

Efni

Inngangur	2
1. Atkvæði	2
1.1. Hvað er atkvæði?.....	3
1.2. Líkön sem lýsa atkvæðum.....	4
1.3. Íslensk atkvæði.....	4
1.4. Gildi atkvæðarannsókna.....	5
2. Vélræn nálgun – aðferðir	6
2.1. Hljóðritun.....	6
2.2. Atkvæðaskipting.....	7
2.3. Takmarkanir.....	9
3. Athuganir	10
3.1. Tíðni mismunandi atkvæðagerða í íslensku.....	10
3.2. Dreifing samhljóða á stuðul, kjarna og hala.....	13
Lokaorð	15
Viðauki: Dæmi um úttak forrits	16
Textar	20
Heimildir	20

Inngangur

Atkvæðið sem grunneining máls kemur víða við sögu í umfjöllun um tungumál – bæði í málvísindalegum rannsóknum og á hagnýtum vettvangi. Sem dæmi um svið þar sem atkvæði gegna hlutverki má nefna samanburð tungumála, athuganir á máltöku barna, þróun forrita sem skynja og skilja talað mál og rannsóknir á fornum kveðskap. Einn geiri atkvæðarannsókna, hljóðskipunarfræði (e. *phonotactics*), snýr að því eftir hvaða lögmálum máhljóðin skipa sér í atkvæði. Vissa eiginleika hljóðskipunar má rannsaka með vélrænum aðferðum þó því fylgi ákveðnar takmarkanir en hér er ætlunin að tengja íslenska atkvæðið slíkum aðferðum.

Málfræðingar hafa oft nálgast hljóðskipun með algildum reglum og hömlum, svo sem að tiltekið hljóð geti aðeins staðið fremst í atkvæði eða orðmynd. Aðrir hafa þó bent á að tölfræðilegar tilhneigingar geti einnig veitt upplýsingar um málið og að ef tíðni ákveðins hljóðskipunarmynsturs sé meiri en ætla mætti af tilviljanakenndri dreifingu sé um að ræða markverðan þátt í lýsingu málsins. Þátt sem hamlnalíkani er hætt við að horfa fram hjá (Kessler 1997).

Hér verður gerð tilraun til að kanna tölfræðilega dreifingu íslenskra máhljóða með tilliti til atkvæðagerðar. Við slíka athugun er fýsilegt, og jafnvel nauðsyn, að nálgast athugunina með vélrænum hætti þar sem handavinna við annars konar nálgun væri yfirþyrmandi. Skrifað hefur verið forrit sem hljóðritar texta og skiptir í atkvæði. Leitast er við að sýna fram á að vélræn nálgun við athuganir á íslenskum atkvæðum sé hentug frá málfræðilegu sjónarhorni, sem og tæknilegu.

Í fyrsta kafla er fjallað um atkvæði almennt, hvernig atkvæði er skilgreint (1.1), hvaða líkön eru notuð til að lýsa atkvæðagerð (1.2) og hugmyndum um hvernig íslenska atkvæðið fellur að þeim líkönum (1.3). Loks er fjallað um hagnýtt gildi atkvæðarannsókna (1.4). Annar kafli fjallar um aðferðir sem beitt er við vélræna hljóðritun (2.1) og atkvæðaskiptingu (2.2). Þá er takmörkunum þessara aðferða gerð skil (2.3). Þriðji kafli fjallar um tilraunir til að byggja atkvæðarannsóknir á aðferðum sem lýst er í kaflanum á undan. Algengustu atkvæðagerðir íslensku (CV, CVC o.s.frv.) eru skoðaðar (3.1) og gerð er grein fyrir dreifingu samhljóða á stuðul, kjarna og hala (3.2).

1. Atkvæði

Hugtakið **atkvæði** er fyrirbæri sem flestir Íslendingar telja sig kunna nokkur skil á og telja jafnvel að það sé svo einfalt að umfjöllun um það sé ástæðulaus. Að hluta til er það rétt að atkvæði er einfalt hugtak því fæstir eiga í erfiðleikum með að tilgreina hve mörg atkvæði eru í tiltekinni

orðmynd eða segð. Málið vandast þó þegar kemur að því að fella atkvæðagerð inn í málfræðilegt líkan en þá fer jafnvel fræðimenn að greina á um hvað rétt reynist.

