



# Greining á orðanotkun í lesskilnings- og náttúruvísindahlutum PISA 2018: Samanburður á íslensku þýðingunni og enska textanum

Auður Pálsdóttir og Sigríður Ólafsdóttir

► Abstract    ► Um höfundana    ► About the authors    ► Heimildir

Markmið þessarar rannsóknar var að kanna samræmi í orðtíðni í íslenskum og enskum textum á lesskilnings- og náttúruvísindahluta PISA-prófanna 2018. Ef þýðing texta í alþjóðlegu prófi er þyngri eða léttari en sami texti á upprunalega málinu getur það haft áhrif á skilning og skekkt samanburð milli tungumála.

Greindir voru tveir textar úr lesskilningshluta PISA 2018 og tveir úr náttúruvísindahlutanum. Notaður var orðtíðnilisti *Íslenskrar risamálheildar* og enskur orðtíðnilisti sem byggist á tveimur málheildum og er aðgengilegur í gegnum hugbúnaðinn *VocabProfile*. Orðin voru flokkuð eftir tíðni í fimm flokka. Ef munur var á tíðniflokki orða á íslensku og ensku var kannað hvort til væri samheiti fyrir íslenska orðið í sama tíðniflokki og það enska og lengd samheita borin saman.

Niðurstöður benda til að hlutfall algengustu orða sé lægra í textum íslensku þýðingarinnar en í ensku frumtextunum og að hlutfall orða í flokki sjaldgæfustu orðanna sé umtalsvert hærra í íslensku textunum en þeim ensku. Þá virðist dreifing orða á milli orðtíðniflokka vera jafnari í ensku en íslensku þýðingunni. Fram komu vísbendingar um ákveðið ósamræmi og ójafnvægi sem fólst í að tveir þriðju hlutar þeirra íslensku orða, sem féllu í annan tíðniflokk en ensku orðin, voru sjaldgæfari en samsvarandi ensk orð. Í ljós kom að fækka hefði mátt orðum í ólíkum orðtíðniflokkum með því að nota íslenskt samheiti í sama orðtíðniflokki og enska orðið og að draga hefði mátt enn frekar úr ósamræminu með því að velja samheiti úr nærliggjandi tíðniflokki. Hlutfall íslenskra samheita sem voru algengari og lengri var yfir 30% í textunum fjórum. Niðurstöðurnar gefa tilefni til að endurskoða þurfi leiðbeiningar OECD og beina því til þýðenda að þeir taki mið af orðtíðnilistum við val á orðum.

**Efnisorð:** PISA, þýðingar, orðaforði, orðtíðni, lesskilningur, náttúruvísindi

## Inngangur

PISA-rannsóknin (e. The Programme for International Student Assessment) byggist á prófum sem lögð eru fyrir 15 ára nemendur í fjölmörgum ríkjum innan og utan OECD (e. Organisation for Economic Co-operation and Development). Prófin eru í þremur hlutum; í lesskilningi, læsi á stærðfræði og læsi á náttúruvísindi, og samin á ensku og frönsku. Prófin eru síðan þýdd á hin ýmsu tungumál þátttökuþjóða. Til að ná fram alþjóðlegum samanburði er grundvallaratriði að

samræmi sé á milli prófa á ólíkum tungumálum og þess vegna eru gerðar ítrustu kröfur um nákvæmni í þýðingum. Ef þýddur texti verður of erfiður hamlar það nemendum að skilja og vinna með textann, umfram þá sem lesa textann á upprunalega tungumálinu. Í lesskilningshluta prófsins reynir til að mynda á hæfni nemenda til að finna, velja og meta upplýsingar í texta, nýta bakgrunnsþekkingu sína og að veita og meta ólík sjónarhorn (LaRusso o.fl., 2016; OECD, 2016). Því getur of þungur texti, með of flóknum orðum, orðið til trafala þannig að lesandinn nái síður að beita viðeigandi lestrartækni.

Þegar PISA-prófin voru lögð fyrir í fyrsta skipti árið 2000 var frammistaða íslenskra ungmenna í lesskilningi á pari við meðaltal OECD en hefur síðan hrakað verulega, samtals um 25 stig sem samsvarar rúmlega hálfu skólaári. Afturförin birtist einnig í því að hlutfall nemenda á lægstu hæfniþrepunum hefur hækkað en lækkað á efstu þrepunum (Menntamálastofnun, 2017; Sigríður Ólafsdóttir og Baldur Sigurðsson, 2017).

Þegar horft er til niðurstaðna PISA í læsi á náttúruvísindi frá árinu 2006 til 2015 hefur frammistöðu íslenskra nemenda farið aftur og gildir það um alla undirþætti, þekkingarsvið og efnissvið náttúruvísinda. Þá stendur mikill meirihluti OECD-ríkja betur að vígi en Ísland á sviði læsis á náttúruvísindi (Menntamálastofnun, 2017).

Líkt og í lesskilningshluta PISA-prófa sýna niðurstöður fyrir læsi á náttúruvísindi fjölgun nemenda í lægstu hæfniþrepum og fækkun í efstu hæfniþrepunum. Hæfniþrep 2 er séð sem lágmarksfærni til að nemandi geti nýtt færni sér til gagns við úrlausnir á verkefnum daglegs lífs sem bíða hans eftir grunnskóla. Hlutfall íslenskra nemenda á hæfniþrepi 2 eða neðar hækkaði úr 47% árið 2006 í 55% árið 2015. Þegar rýnt er nánar í þekkingarsvið læsis á náttúruvísindi má sjá að íslensku nemendurnir eru einum til tveimur árum á eftir jafnöldrum sínum í Noregi, Svíþjóð og Danmörku í eðlis- og efnafræði. Þá benda niðurstöðurnar til þess að íslenskum unglingum hafi farið mest aftur á þekkingarsviði jarðvísinda og stjörnufræði (e. earth and space) (Auður Pálsdóttir, 2017).

Margt bendir til þess að orðaforði sé veigamikill áhrifavaldur á frammistöðu nemenda í lesskilningi og veiti forspá um hvort þeir lendi undir lágmarksviðmiðum í PISA (Arnbak, 2010; Freyja Birgisdóttir, 2016; LaRusso o.fl., 2016; Sigríður Ólafsdóttir, Freyja Birgisdóttir, Hrafnhildur Ragnarsdóttir og Sigurgrímur Skúlason, 2016). Minnkandi áhugi á náttúrufræði er m.a. rakinn til þess að nemendur eigi stöðugt erfiðara með námið því orðaforði og málskilningur er slakur. Til að efla djúpan lesskilning benda rannsóknir til svo ekki verður um villst að efla þurfi hærra stigs orðaforða nemenda (þ.e. lágtíðniorða) og beita þurfi lesskilningsaðferðum í öllum námsgreinum (Auður Pálsdóttir, 2017).

Í samanburði á milli landa eins og í PISA þarf að vera sem best samræmi á hvernig reynir á orðaforða nemenda svo réttmætis sé gætt og prófin mæli það sem þeim er ætlað að mæla. Því var í þessari rannsókn rýnt í orðaval í íslenskri þýðingu PISA 2018.

## Fræðilegur grunnur

OECD gefur út leiðarvísi fyrir þýðendur prófanna (OECD, 2016). Mælt er með að bæði enska og franska útgáfan séu þýddar, hvor í sínu lagi, og þær bornar saman. Þess er þó getið að í sumum löndum sé illframskræmanlegt að þýða bæði tungumálin og þá þurfi að gæta enn meiri nákvæmni. Hér á landi hefur aðeins verið byggt á ensku útgáfunni (Menntamálastofnun, 2016; OECD, 2016). Löggiltir þýðendur eru ráðnir til verksins, þeir eru allir kennarar og hafa kennt eða starfa við kennslu í grunnskólum og framhaldsskólum og er ferlið sambærilegt í öllum þáttökulöndunum. Verkferlið er skýrt á vef Menntamálastofnunar (2016).

