

Lítill lesblinduvillumálheild fyrir íslensku

Svanhvít Lilja Ingólfssdóttir

Háskóli Íslands

svanhviti16@ru.is

Útdráttur

Lesblinda er helsta orsök stafsetningarörðugleika hjá ungum sem öldnum og oft erfitt að eiga við hana. Lesblindir rita oft texta sína með mörgum villum, sem getur reynst hamlandi í samfélagi þar sem stór hluti samskipta fer fram á ritmáli. Ýmiss konar stoðtöl geta gagnast lesblindum við ritun, sér í lagi verkfæri sniðin að þörfum lesblindra. Til þess að greina þarfir á slíkum stoðtölum fyrir íslensku hefur hér verið sett saman lítill villumálheild með upprunalegum og leiðréttum textum þriggja íslenskra málhafa. Villurnar hafa verið flokkaðar eftir villugerð og greining unnið á þeim, til að reyna að skilja betur hvað einkennir villur íslenskra málhafa með lesblindu.

1 Inngangur

Lesblinda (dyslexía) er vandinn við úrvinnslu málhljóða og lýsir sér í erfiðleikum við lestur og skrift. Fólk með lesblindu á sérstaklega erfitt með umskráningu (e. *decoding*), sem er getan til þess að tengja saman stafi og hljóð. Helstu einkennum lesblindu eru hægur og ónákvæmur lestur og margar stafsetningarvillur við ritun texta. Til eru ýmis stoðtöl fyrir fólk sem þarf aðstoð við að skrifa réttan texta, sérstaklega fyrir ensku. Stafsetningarleiðrétting fyrir ensku og fleiri mál er innbyggð í flesta vafra, en hún nær yfirleitt ekki að leiðrétta samhengisháðar villur, þ.e. þegar raunverulegt orð er notað í röngu samhengi. Í sínum er boðið upp á ákveðna leiðréttingu fyrir íslensku, en hún takmarkast að mestu leyti við einstök orð.

Vel útfært málarsleiðréttingarforrit fyrir íslensku kæmi lesblindum að miklu gagni, því þeir gætu þá verið öruggari í skriflegum samskiptum sínum á ýmsum vettvangi. Dæmi um slíkt forrit fyrir ensku er málarsráðgjafarforrit-

ið Grammarly¹ sem býður upp á góða tengingu við ýmsa vafra og þar með málarsleiðréttingar á fjölda samfélagsmiðla og vefsvæða.

Til þess að próa megi slíkt stoðtöl fyrir íslensku verður að vinna ýmsa undirbúningsvinnu, til dæmis rannsaka hvernig forrit kæmi íslenskum notendum með lesblindu að mestu gagni, tryggja aðgengi að tólinu á sem fjölbreyttustum vettvangi og skoða hvaða aðferðir hafa reynst vel fyrir önnur mál. Einn liður í því að leysa þetta vandamál felst enn fremur í því að greina sérstaklega þær villur sem koma fram hjá lesblindum einstaklingum sem hafa íslensku að móðurmáli. Hér hefur verið farin sú leið að taka saman litla lesblinduvillumálheild með því að safna og greina villur þriggja íslenskra málhafa með lesblindu. Villurnar sem þar koma fram hafa verið greindar í villuflokka, og sú greining gefur góða hugmynd um það hvaða villur eru algengastar í máli þessara einstaklinga. Þessi litla villumálheild getur vonandi einnig verið vísbending um hvers konar villur er mikilvægast að leiðrétta í íslensku, sé hugað sérstaklega að lesblindum notendum. Þetta er efnið sem verður rakið í þessari grein. Finna má umfjöllun um fyrri rannsóknir í kafla 2, lýsingu á gögnunum og vinnslu þeirra í kafla 3, samantekt á helstu niðurstöðum í kafla 4 og að lokum umræðu um niðurstöðurnar í kafla 5.

2 Fræðilegur bakgrunnur

Rannsóknir á lesblindu hér á landi hafa að miklu leyti snúist um lestur barna, en meðal helstu heimilda um stafsetningarörðugleika lesblindra er Lesvefur Háskóla Íslands (2019), þar sem finna má fjölbreyttar upplýsingar um læsi og skrift. Að auki hefur efnið verið rannsakað fyrir önnur tungumál, meðal annars með þróun villumálheilda, eins og rakið verður hér í framhaldi. Fyrst er þó farið stuttlega yfir skilgreiningar á hugtakinu

¹<http://grammarly.com>

„lesblinda“, áhrif hennar á líf fólks með lesblindu og mismunandi flokkunarkerfi fyrir lesblinduvillur.

