

Rannsóknaraðferðir

Við gagnaöflun í rannsóknum standa margir möguleikar til boða og markmiðið með þessu skjali er að hjálpa nemendum að velja rétta aðferð. Þessu skjali er alls ekki ætlað að vera tæmandi, heldur aðeins að hjálpa rannsakanda að stíga fyrstu skrefin. Hér verður því aðeins tæpt á nokkrum möguleikum. Einnig skal tekið fram að margir hafa reynslu af þeim aðferðum og tólum sem rædd verða hér á eftir og því er um að gera að leita ráða hjá þeim.

Allar aðferðir við gagnasöfnun hafa sína kosti og galla. Hér verður fjallað um tvær leiðir, leitir í málheildum og ýmiss konar máltilraunir, og verður að meta það eftir rannsóknarefninu hvaða leið er farin. Þess vegna er líka mikilvægt að gera sér grein fyrir þessum kostum og göllum þegar gögnin eru greind.

Kosturinn við málheildir er sá að þær skila náttúrulegum gögnum, þ.e. setningum sem einhver hefur sagt/skrifað. Núorðið standa margir möguleikar til boða og ber þá einkum að nefna *Risamálheildina* (Steinþór Steingrímsson o.fl. 2018) og *Sögulega íslenska trjábankann* (IcePaCH, Wallenberg o.fl. 2011). Risamálheildin er mun stærri en IcePaCH býður upp á miklu hnitmiðaðri leitir eins og nánar verður fjallað um hér að neðan. Málheildirnar bjóða hins vegar upp á það að setningarnar séu skoðaðar í samhengi og því er hægt að taka tillit til þess við greiningu gagnanna.

Gallinn við málheildir er hins vegar sá að þær eru aldrei tæmandi. Ákveðnar formgerðir geta tengst tilteknu málsniði sem ekki finnst í textum málheildarinnar eða að þær eru svo sjaldgæfar að tilviljun ræður því að þær hafa ekki ratað inn í þessa texta. Af þessu leiðir að þótt einhver setningagerð finnst ekki í málheildinni er ekki hægt að álykta sem svo að hún sé útilokuð í málinu og sama má segja um samhengið sem setningagerðirnar koma fram í. Einnig er erfitt að taka tillit til hljóðkerfisfræðilegra þátta í málheildum sem byggjast algjörlega á ritmáli eða upptökum sem hafa verið skráðar en ekki er hægt að hlusta á.

Til þess að fylla upp í göt í málheildargögnunum má beita ýmiss konar tilraunum. Þannig er hægt að kanna og bera saman setningagerðir án þess að þurfa að treysta á að setningagerðin hafi verið fest á blað í öllu mögulegu samhengi sem þarf til þess að átta sig á eðli hennar. Einnig gerir það samanburðinn nákvæmari þar sem hægt er að nota lágmarkspör og kanna hvort tiltekin setningagerð sé tæk yfirleitt og þá í hvaða samhengi.

Gallinn við máltilraunir er sá að upplýsingarnar sem fást takmarkast af hönnun tilraunarinnar og því er engin tilraun tæmandi. Einnig geta tilraunir verið tímafreakar þar sem vanda þarf til verka svo óskyld atriði hafi ekki áhrif á niðurstöðurnar. Tryggja verður að þátttakandi í tilraun túlki setningarnar út frá því samhengi sem skiptir máli og önnur atriði hafi ekki áhrif á þátttakandann, þ.e. einstök orð, merking setningarinnar, orðalag og fleira. Auk þess þarf að tryggja að nægur fjöldi þátttakanda fái st og gæta að ýmsu í uppsetningu og framkvæmd tilraunarinnar. Það getur alltaf komið fyrir að einhver atriði spilli fyrir með einu eða öðru móti, þar á meðal atriði sem rannsakandi hafði ekki leitt hugann að.

Eins og sjá má hafa báðar leiðir ákveðna kosti og galla þannig að spurningin er ekki hvor leiðin sé betri en hin, heldur hvor þeirra henti verkefninu sem verið er að vinna og eðli rannsóknarspurninganna. Málheildarannsóknir henta sumum rannsóknum en öðrum ekki. Það sama er að segja um tilraunir. Í mörgum tilvikum þarf þó báðar aðferðir til þess að ná fram skýrri heildarmynd af fyrirbærinu sem verið er að kanna.