Sterkasta forspá um frammistöðu nemenda í lesskilningshluta PISA er orðaforði þeirra (Arnbak, 2010; Freyja Birgisdóttir, 2016). Skilningur nemenda á orðum er forsenda þess að þeir geti ráðið

í merkingu textans og unnið með hann. Í íslenskri rannsókn kom í ljós að 98% orðskilningur er nauðsynlegur fyrir góðan lesskilning, og ef nemendur skilja ekki tvö orð eða fleiri af 100 ná þeir ekki að átta sig nægilega vel á innihaldi hans (Erla Lind Þórisdóttir, 2017). Er það í samræmi við niðurstöður rannsókna með nemendum sem fást við enskan texta (Hu og Nation, 2000; Laufer og Ravenhorst-Kalovski, 2010). Því er ljóst að hvert orð í texta skiptir máli, þ.e. hvort hann skilst eða ekki.

Í leiðbeiningum OECD (2016) eru sett almenn viðmið um hvernig þyngd orða tengist tíðni þeirra, lengd eða sérhæfingu án tillits til tungumáls. Um þýðingar milli mála segir að velja eigi sambærileg orð að þyngd og sérhæfingu í merkingu og þannig skuli forðast að velja flóknari eða einfaldari orð en eru í upphaflegu útgáfunni. Á það við í þýðingu á fyrirmælum, textum og svarmöguleikum. Til að meta hvort orð eru flókin eða einföld segir í leiðbeiningum OECD að lengri orð hafi tilhneigingu til að vera flóknari, tæknilegri og/eða óhlutlægri en stutt orð. Þýðendur eigi að forðast að velja erfið orð með því að nota orð eða orðasambönd sem innihalda algengari orð. Á hinn bóginn eigi ekki að þýða orð og orðasambönd sem tilheyra daglegu lífi með tæknilegri orðum eða orðum sem frekar tilheyra ritmáli (OECD, 2016, bls. 11). Í leiðbeiningum OECD segir:

It is no coincidence that average word length and sentence length almost always feature among indices of complexity used in readability formulas, regardless of the language for which the formulas have been developed. Longer words tend to be less frequent, more technical and/or more abstract than short words ... As a general rule, translators should try their best to avoid translating difficult words in the text by using easier words, or paraphrases that use more common terms. Conversely, everyday terms used in the source materials should not be translated as more 'technical', 'scientific' or 'literary' expressions. (OECD, 2016, bls. 11)

Þannig er í leiðarvísi OECD ekkert fjallað um samsett orð sem eru eðlilega lengri en ósamsett orð. Íslensk samsett orð innihalda iðulega algeng orð eða orðhluta og eru því gegnsæ og auðskilin. Sem dæmi má nefna orðið nafnlaus en samsvarandi enska orðið anonymous er af grískum uppruna og því líklegra að það sé torskildara en íslenska orðið.

Til að bera saman áreiðanleika þýðinga hafa rannsakendur gert tölfræðilega úrvinnslu á svörum nemenda, þar sem borin hefur verið saman dreifing stiga eftir einstökum prófum og prófþáttum (e. DIF, difference item functioning) (Chen og Jiao, 2014; Grisay, de Jong, Gebhardt, Berezner og Halleux-Monseur, 2007; Oliveri og Ercikan, 2011). Örfáar rannsóknir hafa beinst að því hvort samræmi sé í orðavali frumtexta PISA og þýðinga (Hopfenbeck o.fl., 2018). Ekki er vitað um samanburð á tíðni orða í þýðingum og upprunalegum texta PISA-prófa út frá málheildum, þ.e. orðasöfnum sem byggjast á fjölda margs konar texta.

Arffman (2010) bar saman finnsku þýðinguna og enska textann í lesskilningshluta prófsins sem lagt var fyrir árið 2000. Hún komst að því að finnsku orðin voru að jafnaði lengri en þau ensku. Á hinn bóginn var finnski textinn iðulega óformlegri og því auðveldari en sá enski. Ástæðan fyrir því reyndist vera að í finnsku er lítið sem ekkert af erlendum orðum en aftur á móti eru mörg orð í ensku af erlendum uppruna. Þó kom í ljós að ýmist of erfiður eða of léttur texti í samanburði við upprunalega enska textann varð til þess að jafnvægi náðist að einhverju leyti, hvor skekkjan vóg upp á móti hinni. Arffman dró þá ályktun að nauðsynlegt væri að veða og meta hvert og eitt orð í báðum málum til að fá nákvæman samanburð á upprunalega textanum og þýðingunni.

Ásdís Bergþórsdóttir (2016) greindi íslensk orð í náttúruvísindahluta PISA-prófsins árið 2015 og bar saman við upphaflegu ensku útgáfuna. Aðferðin sem hún beitti við að meta flækjustig íslenskra orða fólst í að skoða hve oft orðin birtust á leitarsíðunni Google.com. Hún komst að því að íslensku orðin voru iðulega of flókin miðað við þau ensku. Þá benti hún á dæmi um rangar þýðingar á orðum.

Sterk tengsl á milli tíðni orða og þess hve líklegt er að einstaklingar þekki orðin hafa komið fram í fjölmörgum erlendum rannsóknum (Baayen, Wurm og Aycock, 2007; Balota, Yap og Cortese, 2006; Gardner, Rothkopf, Lapan og Lafferty, 1987; Meunier og Segui, 1999; Oldfield og Wingfield, 1965). Því algengari sem orðin eru þeim mun líklegri er að einstaklingar þekki og skilji þau orð.

Tíðni orða er hægt að finna í orðtíðnilistum sem byggja á orðasöfnum, svokölluðum málheildum. Síðustu tvo áratugi hafa orðið miklar tæknilegar framfarir í tölvuvinnslu stórra málheilda sem byggja á tungumáli samtímans (Biber og Reppen, 2015; Biber, Reppen og Friginal, 2010; Martinez og Schmitt, 2015; Nation, 2016; Reppen, 2010). Orðum er raðað eftir því hve oft þau koma fyrir í málheildinni, það algengasta er númer eitt og svo koll af kolli. Slíkar upplýsingar eru einkar mikilvægar til að finna orð sem mikilvægt er að nemendur þekki (Cobb, 2007; Dale og O'Rourke, 1981; Hiebert, Scott, Castaneda og Spichtig, 2019). Orðtíðnilistar hafa þó þær takmarkanir að þar er að finna einstök orð en ekki orðasambönd og ekki er tekið tillit til margvíslegrar merkingar sem sama orð í sama orðflokki getur haft.

Í hverju tungumáli eru fá orð sem eru mjög algeng bæði í tali og riti, svokölluð hátíðniorð, en ótal mörg orð sem koma sjaldan fyrir, svokölluð lágtíðniorð (Friðrik Magnússon, 1988; Nation, 2001). Rannsóknir hafa leitt í ljós að nemendur þekkjja lágtíðniorðin í mismiklum mæli, en þekking á slíkum orðum hefur mikil áhrif á lesskilning þeirra (Biemiller, 2008; Roessingh, 2016). Af því má ráða að hlutfall lágtíðniorða þarf að vera sambærilegt á hinum ýmsu tungumálum til að samanburður á milli ríkja sé réttmætur (sjá t.d. Romanko, 2017).

