LÍNO-2 miðað við aldurshóp eða íslenskufærni.

Við túlkun niðurstaðna er mikilvægt að hafa í huga takmarkanir hverrar rannsóknar. Má þar fyrst nefna samsetningu nýju málheildarinnar, málheildar fyrir íslenskan námsorðaforða (MÍNO), en aðeins var hægt að nýta texta sem aðgengilegir voru í stafrænu formi. Hér ber að hafa í huga að mun meira magn er til af enskum texta en íslenskum og því er íslenska málheildin töluvert minni en enska fyrirmyndin. Á sama tíma var leitast við að tryggja samræmi í hlutfalli ólíkra texta í málheildinni. Hafa þarf í huga að LÍNO-2 er hvorki endanlegur né tæmandi listi yfir íslenskan námsorðaforða í lagi 2. Notkunarsvið og tíðni einstakra orða breytast með tíð og tíma.

Hinum nýja lista yfir íslenskan námsorðaforða (LÍNO-2) er ætlað að vera lifandi verkfæri sem vafalítið mun taka breytingum eftir því sem málnotkun þróast í samfélaginu og rannsóknum á íslenskum námsorðaforða nemenda fleytir fram. Í þessu ljósi verða MÍNO og LÍNO-2 aðgengilegir öllum á netinu sem veitir hverjum sem er, jafnt kennurum, námsbókahöfundum, rannsakendum og jafnvel þýðendum tækifæri til að vanda vinnu sína enn markvissar, námi nemenda til heilla.

Icelandic academic word list

In Iceland and elsewhere, there is an increased awareness of the key role that vocabulary plays in learners' literacy development (Auður Pálsdóttir, 2017; Beck et al., 2013; Roessingh, 2016; Sigríður Ólafsdóttir & Baldur Sigurðsson, 2017). The academic words vused across different subject areas (Tier 2) are pivotal for students' academic procedures (DiCerbo et al., 2014; Lawrence et al., 2016).

The purpose of this research was to develop a Tier 2 *Icelandic academic word list* (i. listi yfir íslenskan námsorðaforða, LÍNO-2). Such a list is important since it provides information about Icelandic words that learners need to know for active participation in school, their reading comprehension, discussion and writing skills.

In the Icelandic language, the number of words is countless, and therefore, it is important to know what words lie at the heart of academic success. Thus, it is v to research which words and how many words are useful to work with at each education level. Words that go beyond the most common words of the language (Tier 1) are called academic vocabulary (Beck et al., 2013; Roessingh, 2016). A part of these words is taught systematically within the various school subjects. Those are Tier 3 words. However, the words in Tier 2 are often overlooked by teachers and neglected in school activities. Words in this category are general, high-utility academic words used across various fields of study. They play a key role when discussing complex issues. The learners' poor knowledge of these words is one of the main reasons why they have difficulty understanding a written text, which in turn, reduces their ability for academic progress.

A new Icelandic corpus (Málheild fyrir íslenskan námsorðaförða, MÍNO) was compiled containing selected student and academic texts from Tagged Icelandic Corpus (MÍM, Stofnun Árna Magnússonar í íslenskum fræðum, n.d.-a) and Icelandic Gigaword Corpus (IGC, Stofnun Árna Magnússonar í íslenskum fræðum, n.d.-b). MÍNO has a total of 31,680,235 running words. All the texts are current texts published this century. Words that appear 100 times or more in the corpus are sorted by frequency. The total word frequency list contains 10,314 words (lemmas, Tiers 1, 2 and 3). Objective and subjective approaches were used to categorise the words of the corpus, i.e., to distinguish between words belonging to Tier 1 and Tier 2, as well as words belonging to Tier 2 and Tier 3 (Dang, 2020). The Icelandic academic word list (LÍNO-2) contains 2,294 Tier 2 words (lemmas) from the new Icelandic language corpus MÍNO (Beck et al., 2013).

The words on LÍNO-2 were divided into five bands depending on their frequency in MÍNO. The first band included words from the most frequent 1000 words. The second band included words from the next 1000 etc. (Dang, 2020; Nation, 2001). The frequency bands are a useful reference when choosing words to teach the youngest learners. With the increased student age, the words from lower frequency bands are included step by step. LÍNO-2 is a valuable contribution to teaching Icelandic as a first and a second language.