En óháð öllum rannsóknaraðferðum er alltaf mikilvægt að afmarka viðfangsefnið vel og setja fram skýra rannsóknarspurningu, t.d. *Getur X farið á undan Y í setningagerð A ef X táknar nýjar upplýsingar?* Ef það er ekki gert getur rannsakandi hæglega misst tókin á viðfangsefninu. Engin rannsókn getur svarað öllum spurningum sem eru mikilvægar en rannsóknin verður að svara því sem spurt er um. Og vert er að nefna að neikvæðar niðurstöður eru líka niðurstöður. Þótt niðurstöðurnar styðji ekki tilgátuna sem sett var fram geta þær samt verið gagnlegar. Þær geta þá ýmist bent til þess að tilgátan sé röng eða að aðferðarfræðin hafi ekki verið heppileg.

1. Málheildir

Í málheildarrannsóknum á íslensku koma einkum tvær málheildir til greina, *Risamálheildin* og *IcePaCH* og þær hafa báðar sína kosti og galla.

Risamálheildin er mjög stór og er því auðvelt að afla mikilla gagna úr henni. Undir flípanum *útvíkkuð leit* á vefsíðunni er hægt að leita eftir orðflokkum, orðmyndum, uppflöttiorðum, beyingarmyndum og þess háttar. Þar að auki er hægt að skilgreina orðin í kring með sama hætti og einnig hversu mörg orð mega vera á milli. Gallinn er hins vegar sá að eingöngu er hægt að leita eftir eiginleikum orðanna og innbyrðis röð þeirra. Því geta niðurstöðurnar innihaldið orðastrengi sem passa við leitaraskilyrðin en eru ekki dæmi um þá setningagerð sem verið er að kanna. Því getur umtalsverður tími farið í það að hreinsa burt niðurstöður sem skipta ekki máli.

Trjábankar eru gríðarlega gagnlegir í setningafræðilegum rannsóknum þar sem þeir gera okkur kleift að gera mun hnitmiðaðri leitir en hefðbundnar málheildir á borð við *Risamálheildina*, t.d. er hægt að leita eftir ákveðnum setningafræðilegum formgerðum. Þegar þetta er skrifað er sögulegi trjábankinn *IcePaCH* helsti kosturinn hvað þetta varðar.¹ Þá er vert að nefna að *IcePaCH* inniheldur dæmi úr nútímaíslensku og gagnast því einnig við rannsóknir á íslensku nútímamáli.

Í *IcePaCH* má leita í PaCQL-leitarvélinni (Anton K. Ingason 2016). Hana má finna á *Treebank Studio* (treebankstudio.org) og þar eru skýrar leiðbeiningar um PaCQL-málið. Til að nota *IcePaCH* er einnig nauðsynlegt að kynna sér vel þau merki sem notuð eru og þau má finna á [linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus_\(IcePaHC\)](http://linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus_(IcePaHC)).

Leitarniðurstöður má skoða með tvennum hætti, ýmist á vefformi þar sem sjá má tíðni formgerðarinnar sem leitað er að og allar setningar sem innihalda formgerðina og upplýsingar um textann. Þá birtist bæði setningin sjálf og formgerðargreiningin. Til þess að vinna áfram með leitarniðurstöður má stilla leitarvélina þannig að hún skili niðurstöðum á TSV-formi og er þá hægt að hala þeim niður, vinna með þær áfram og greina í t.d. Excel, R og SPSS, þ.e. til þess að hreinsa leitarniðurstöður ef þörf krefur og greina þær tölfræðilega.

¹ Hér má einnig benda á *Samtímalega íslenska trjábankann* (IceConTree, Þórunn Arnardóttir o.fl. 2020). *IceConTree* inniheldur texta úr *Risamálheildinni* og hentar vel í rannsóknum á nútímamáli. Því er hægt að fá talsvert meira af gögnum en hægt er að fá úr *IcePaCH*. Textinn hefur verið þáttaður að fyrirmynd *IcePaCH* og því er nauðsynlegt að kynna sér þær merkingar sem notaðar voru þar. Þegar þetta er ritað virðist því miður ekki mögulegt að leita í *IceConTree* með PaCQL og yrði því að nota önnur verkfæri, t.d. *CorpusSearch 2* (Randall 2005).