Til að skilja betur frammistöðu 15 ára íslenskra nemenda og breytingar þar á er mikilvægt að rýna í niðurstöður PISA-prófanna hér á landi (Auður Pálsdóttir, 2017; Sigríður Ólafsdóttir og Baldur Sigurðsson, 2017). Ein ástæða neikvæðrar þróunar á Íslandi er sennilega sú að bóklestur barna og unglunga hefur farið minnkandi síðustu áratugi (Freyja Birgisdóttir, 2016; Rannsóknir og greining, 2016). Lesskilningur eykst með ástundun lestrar og því er hrakandi færni barna og unglunga til að nýta sér ritadán texta óhjákvæmileg afleiðing þess að þau lesa minna en áður. Þó má gera ráð fyrir að sama eigi við ungmenni annars staðar í heiminum, að vegna aukinnar tölvunotkunar og samhliða sífjölbreytilegra framboði afþreyingarefnis verji þau minni tíma í lestur. Íslenskan er þó í hópi þeirra tungumála sem standa sérstaklega illa í stafrænum heimi því tiltölulega lítið er til af íslensku efni á veraldarvefnum. Íslenskt tungumál er einmitt eitt þeirra mála sem eiga á hættu að deyja út með komandi kynslóðum ef ekki verður lögð áhersla á að gera þau gjaldgeng í stafrænum heimi (Eiríkur Rögnvaldsson, Kristín M. Jóhannsdóttir, Sigrún Helgadóttir og Steinþór Steingrímsson, 2012).

Slæm staða íslenskra nemenda í samanburði við jafnaldra þeirra erlendis í lesskilningi og læsi á náttúruvísindi í PISA kallar á nánari rýni í nákvæmni þýðingar frá ensku yfir á íslensku. Markmið þessarar rannsóknar var því að kanna samræmi í orðtíðni í íslenskum og enskum textum lesskilnings- og náttúruvísindahluta PISA-prófanna 2018. Eins og áður segir er í leiðbeiningum OECD (2016) um þýðingar PISA-prófanna tekið fram að þyngd orða og sérhæfing í merkingu sé mikilvægt viðfangsefni þýðenda, en jafnframt bent á að lengri orð hafi tilhneigingu til að vera sjaldgæfari og því þurfi þýðendur að forðast að þýða erfið orð með auðveldari orðum eða umorða erfið orð með auðveldara orðalagi (bls. 11). Í þessu ljósi þótti fróðlegt að skoða sérstaklega íslensku orðin sem féllu ekki í sama tíðniflokk og samsvarandi enska orðið og bera saman lengd íslensku orðsins við samheiti þess í sama eða nærliggjandi tíðniflokki og enska orðið.

Mótaðar voru fjórar rannsóknarspurningar sem öflun og greining gagna snerist um:

Hvert er hlutfall orða í fimm tíðniflokkum í íslensku og ensku í lesskilnings- og náttúruvísindatextum PISA-prófanna 2018?

Hversu hátt hlutfall þýddra íslenskra orða fellur í annan tíðniflokk en samsvarandi ensk orð í upprunalegu textunum?

Fyrir hversu hátt hlutfall þeirra orða hefði mátt velja samheiti sem fellur í sama eða nærliggjandi tíðniflokk og samræmist þannig betur enska orðinu?

Hversu oft hlutfallslega er algengara íslenska orðið lengra en sjaldgæfara samheiti þess?

Rannsókn af þessum toga og svör við spurningum hennar getur verið hvati fyrir aðrar þjóðir í þeirri viðleitni að samanburður á milli landa verði réttmætari.

## Aðferð

Valdir voru af handahófi fjórir textar úr PISA-prófinu árið 2018, tveir textar úr lesskilningshlutanum (textinn R055 um kóngulær og R406 um brúðumeistara) og tveir textar úr náttúruvísindahlutanum (textinn S527 um aldaða risaeðla og S326 um mjólk) og orðaval í íslensku þýðingunni var borið saman við ensku textana. Greiningin hófst í lok mars 2018, þ.e. á svipuðum tíma og prófin voru lögð fyrir á Íslandi. Aðgengi að PISA-prófinu 2018 fékkst hjá Menntamálastofnun.

## Val á málheildum og orðtíðnilistum

Greining þessara texta PISA-prófsins byggir á orðtíðnilistum sem koma úr íslenskum og enskum málheildum. Þrjár umfangsmiklar málheildir hafa fengist úr rannsóknum á íslensku máli. Úr hverri þeirra hafa verið unnir orðtíðnilistar þar sem hvert orð er sett fram sem svokallað flettiorð, eða lemma, þ.e. ein lemma er hvert orð í öllum þeim myndum sem það birtist í textanum (Friðrik Magnússon, 1988, bls. 2).

Íslensk orðtíðnibók (Friðrik Magnússon og Stefán Briem, 1991) byggir á textasafni úr íslenskum og þýddum sögum, ævisögum, fræðilegum textum, ásamt barna- og unglingabókum, samtals um fimm hundruð þúsund lesmálsorðum. Fjöldi lesmálsorða í texta er í raun fjöldi orðanna í textanum, alveg sama hvort sömu orðin koma oft en einu sinni fyrir og hvort um mismunandi beygingarmyndir er að ræða.

Í Markaðri íslenskri málheild má finna um 25 milljónir lesmálsorða úr fjölbreyttum tölvutækum textum sem safnað var á árunum 2006–2010 (Stofnun Árna Magnússonar í íslenskum fræðum, e.d.). Meirihlutinn kemur úr útgefnum bókum, dagblöðum og tímaritum, en innan við 15% eru textar frá opinberum aðilum eins Alþingi, ráðuneytum og úr dómum.

Til að renna styrkari stöðum undir rannsóknir á íslensku máli var ráðist í gerð nýrrar málheildar sem kallast *Íslensk risamálheild*. Hún er safn um 1300 milljón lesmálsorða þar sem stærstur hlutinn kemur úr margs konar vefmiðlum, prentuðum blöðum, ljósvakamiðlum og opinberum textum. Um það bil 86% textanna eru frá því eftir árið 2000, og rúmlega 94% frá því eftir árið 1980. Í þessari rannsókn var byggt á Risamálheildinni en sóttur var rafrænn orðtíðnilisti sem aðgengilegur er til niðurhals á vefnum malfong.is og unninn var með sérsmíðuðu forriti upp úr gögnum Risamálheildarinnar (Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson og Jón Guðnason, 2018; Stofnun Árna Magnússonar í íslenskum fræðum, 2017, e.d.).

Til að greina tíðni orða í ensku textunum var notaður orðtíðnilisti sem byggist á tveimur málheildum og aðgengilegur er í gegnum hugbúnaðinn *VocabProfile* (Cobb, e.d.). Þessar málheildir voru valdar því þær samræmdust best Risamálheildinni og orðtíðnilistinn samanstendur af lemmum. Annars vegar var það enska málheildin *The New General Service List* (NGSL; Browne, Culligan og Phillips, 2013) sem byggir á 273 milljón lesmálsorðum úr *Cambridge English Corpus* (CEC), með töluðu og rituðu nútímamáli í dagblöðum, á vefmiðlum, í bókum, tímaritum, útvarpi, skólum, háskólum og á vinnustöðum. Hins vegar var það *The New Academic Word List* (NAWL) sem felur í sér vinnslu á 288 milljón lesmálsorðum sem koma úr fagbókum á

toppsölulistum og úr fræðilegum erindum. Þessum tveimur málheildum er sérstaklega ætlað að vinna saman.

PISA er ætlað að varpa ljósi á færni nemenda við lok grunnskóla til að taka virkan þátt í samfélaginu og ná árangri í námi (OECD, 2017). Bæði íslenski orðtíðnilistinn og sá enski samanstanda af orðum (þ.e. lemmum) sem ætla má að gegni stóru hlutverki á þeirri vegferð. Orðin tilheyra þó einna helst heimi hinna fullorðnu. Þá virðist vanta í Risamálheildina orðaforða sem er að finna t.d. í námsefni og textum sem sérstaklega eru ætlaðir ungu fólki en eins og áður segir inniheldur önnur enska málheildin tungumál sem notað er í skólum.

## Greining gagna

Þegar ensku textarnir voru keyrðir í gegnum hugbúnaðinn *ν* fékkst samstundis sundurliðun á orðum byggð á tíðni þeirra. Orð sem féllu í hóp algengustu 1000 orðanna fengu bláan lit, næstu 1000 grænan, næstu 1000 bleikan, næstu 1000 gulan og orð sem voru sjaldgæfari eða fundust ekki í málheildinni fengu rauðan lit.