2.1 Lesblinda, skilgreiningar og áhrif

Til er fjöldi skilgreininga á lesblindu, en algengast er að henni sé lýst sem óvæntum lestrarerfiðleikum sem ekki er hægt að rekja til þátta á borð við greind, félagslegar aðstæður eða kennsluhætti. Alþjóðlegu lesblindusamtökin (International Dyslexia Association) skilgreina lesblindu sem: „sér-tækan námsörðugleika af taugalíffræðilegum uppruna. Einkennin birtast í erfiðleikum með nákvæman, sjálfvirkan lestur og slakri færni í stafsetningu og umskráningu. Erfiðleikana má í flestum tilfellum rekja til vanda við úrvinnslu málhljóða sem er ekki í samræmi við aðra vitsmunalega hæfni einstaklingsins og koma fram þrátt fyrir vandaða lestrarkennslu. Dyslexía getur leitt til erfiðleika með lesskilning og ónógrar lestrarreynslu, sem aftur getur komið niður á þróun orðaforða og öflun þekkingar í gegnum lestur (Lyon o.fl., 2003).²

Lesblinda getur hamlað einstaklingum í námi, starfi og í daglegu lífi, og þótt skilningur á eðli lesblindu hafi aukist á undanförunum áratugum finna lesblindir gjarnan fyrir vanmætti og kvíða (Sako, 2016). Þótt þetta eigi sérstaklega við í skólakerfinu, þar sem mikil áhersla er lögð á lestur og skrift, má einnig nefna breytt samskiptamynstur fólks á undanförunum áratug sem mögulegan nýjan áhrifaþátt. Sífellt stærri hluti samskipta fólks, bæði í vinnu og utan hennar, fer fram í rituðu máli, svo sem í tölvupósti, textaskilaboðum og á samfélagsmiðlum. Í bandarískri rannsókn á því hvernig lesblindir nota samfélagsmiðla (Reynolds og Wu, 2018) var rætt við nokkra einstaklinga sem langflestum þótti mun erfiðara að skrifa texta inn á Facebook heldur en að lesa texta, og sumir forðuðust með öllu að tjá sig þar. Þeir kölluðu eftir leiðréttingartóli sem hjálpaði þeim að skrifa réttan texta.

Áhrif þessara breyttu aðstæðna á líðan les-

²Dyslexia is a specific learning disability that is neurobiological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge. (Þýðing: lesvefurinn.hi.is)

blindra á Íslandi hafa ekki verið könnuð að ráði, en Félag lesblindra á Íslandi stendur nú fyrir fjáröflun til þess að rannsaka kvíða hjá börnum með lesblindu hér á landi. Félagið hefur einnig staðið fyrir vitundarvakningu hjá kennurum um áhrif lesblindu á streitu hjá grunnskólabörnum (Félag lesblindra á Íslandi, 2019).

Annað sem getur háð lesblindum er að eiga erfitt með vefleit (t.d. á Google), bæði vegna þess að þeim getur reynst erfitt að orða leitarfyrirspurnina þannig að hún skili árangri og svo vegna þess að hún kann að innihalda stafsetningarvillur (Morris o.fl., 2018). Þetta er vandamál sem mætti takast á við með málátæknilausnum á borð við leitartillögur og spáskrift.

Á Lesvef Háskóla Íslands kemur fram að erfiðara geti verið að vinna bug á stafsetningarvandinum en lesvandinum, vegna undirliggjandi umskráningarerfiðleikanna, sem koma einmitt sterkt fram í stafsetningu. Erfiðara reynist að festa stafsetningu orðs í langtímaminninu ef umskráningu er ábótavant, sérstaklega ef stafsetning orðs er ekki í samræmi við framburð þess. Hægt er að ná góðum árangri í lestri með skilvirkri þjálfun, en stafsetningu getur verið ábótavant fram eftir aldri. Þetta er því vandi sem getur fylgt fólki á öllum ævistigum.

2.2 Einkenni og flokkun lesblinduvillna

Oft er talað um skrifblindu sem eitt af einkennum lesblindu, og á hún þá sérstaklega við um erfiðleika lesblindra við skrift. Sjá má viss einkenni á villunum sem lesblindir gera þegar þeir skrifa texta, og ekki síst dreifingu þeirra. Samkvæmt því sem fram kemur á Lesvefnum eru eftirfarandi einkenni algengust í stafsetningu lesblindra, og má gera ráð fyrir því að þetta eigi sérstaklega við um íslenska málhafa með lesblindu:

- erfiðleikar með að sundurgreina hljóð orða
- brottfall stafa (*stona* í stað *stofna*)
- stafavíxl (*gagna* í stað *ganga*).
- að eiga erfitt með að greina stutt og langt sérhljóð (einfaldan/tvöfaldan samhljóða, s.s. *baka* í stað *bakka*)
- orð rituð eftir framburði (*jeg* í stað *ég*)
- að eiga erfitt með að greina mun á hörðum og linum samhljóðum

- að rugla saman hljómlíkum stöfum (grannir og breiðir sérhljóðar, b/d, f/v, n/m o.fl.)