2. Tilraunir

Við hönnun málfræðilegra tilrauna verður alltaf að spyrja hvað muni henta fyrirbærinu sem kannað er og þeim spurningum sem tilraunin á að svara. Hér verður fjallað um nokkrar rannsóknaraðferðir og möguleg verkfæri sem nýta má við framkvæmd þeirra og úrvinnslu. Það er gott viðmið þegar tilraun er hönnuð að nota tvær prófsetningar fyrir hvert atriði, þrjár uppfyllingarsetningar fyrir hverja prófsetningu og að þátttakendur séu a.m.k. 25 - 50.

Dómapróf (e. *grammaticality/acceptability judgement task*) eru ein algengasta tegund tilrauna í málvísindum, enda henta þau mjög mörgum fyrirbærum, eru ódýr og einföld í framkvæmd og svara flestum spurningum sem setningafræðingar hafa áhuga á. Upplýsingarnar sem þau skila af sér eru einfaldlega mat þátttakanda á setningum í gefnu samhengi. Þetta getur því sagt okkur hvort setningin er tæk eða ekki. Dómapróf eru hentug þar sem þau segja okkur ekki bara hvað er mögulegt heldur líka hvað er útilokað. Hins vegar er ekki víst að sjáanlegur munur milli tveggja valmöguleika komi fram þannig að dómapróf segja okkur ekki endilega hvaða setningagerð er sjálfgefin eða ómörkuð samanborið við aðrar setningagerðir.

Í dómaprófum er góð regla að hverri setningu fylgi inngangssetning sem á að tryggja að þátttakandi túlki og meti prófsetninguna í réttu samhengi. Þátttakandi metur svo setninguna á þeim kvarða sem rannsakandi hefur valið, hvort sem það er *-?–√, *já/nei*, eða Likert-kvarði (t.d. 1–5, 1–7 eða -3–3).

Með *flýtidómaprófi* (e. *speeded acceptability task*) má fá frekari upplýsingar með því að blanda saman dómaprófi og sjálfstýrðum lestri (e. *self-paced reading*). Ólíkt hefðbundnum dómaprófum sér þátttakandi ekki setninguna í heild heldur aðeins eitt orð (eða meira) í einu og sér þá ekki setninguna þegar hún er metin. Með þessu móti fást upplýsingar um úrvinnslu setningarinnar og þannig er m.a. hægt að gera greinamun á setningum sem hljóta sambærilega dóma ef einhverjar tafir verða í úrvinnslu einnar setningar sem ekki eiga sér stað í hinni. Þetta getur bent til þess að einhver ferli eigi sér stað í annarri setningunni en ekki í hinni og valdi því þessum tögum. Þetta getur einnig gefið vísbendingar um það hvort ein setningagerð er leidd af annarri eða ekki.

Bæði hefðbundin dómapróf og flýtidómapróf má gera á netinu og eru ýmis verkfæri í boði til þess. Fyrir hefðbundin dómapróf má nota Google Forms en niðurstöðurnar má vinna áfram í Google Sheets eða Excel og er svo hægt að greina þær í SPSS eða R.

Við flýtidómapróf þarf önnur tól. Þar má nefna Ibex Farm, sem er sérstaklega hannað til að framkvæma tilraunir í gegnum netið. Ítarlegar upplýsingar er að finna á netinu (<https://spellout.net/ibexfarm/>) um það hvernig tilraunin er sett upp en hægt er að hala niður sniðmáti sem má breyta og laga að tilrauninni. Þótt uppsetningin krefjist forritunar er reynsla af forritun ekki nauðsynleg þar sem leiðbeiningar og sniðmátið leiða notandann í gegnum ferlið.

Annar möguleiki er OpenSesame. Lundquist o.fl. (2019) notuðu þetta forrit og hafa gert aðgengilegan kóðann sem þau notuðu við uppsetningu tilraunarinnar og úrvinnslu gagnanna og er því hægt að hafa hann sem fyrirmynd. OpenSesame býður einnig upp á tilraunir í gegnum netið og það getur auðveldað rannsakendum að finna þátttakendur.