The value of LINO-2 is fourfold. First, the value is for teachers and students in Iceland as an effective tool to enhance student Icelandic vocabulary, reading comprehension and productive language skills, i.e., proficiency in discussing and writing. Second, LÍNO-2 is valuable for developing teaching materials for different age groups and groups with diverse Icelandic language proficiency. Third, LÍNO-2 is important for assessing student status or learning progress of receptive and productive vocabulary, reading comprehension and writing skills. Fourth, the value of LINO-2 is significant for research aiming at keeping Icelandic words alive with future generations.

This research is the first of several projects funded by the University of Iceland Research Fund (í. Rannsóknasjóður Háskóla Íslands), the Icelandic Language Technology Fund (í. Markáætlun um tungu og tækni) and the Icelandic language fund (í. Íslenskusjóðurinn), where Icelandic academic vocabulary is the core subject.

Key words: Vocabulary, academic vocabulary, language corpus, word frequency, literacy, reading comprehension, writing proficiency, Icelandic language

Um höfunda

Auður Pálsdóttir, PhD, (audurp@hi.is), er dósent við Menntavísindasvið Háskóla Íslands. Meginrannsóknarsvið hennar eru samfélagsgreina- og náttúrugreinamenntun, mat á skólastarfi, kennslufræði samfélags- og náttúrugreina, þróun námsorðaforða, greining námsefnis og sjálfbærnimenntun.

Sigríður Ólafsdóttir, PhD, (sol@hi.is), er dósent við Menntavísindasvið Háskóla Íslands. Meginrannsóknarsvið hennar beinast að málþroska og læsi, þróun orðaforða og lesskilnings hjá nemendum sem eiga íslensku sem móðurmál og annað mál.

About the authors

Auður Pálsdóttir, PhD, (audurp@hi.is) is an associate professor at the University of Iceland, School of Education. Her main research interests include social science and science education, school self-evaluation, pedagogy of science and social science teaching, vocabulary development and school textbook analysis, and sustainability education.

Sigríður Ólafsdóttir, PhD, (sol@hi.is) is an associate professor at the University of Iceland, School of Education. Her main research interests include language development and literacy and the development of vocabulary and reading comprehension among Icelandic first and second-language learners.

Heimildir

- Anton K. Ingason, Sigrún Helgadóttir, Hrafn Loftsson og Eiríkur Rögnvaldsson. (2008). A mixed method lemmatization algorithm using Hierarchy of linguistic identities (HOLI). Í B. Nordström og A. Ranta (ritstjórar), *Advances in Natural Language Processing. GoTAL 2008. Lecture Notes in Computer Science*, (5221). Springer. https://doi.org/10.1007/978-3-540-85287-2_20
- Auður Pálsdóttir. (2017). Læsi á náttúruvísindi. Vísbendingar úr PISA 2015. Í *Helstu niðurstöður PISA 2015* (bls. 36–42). Menntamálastofnun. https://mms.is/sites/mms.is/files/helstu_nidurstodur_pisa_2015_prent_-_loka.pdf
- Auður Pálsdóttir. (2019). Læsi á náttúruvísindi. Í *PISA 2018: Helstu niðurstöður á Íslandi* (bls. 96–108). Menntamálastofnun. https://mms.is/sites/mms.is/files/pisa_2019_0.pdf
- Auður Pálsdóttir og Sigríður Ólafsdóttir. (2019). Greining á orðanotkun í lesskilnings- og náttúruvísindahlutum PISA 2018: Samanburður á íslensku þýðingunni og enska textanum. *Sérrit Netlu 2019 – Alþjóðlegar menntakannanir*. https://netla.hi.is/serrit/2019/altjodlegar_menntakannanir/02.pdf
- August, D. og Shanahan, T. (ritstjórar). (2006). *Developing literacy in second-language learners: Report of the National literacy panel on language minority children and youth*. Lawrence Erlbaum Associates.
- Ársæll Sigurðsson. (1940). Algengustu orðmyndir málsins og stafsetningarkennslan. *Menntamál*, (13), 8–42.
- Bailey, A. L., Butler, F. A., Stevens, R. og Lord, C. (2007). Further specifying the language demands of school. Í A. L. Bailey (ritstjóri), *The language demands of school: Putting academic English to the test* (bls. 103–156). Yale University Press.
- Baumann, J. F. og Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent & Adult Literacy*, 54(1), 4–12. <https://doi.org/10.1598/JAAL.54.1.1>
- Beck, I. L., McKeown, M. G. og Kucan, L. (2013). *Bring words to life: Robust vocabulary instruction*. Guilford Press.
- Birna Arnbjörnsdóttir og Hafdís Ingvarsdóttir. (2018). Language development across the life span. English in Iceland: From input to output. Í Birna Arnbjörnsdóttir og Hafdís Ingvarsdóttir (ritstjórar), *Language development across the life span: The impact of English on education and work in Iceland*. Springer.
- Brezina, V. og Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New general service list. *Applied Linguistics*, 36(1), 1–22. <https://doi.org/10.1093/applin/amt018>