Margs konar flokkunarkerfi fyrir lesblinduvillur hafa verið þróuð í gegnum árin, og engin ein aðferð sem hefur verið tekin upp og gildir um öll tungumál. Þetta má að skýra með því að nákvæm skipting í villuflokka fer að hluta til eftir tungumálinu sem á í hlut. Ritkerfi tungumála eru misgegnisæ, þ.e. mismikið samræmi er milli málhljóða orðs og stafsetningar þess. Dæmi um tungumál þar sem þetta misræmi telst mikið (e. *phonetically obscure languages*) eru enska, franska og danska, en á hinn endann (e. *phonetically transparent languages*) falla til dæmis finnska, tékkneska og spænska. Hér er stundum einnig notað hugtakið „ritháttardýpt“ (e. *orthographic depth*), og teljast þá síðarnefndu málin þrjú vera með grunnan rithátt, þar sem samræmi milli ritháttar og málhljóða er mikið. Íslenska á frekar heima með þessum málum, þar sem tiltölulega mikið samræmi er milli málhljóða og rittákna, þótt viss einkenni flæki málið, svo sem samlögun ýmissa málhljóða sem og reglur um notkun „i/y“ og „hv/kv“, svo þekktustu dæmin séu tekin.

Ein leið til að flokka villur í stafsetningu er í regluvillur annars vegar og hljóðrænar villur hins vegar, og það er til dæmis gert á fyrrnefndum Lesvef Háskóla Íslands. Regluvillur eru hinar dæmigerðu stafsetningarvillur, þar sem stafsetningarreglur eru brotnar og þær eru óháðar nokkrum hljóðkerfisvanda. Hljóðrænu villurnar skiptast aftur í hljóðréttar (e. *phonetic*) og hljóðrangar (e. *dysphonetic*) villur. Hljóðréttar villur eru þær villur þar sem orðið er skrifað upp eins og það hljómar, en stafsetningin er ekki í samræmi við hefð (svo sem „vestla“ í stað „versla“). Hljóðrangar villur eru villur þar sem orð er rangt ritað, en þó án þess að samræmi sé við hljóð orðanna (svo sem „stona“ í stað „stofna“).

Ýmsar rannsóknir benda til þess að hljóðrænar villur séu algengari meðal lesblindra í tungumálum með mikla ritháttardýpt, eins og ensku (Wimmer o.fl., 2000). Þetta er þó ekki einhlítt, því rannsóknir á málum með grynnri rithátt, eins og tékknesku og bosníska, hafa gefið til kynna að hljóðrænar villur hjá lesblindum málhöfum séu síst fátíðari þar (Caravolas og Volín (2002) og Duranovic (2016)).

Önnur leið til flokkunar er að líta á annars vegar óþekkt orð (non-word errors) og hins vegar þekktar beygingarmyndir (real-word errors). Þá er

miðað við hvort orðmyndin finnst í gagnagrunnum/listum yfir þekkt orð eða orðmyndir. Þessi flokkun er gagnleg við þróun málfarsleiðréttingarforrita, því nýta má ýmis fyrirbyggjandi gögn til að greina og jafnvel leiðrétta slíkar villur sjálfkrafa. Þessi flokkun er til dæmis nefnd í verkáætlun máltækniáttaks ríkisstjórnarinnar 2018–2022, í hlutanum þar sem fjallað er um gerð málrýna (Anna B. Nikulásdóttir o.fl., 2017).

Þessu tengd eru ruglingsmengi, en það eru orðapör sem hljóma eins í framburði en ruglast oft í ritmáli, og eru algengar villur hjá lesblindum (Steinunn Rut Friðriksdóttir og Anton Karl Ingason, 2020). Þetta eru svokallaðar samhengisháðar villur, þar sem málfræðilegt samhengi segir til um hvort orðið er rétt eða rangt ritað.

Fleiri aðferðir til flokkunar eru til, svo sem í yfirflokkana hljóðkerfisvillur, stafsetningarvillur og málfræðivillur, sem síðan greinast í undirflokk (Tops o.fl., 2012). Til viðbótar við framangreint má einnig nefna að lesblindir gera einfaldlega fleiri villur en aðrir málnotendur (sjá t.d. Tops o.fl. (2012)). Það orsakast líka að hluta til af því að hinn lesblindi kemur síður auga á þær þegar hann rennir yfir texta sinn.

Hér hefur verið farið yfir ýmsar leiðir til að flokka þær villur sem koma fram hjá lesblindum. En til þess að greina megi þessar villur þarf að safna saman fjölbreyttum textum sem lesblindir hafa skrifað og skrásetja villurnar. Til eru málheildir með textum frá mismunandi hópum málnotenda (s.s. almennum málnotendum, börnum, lesblindum, fólki með íslensku sem erlent mál), þar sem villur hafa verið greindar og leiðréttar.