Myndunarpróf (e. *production tasks*) snúast um það að fá þátttakanda til að mynda setningu eða setningarhluta frekar en að dæma setningu. Ólíkt dómaprófum fáum við ekki upplýsingar um

Það hvað er tækt eða ótækt heldur sjáum við hvað er sjálfgefið miðað við samhengi. Til eru ýmsar leiðir til að framkvæma slík próf og verða nokkrar þeirra ræddar hér.

Einföldustu myndunarprófin eru eyðufyllingarpróf (e. *cloze tasks*) en þar fær þátttakandi setningu með eyðu sem þarf að fylla inn í. Einnig er hægt að láta samhengi fylgja með, hvort sem það er skrifað, lesið upp af stjórnanda tilraunarinnar og/eða í formi myndar. Próf af þessu tagi eru mjög hentug þegar fallmörkun er rannsökuð, sbr. rannsókn Ástu Svavarsdóttur (1982) á þágufallshneigð. Það er samt sem áður vert að nefna að niðurstöðurnar sýna e.t.v. fyrst og fremst hvaða form eru sjálfgefin.

Framköllunarpróf (e. *elicitation task*) miða að því að fá þátttakanda til að mynda ákveðna setningagerð. Þá mætti nota myndir og samhengi til þess að fá þátttakanda til þess að búa til setningu sem tilheyrir ákveðinni tegund formgerða. Þetta gæti verið saga, spurningar og svör eða villa sem stjórnandi gerir í lýsingu myndarinnar og biður þátttakanda um að leiðrétta. Próf af þessu tagi eru mjög algeng í máltökurannsóknum; sjá t.d. Sigríði Sigurjónsdóttur (1992).

Í umorðunarprófi fær þátttakandi að sjá setningu og svo upphaf næstu setningar sem hann á að klára. Hér mætti líta til Lundquist o.fl. (2019) en þau gerðu einmitt þess háttar tilraun. Því má nota þá rannsókn sem fyrirmynd þegar kemur að uppsetningu og framkvæmd tilraunarinnar og úrvinnslu gagnanna. Þessi grein getur því hjálpað rannsakandanum í gegnum allt ferlið.

Út frá þeirri fyrirmynd myndi þátttakandi fyrst sjá setningu í ákveðinn tíma áður en hún hverfur og þátttakandi sér upphaf næstu setningar þar sem einhver liður hefur verið kjarnafærður. Þátttakandi er þá beðinn um að klára setninguna og nota sömu orð og í fyrri setningunni.

Fyrir tilraunir af þessu tagi mundi OpenSesame gagnast vel (sjá Lundquist o.fl. 2019). Þar sem umorðun og framköllunarprófin geta falið í sér munnleg svör þyrfti einnig forrit til að vinna úr upptökunum, t.d. ELAN eða Praat. ELAN gæti verið einfaldari kostur þar sem hægt væri að fylgja nákvæmlega aðferðum Lundquist o.fl. en það væri líka hægt að laga þær að Praat. ELAN býður einnig upp á að vinna með myndbandsupptökur. Svo mætti nota R (í gegnum RStudio) eða SPSS til að greina niðurstöðurnar tölfræðilega.

Vert er að nefna að í hljóðfræðilegum athugunum verður að huga að hljóðgæðum. Þar skiptir mestu máli að hljóðneminn sé svo góður að hann nemi tíðni allt að 40 kHz. Til þess að greina upptökurnar þyrfti að nota Praat þar sem ELAN hefur ekki öll þau greiningartól sem standa til boða í Praat. Praat og ELAN eru fyrst og fremst hugsuð til þess að vinna úr upptökum frekar en að búa þær til og því þyrfti annað forrit til þess að gera upptökurnar. Úr nógu er að velja og hægt er að finna ókeypis forrit sem duga vel bæði fyrir tölvu og snjallsíma. Upptakan yrði svo færð inn í Praat og unnið með hana þar. Ef tilraunin er gerð í gegnum OpenSesame er hægt að gera upptökuna í gegnum það (sbr. Lundquist o.fl. 2019).