Orðin í íslensku þýðingunni voru borin saman við orðtíðnilista Risamálheildarinnar og flokkuð á sama hátt í fimm tíðniflokka. Þannig fengust upplýsingar um hversu hátt hlutfall orða féll í hvern tíðniflokk á báðum málum og hvort þýdd íslensk orð tilheyrðu sama tíðniflokki og ensku orðin.

Enska hugbúnaðarforritið og Risamálheildin greina tíðni einstakra orða en ekki orðasambanda. Því var farin sú leið að tíðni einstakra íslenskra orða var fundin og ef það voru fleiri en eitt samsvarandi ensk orð var skráð tíðni þess orðs sem lenti í sjaldgæfasta tíðniflokknum. Til að meta gildi samanburðarins var hugað að margræðni orða og þau ávallt rýnd í samhengi textanna og samkvæmt orðflokki.

Íslensk orð sem féllu ekki í sama orðtíðniflokk og samsvarandi ensk orð voru skráð í lista og leitað að samheitum þeirra á vefsíðunum [malid.is](http://malid.is) og [snara.is](http://snara.is) til að athuga hvort samheiti væri til í sama eða nærliggjandi tíðniflokki og enska orðið, og hæfðu samhengi textanna. Fyrirnefnda vefsíðan inniheldur Íslensk nútímamálsorðabók (Halldóra Jónsdóttir og Þórdís Úlfarsdóttir, 2018) sem gefur samheiti orða.

Að lokum var borin saman lengd íslensku samheitanna, þ.e. fjöldi bókstafa í hverju orði, til að kanna í hvaða tilvikum algengara samheitið væri lengra en það sjaldgæfara.

## Niðurstöður

Eins og áður kom fram var markmið þessarar rannsóknar að kanna samræmi í orðtíðni í íslenskum og enskum textum lesskilnings- og náttúruvísindahluta PISA-prófanna 2018.

Hér eru fyrst kynntar niðurstöður greiningar á hlutfalli orða í tíðniflokkunum fimm í textunum fjórum. Þá er gerð grein fyrir fjölda og hlutfalli þýddra orða sem falla í annan tíðniflokk en samsvarandi ensk orð í upprunalegu textunum. Næst er fjallað um fjölda og hlutfall þeirra orða sem hefði mátt velja samheiti fyrir og samræmdist þannig betur tíðniflokki enska orðsins. Þar á eftir er tilgreint að hve miklu leyti algengara íslenska orðið er lengra en sjaldgæfara samheiti þess. Í lokin er samantekt á niðurstöðum.

## Hlutfall orða í hverjum orðtíðniflokki

Greining á orðum byggðist á tíðni þeirra og var þeim skipt upp í fimm tíðniflokka. Þannig fengust upplýsingar um hlutfall orða í hverjum tíðniflokki. Niðurstöður má sjá í töflu 1 fyrir

íslensku textana og töflu 2 fyrir ensku textana. Í hnotskurn sýna þær að hlutfall algengustu orða var lægra í öllum textum íslensku þýðingarinnar en ensku útgáfunnar og á sama hátt að hlutfall orða í flokki sjaldgæfustu orðanna var umtalsvert hærra í íslensku textunum.

**Tafla 1.** Fjöldi og hlutfall orða sem féll í hvern tíðniflokk í íslensku PISA-textunum 2018.

Heiti PISA-texta 2018 í íslensku þýðingu	Heildar-fjöldi orða í texta*	Orðtíðniflokkar: Fyrstu þúsund algengustu orðin (blár) o.s.frv.									
		Blár (≤1000)		Grænn		Bleikur		Gulur		Rauður (4000<)	
		fj.	%	fj.	%	fj.	%	fj.	%	fj.	%
Kóngulær (R055)	254	180	70,9	11	4,3	6	2,4	4	1,6	53	20,9
Brúðumeistarin (R406)	487	374	76,8	18	3,7	5	1,0	8	1,6	82	16,8
Mjól (S326)	131	93	71,0	3	2,3	10	7,6	4	3,1	21	16,0
Aldaði risaeðlanna (S527)	174	111	63,8	9	5,2	8	4,6	9	5,2	37	21,3

\*Tölustafir töldust ekki með

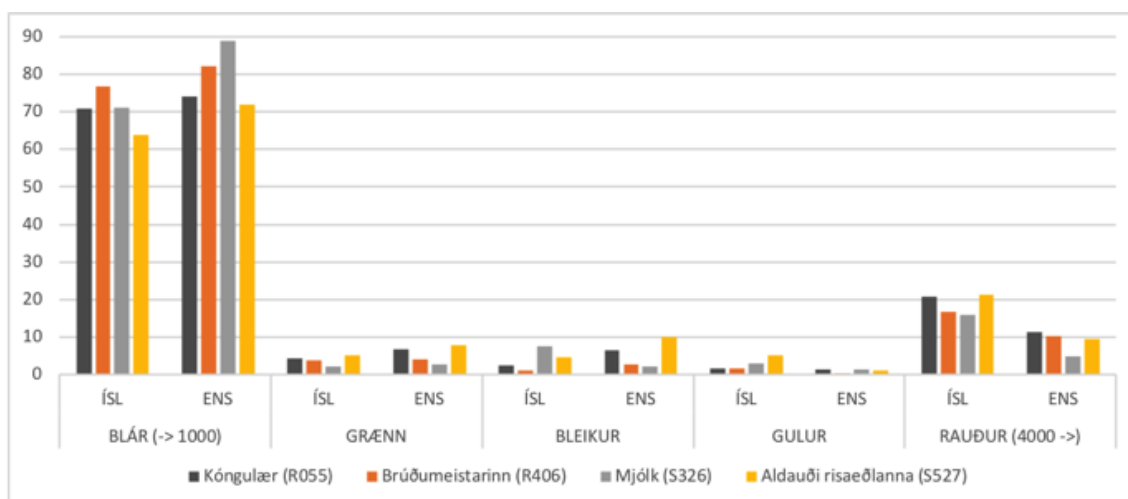
**Tafla 2.** Fjöldi og hlutfall orða sem féll í hvern tíðniflokk í ensku PISA-textunum 2018.

Heiti PISA-texta 2018 í ensku útgáfu	Heildar-fjöldi orða í texta*	Orðtíðniflokkar: Fyrstu þúsund algengustu orðin (blár) o.s.frv.									
		Blár (≤1000)		Grænn		Bleikur		Gulur		Rauður (4000<)	
		fj.	%	fj.	%	fj.	%	fj.	%	fj.	%
Drugged spiders drop stitches (R055)	265	196	74,0	18	6,8	17	6,4	4	1,5	30	11,3
The master of Kokeshi dolls (R406)	546	449	82,2	23	4,2	15	2,7	2	0,4	57	10,4
Milk (S326)	143	127	88,8	4	2,8	3	2,1	2	1,4	7	4,9
Extinction of the dinosaurs (S527)	192	138	71,9	15	7,8	19	9,9	2	1,0	18	9,4

\*Tölustafir töldust ekki með

Eins og sjá má í töflu 1 og 2 var hlutfall algengustu orða (blár flokkur) lægra í þessum fjórum íslensku þýðingum en í ensku textunum. Þá var hlutfall orða í flokki sjaldgæfustu orðanna (rauður flokkur) umtalsvert hærra í íslensku textunum en þeim ensku. Í milliflokkunum þremur (grænn, bleikur og gulur) var hlutfallið hærra í ensku textunum sem sýnir að dreifing orða á milli orðtíðniflokka var jafnari en í íslensku þýðingunni.

Þennan mun á milli tungumálanna má glögglega sjá á mynd 1 sem sýnir hlutfall orða í hverjum tíðniflokki hvers texta, á íslensku og ensku.