1.1. Hvað er atkvæði?

Jafnt fræðimenn sem alþýða manna líta svo á að sérhljóð séu grundvallareining atkvæða. Vanalegt er að líta svo á að sérhljóð segðar séu jafnmörg og atkvæðin. Í íslensku er þessu yfirleitt svona farið þó að við vissar aðstæður megi líta svo á að samhljóð verði atkvæðisbær. Þannig geta nefhljóð orðið atkvæðisbær þar sem sérhljóð fellur brott vegna áhersluleysis, svo sem þegar segðin *hljólinu* er borin fram *hljól'nu* (Kristján Árnason 2005:238).

Það eru þó ekki allir fræðimenn sammála um hvernig túlka beri atkvæðaskiptingu í segð eins og *hljól'nu* og málið vandast enn í erlendum málum þar sem atkvæðisbær samhljóð, eða hvað svo sem rétt er að nefna þau, eru algengari en gerist í íslensku. Til dæmis hefur verið bent segðir eins og:

[ts'ktsk^wts']

Þetta er dæmi um IPA hljóðritaða segð úr tungumálinu nuxálk, sem einnig hefur verið nefnt *bella coola* eftir bæjarfélagi í Kanada þar sem málið er talað af nokkrum innfæddum. Á íslensku merkir segðin ‘hann kom’ en hún er sérstök fyrir þær sakir að í henni er ekkert sérhljóð. Fræðimenn greinir á um hvernig túlka beri segð af þessu tagi og hún hefur verið talin 0, 2, 3, 5 eða 6 atkvæði – allt eftir því hvaða greiningaraðferð er notuð (Bagemihl 1991)¹.

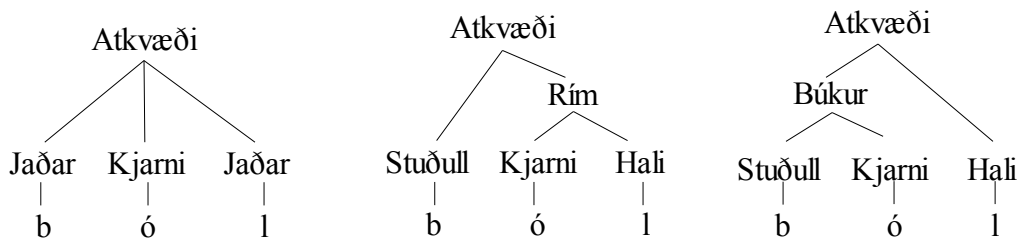
Ef vel er að gáð koma slíkar vandræðasegðir ekki aðeins fyrir í framandi málum fjarlæggra heimshluta. Í íslensku er til að mynda algengt að einhver tjái merkingu eins og ‘heyrdú nú mig’ eða ‘ég er að hugsa málið’ með einhvers konar *hmmm*. Í ensku er beðið um þögn með segð sem í rithætti er eitthvað í ætt við *shh* og óskað er eftir athygli með hljóði á borð við *psst*.

Jaðartilvik af þessu tagi eru þó svo sjaldgæf í íslensku að ekki þarf að hafa miklar áhyggur af því að þær spilli kenningakerfum sem lýsa málinu. Rökrétt virðist að nefna sérhljóðalausar segðir undantekningar í íslensku og ganga út frá því að sérhljóð séu í raun grundvallareiningar atkvæða. Þetta þarf ekki að hafa veruleg áhrif á heildarmynd málvísinda – þó ef til vill megi nota dæmi af þessu tagi sem rök gegn algildum hömlum í mállýsingu. Hér verður gengið út frá því að atkvæði í íslenskri segð séu að jafnaði jafnmörg og sérhljóðin.

¹ Tilvitnað eftir *Wikipedia*. Sótt 5. desember 2006 af <<http://en.wikipedia.org/wiki/Syllable>>