- Brezina, V. og Gablasova, D. (2017). How to produce vocabulary lists? Issues of definition, selection and pedagogical aims. A response to Gabriele Stein. *Applied Linguistics*, 38(5), 764–767. <https://doi.org/10.1093/applin/amx022>
- Browne, C., Culligan, B. og Phillips, J. (2013a). *New academic word list 1.0. New general service list*. <https://www.newgeneralservicelist.com/new-general-service-list-1>
- Browne, C., Culligan, B. og Phillips, J. (2013b). *New general service list project. New general service list*. <http://www.newgeneralservicelist.org/>
- Calderón, M., August, D., Slavin, R., Duran, D., Madden, N. og Cheung, A. (2005). Bringing words to life in classrooms with English-language learners. Í E. H. Hiebert og M. L. Kamil (ritstjórar), *Teaching and learning vocabulary: Bringing research to practice* (bls. 115–136). Lawrence Erlbaum.
- Collier, V. P. og Thomas, W. P. (2009). *Educating English learners for a transformed world*. Fuente Press.
- Coxhead, A. (1998). *An academic word list* (English Language Institute occasional publication no. 18). Victoria University of Wellington.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. https://www.victoria.ac.nz/_data/assets/pdf_file/0020/1626131/Coxhead-2000.pdf
- Coxhead, A. (2020). Academic vocabulary. Í S. Webb (ritstjóri), *Routledge handbook of vocabulary studies* (bls. 97–110). Routledge.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Multilingual Matters.
- Dang, T. N. Y. (2020). Corpus-based wordlists in second language vocabulary research, learning, and teaching. Í S. Webb (ritstjóri), *Routledge handbook of vocabulary studies* (bls. 288–303). Routledge.
- DiCerbo, P. A., Anstorm, K. A., Baker, L. L. og Rivera, C. (2014). A review of the literature on teaching academic English to English language learners. *Review of Educational Research*, 84(3), 446–482. <https://doi.org/10.3102/0034654314532695>
- Erla Lind Þórisdóttir. (2017). *Orðaþröskuldur íslenskra grunnskólanema á miðstigi: Hlutfall þekktra orða í náttúrufræðitexta og lesskilningur [meistararitgerð, Menntavísindasvið Háskóla Íslands]*. Skemman. <http://hdl.handle.net/1946/28485>
- Freyja Birgisdóttir. (2016). Orðaforði og lestrarfærni: Tengsl við gengi nemenda á lesskilningshluta PISA. *Sérít Netlu 2016 – Um læsi*. https://netla.hi.is/serrit/2016/um_laesi/04_16_laesi.pdf
- Friðrik Magnússon. (1988). Hvað er títt? Tíðnikönnun Orðabókar Háskólans. *Orð og tunga*, (1), 1–49. <https://timarit.is/page/6760074#page/n12/mode/2up>
- Friðrik Magnússon, Stefán Briem og Jörgen Pind. (1991). *Íslensk orðtíðnibók*. Orðabók Háskólans.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241–265. <https://doi.org/10.1093/applin/amm010>
- Gardner, D. og Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Gottlieb, M. og Ernst-Slavit, G. (2013). *Academic language in diverse classrooms: Promoting content and language learning. English language arts, grades 3–5*. Corwin.
- Hu, M. og Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 23, 403–430.
- Hyland, K. (2016). General and specific EAP. Í K. Hyland og P. Shaw (ritstjórar), *Routledge handbook of english for academic purposes* (bls. 17–29). Routledge.
- Kristín Bjarnadóttir (ritstjóri). (e.d.). *Beýgingarlýsing íslensks nútímamáls*. Stofnun Árna Magnússonar í íslenskum fræðum. <https://bin.arnastofnun.is/forsida/umleit/flettioerd/>
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? Í C. Laurén og M. Nordemann (ritstjórar), *Special languages: From humans to thinking machines* (bls. 316–323). Multilingual Matters.
- Laufer, B. (1992). How much lexis is needed for reading comprehension? Í P. J. L. Arnaud og H. Béjoint (ritstjórar), *Vocabulary and applied linguistics* (bls. 126–132). Macmillan.
- Laufer, B. (2020). Evaluating exercises for learning vocabulary. Í S. Webb (ritstjóri), *The Routledge handbook of vocabulary learning* (bls. 351–368). Routledge. <https://doi.org/10.4324/9780429291586>
- Laufer, B. og Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.