**Mynd 1.** Hlutfall orða í orðtíðniflokkunum fimm eftir textunum fjórum á íslensku og ensku.

Greining á þeim orðum sem féllu í ólíkan tíðniflokk á íslensku en ensku er kynnt í töflu 3. Fram komu vísbendingar um ákveðið ójafnvægi sem fólst í að tveir þriðju hlutar þeirra íslensku orða sem féllu í ólíkan tíðniflokk en ensku voru sjaldgæfari en ensku orðin sem þau stóðu fyrir.

**Tafla 3.** Fjöldi og hlutfall orða sem féll í ólíkan tíðniflokk á íslensku en ensku í PISA 2018.

Heiti textanna í PISA-prófinu 2018	Heildarfjöldi orða í íslensku þýðingunni	Fjöldi orða sem féll í ólíkan tíðniflokk á íslensku en ensku (%)	Íslensk orð í sjaldgæfari tíðniflokki en ensk	Íslensk orð í algengari tíðniflokki en ensk
Kóngulær (R055)	254	60 (23,6%)	36	24
Brúðumeistarinn (R406)	487	74 (15,2%)	50	24
Mjólk (S326)	131	67 (51,1%)	43	24
Aldauði risaeðlanna (S527)	174	36 (20,7%)	32	4
SAMTALS	1046	237 (22,7%)	161	76

Í töflu 3 má sjá að hlutfall þeirra íslensku orða sem lentu í öðrum tíðniflokki en samsvarandi ensk orð var 22,7%. Af þessum orðum voru 68% sem lentu í sjaldgæfari tíðniflokki (161 orð) en þau sem lentu í algengari tíðniflokki (76 orð). Þetta er vísbending um ákveðið ósamræmi.

Dá var leitað að íslenskum samheitum sem gætu leyst hin af hólmi og væru í sama eða nærliggjandi tíðniflokki og enska orðið. Í töflu 4 má sjá fjölda mögulegra samheita á íslensku sem hefði mátt nota og falla betur að tíðni samsvarandi enskra orða. Í ljós kom að bæta hefði mátt samræmið á íslensku þýðingunum við ensku textana með því að velja samheiti sem falla betur að tíðniflokki enska orðsins.



**Tafla 4.** Fjöldi samheita á íslensku sem gætu fallið í sama eða nærliggjandi tíðniflokk og sambærileg ensk orð í PISA 2018.

Heiti textanna í PISA- prófinu 2018	Fjöldi orða í ólíkum tíðniflokki á íslensku en ensku	Unnt að breyta í íslensk samheiti í sama tíðniflokki og enska orðið	Unnt að breyta í íslensk samheiti í nærliggjandi tíðniflokki og enska orðið
Kóngulær (R055)	60	11 <i>(9 í algengari 2 í sjaldgæfari)</i>	4 <i>(4 í algengari)</i>
Brúðumeistarinn (R406)	74	11 <i>(10 í algengari 1 í sjaldgæfari)</i>	2 <i>(1 í algengari 1 í sjaldgæfari)</i>
Mjólk (S326)	67	9 <i>(8 í algengari 1 í sjaldgæfari)</i>	6 <i>(4 í algengari 2 í sjaldgæfari)</i>
Aldauði risaeðlanna (S527)	36	2 <i>(2 í sjaldgæfari)</i>	1 <i>(1 í algengari)</i>
<b>SAMTALS:</b>	<b>237</b>	<b>33</b> <b><i>(29 í algengari 6 í sjaldgæfari)</i></b>	<b>13</b> <b><i>(10 í algengari 3 í sjaldgæfari)</i></b>

Eins og sjá má í töflu 4 hefði í mörgum tilvikum verið hægt að finna gjaldgeng orð sem falla betur að tíðniflokki samsvarandi enskra orða. Til dæmis hefði í sögunni um kóngulær vel mátt nota 11 íslensk samheiti í sama tíðniflokki og enska orðið og fækka þannig orðum í ólíkum tíðniflokki úr 60 (23,6%) niður í 49 orð (19,3%). Af þessum 11 íslensku samheitum hefðu níu orð fallið í algengari orðtíðniflokk en tvö í sjaldgæfari. Í viðbót hefði mátt velja fjögur íslensk samheiti sem voru í nærliggjandi tíðniflokki og enska orðið sem notað var í þýðingunni. Það hefði bætt samræmið á íslensku þýðingunum við ensku textana enn frekar.

Eins og áður kemur fram segir í leiðarvísi OECD (2016) að lengri orð hafi tilhneigingu til að vera sjaldgæfari en styttri orð. Því var ákveðið að skoða sérstaklega íslensku orðin sem féllu ekki í sama tíðniflokk og samsvarandi ensk orð og bera saman lengd íslenska orðsins við samheiti þess í sama tíðniflokki og enska orðið, þ.e. þegar samheiti fannst. Niðurstöður þessarar greiningar má sjá í töflu 5 sem sýnir að yfir þriðjungur orða var algengari og lengri en samheiti þeirra.

**Tafla 5.** Fjöldi (og hlutfall) íslenskra samheita sem eru algengari og lengri en orðin sem voru valin í íslensku þýðingu PISA 2018.

Lesskílningshluti			Náttúruvísindahluti
Kóngulær	Brúðumeistarinn	Mjólk	Aldauði risaeðlanna
4 af 11 (36%)	5 af 15 (33,3%)	6 af 15 (40%)	1 af 3 (33,3%)

Eins og fram kemur í töflu 4 voru 11 orð í textanum um kóngulær sem mátti finna samheiti fyrir í sama tíðniflokki og enska orðið. Í töflu 5 má sjá að af þeim voru fjögur orð bæði algengari og lengri en orðin sem höfðu verið valin í þýðingunni. Þar kemur einnig fram að hlutfall algengari og lengri samheita var yfir 30% í textunum fjórum. Þetta þýðir að viðmið um að lengri orð hafi tilhneigingu til að vera sjaldgæfari en styttri orð dugar ekki ein sér við þýðingar úr ensku yfir á íslensku. Þetta má líka sjá í eftirfarandi dæmum.

Dæmi 1: (ísl) ... höggva tré ... (ens) ... cut down the trees ...

Orðið *höggva* fellur í sjaldgæfari tíðniflokk en *cut*. Samheitið *saga* fellur í sama tíðniflokk og *cut* og er algengara og styttra en samheitið. Algengara íslenska orðið er því styttra en það sjaldgæfara, sem er í samræmi við leiðarvísi OECD.

Dæmi 2: (ísl) ... langt *skeið* myrkurs ... (ens) the long *period* of darkness ...

Orðið *skeið* fellur í sjaldgæfari tíðniflokk en *tímabil*. Samheitið *tímabil* fellur í sama tíðniflokk og *period*. Orðið *tímabil* er lengra og algengara en samheitið.

Dæmi 3: (ísl) ... uppgötvuðu *eftirfarandi* staðreyndir ... (ens) ... the *following* facts ...

Orðið *eftirfarandi* fellur í sjaldgæfari tíðniflokk en *following*. Samheitið *þessar* fellur í sama tíðniflokk og *following*, sem er bæði algengara og styttra en *eftirfarandi*.

## Samantekt

Af niðurstöðum þessarar rannsóknar má sjá að hlutfall algengustu orða var lægra í öllum textum íslensku þýðingarinnar sem hér voru skoðaðir en í ensku frumtextunum. Hlutfall orða í flokki sjaldgæfustu orðanna var umtalsvert hærra í íslensku textunum en þeim ensku. Þá virðist dreifing orða á milli orðtíðniflokka hafa verið jafnari í ensku en íslensku þýðingunum.

Hlutfall þýddra íslenskra orða sem féllu í annan tíðniflokk en samsvarandi ensk orð í upprunalegu textunum var á bilinu 15,2–51,1%. Fram komu vísbendingar um ákveðið ójafnvægi sem fólst í að tveir þriðju hlutar þeirra íslensku orða, sem féllu í ólíkan tíðniflokk en ensku orðin, voru sjaldgæfari en samsvarandi ensk orð.