Villumálheildir með textum lesblindra koma að góðu gagni þegar greina þarf hversu áberandi hvert þessara einkenna er hjá lesblindum, og þar með komast að því hver þeirra beri að leggja mesta áherslu á við þróun stoðtöla fyrir lesblinda. Slíkar lesblinduvillumálheildir hafa verið þróaðar fyrir ýmis tungumál, svo sem ensku (Pedler, 2007), spænsku (Rello o.fl., 2014) og arabísku (Alamri og Teahan, 2017), og meðal annars verið notaðar í rannsóknum á lesblindu og í máltækni til þjálfunar á leiðréttingarforritum.

3 Aðferðafræði

Eins og fram hefur komið hefur hér verið tekið saman villumálheild með lesblinduvillum. Hér verður gerð þessarar málheildar lýst, gögnum sem liggja henni til grundvallar og aðferðunum sem

notaðar voru til skráningar og greiningar á villunum.

3.1 Gögn

Útbúin var lítil villumálheild með textum frá þremur málhöfum. Textarnir eru til bæði í upprunalegri útgáfu og yfirlæsni, þar sem villur hafa verið leiðréttar. Málheildin er alls rúm 3000 orð. Málhafarnir þrír eru annaðhvort greindir með lesblindu eða sýna skýr einkenni hennar, en hafa að öðru leyti mismunandi bakgrunn eins og lesa má nánar um í framhaldinu. Textarnir eru frá tímabilinu 2015–2020 og eru af ýmsum toga. Þetta eru meðal annars tölvupóstskeyti, vefsíðu- og kynningartextar, ferilskrár, starfsumsóknir og skóla- verkefni. Leiðréttingarnar voru unnar jafnóðum af höfundu (SLI).

3.1.1 Málhafi 1

Málhafi 1 er íslensk kona fædd árið 1961, með háskólapróf. Hún hefur ekki fengið formlega lesblindugreiningu, þar sem slíkt tíðkaðist ekki í skólum þegar hún var barn. Hún sýnir þó augljós einkenni lesblindu; meðal annars les hún orð stundum upp á annan hátt en þau eru skrifuð, svo víxlar hún oft stöfum innan orða þegar hún skrifar, sleppir stöfum innan úr eða aftan af orðum og á erfitt með að greina hvenær skrifa skal eitt eða tvö orð. Skortur á lesblindugreiningu hamlaði henni í skólagöngu því hún hélt að greindarskortur væri ástæða þess að hún ætti erfitt með réttitun og lestur.

3.1.2 Málhafi 2

Málhafi 2 er íslensk kona fædd árið 1991, með háskólapróf. Hún var greind með ADHD og lesblindu snemma á grunnskólaaldri og hefur notið þeirrar aðstoðar sem boðist hefur í skólakerfinu vegna lesblindunnar; meðal annars hefur hún fengið lengri tíma til að leysa próf. Hún gerir margar og margs konar villur í ritun texta og á erfitt með umskráningu hljóða yfir í stafi. Þetta lýsir sér til dæmis í því að jafnvel þótt orð sé borið fram hægt og með skýrum skilum milli orðhluta hjálpar það henni ekki að ná að stafsetja það rétt. Hún kvartar einnig yfir að eiga í erfiðleikum með vefleit, þ.e. að finna árangursríkustu leitarfyrirspurnina hverju sinni.

3.1.3 Málhafi 3

Málhafi 3 er íslenskur karl fæddur 1961. Hann var ekki greindur með lesblindu fyrr en á fertugsaldri

og lýsir sögu sinni á svipaðan hátt og málhafi 1. Hann gerir ýmiss konar ritvillur, bæði innsláttarvillur og villur vegna umskráningar hljóða. Hann hefur kynnt sér lesblindu vel og starfar fyrir Félag lesblindra á Íslandi. Hann nýtir sér ýmis stöðtöl til að lesa og skrifa texta, svo sem talgreiningu, leiðréttingarforrit og talgervingu.

3.2 Gerð villumálheildarinnar

Upprunalegu og yfirlæsnu textarnir voru lesnir saman, málsgrein á móti málsgrein. Til einföldunar voru eingöngu teknar með samhlíða málsgreinar, þ.e. ef málsgreinum hafði verið skipt upp í tvær var þeim sleppt, sem og ef tveimur málsgreinum var skeytt saman í eina. Um þetta voru nokkur dæmi, svo sem þegar tilvísunarsetning var slitin frá aðalsetningunni og rituð sem heil málsgrein (málsgrein sem byrjar á „Sem“), en hafði verið leiðrétt yfir í eina málsgrein í yfirlæsnu útgáfunni.

Þess skal einnig getið að þegar textar málhafa 1 og 2 voru lesnir yfir leyfði yfirlæsari sér á nokkrum stöðum að gera leiðréttingar umfram það sem nauðsynlegt var, til dæmis til að veita viðbótarupplýsingar eða lagfæra stíl. Þetta var gert vegna þess að textarnir voru til raunverulegrar notkunar og oft þess eðlis að æskilegt var að leiðrétta meira en augljósustu villur til þess að þeir skiluðu tilætluðum árangri (t.d. í atvinnuumsóknum). Textar málhafa 3 voru hins vegar eingöngu leiðréttir í rannsóknarskyni og ættu því einungis að innihalda leiðréttingar á raunverulegum villum en ekki stíl.