- Lawrence, J. F., Francis, D., Paré-Blagoev, J. og Snow, C. E. (2016). The poor get richer: Heterogeneity in the efficacy of school-level intervention for academic language. *Journal of Cognitive Neuroscience*, 10(4), 767–793. <https://doi.org/10.1080/19345747.2016.1237596>
- Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. Í R. K. Wagner, A. E. Muse og K. R. Tannenbaum (ritstjórar), *Vocabulary acquisition: Implications for reading comprehension* (bls. 52–77). The Guilford Press.
- Nagy, W. E. og Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91–108. <https://doi.org/10.1002/RRQ.011>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P. (2006). How large vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. https://www.wgtn.ac.nz/__data/assets/pdf_file/0005/1857641/about-bnc-coca-vocabulary-list.pdf
- Nation, I. S. P. (2016). *Making and using word lists for language teaching and learning*. John Benjamins.
- Nippold, M. A. og Sun, L. (2008). Knowledge of morphologically complex words: A developmental study of older children and young adolescents. *Language, Speech, and Hearing Services in Schools*, 39(3), 365–373. [https://doi.org/10.1044/0161-1461\(2008/034\)](https://doi.org/10.1044/0161-1461(2008/034))
- OECD. (2016). *PISA 2018 translation and adaptation guidelines. First meeting of the PISA 2018 national project managers*. <https://www.oecd.org/pisa/pisaproducts/PISA-2018-TRANSLATION-AND-ADAPTATION-GUIDELINES.pdf>
- Oller, D. K., Pearson, B. Z. og Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(2), 191–230. <https://doi.org/10.1017/S0142716407070117>
- Peters, E. (2020). Factors affecting the learning of single-word items. Í S. Webb (ritstjóri), *Routledge handbook of vocabulary studies* (bls. 125–142). Routledge.
- Quigley, A. (2018). *Closing the vocabulary gap*. Routledge.
- Roessingh, H. (2016). Academic language in K-12: What is it, how is it learned, and how can we measure it? *BC TEAL Journal*, 1(1), 67–81. <https://doi.org/10.14288/bctj.v1i1.235>
- Roessingh, H. og Douglas, S. (2013). Raising the lexical bar: The potential of teacher talk to support learning academic vocabulary. Í M. C. G. Anderson (ritstjóri), *Teaching and leading in diverse schools* (bls. 99–124). The Federation of North Texas Universities.
- Roessingh, H. og Elgie, S. (2009). Early language and literacy development among young English language learners: Preliminary insights from a longitudinal study. *TESL Canada Journal*, 26(2), 24–45. <https://doi.org/10.18806/tesl.v26i2.413>
- Roessingh, H. og Kover, P. (2002). Working with younger-arriving ESL learners in high school English: Never too late to reclaim potential. *TESL Canada Journal*, 19(2), 1–19. <https://doi.org/10.18806/tesl.v19i2.926>
- Roessingh, H., Elgie, S. og Kover, P. (2015). Using lexical profiling tools to investigate children's written vocabulary in grade 3: An exploratory study. *Language Assessment Quarterly*, 12(1), 67–86. <https://doi.org/10.1080/15434303.2014.936603>
- Roessingh, H., Kover, P. og Watt, D. (2005). Developing cognitive academic language proficiency: The journey. *TESL Canada Journal*, 23(1), 1–27. <https://doi.org/10.18806/tesl.v23i1.75>
- Schmitt, N. og Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Schmitt, N. og Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145–171. <https://doi.org/10.2307/3588328>
- Schmitt, N., Jiang, X. og Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Sigríður Ólafsdóttir og Auður Pálsdóttir. (2023). *Listi yfir íslenskan námsorðaförða: LÍNO-2. CLARIN-IS*. <http://hdl.handle.net/20.500.12537/307>
- Sigríður Ólafsdóttir og Baldur Sigurðsson. (2017). Hnignandi frammistaða íslenskra nemenda í lesskilningshluta PISA frá 2000 til 2015. Leiðir til að snúa þróuninni við. *Netla – Vefritmarit um uppeldi og menntun*. <http://netla.hi.is/greinar/2017/ryn/16.pdf>