Aðferðir sniðnar að einstökum verkefnum

Hér á eftir verður fjallað um nokkrar hugmyndir að ákveðnum verkefnum til þess að gefa skýrari dæmi um það hvernig megi velja aðferðafræði. Ræddar verða aðferðir í tengslum við nokkur ólík rannsóknarverkefni.

Tilbrigði í fallmörkun tveggja andlaga sagna

Söguleg þróun: Hér þyrfti að leita í IcePaCH. Dæmi um einfalda leit væri eftirfarandi:

(1) IP.* doms NP-OB2
NP-OB2 idoms .*-D
NP-OB1 idoms .*-A

*Setning inniheldur óbeint andlag
Óbeint andlag er í þágufalli
Beint andlag er í þolfalli*

meta:
text textid
text year
text century

*Upplýsingar um dæmin
Úr hvaða texta dæmið kemur
Hvaða ár hann kom út
Á hvaða öld hann kom út*

Þessi leit skilar 4588 dæmum. Þessa leit mætti þrengja, t.d. með því að horfa til ákveðinna sagna eða tímabila eða einskorða leitina við fornöfn eða fulla nafnliði.

Til þess að skilja á milli fornafna og fullra nafnliða mætti t.d. bæta við eftirfarandi skipunum við leitina hér að ofan.

(2) NP:Full
NP-OB[12] idoms N-.*

NP:PRO
NP-OB[12] idoms PRO-.*

Til þess að afmarka ákveðin tímabil er hægt að vinna með gögnin, t.d. í Excel, með því að flokka dæmin eftir árum eða öldum og jafnvel afrita þau í annað skjal sem unnið er með áfram. Til þess að gera það er leitarvélin stillt þannig að frálagið sé á TSV-formi og niðurstöðunum er svo halað niður.

Samtímaleg tilbrigði: Mögulegt væri að prófa þetta með einhvers konar myndunarprófi. Fyrsta skref væri að athuga hvaða sagnir er hægt að prófa, þ.e. hvaða tveggja andlaga sagnir sýna tilbrigði í fallmörkun. Þegar sagnirnar hafa verið valdar er hægt að hanna tilraunina.

Einn möguleiki væri að kanna þetta með eyðufyllingum. Þátttakandi myndi þá sjá setningu þar sem eyða stendur í stað andlaganna tveggja. Þátttakanda eru þá sýnd andlögin í nefnifalli og hann fyllir í eyðurnar og notar það form sem samræmist best hans máltilfinningu. Einnig væri hægt að sýna þátttakanda myndir og biðja hann beðinn um að lýsa því sem er að gerast á myndinni í einni setningu. Þriðji möguleikinn er að láta þátttakanda umorða setningu (sjá umræðu hér að ofan).

Óbein njótandaandlög (benefaktíf andlög) í tímans rás

Hér liggur beint við að leita dæma í IcePaCH. Sem dæmi um einfalda leit mætti nefna eftirfarandi:

(1) IP.* idoms NP-OB2
NP-OB2 hassister NP-OB1

*Setning inniheldur óbeint andlag
Beint andlag er til staðar*

Þessi leit skilar 4680 dæmum og inniheldur einnig hefðbundnar tveggja andlaga sagnir. Því þarf að leita leiða til að hreinsa til í gögnunum. Engin bein leið virðist vera til að gera það í PaCQL en þó eru ýmsar krókaleiðir mögulegar. Einn möguleiki er að taka hefðbundnartveggja andlaga sagnir út fyrir sviga, mögulega með eftirfarandi viðbótum:

(2) Ben:No

VB.* idomslemma gefa|kynna|veita

Sagnir til útilokunar

Ben:yes

VB.* idomslemma .*

Aðrar sagnir

Með þessu móti er hægt að fara yfir niðurstöður leitarinnar og bæta svo sögnum við í útilokunarhópin og er leitin svo endurtekin. Að því loknu má bæta við eftirfarandi línunum:

(3) meta:

text year

text century

Upplýsingar um dæmin

Hvaða ár hann kom út

Á hvaða öld hann kom út

Leitarvélina er síðan stillt á TSV-frálag og niðurstöðunum svo halað niður og er hægt að vinna þær áfram, t.d. í Excel, þar sem hægt er að fjarlægja hefðbundnu tveggja andlaga sagnirnar.