1.2. Líkön sem lýsa atkvæðum

En hvaða líkön eru þá notuð til að lýsa atkvæðum? Í íslensku og skyldum málum er yfirleitt gert ráð fyrir að sérhljóð myndi **kjarna** (e. *nucleus*) atkvæðis. Vandinn vex þó þegar gera skal grein fyrir því hvernig nærliggjandi samhljóðar raðast utan á kjarnann. Til eru hugmyndir um að undanfarandi samhljóð, nefnd **stuðull** (e. *onset*), séu nátengdari kjarnanum en þau eftirfarandi og saman myndi stuðull og kjarni **búk** (e. *body*) atkvæðis. Aðrir telja að eftirfarandi samhljóð, **hali** (e. *coda*), tengist kjarnanum sterkari böndum og myndi með honum svonefnt **rím** (e. *rhyme*) en stuðull. Loks eru þeir sem aðhyllast flata atkvæðabyggingu þar sem kjarninn er ekki spyrstur sérstaklega við annan **jaðar** (e. *margin*) atkvæðisins. Hugtök eins og stuðull, kjarni og hali eru nefnd **stofnhlutar atkvæða**. Dæmi um hvernig íslenska orðinu *ból* er lýst með ofangreindum líkönum er sýnt á mynd 1. Bent hefur verið á að þessar ólíku hugmyndir um atkvæðabyggingu séu tækifæri fyrir tölvumálvísindi (e. *computational linguistics*) því hægt sé að meta með tölfræðilegum aðferðum hvort meiri fylgni er milli hljóðfræðilegra eiginleika kjarnasérhljóðs og stuðuls þess annars vegar eða hala þess hins vegar (Kessler 1997).



Mynd 1: Líkön sem lýsa atkvæðum

Einkvæða orðið *ból* myndar svonefnt CVC atkvæði (e. *consonant-vowel-consonant syllable*) en það hve mörg samhljóð skipa sér sitt hvoru megin við kjarnasérhljóð er annað verkefni atkvæðalýsingar. Í íslensku eru til atkvæði eins og CV, CVC og CCV svo dæmi séu nefnd. Talað er um að atkvæði sem enda á sérhljóðum séu **opin atkvæði** en þau sem endi á samhljóðum séu **lokuð**. Í þriðja kafla (3.2) verður tíðni ólíkra atkvæðagerða í íslensku athuguð.

1.3. Íslensk atkvæði

Kristján Árnason hefur fjallað nokkuð ýtarlega um stofnhlutagerð íslenskra atkvæða (2005:180-222) og verður gengið út frá niðurstöðum hans hér. Lýsing á stofnhlutagerð íslenskra atkvæða byggir í meginatriðum á því að kjarni og hali myndi saman rím sem svo myndar atkvæði með stuðli, það er **stuðull-rím líkaninu**. Þegar skipað er í atkvæði er frumþörf atkvæðisins að taka með

sér stuðul svo að þar sem eitt samhljóð er milli sérhljóða er litið svo á að það tilheyri síðara atkvæðinu, svo sem *hes.tur* en ekki *hest.ur* eins línuskiptingarreglur kunna að benda til. Séu fleiri en eitt samhljóð milli sérhljóða fara þau sem eftir eru í hala fyrra atkvæðis. Kjarninn er eini nauðsynlegi hluti atkvæðisins og mörg dæmi eru um atkvæði sem eru annaðhvort án stuðuls eða hala, svo sem *án, ís, út og já, þú, sé*. Sum atkvæði eru raunar aðeins kjarni, eins og orðið *á* í merkingunni vatnsfall.

Það óvanalegasta í lýsingu íslenska atkvæðisins er að öllum líkindum sú staðreynd að gert er ráð fyrir að sérhljóð geti tekið eitt samhljóð með sér inn í kjarna. Slíkir kjarnar eru þá nefndir **samhljóðakjarnar** en hinir sem eru aðeins skipaðir sérhljóði **sérhljóðakjarnar**. Með þeim hætti má skýra lengd íslenskra sérhljóða með algildri reglu: Sérhljóð í sérhljóðakjarna eru löng í atkvæði sem ber áherslu, en önnur sérhljóð eru stutt. Nánar er rætt um hvernig hljóð skipa sér í stuðul, kjarna og hala í kafla um vélræna atkvæðaskiptingu (2.2).

Í hnotskurn fylgir íslenska atkvæðið stuðull-rím líkaninu með því tilbrigði þó að samhljóð geta staðið með sérhljóði í kjarna. Meginreglan um hvernig samhljóð skipa sér í atkvæði er að stuðull hefur forgang, en hljóð sem eftir standa tilheyra hala undanfarandi atkvæðis.

1.4. Gildi atkvæðarannsókna

Rannsóknir á atkvæðum geta komið að gagni á margvíslegum sviðum. Fyrst ber að nefna að þekking á atkvæðum tungumála er áhugaverð frá sjónarhóli samanburðarmálvísinda (e. *comparative linguistics*). Atkvæði hegða sér með allólíkum hætti í mismunandi málum og þó að íslenska bjóði til dæmis upp á tiltölulega flókna samhljóðaklasa og hafi fjölbreyttar gerðir opinna og lokaðra atkvæða hafa sum mál til að mynda aðeins opin atkvæði. Havaíska er slíkt mál en þar eru halar ekki leyfðir (Lyovin 1997:260)² og öll atkvæði eru af taginu CV að frátöldum fáeinum tökuorðum (Elbert og Pukui 1979:13)³.