- Sigríður Ólafsdóttir og Baldur Sigurðsson. (2019). Rýnt í niðurstöður lesskilningshluta PISA 2018 og tækifæri til framfara. Í *PISA 2018: Helstu niðurstöður á Íslandi* (bls. 38–61). Menntamálastofnun. https://mms.is/sites/mms.is/files/pisa_2019_0.pdf
- Sigríður Ólafsdóttir og Hrafnhildur Ragnarsdóttir. (2010). Íslenskur orðaforði íslenskra grunnskólanema sem eiga annað móðurmál en íslensku. *Ráðstefnurit Netlu – Menntakvika 2010*. <https://netla.hi.is/serrit/2010/menntakvika2010/alm/023.pdf>
- Sigríður Ólafsdóttir, Auður Pálsdóttir og Starkaður Barkarson. (2023a). *Málheild fyrir íslenskan námsorðaforða: MÍNO. CLARIN-IS*. <http://hdl.handle.net/20.500.12537/299>
- Sigríður Ólafsdóttir, Auður Pálsdóttir og Starkaður Barkarson. (2023b). *Orðiðnilisti málheildar fyrir íslenskan námsorðaforða: MÍNOfreq. CLARIN-IS*. <http://hdl.handle.net/20.500.12537/307>
- Sigríður Ólafsdóttir, Barbara Laster og Kristján K. Stefánsson. (2020). Adolescent learning of academic vocabulary in Iceland. *Journal of Adolescent and Adult Literacy*, 64(1), 79–87. <https://doi.org/10.1002/jaal.1055>
- Sigríður Ólafsdóttir, Freyja Birgisdóttir, Hrafnhildur Ragnarsdóttir og Sigurgrímur Skúlason. (2016). Orðaforði og lesskilningur nemenda með íslensku sem annað mál: Áhrif aldurs við komuna til Íslands. *Sérarit Netlu 2016 – Um læsi*. https://netla.hi.is/serrit/2016/um_laesi/03_16_laesi.pdf
- Sigríður Ólafsdóttir, Freyja Birgisdóttir, Hrafnhildur Ragnarsdóttir og Sigurgrímur Skúlason. (2017). Íslenskur orðaforði og lesskilningur nemenda með íslensku sem annað mál: Tengsl við móðurmálskennslu. *Glæður – Tímarit sérkennara*, 27, 35–51.
- Sigrún Helgadóttir, Ásta Svavarsdóttir, Eiríkur Rögnvaldsson, Kristín Bjarnadóttir og Hrafn Loftsson. (2012). The Tagged Icelandic corpus (MÍM). *Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages -SaLTMiL 8 – AfLaT2012* (bls. 67–72). Ísland, Tyrklandi.
- Snow, C. (2014). Input to interaction to instruction: Three key shifts in the history of child language research. *Journal of Child Language*, 41(1), 117–123. <https://doi.org/10.1017/S0305000914000294>
- Snow, C. (2016). The role of vocabulary versus knowledge in children's language learning: A fifty-year perspective. *Journal of the Study of Education and Development*, 40(1), 1–18. <https://doi.org/10.1080/02103702.2016.1263449>
- Staples, S., Biber, D. og Reppen, R. (2018). Using corpus-based register analysis to explore the authenticity of high-stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. *The Modern Language Journal*, 102(2), 310–332. <https://doi.org/10.1111/modl.12465>
- Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson og Jón Guðnason. (2018). *Risamálheild: A very large Icelandic text corpus. Proceedings of LREC 2018*. <http://www.lrec-conf.org/proceedings/lrec2018/summaries/746.html>
- Stofnun Árna Magnússonar í íslenskum fræðum. (e.d.-a). *Mörkuð íslensk málheild*. <http://mim.arnastofnun.is/>
- Stofnun Árna Magnússonar í íslenskum fræðum. (e.d.-b). *Risamálheildin*. <http://igc.arnastofnun.is/is/index.html#About>
- Vilkaitè-Lozdiènè, L. og Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness: High-, mid-, and low-frequency words. Í S. Webb (ritstjóri), *Routledge handbook of vocabulary studies* (bls. 81–96). Routledge.
- West, M. (1953). *The general service list of English words*. Longman Green.
- Zipf, G. (1935). *The psychobiology of language: An introduction to dynamic philology*. M.I.T. Press.
- Þórhildur Oddsdóttir. (2012). Orð eru til alls fyrst. *Milli mála*, 4, bls. 257–292.



Auður Pálsdóttir og Sigríður Ólafsdóttir. (2023). Íslenskur námsorðaforði
 Netla – Vef tímarit um uppeldi og menntun. Menntavísindasvið Háskóla Íslands.
 Sótt af <http://netla.hi.is/greinar/2023/alm/09.pdf>
 DOI: <https://doi.org/10.24270/netla.2023/9>