Villumálheildin var sett upp í forritinu SketchEngine³, sem nota má til að halda utan um og vinna með málheildir. Allar villur voru merktar með xml-tögunum <err> og <corr>, þar sem fyrra tagið er sett utan um villuna og það seinna utan um leiðréttinguna, ásamt eigindi sem stendur fyrir villuflokk:

```
Maðurinn hennar var búinn að <err
type="omission">stykja</err> <corr
type="omission">styrkja</corr>.
```

Ekki er gert ráð fyrir fleiri en einni villu í hverju orði, og er þá flokkuð sú villa sem þykir vega þyngra / vera alvarlegri.

SketchEngine býður ekki upp á viðmót til þess að merkja villur í villumálheild, og því var útbúin Python-skrifta með einföldu skipanalínuviðmóti þar sem hægt var að merkja hverja villu eftir

³<https://www.sketchengine.eu/>

villuflokkir og vista málsgreinina í skrá, sem svo var hægt að hlaða upp í SketchEngine. Til þess að finna hvar munur var á upprunalegu og yfirlesnu málsgreinunum var notast við Python-kóðasafnið *diffli*, sem merkir við mun á texta. Skriftan er ekki fullkomin og einn galli á henni er að hún flokkar aðeins eitt orð í einu, svo ef gerðar hafa verið miklar breytingar í leiðréttingunni, t.d. setningaliður færður til, verður til ein „villa“ fyrir hvert orð. Þetta væri þó vel hægt að lagfæra með smá yfirlegu og fínþússun. Þessi kóði verður gerður aðgengilegur ásamt villumálheildinni.⁴

3.3 Villuflokkar

Megináskorunin við gerð málheildarinnar fólst í því að skilgreina villuflokkana svo villumálheildin nýttist sem best í þeim tilgangi sem hér hefur verið lýst. Þar sem lesblinda er umfjöllunar-efni þessarar greinar er eðlilegast að sérstaklega sé litið til villna sem helst einkenna mál lesblindra, en auk þess koma upp ótal aðrar villur, svo sem í greinarmerkjasetningu, við notkun orðasambanda, í setningarskipan og fleira. Þess vegna er ekki nóg að líta á eingöngu hljóðrænar villur, stafavíxl og brottfall.

Á endanum var notast við villukerfi sem styðst að nokkru leyti við flokkun Rello o.fl. (2014). Þar er notast við hugtakið „einfaldar villur“ yfir flokk fjögurra villutegunda sem einmitt eru algengar hjá lesblindum og hafa flestar verið nefndar hér að framan. Þetta eru:

- brottfall stafa (*klipa* í stað *klippa*)
- viðbót stafa (*njóta* í stað *njóta*)
- stafavíxl (*eining* í stað *einnig*)
- stöfum skipt út (*bagga* í stað *bakka*)

Ákveðið var að greina þessar villur frekar með málþæknisjónarmið í huga, í samhengisháðar og ósamhengisháðar villur. Einnig voru sérstaklega merktar vissar villutegundir sem vitað er að reynast lesblindum erfðar, eins og y/ý-villur (*hytta* í stað *hitta* eða *íta* í stað *yta*) og hljóðréttar villur (*vestla* í stað *versla*). Þessar villur eru einnig greindar í samhengisháðar og ósamhengisháðar villur.

⁴Þ.e. aðgengilegur til kennara. Ekki hefur verið óskað eftir heimild frá eigendum textanna til opinberrar birtingar á málheildinni en það má vel skoða.

Að öðru leyti var notast við villuflokka svipaða þeim sem taldir eru upp í verkáætlun Málþæknisáætlunar (bls. 90–91). Meðal annarra flokka eru hljóðréttar villur, málfræðivillur, greinarmerkjavillur og villur í hástafa- og bilanotkun, villur þar sem orð vantar eða er ofaukið, villur í stíl og meðferð orðasambanda og fleira. Helstu yfirflokkar villna eru eftirfarandi:

- Einfaldar villur
- Hljóðrænar villur
- Málfræðivillur
- Málnotkunarvillur
- Greinarmerki, bil, hástafir og fleira
- Stílfræðilegar villur eða betrubætur