Félagsleg málvísindi (e. *sociolinguistics*) og sálfræðileg málvísindi (e. *psycholinguistics*) hafa ekki látið atkvæðið ósnert. Rannsóknir hafa þótt gefa til kynna að heilahvel karla og kvenna meðhöndli atkvæði með mismunandi hætti og að munur sem mælist á heilastarfsemi kynjanna við skynjun atkvæða komi ekki fram við annars konar hljóðáreiti (Obleser o.fl. 2004). Einnig má nefna að við athuganir á máltöku barna koma atkvæði mjög við sögu. Nýir málhafar mynda í fyrstu aðeins stök

2 Tilvitnað eftir *Wikipedia*. Sótt 5. desember 2006 af <http://en.wikipedia.org/wiki/Hawaiian_language>

3 Tilvitnað eftir *Wikipedia*. Sótt 5. desember 2006 af <http://en.wikipedia.org/wiki/Hawaiian_language>

sérhljóð og samhljóð en þegar kemur að því að setja hljóðin saman eru CV atkvæði það fyrsta sem barn nær tökum á. Við 10 mánaða aldur geta svo flóknari atkvæði byrjað að myndast, til dæmis CVC (Sigríður Sigurjónsdóttir 2002). Þetta á einnig við um önnur tungumál en íslensku (Levelt 1999) og getur verið fróðlegt að skoða í samanburði á máltöku ólíkra tungumála.

En atkvæði koma ekki aðeins við sögu í fræðilegum rannsóknum. Skilningur á atkvæðabyggingu tungumáls kemur einnig til góða á hagnýtum sviðum eins og í tungutækni. Ef ætlunin er að forrita formendatalgervil, það er talgervil sem ekki byggir á upptökum, eða hugbúnað sem skilur mælt mál er þekking á atkvæðum gagnleg. Sumir hafa raunar haldið því fram að eðlilegra sé að byggja slíkan hugbúnað á atkvæðum en einstökum hljóðöllum (King 1998). Þá þarf að setja málfræðilegar reglur fram með forritunarmálum. Í öðrum kafla er skoðað dæmi um einfalda brú milli málfræðilegs líkans og tölvunarfræðilegrar útfærslu.

2.Vélræn nálgun – aðferðir

Sú aðferð sem hér verður beitt við að rannsaka íslensk atkvæði felst í því að láta forrit hljóðrita texta og skipta honum í atkvæði. Sama forrit er svo notað til að safna tölfræði um ýmsa eiginleika atkvæða í þeim textum sem unnið er með. Við hljóðritunina var að miklu leyti stuðst við bók Eiríks Rögnvaldssonar (1989) en þó var ný hljóðritunarhefð til samræmis við IPA hljóðritunarstafrófið tekin fram yfir eldri rithátt þar sem misræmi er. Við atkvæðaskiptingu var alfarið byggt á bók Kristjáns Árnasonar (2005:180-222) og henni fylgt eins nákvæmlega og unnt var. Forritið var skrifað í C# forritunarmálinu og er rúmlega 1100 línur af kóða.

2.1.Hljóðritun

Þróun hugbúnaðar sem hljóðritar íslenskan texta er ekki með öllu vandalaus. Einn vandi sem kemur strax upp er hvernig taka skuli tillit til ólíkra eiginleika einstakra orða og samfellds máls. Ákvörðun var tekin um að hljóðrita hvert orð án tillits til umhverfis og stilla brotfullum og samlögnum í hóf. Slík nálgun er mun einfaldari en ef litið væri á textann sem samfellt mál og þannig má forðast snúna ákvarðanatöku vegna fjölbreytilegra tilbrigða í tali er varða hlé og áherslumynstur. Þó að þar með sé ef til vill verið að lýsa ofvandaðri íslensku verða færð rök fyrir því að sá galli aðferðarinnar dragi ekki um of úr gildi niðurstaðna (2.3).

Hugsunin á bak við forritið sem hljóðritar, og ákveðið hefur verið að nefna **hljóðald**, er ekki flókin. Flestar aðgerðir snúast um að skipta einu út fyrir annað og með því að gera það í ákveðinni röð

breytist hefðbundin stafsetning smám saman í hljóðritaðan IPA streng. Meginhluti hljóðritunarinnar fer fram áður en skipt er í atkvæði en henni er svo lokið þegar atkvæðaskipting er ljós.