Þá kom í ljós að fækka hefði mátt orðum í ólíkum orðtíðniflokkum þessara fjögurra texta með því að nota íslenskt samheiti í sama orðtíðniflokki og enska orðið í 33 tilvikum. Til viðbótar hefði mátt nota 13 samheiti úr nærliggjandi orðtíðniflokki. Þetta þýðir að ef notuð hefðu verið íslensk samheiti úr sama tíðniflokki og enska orðið hefði hlutfall orða í ólíkum tíðniflokki lækkað úr 22,7% í 19,5%. Þá hefði mátt draga enn frekar úr ósamræminu með því að velja samheiti úr nærliggjandi tíðniflokki.

Hlutfall algengari og lengri samheita var yfir 30% í textunum fjórum. Þetta þýðir að viðmið um að lengri orð hafi tilhneigingu til að vera sjaldgæfari dugar skammt við þýðingar úr ensku yfir á íslensku.

Í hnotskurn benda niðurstöður þessarar rannsóknar til að bæta hefði mátt þýðingu íslensku textanna í PISA-prófinu 2018 með því að taka mið af orðtíðnilistum við val á orðum. Þannig hefði samræmi íslensku þýðingarinnar og enska frumtextans orðið betra og réttmæti mælingarinnar í PISA-rannsókninni verið betur tryggt.

## Umræður

Markmið þessarar rannsóknar var að kanna samræmi í orðtíðni í íslenskum og enskum textum lesskilnings- og náttúruvísindahluta PISA-prófanna 2018.

Niðurstöður sýna að hlutfall algengustu orða í þeim textum sem hér voru rannsakaðir var lægra í íslensku þýðingunum en í ensku frumtextunum. Þá var hlutfall orða í flokki sjaldgæfustu orðanna umtalsvert hærra í íslensku textunum en þeim ensku og dreifing orða á milli orðtíðniflokka var meiri og jafnari í ensku textunum en þeim íslensku. Einnig kom í ljós að tveir þriðju hlutar þeirra íslensku orða sem féllu í ólíkan tíðniflokk en ensku orðin voru sjaldgæfari en þau ensku í þeim textum sem hér voru teknir til greiningar. Þetta leiðir í ljós að fleiri lágtíðniorð voru í íslensku textunum en þeim ensku. Niðurstöður gefa einnig ástæðu til að ætla að draga hefði mátt úr ósamræminu með því að velja íslensk samheiti úr sama eða nærliggjandi tíðniflokki.

Fjölmargar rannsóknir á ensku hafa leitt í ljós að tengsl eru á milli tíðni orða og þess hversu líklegt er að nemendur þekki orðin (Baayen o.fl., 2007; Balota o.fl., 2006; Gardner o.fl., 1987; Meunier og Segui, 1999; Oldfield og Wingfield, 1965). Niðurstöður þessarar rannsóknar vekja upp þá spurningu hvort íslenski textinn hafi verið hlutfallslega of þungur, þ.e. með of mörgum sjaldgæfum orðum, miðað við enska textann, orðum sem eru engu að síður mjög mikilvæg en nemendur þekkja í mismiklum mæli (Biemiller, 2008; Roessingh, 2016). Orðskilningur íslenskra nemenda kann því að hafa orðið til trafala við úrlausn þessara hluta PISA-prófsins 2018 og þannig hafi hallað á íslenska nemendur umfram þá sem lásu textana á ensku. Meira hafi því reynt á orðaforða íslenskra nemenda en hæfni þeirra til að finna, velja og meta upplýsingar í texta, nýta bakgrunnsþekkingu sína og að veга og meta ólík sjónarhorn sem PISA-prófin snúast einmitt um (LaRusso o.fl., 2016; OECD, 2016). Þótt óumflýjanlegt sé að einhver orð á íslensku og ensku falli í ólíka tíðniflokka ættu þýðendur að leitast við að hafa líkt hlutfall orða í orðtíðniflokkum þýðingar og frumtexta, eða gæta jafnvægis á milli fjölda þýddra orða sem eru sjaldgæfari og algengari en orðin í frumtextanum.

Íslenskir nemendur hafa staðið sig einna verst í alþjóðlegum samanburði hvað varðar frammistöðu í lesskilningi og læsi á náttúruvísindi eftir að grunnskóla lýkur samkvæmt niðurstöðum PISA árið 2015, en þróunin hefur verið neikvæð frá árinu 2000 (Auður Pálsdóttir, 2017; Menntamálastofnun, 2017; Sigríður Ólafsdóttir og Baldur Sigurðsson, 2017). Fyrstu niðurstöður PISA 2018 eru áætlaðar í lok árs 2019 og geta niðurstöður þessarar rannsóknar verið innlegg í túlkun þeirra.

Þar sem hlutfall algengari og lengri samheita var meira en þriðjungur orða, sem féllu í ólíkan tíðniflokk en á ensku í textunum fjórum, má ætla að almenn viðmið OECD (2016) um að lengri orð hafi tilhneigingu til að vera sjaldgæfari sé hugsanlega misvísandi fyrir íslenska þýðendur. Þess ber að geta að löng orð í íslensku eru gagnsærri en ensk orð og það skiptir væntanlega máli í þessu sambandi. Langt íslenskt orð getur verið auðskiljanlegt þótt lesandi hafi aldrei séð það áður, en slíkt er ólíklegt með enskt orð.

Aftur á móti benda niðurstöðurnar til að mikilvægt sé að þýðendur noti orðtíðnilista sem byggja á málheildum til að gæta samræmis í orða vali við upphaflegu útgáfuna.

Rannsóknin hefur takmarkanir líkt og aðrar. Hér var unnið með fjóra texta sem eru aðeins hluti þeirra viðfangsefna sem 15 ára nemendur tókust á við árið 2018 og ekki fengu allir þátttakendur sömu textana. Þá þarf líka að hafa í huga að ekki er nógu mikið vitað um sambandið á milli tíðni íslenskra orða í Risamálheildinni og orðaforða Íslendinga á misjöfnum aldri. Spyrja þarf hvort orðaforði aukist samhliða orðtíðni með hækkandi aldri og hvað megi gera ráð fyrir að hátt hlutfall 15 ára unglunga þekki orð í hverjum tíðniflokki. Orðtíðnilistar hafa þó þær takmarkanir að þar er ekki tekið tillit til margræðni orða. Aðeins er orðflokkur gefinn (t.d. nafnorð, sagnorð, forsetning) en ekki er greint á milli mismunandi merkingar sem sama orð í sama orðflokki getur haft. Dæmi um slíkt er orðið skeið sem kynnt er í niðurstöðum. Í kvenkyni merkir orðið mataráhald en í hvorugkyni tímabil (sem á við í samhengi textans). Þá þarf að hafa í huga að orðtíðnilistarnir sem unnið var með byggja á málheildum sem eru ekki að öllu leyti sambærilegar.

Þótt niðurstöður þessarar rannsóknar gefi vísbendingar um að endurskoða þurfi leiðbeiningar fyrir þýðendur alþjóðlegra prófa kalla þær einnig á að unninn verði íslenskur hugbúnaður á borð við *VocabProfile*. Slíkt verkfæri, sem lægi aðgengilegt á vef, fæli ekki aðeins í sér mikinn tímasparnað í vinnu þýðenda við að rýna í fjölbreytni orða út frá orðtíðnilistum. Slíkt verkfæri myndi líka gefa kennurum tækifæri til að fylgjast með framförum nemenda sinna og að finna orð sem gætu reynst nemendum þeirra erfið, en einnig nemendum sjálfum og íslenskum almenningi sem rýna vill í eigin skrif.