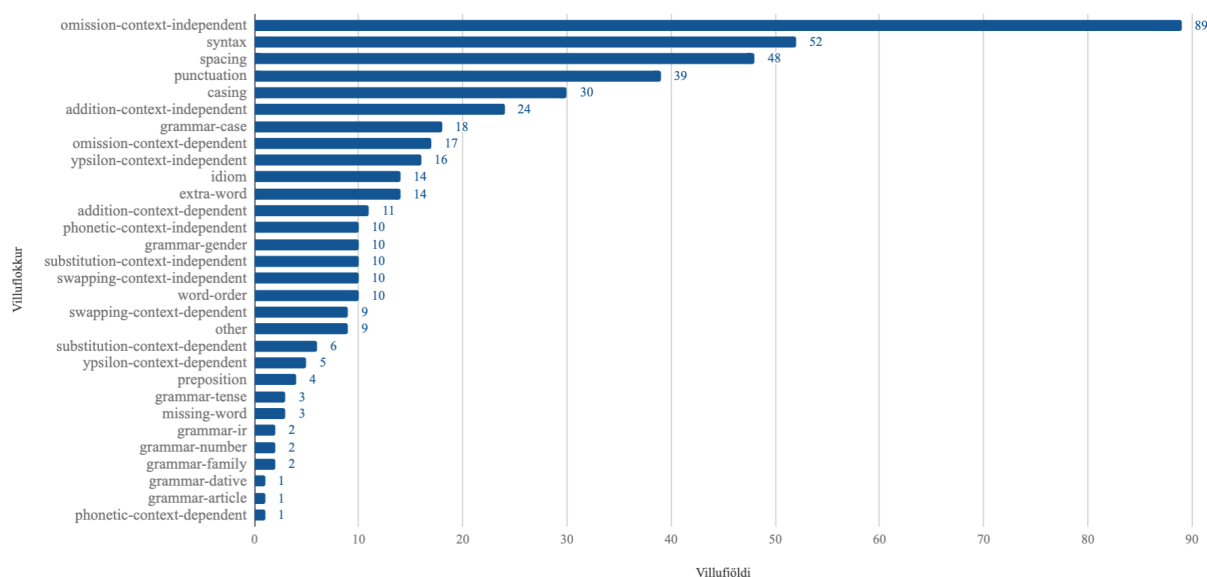
Nánari skiptingu má finna í næsta kafla þar sem niðurstöðurnar eru tífundaðar. Ýmislegt í þessari flokkun mætti endurskoða og vinna betur, sérstaklega með því að skilgreina nánar yfir- og undirflokk, og samræma skráningu á samhengisháðum og ósamhengisháðum villum.

Enn fremur er sitthvað sem mætti betrubæta við flokkunina sjálfa, en fjöldi vafaatriða kom þar fram, til dæmis þegar villa getur fallið í tvo eða jafnvel þrjá flokka (nefna má sem dæmi villuna *baga* í stað *baka*, sem gæti flokkast bæði sem einföld og hljóðrétt villa, og y-villur, sem gætu talist einfaldar villur en eru hér flokkaðar sem y-villur). Rétt eins og við á um villuskráninguna er þó einfalt að breyta flokkunarskriftunni með uppfærðri villuflokkun eftir því sem við á, og flokka villurnar aftur, enda málheildin ekki stór.

4 Niðurstöður

Í eftirfarandi framsetningu á villuflokkunum sem greindir voru í málheildinni var ákveðið að sleppa úr þeim flokkum sem gætu skekkt niðurstöðurnar. Hér er átt við stílfræðilegar viðbætur og leiðréttingar sem höfundur gerði án þess að um raunverulegar villur væri að ræða.

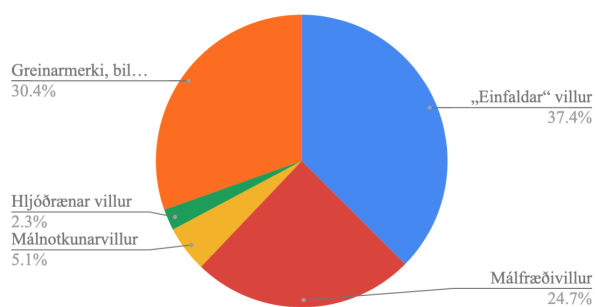
Fyrst ber að skoða grunnskiptingu villna í flokkana sem greindir voru. Hana má sjá á mynd 1. Af 470 villum sem greindar voru eru ósamhengisháðar brottfallsvillur algengastar, eða 89 talsins. Setningafræðilegar villur eru þarna ofarlega á blaði en sökum þess hvernig flokkunarskriftan var hönnuð (hvert orð er ein villa) eru þessar villur hlutfallslega of margar, og ætti því að taka með fyrirvara.



Mynd 1: Tíðni hvers villuflokks. Heildarfjöldi villna er 470.

Villur í notkun bila, greinarmerkja og hástafa eru einnig nokkuð algengar, eða 117 samtals. Þar á eftir má sjá fleiri flokka einfaldra villna, sem og málfræðivillur, orðnotkunarvillur og y-villur. Sumir villuflokkarnir voru lítið notaðir og mætti hæglega einfalda flokkunina með hliðsjón af þessum niðurstöðum eftir þörfum. Hér er til dæmis átt við málfræðivillur eins og þágufallshneigð (1 villa), villur í beygingu frændsemisorða (2 villur) og ranga notkun tölu eða greinis (2 og 1 villa).