Þegar inntak hefur verið hreinsað af greinarmerkjum og öðrum táknum sem ekki er hægt að hljóðrita er byrjað á að skipta út táknum. Þannig verður til dæmis ‘é’ að ‘je’, ‘hv’ að ‘kv’ og grönnum sérhljóðum er skipt út fyrir breiða á undan ng og nk eftir því sem við á. Þá eru ýmis brottföll framkvæmd og reynt að miða við almennan framburð en valkvæðum brottföllum er iðulega sleppt. Þegar farið hefur verið í gegnum öll helstu einföld atriði hljóðritunar er reynt að giska á vandasamari atriði út frá umhverfi en erfiðast reyndist að túlka bókstafinn ‘g’ og velja á milli hljóðanna ‘f’ og ‘v’. Hljóðaldið giskar þó á rétt hljóð í allflestum tilvikum.

Á ákveðnu stigi í hljóðrituninni er reiknirit sem skiptir í atkvæði keyrt (2.2) en að því loknu er aðblástur settur inn með tilliti til atkvæðagerðar, þ.e. á hörðum hljóðum á eftir stuttu sérhljóði (Kristján Árnason 2005:207). Þegar hljóðritun er lokið er hljóðstrengurinn lagaður til og settur í gegnum lítið undirforrit sem varpar þeim hljóðkóðum sem forritið notast við yfir í samsvarandi IPA hljóðritunartákn.

Úttak hljóðaldsins er að mestu til samræmis við hefðbundna íslenska hljóðritun að því undanskildu að lengd hljóða er hvergi táknuð sérstaklega. Atkvæðaskiptingin merkir kjarna atkvæða og í því felst hvar löng sérhljóð geta komið fram en þar sem lengd sérhljóða er háð áherslu og er ekki merkingargreinandi í nútímaíslensku er látið þar við sitja. Tvírituð samhljóð eru látin haldast sem slík enda eðlilegra að lýsa þeim sem tvöföldum samhljóðum en löngum, ekki síst þegar fengist er við atkvæðagerð, því þau skipta sér milli atkvæða eins og um tvö hljóð væri að ræða (Kristján Árnason 2005:196).

2.2. Atkvæðaskipting

Þegar hljóðaldið (2.1) kallar á forritið sem skiptir í atkvæði, nefnum það **atkveðald**, er strengurinn sem unnið er með kominn langleiðina að því marki að vera IPA jafngildi þess inntaks sem forritið fékk frá notanda. Atkvæðaskipting og greining atkvæða í stofnhluta fer fram í nokkrum skrefum.

Atkveðaldið vinnur með eitt orð í einu og byrjar á að búta það niður þannig að hvert ‘atkvæði’ hefjist á sérhljóða að frátölu því fyrsta sem fær sjálfkrafa upphafssamhljóða orðsins sem stuðul, til dæmis *hest.ur*, *fótb.olt.i*, *esj.a*, *sebr.ah.est.ur*. Þessi skipting er augljóslega gölluð en næsta skref forritsins er að framfylgja þeirri reglu að stuðull hafi forgang í íslenskum atkvæðum (1.3). Fyrir

hvert sérhljóð sem nú stendur í upphafi atkvæðis athugar forritið hvort stuðull, þ.e. samhljóði, sé til staðar í atkvæðinu á undan og flytur hann yfir til sérhljóðsins ef svo er. Atkvæðaskiptingin verður þá *hes.tur, fót.bol.ti, es.ja, seb.ra.hes.tur*. Nokkuð hefur unnist með þessu, hestur og fótbolti skiptast nú rétt, en nú þarf að taka tillit til þess að samhljóðaklasarnir *s, p, t, k + v, j, r* auk *br, fr* mynda ætíð saman stuðul. Þetta eru klasarnir sem eru undantekning frá lengdarreglu íslenskra sérhljóða en íslensk sérhljóð eru löng í áhersluatkvæðum ef á eftir fer ekki fleiri en eitt samhljóð, eða einn þessara samhljóðaklasa. Atkvæðaskiptingin verður því á endanum *e.sja* og *se.bra.hes.tur*.