Í rannsóknnum er grundvallaratriði að þær mæli það sem þeim er ætlað að mæla, þ.e. að gæta þarf að réttmæti rannsókna svo hægt sé að draga trúverðugar ályktanir af niðurstöðum þeirra. Ef til vill er ástæða til að benda á að ef Íslendingar spara sér að þýða úr frönsku verður að gera einhverjar aðrar ráðstafanir til að tryggja réttmæti þýðingarinnar. Einhver mikilvæg ástæða hlýtur að vera fyrir því að þýða verkefni úr tveimur málum. Í samanburðarrannsóknnum eins og PISA þarf að

tryggja að prófin geri sams konar kröfur til þátttakenda sem taka prófin á ólíkum tungumálum þannig að jafnmikið reyni á ólíka færniþætti sem mældir eru. Ef þáttur orðaforða er þyngri hjá einum hópi þátttakenda en öðrum dregur það úr réttmæti samanburðar á milli ríkja. Því ætti notkun flokkaðra orðtíðnilista sem byggjast á málheildum að vera eftirsóknarverður hluti af vönduðum vinnubrögðum við þýðingar sem lagðar eru til grundvallar svo mikilvægri rannsókn sem PISA-prófin fela í sér.

### **Analysis of word use in the reading literacy and science literacy parts of PISA 2018: Comparison of the Icelandic translation and the original English version**

Icelandic learners' performance in the reading and science literacy parts of PISA has declined from 2000 to 2015, and the drop in mean scores is one of the most dramatic among participating countries. The percentage of Icelandic participants in the highest proficiency levels has fallen, and the percentage in the lowest levels has risen.

PISA tests are written in two parallel source versions, English and French, and then translated into other languages. OECD publishes guidelines (2016) for translators in which it is stated that translators should avoid simplifying or complicating the vocabulary and the syntax. Due to the direct relationship between word understanding and text comprehension (Laufer & Ravenhorst-Kalovski, 2010), it is of high importance that translated words be carefully chosen. If there is a higher number of difficult words in one language than another, the readers may have more difficulty in applying the requested reading strategy. Such bias may affect the validity of the measurement.

Studies have demonstrated a strong relationship between the extent to which words are known by individuals and word frequency (Baayen, Wurm & Aycocock, 2007; Balota, Yap & Cortese, 2006; Gardner, Rothkopf, Lapan & Lafferty, 1987; Meunier & Segui, 1999; Oldfield & Wingfield, 1965).

In the PISA 2018 translation and adaptation guidelines (OECD, 2016, p. 11) it is stated that „longer words tend to be less frequent, more technical and/or more abstract than short words“. Nonetheless there is no requirement that translators refer to word frequency lists as an effort to match the frequency of translated words with the words in the original version.

The purpose of this research was to compare the alignment of word frequency in Icelandic translated texts and the original English versions of PISA 2018.

Two text parts were randomly selected from the reading literacy section and two from the natural science section of PISA 2018. Information about the frequency of Icelandic words was obtained from a frequency list based on the Icelandic Gigaword Corpus (Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson & Jón Guðnason, 2018; Stofnun Árna Magnússonar í íslenskum fræðum, 2017). The software *VocabProfile* (Cobb, n.d.) was used for the English words, based on two corpuses: the New General Service List and the New Academic Word List. The words were grouped into frequency bands, the most common 1000 words in each band for the most frequent 4000 words, and less frequent words together in one band. When translated Icelandic words did not fall into the same frequency band as the corresponding English words, appropriate Icelandic synonyms were looked for, and the length of the synonyms was compared.

Results of the study indicate that the share of words in the highest frequency band was lower in the Icelandic translated texts than in the English versions, and the share

of Icelandic words in the lowest frequency band was higher, in all four analysed texts. Additionally, the English words were more evenly distributed between the five frequency bands than the Icelandic words. Furthermore, among the words that did not belong to the same frequency band in Icelandic as in English, the proportion of Icelandic words of lower frequency was higher. If the Icelandic translators had made use of synonyms in the same or adjacent frequency band to the English corresponding words, a better equilibrium between the languages could have been obtained. More than 30% of more frequent Icelandic words were longer than their less frequent synonyms, which suggests that for Icelandic words it is not a reliable rule that longer words tend to be less frequent than shorter words.

Our findings are an indication that the PISA translation guidelines should include a requirement that translators make use of word frequency lists when choosing words for their translations, so as to make the comparison between countries more valid. If word knowledge is more challenging in one country than another, the impact of the different proficiency factors to be measured is not the same, which may affect the validity of the study. The findings should contribute to the interpretation of PISA results in reading and science literacy for 2018, at least when comparing participants who took the tests in Icelandic and English.

**Key words:** PISA, translations, vocabulary, word frequency, reading literacy, science literacy

### Um höfundana

Auður Pálsdóttir (audurp@hi.is) er lektor við Menntavísindasvið Háskóla Íslands. Meginrannsóknarsvið hennar eru náttúrufræðimenntun, mat á skólstarfi, kennslufræði samfélags- og náttúrugreina og sjálfbærni menntun.

Sigríður Ólafsdóttir (sol@hi.is) er lektor við Menntavísindasvið Háskóla Íslands. Meginrannsóknarsvið hennar eru orðaforði, lesskilningur og ritunarfærni nemenda í móðurmáli og öðru máli.

### About the authors

Auður Pálsdóttir (audurp@hi.is) is an assistant professor at The School of Education. Her main research interests include science education, school self-evaluation, pedagogy of science and social science teaching and sustainability education.

Sigríður Ólafsdóttir (sol@hi.is) is an assistant professor at The School of Education. Her main research interests include vocabulary studies, vocabulary acquisition and development, reading comprehension, and writing skills among first and second language learners.

## Heimildir

- Arffman, I. (2010). Equivalence of translations in international reading literacy studies. *Scandinavian Journal of Educational Research*, 54(1), 37–59. doi:10.1080/00313830903488460
- Arnbak, A. (2010). To what extent do basic skills predict students' PISA reading score? Í N. Egelund (ritstjóri), *Northern Lights on PISA 2009 – focus on reading* (bls. 23–43). Kaupmannahöfn: Nordic Council of Ministers.
- Auður Pálsdóttir. (2017). Læsi á náttúruvísindi. Vísbendingar úr PISA 2015. Í Menntamálastofnun, *Helstu niðurstöður PISA 2015* (bls. 36–42). Sótt af [https://mms.is/sites/mms.is/files/helstu\\_nidurstodur\\_pisa\\_2015\\_prent\\_-\\_loka\\_0.pdf](https://mms.is/sites/mms.is/files/helstu_nidurstodur_pisa_2015_prent_-_loka_0.pdf)