Þegar villurnar hafa verið greindar í yfirflokkana sem nefndir eru í kafla 3.3 (að undanskildum stílvillum sem hér er sleppt), fæst eftirfarandi skipting.



Mynd 2: Skipting villna í yfirflokk.

Hér er við hæfi að líta nánar á „einföldu“ villurnar (brottfall og viðbót stafa, stafavíxl og stöfum skipt út), því þar er flokkunin ítarlegust. Í töflu 1 má sjá að villur þar sem stafir falla brott úr orðum eru tíðastar, en einnig er algengt að stöfum sé bætt við. Hér er ekki gerður greinarmunur á sam-

Villuflokkur	Villufjöldi
Brottfall	106
Viðbót	35
Staf/stöfum skipt út	16
Stafavíxl	19

Tafla 1: Dreifing einfaldra villna.

hengisháðum/ósamhengisháðum villum, en eins og sjá mátti á mynd 1 voru ósamhengisháðar brottfallsvillur algengasti flokkurinn. Skipting einföldu villnanna í samhengisháðar og ósamhengisháðar er 24,4% samhengisháðar villur (raunveruleg orð) og 75,6% ósamhengisháðar (orðleysisvillur).

5 Umræða

Eins og sjá má eru einföldu villurnar stærsti flokkurinn, en það eru einmitt þær villur sem oftast eru nefndar sem dæmigerðar lesblinduvillur. Sérstaklega algengt er brottfall eins eða fleiri stafa úr orðum. Oft vantar orðhluta innan úr orði, t.d. *hugbúnafræði* í stað *hugbúnaðarfræði*, eða aftan á orð (*reyn* í stað *reyna*). Næstalgengasta tegund einfaldra villna reyndist vera viðbót stafa í orð, sem stundum virðist handahófskennd en felst líka í mörgum tilfellum í tvöföldun samhljóða (*listsköpunn*, *oppnað*, *alldrei*, *þjónustuverri*).

Fátíðari voru útskiptingar stafa og stafavíxl, en stafavíxl eru þó einmitt í daglegu tali oft talin dæmigerðasta villan hjá lesblindum, nokkurs konar erkitýpa lesblinduvillnanna. Þau komu þó

fyrir, til dæmis í orðum eins og *fyrir* í stað *fyrri*, *kveður* í stað *kveður* og *þaning* og *eining* í stað *þannig* og *einnig*.

Y-villur voru 21 samtals, að meirihluta ósamhengisháðar (*hytta*, *ítt*) en einnig samhengisháðar, s.s. *sínir* og *bíður*). Hér eru bæði taldar villur þar sem y/y er ritað þar sem ætti að vera i/í og öfugt.

Hljóðréttar villur voru skráðar 10, en hugsanlega mætti flokka fleiri villur sem hljóðréttar sem til dæmis voru skráðar sem brottfallsvillur (t.d. *búin* í stað *búinn*). Dæmi um hljóðréttar villur í gögnunum eru *vestla* í stað *versla*, *stafróðröð* í stað *stafrófsröð* (ekki alveg nákvæm uppskrift þó), *stæðstu* í stað *stærstu* og *allskis* í stað *alls kyns*.

Reglurvillur, s.s. málfræðivillur þar sem málfræðireglur eru brotnar (og sem y-villurnar falla einnig undir) komu einnig fyrir, sérstaklega þar sem notað var rangt fall eða rugl var á málfræðilegu kyni.

Talsvert var um að notkun greinarmerkja (komma, punkta og gæsalappa) væri ábótavant, og einnig voru hástafir stundum notaðir í stað lágstafa og öfugt. Orðskiptingavillur flokkast sem villur í notkun bila, bæði þegar bil vantar milli orða (*ótalnámskeið*) og öfugt, þegar orð eru slitin sundur þar sem það á ekki við (*ferðamanna garðurinn*).

Á heildina litið var dreifing villuflokka nokkuð jöfn yfir alla málhafana þrjá. Einföldu villurnar voru algengastar í textum þeirra allra, og þá sérstaklega villur vegna brottfalls stafa eins og áður er getið. Í textum þeirra allra koma fyrir allir flokkar einfaldra villna, sem og y-villur og hljóðréttar villur. Þetta er einmitt dæmigert fyrir lesblinda málnotendur og í samræmi við það sem komið hefur fram í rannsóknum.

Þessi málheild er vitaskuld mjög lítil og nær bara yfir texta þriggja einstaklinga, og ber því ekki endilega að líta svo á að hún sé lýsandi fyrir texta lesblindra á íslensku. Þau líkindi sem má sjá með villunum sem þessir þrír einstaklingar gera eru þó það mikil að málheildina má í það minnsta hafa til hliðsjónar við greiningu á því hvaða villur eru algengastar í textum lesblindra móðurmálshafa á íslensku.