Notkun forsetningaliða í stað óbeins andlags

Málheildarrannsókn: Hér væri hægt að athuga sögulega þróun með leit í IcePaCH. Eflaust væri einfaldast að leita út frá ákveðnum sögnum. Sem dæmi mætti nefna einfalda leit þar sem leitað er að dæmum með sögninni *gefa*:

(1) IP.* idoms VB.*

VB.* idomslemma gefa

Setning inniheldur sögn

Sögnin er gefa

IO:NP

VB.* hassister NP-OB2

Óbeint andlag er til staðar

IO:PP

VB.* hassister PP

PP idoms NP.*

Forsetningaliður er til staðar

Forsetningaliður inniheldur nafnlið

meta:

text year

text century

Ártal

Öld

Þar sem IcePaCH gerir ekki greinamum á forsetningaliðum af því tagi sem leitað er að og forsetningaliðum almennt, verður að fara í gegnum niðurstöðurnar og hreinsa. Þetta er hægt að gera með því að stilla leitarvélina þannig að hún skili niðurstöðum á TSV-formi og er þá hægt að hreinsa niðurstöðurnar, t.d. í Excel.

Tilraunir: Hvað tilraunir varðar eru ýmsar leiðir færar. Í fyrsta lagi er mögulegt að útbúa dómapróf. Það gæti verið með hefðbundnum hætti eða flýtidómapróf. Þannig myndi tilraunin

einnig skila gögnum sem tengjast setningarúrvinnslu til viðbótar við dóma en það gæti hugsanlega sagt okkur eitthvað um samband þessara formgerða.

Annar möguleiki væri að kanna þetta með myndunarprófi. Annars vegar væri hægt að sýna þátttakanda mynd og biðja hann um að lýsa því sem gerist á myndinni. Stjórnandi tilraunarinnar gæti reynt að búa til samhengi sem gæti auðveldað þátttakanda að fá forsetninguna upp í hugann. Hins vegar væri mögulegt að nota umorðunarpróf þar sem þátttakandi heyrir eða les stutta sögu sem hann þarf svo að endursegja í einni setningu. Sagan yrði þá samin þannig að hún myndi kalla fram tiltekna formgerð. Hönnun og framkvæmd myndunarprófa gæti þó verið mjög flókin í þessu tilviki þar sem það er ekki öruggt að þátttakandi myndi velja setningu með forsetningarlið í stað óbeins andlags ef báðir möguleikar koma til greina. Því þyrfti að huga sérstaklega að merkingarlegum mun á milli formgerðanna tveggja (t.d. hvort mark/viðtakandi sé lifandi eða ekki) til þess að tryggja að forsendur formgerðanna séu fyrir hendi.

Umröðun andlaga

Málheildarrannsókn: Við þetta mætti bæta sögulegum upplýsingum út frá leit í IcePaCH. Viðeigandi skipanir til að leita að umröðun eru í leiðbeiningum fyrir Treebank Studio. Þar má einnig aðlaga skipanirnar svo að leitað sé að umröðun fyrir allar sagnir og unnt sé að greina á milli umröðunar og færslu þungs nafnliðar. Niðurstöðurnar má kanna bæði út frá því hver sögnin er og þeirri tegund upplýsinga sem andlögin fela í sér.

Tilraunir: Í fyrsta lagi væri mögulegt að útbúa hefðbundið dómapróf eða flýtidómapróf. Í seinna tilvikinu myndi tilraunin einnig gefa vísbendingar um úrvinnslu setningarinnar þar sem tafir benda til erfiðleika í úrvinnslu. Til dæmis, ef umröðun er möguleg er munur á viðbragðstímum milli umröðunar og hefðbundinnar orðaraðar? Hvaða áhrif sjást í þolmynd? Er munur á viðbragðstímum meiri ef beint andlag er gert að frumlagi ef sögn leyfir umröðun eða ekki?