- Ásdís Bergþórsdóttir. (2016). *A brief report on the Icelandic translation of the PISA science test*. Sótt af <https://notendur.hi.is/asb23/report.pdf>
- Baayen, R. H., Wurm, L. H. og Aycocock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, 2(3), 419–463. doi:10.1075/ml.2.3
- Balota, D.A., Yap, M.J. og Cortese, M.J. (2006). Visual word recognition: The journey from features to meaning (a travel update). Í M.J. Traxler og M.A. Gernsbacher (ritstjórar), *Handbook of psycholinguistics* (2. útgáfa, bls. 285–375). doi:10.1016/B978-012369374-7/50010-9
- Biber, D. og Reppen, R. E. (2015). *Handbook of corpus linguistics*. Cambridge: Cambridge University Press.
- Biber, D., Reppen, R. og Friginal, E. (2010). Research in corpus linguistics. Í R. B. Kaplan (ritstjóri), *The Oxford handbook of applied linguistics* (bls. 548–567). Oxford: Oxford University Press.
- Biemiller, A. (2008). *Words worth teaching*. Columbus, OH: SRA/McGraw-Hill.
- Browne, C., Culligan, B. og Phillips, J. (2013). *The new general service list*. Sótt af <http://www.newgeneralservicelist.org/>
- Chen, Y.-F. og Jiao, H. (2014). Exploring the utility of background and cognitive variables in explaining latent differential item functioning: An example of the PISA 2009 reading assessment. *Educational Assessment*, 19(2), 77–96. Sótt af <http://www.tandfonline.com/doi/abs/10.1080/10627197.2014.903650>
- Cobb, T. (e.d.). *Range for texts v.3* [tölvuhugbúnaður]. Sótt af <https://www.lexutor.ca/cgi-bin/range/texts/index.plht>
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, 11(3), 38–64. doi:10.125/44117
- Dale, E. og O'Rourke, J. (1981). *The living word vocabulary: A national vocabulary inventory*. Chicago: World Book – Childcraft International.
- Eiríkur Rögnvaldsson, Kristín M. Jóhannsdóttir, Sigrún Helgadóttir og Steinþór Steingrímsson. (2012). Íslensk tunga á stafrænni öld / Icelandic language in the digital age. Í G. Rehm og H. Uszkoreit (ritstjórar), *Hvítbókaröð / White Paper Series*. Sótt af <http://www.meta-net.eu/whitepapers/e-book/icelandic.pdf>
- Erla Lind Þórisdóttir. (2017). *Orðapröskuldur íslenskra grunnskólanemenda á miðstigi: Hlutfall þekkttra orða í náttúrufræðitexta og lesskilningur* (óútgefin meistararitgerð). Sótt af <http://hdl.handle.net/1946/28485>
- Freyja Birgisdóttir. (2016). Orðaforði og lestrarfærni: Tengsl við gengi nemenda á lesskilningshluta PISA. *Sérrit Netlu 2016 – Um læsi*. Sótt af [http://netla.hi.is/serrit/2016/um\\_laesi/04\\_16\\_laesi.pdf](http://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. Sótt af [https://timarit.is/view\\_page\\_init.jsp?pageId=6760074](https://timarit.is/view_page_init.jsp?pageId=6760074)
- Friðrik Magnússon og Stefán Briem. (1991). *Íslensk orðtíðnibók*. Reykjavík: Orðabók Háskólans.
- Gardner, M. K., Rothkopf, E. Z., Lapan, R. og Lafferty, T. (1987). The word frequency effect in lexical decision: Finding a frequency-based component. *Memory and Cognition*, 15(1), 24–28. doi:10.3758/BF03197709
- Grisay, A., de Jong, J. H. A. L., Gebhardt, E., Berezner, A. og Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, 8(3), 249–266. Sótt af <https://www.ncbi.nlm.nih.gov/pubmed/17804893>
- Halldóra Jónsdóttir og Þórdís Úlfarsdóttir (ritstjórar). (2018). *Íslensk nútímamálsorðabók*. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum. Sótt af <http://islenskordabok.arnastofnun.is/>
- Hiebert, E. H., Scott, J. A., Castaneda, R. og Spichtig, A. (2019). An analysis of features of words that influence vocabulary difficulty. *Education Sciences*, 9(1). doi:10.3390/educsci9010008
- Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J. og Baird, J. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353. doi:10.1080/00313831.2016.1258726
- Hu, M. og Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430. Sótt af <https://www.victoria.ac.nz/lals/about/staff/paul-nation-pubsdate-2000s>
- LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., . . . Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness*, 9(2), 201–222. doi:10.1080/19345747.2015.1116035

- 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.
- Martinez, R. og Schmitt, N. (2015). Vocabulary. Í D. Biber og R. Reppen (ritstjórar), *The Cambridge handbook of English corpus linguistics* (bls. 439–459). Cambridge: Cambridge University Press.
- Menntamálastofnun. (2016). *Um þýðingarferli PISA*. Sótt af <https://mms.is/frettir/um-thydingarferli-pisa>
- Menntamálastofnun. (2017). *Helstu niðurstöður PISA 2015*. Sótt af [https://mms.is/sites/mms.is/files/helstu\\_nidurstodur\\_pisa\\_2015\\_prent\\_-\\_loka\\_0.pdf](https://mms.is/sites/mms.is/files/helstu_nidurstodur_pisa_2015_prent_-_loka_0.pdf)
- Meunier, F. og Segui, J. (1999). Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language*, 41(3), 327–344. doi:10.1006/jmla.1999.2642
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2016). *Making and using word lists for language teaching and learning*. Amsterdam: John Benjamins Publishing Company.
- OECD. (2016). *PISA 2018 translation and adaptation guidelines. Prag: First meeting of the PISA 2018 national project managers*. Sótt af <https://www.oecd.org/pisa/pisaproducts/PISA-2018-TRANSLATION-AND-ADAPTATION-GUIDELINES.pdf>
- OECD. (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving, revised edition*. París: PISA, OECD publishing.
- Oldfield, R. C. og Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4), 273–281. doi:10.1080/17470216508416445
- Oliveri, M. E. og Ercikan, K. (2011). Do different approaches to examining construct comparability in multilanguage assessments lead to similar conclusions? *Applied Measurement in Education*, 24(4), 349–366. doi:10.1080/08957347.2011.607063
- Rannsóknir og greining. (2016). *Ungt fólk 2016: Lýðheilsa ungs fólks í Reykjavík. Niðurstöður rannsókna meðal nemenda í 8., 9. og 10. bekk árið 2016*. Sótt af [http://reykjavik.is/sites/default/files/yms\\_skjol/skjol\\_utgefid\\_efni/reykjavik\\_heild\\_2016\\_.pdf](http://reykjavik.is/sites/default/files/yms_skjol/skjol_utgefid_efni/reykjavik_heild_2016_.pdf)
- Reppen, R. (2010). *Using corpora in the language classroom*. New York: Cambridge University Press.
- 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. Sótt af <https://ojs-o.library.ubc.ca/index.php/BCTJ/article/view/235>
- Romanko, R. (2017). Measuring the vocabulary burden of popular English songs. *Vocabulary Learning and Instruction*, 6(2). doi:10.7820/vli.v06.2.2187-2759
- Sigríður Ólafsdóttir og Baldur Sigurðsson. (2017). Hnignandi frammistaða íslenskra nemenda í lesskilningshluta PISA frá 2000 til 2015: Leiðir til að snúa þróuninni við. *Netla – Vef tímarit um uppeldi og menntun*. Sótt af <http://netla.hi.is/greinar/2017/ryn/16.pdf>
- Sigríður Ólafsdóttir, Freyja Birgisdóttir, Hrafnhildur Ragnarsdóttir og Sigurgrímur Skúlason. (2016). Orðaforði og lesskilningur hjá börnum með íslensku sem annað mál: Áhrif aldurs við komuna til Íslands. *Sérrit Netlu 2016 – Um læsi*. Sótt af [http://netla.hi.is/serrit/2016/um\\_laesi/003.pdf](http://netla.hi.is/serrit/2016/um_laesi/003.pdf)
- Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson og Jón Guðnason. (2018). Risamálheild: A very large Icelandic text corpus. *Ráðstefnurit LREC 2018*, 4361–4366. Sótt af <http://www.lrec-conf.org/proceedings/lrec2018/pdf/746.pdf>
- Stofnun Árna Magnússonar í íslenskum fræðum. (2017). *Risamálheildin*. Sótt af <http://malfong.is/?pg=rnh>
- Stofnun Árna Magnússonar í íslenskum fræðum. (e.d.). *Mörkuð íslensk málheild*. Sótt af <https://arnastofnun.is/is/textasofn>



Auður Pálsdóttir og Sigríður Ólafsdóttir (2019).

Greining á orðanotkun í lesskilnings- og náttúruvísindahlutum PISA 2018: Samanburður á íslensku þýðingunni og enska textanum.

Netla – Vef tímarit um uppeldi og menntun. Menntavísindasvið Háskóla Íslands.

Sótt af [http://netla.hi.is/serrit/2019/altjodlegar\\_menntakannanir/02.pdf](http://netla.hi.is/serrit/2019/altjodlegar_menntakannanir/02.pdf)

DOI: <https://doi.org/10.24270/serritnetla.2019.31>