Hægt er að velta fyrir sér hvernig nýta megi niðurstöður þessarar athugunar til þróunar á leiðréttingarforriti sem sniðið er að lesblindum. Það vekur til dæmis eftirtekt hversu margar villur eru í málheildinni. Í um 3000 orða textasafni voru greindar 470 villur, og þótt sú tala segi ekki allt er hún ágætis vísbending um að huga þurfi að mörgu

við hönnun stoðtóla fyrir lesblinda. Algenzt er að fleiri en ein villa sé í hverri málsgrein og afar fáar málsgreinar eru villulausar. Margar villur í einni málsgrein geta gert leiðréttingarforriti mjög erfitt fyrir að átta sig á því hvað notandi vill segja og leiðrétt málsgreinina í samræmi við það.

Einföldu villurnar, sem eru algengastar, geta verið allt frá því að vera mjög einfaldar að leysa (t.d. með því að skoða stafafjarlægð í orðleysisvillum) yfir í að vera frekar snúnar (samhengisháðar villur í sama orðflokki og rétta orðið). Hér gæti þurft að beita sérstökum aðferðum, eins og að nýta ruglingsmengjamálheild.

Málfræðivillur, málfræðivillur, stílvillur og greinarmerkjavillur virðast vera svipaðs eðlis og villur sem aðrir málnotendur gera, þótt þær séu greinilega algengari hjá lesblindum eins og aðrar villur. Því ætti ekki að þurfa sérstakar aðferðir til að leiðrétt þessar villur hjá lesblindum umfram aðra.

Í raun er líklegt að helsta áskorunin við þróun málfræðileiðréttingarhugbúnaðar fyrir lesblinda felist í að leysa það vandamál að leiðrétt málsgreinar sem innihalda margar villur, en ekki bara eina eða tvær. Það verður spennandi máltæknaverkefni til að kljást við á næstu árum.

Heimildaskrá

- Maha Alamri og William J Teahan. 2017. <https://doi.org/10.18653/v1/W17-1309> A New Error Annotation for Dyslexic texts in Arabic. í *Proceedings of the Third Arabic Natural Language Processing Workshop*, síður 72–78, Valencia, Spain. Association for Computational Linguistics.
- Anna B. Nikulásdóttir, Jón Guðnason og Steinþór Steingrímsson. 2017. *Language Technology for Icelandic 2018-2022*. Ministry of Education, Science and Culture.
- M. Caravolas og Jan Volín. 2002. <https://doi.org/10.1002/dys.206> Phonological Spelling Errors among Dyslexic Children learning a Transparent Orthography: The Case of Czech. *Dyslexia (Chichester, England)*, 7:229–45.
- Mirela Duranovic. 2016. <https://doi.org/10.1177/0022219416645814> Spelling Errors of Dyslexic Children in Bosnian Language With Transparent Orthography. *Journal of Learning Disabilities*, 50.
- Félag lesblindra á Íslandi. 2019. <https://www.lesblindir.is/fjaroflun/rannsokn-a-kvida-lesblindra-barna/> Rannsókn á kvíða lesblindra barna.

- Lesvefur Háskóla Íslands. 2019.
<https://www.lesblindir.is/fjaroflun/rannsokn-a-kvida-lesblindra-barna/> Lesvefurinn um læsi og lestrarerfiðleika.
- G Reid Lyon, Sally E Shaywitz og Bennett A Shaywitz. 2003. A definition of dyslexia. *Annals of dyslexia*, 53(1):1–14.
- Meredith Ringel Morris, Adam Fourney, Abdullah Ali og Laura Vonessen. 2018. <https://doi.org/10.1145/3173574.3173609> Understanding the needs of searchers with dyslexia. í *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, New York, NY, USA. Association for Computing Machinery.
- Jennifer Pedler. 2007. Doktorsritgerð, London University.
- Luz Rello, Joaquim Llisterri og Ricardo Baeza-Yates. 2014. Dyslist: An annotated resource of dyslexic errors.
- Lindsay Reynolds og Shamoei Wu. 2018. “I’m Never Happy with What I Write”: Challenges and strategies of people with dyslexia on social media.
- Enkeleda Sako. 2016. <https://doi.org/10.26417/ejis.v2i2.p233-241> The Emotional and Social Effects of Dyslexia. *European Journal of Interdisciplinary Studies*, 2(2):233–241.
- Steinnun Rut Friðriksdóttir og Anton Karl Ingason. 2020. A New Error Annotation for Dyslexic texts in Arabic. í *Proceedings of ICAART 12 (International Conference on Agents and Artificial Intelligence)*, Valletta, Malta. SCITEPRESS.
- Wim Tops, Maaïke Callens, Evi Bijn og Marc Brysbaert. 2012. <https://doi.org/10.1177/0022219412468159> Spelling in Adolescents With Dyslexia: Errors and Modes of Assessment. *Journal of learning disabilities*.
- Heinz Wimmer, Heinz Mayringer og Karin Landerl. 2000. <https://doi.org/10.1037/0022-0663.92.4.668> The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology*, 92:668–680.