Annar möguleiki væri að nota myndunarpróf þó slík próf geti reynst erfið vegna þess að þátttakandi á sennilega eftir að velja hefðbundna röð frekar en umröðun. Því er nauðsynlegt að rannsakandi kynni sér mjög vel það sem vitað er um skilyrði fyrir umröðun. Í myndunarprófi væri hægt að sýna þátttakanda mynd og biðja hann um að lýsa því sem gerist á myndinni en stjórnandi tilraunarinnar gæti reynt að setja fram samhengi sem myndi auðvelda þátttakanda að nota umröðun. Einnig væri mögulegt að nota umorðunarpróf þar sem stjórnandi tilraunarinnar segir þátttakanda stutta sögu sem þátttakandi þarf að endursegja í einni setningu. Sagan yrði þá samin þannig að hægt væri að endursegja hana með umröðun.

Samspil setningafræði og hljóðkerfisfræði

Fyrir þetta verkefni lægi beinast við að gera tilraun. Þar mætti láta þátttakendur lesa upp prófsetninguna og upptökurnar yrðu svo greindar í Praat eða ELAN (sjá t.d. fyrirmyndir hjá Dehé 2006; 2008). Þar sem gæði upptökunnar skipta máli er illmögulegt að gera slíka tilraun í gegnum netið en hana mætti framkvæma á eftirfarandi hátt:

Setningarnar (bæði próf- og samhengissetningarnar) má setja upp í Powerpoint, Keynote eða Google Slides og þátttakanda eru þá sýndar setningarnar á fartölvu eða spjaldtölvu. Þátttakandi les upp prófsetninguna og upplesturinn er tekinn upp með því upptökuforriti sem rannsakandi

kýs. Einnig væri mögulegt að nota OpenSesame við framkvæmd rannsóknarinnar og þá væri hægt að taka upp svör þátttakanda með OpenSesame. Að tilrauninni lokinni er hægt að vinna upptökurnar og greina þær með Praat (sjá leiðbeiningar fyrir byrjendur á síðunni <https://www.gouskova.com/2016/09/03/praat-tutorial/>).

Hér væri einnig mögulegt að gera málheildarrannsókn þótt slík rannsókn gæti verið flókin í framkvæmd. Ein leið væri að leita í þeim hlutum Risamálheildarinnar sem til eru hljóðskrár fyrir, t.d. Alþingisræður eða ruv.is. Hér má líka skoða málheildir sem eru taldar upp á síðunni malfong.is. Með þessu móti væri þá einnig hægt að fá náttúruleg gögn. Rétt eins og með tilraunagögnin yrði unnið með þessar upptökur og þær greindar með Praat.

Heimildir

- Anton Karl Ingason. 2016. „PaCQL: A new type of treebank search for the digital humanities“. *Italian Journal of Computational Linguistics* 2 (2): 51–66.
- Ásta Svavarsdóttir. 1982. „Þágufallssýki“. *Íslenskt mál og almenn málfræði* 4 (1): 19–62.
- Dehé, Nicole. 2006. „Some notes on the focus-prosody relation and phrasing in Icelandic“. Í *Nordic Prosody IX*, ritstj. Gösta Bruce og Merle Horne, 47–56. Frankfurt am Main: Peter Lang.
- . 2008. „To delete or not to delete: The contexts of Icelandic Final Vowel Deletion“. *Lingua* 118 (5): 732–53.
- Randall, Beth. 2005. *CorpusSearch 2*. <https://sourceforge.net/projects/corpussearch/>.
- Sigríður Sigurjónsdóttir. 1992. „Binding in Icelandic: Evidence from language acquisition“. Los Angeles: University of California, Los Angeles / University of California, Los Angeles.
- Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson, og Jón Guðnason. 2018. „Risamálheild: A very large Icelandic text corpus“. Í *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, ritstj. Nicoletta Calzolari (Conference chair), Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, o.fl. Miyazaki, Japan: European Language Resources Association (ELRA). malheildir.arnarstofnun.is.
- Wallenberg, Joel, Anton Karl Ingason, Einar Freyr Sigurðsson, Kristján Rúnarsson, og Eiríkur Rögnvaldsson. 2011. „Icelandic Parced Historical Corpus (IcePaHC)“. http://www.linguist.is/icelandic_treebank.
- Þórunn Arnardóttir, Anton Karl Ingason, Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson, og Jón Guðnason. 2020. „The Icelandic contemporary treebank (IceConTree)“. <http://hdl.handle.net/20.500.12537/21>.