Þegar atkvæðaskiptingu er lokið merkir atkveðaldið kjarna hvers atkvæðis. Í fleirkvæðum orðum er reglan einfaldlega sú að ef eitt eða fleiri samhljóð standa hægra megin við sérhljóð í atkvæði gengur það samhljóð sem næst stendur sérhljóðinu inn í kjarna. Í einkvæðum orðum er aftur á móti gert ráð fyrir að lokasamhljóð standi út af þegar skipað er í atkvæði og því tekur sérhljóð einkvæðs orðs aðeins samhljóð með sér inn í kjarna ef orðið endar á samhljóðaklasa. Með þessu móti má, eins og áður hefur verið rætt (1.3), setja fram algilda reglu um að löng sérhljóð komi fram í sérhljóðakjörnum atkvæða sem bera áherslu.

Hér fyrir neðan má sjá dæmi um inntak forrits og úttak eftir að hljóðald og atkveðald hafa lokið störfum. Bandstrik skiptir milli atkvæða og kjarni hvers atkvæðis er innan sviga. Fleiri dæmi um úttak forrits eru gefin í viðauka.

Inntak:

Margir sebrahestar sáust á vappi í Esjunni þennan dag og veltu margir fyrir sér hvort þeir hefðu flogið frá Afríku og samhliða undrun sinni var mönnum sérstaklega hugleikið að vita hvort þeim hefðu vaxið vængir sem þeyttu þeim í gegnum norræna neþjuna á leiðinni og allt norður til Íslands.

Úttak:

m(ar)-c(ɪʀ)	v(ɛɿ)-t(ɣ)	(ɔ)x	k ^h v(ɔʀ)t
s(ɛ)-pr(a)-h(es)-t(ar)	m(ar)-c(ɪʀ)	s(am)-ɿ(ɪ)-ð(a)	θ(ei)ṁ
s(au)-(ɣs)t	f(ɪ)-r(ɪʀ)	(ɣn)t-r(ɣṁ)	h(ɛv)-ð(ɣ)
(au)	sj(ɛ)ʀ	s(ɪm)-n(ɪ)	v(ax)-s(tθ)
v(ah)-p(ɪ)	k ^h v(ɔʀ)t	v(a)ʀ	v(aijɪ)-c(ɪʀ)
(i)	θ(ei)ʀ	m(œn)-n(ɣṁ)	s(ɛ)ṁ
(ɛ)-sj(ɣn)-n(ɪ)	h(ɛv)-ð(ɣ)	sj(ɛʀ)s-t(ah)k-l(ɛ)-ɣ(a)	θ(eih)-t(ɣ)
θ(en)-n(an)	fl(ɔi)-j(tθ)	h(ɣk)-l(ei)-c(tθ)	θ(ei)ṁ
t(a)x	fr(au)	(a)θ	(i)
(ɔ)x	(a)-fr(i)-k(ɣ)	v(ɪ)-t(a)	c(ek)-n(ɣṁ)

n(ɔr)-r(ai)-n(a)	l(ei)-ð(im)-n(i)	n(ɔr)-ð(ɣr)
n(ɛ)-pj(ɣ)-n(a)	(ɔ)x	t ^h (i)!
(au)	(at)!t	(is)t-l(an)ts

2.3. Takmarkanir

Þegar metið er hve gagnlegar ofangreindar aðferðir eru við rannsóknir á íslenskum atkvæðum er ýmislegt sem hafa verður í huga. Áður var nefnt að forritið tekur tillit til sumra en ekki allra brottfalla og samlagana sem teljast eðlilegar í málinu. Þegar áhrif þessa eru metin verður að hafa í huga eðli þeirra rannsókna þar sem forritið á að nýtast. Þessar rannsóknir snúast að mestu um að reikna út hlutfallslega tíðni ýmissa mynstra í löngum textum, til dæmis eru um 70.000 lesmálsorð notuð til grundvallar athugunum í þriðja kafla, og flest þeirra eru eðli málsins samkvæmt algeng orð í málinu. Ef við hugsum okkur að við viljum hrekja ágæti forritsins á þeim forsendum að það sniðgangi hið allt að því skyldubundna brottfall að ‘gnt’ verði ‘nt’ í orðum eins *hrygnt* þá er gagnrýnin rökleysa ef ekki er sýnt fram á að þessi galli spilli að einhverju marki þeim hlutföllum og tíðnitölum sem verið er að rannsaka. Reyndin er sú að strengurinn ‘gnt’ kemur aðeins þrisvar sinnum fyrir í þeim 70.000 lesmálsorðum sem notuð voru í athuganir í þriðja kafla sem gefur í versta falli óvissu á fimmta aukastaf. Afleiðing af þessu er að ekki þótti taka því að eltast við öll sjaldgæf brottföll. Sum voru þó tekin með í kjölfar þess að þau komu stöku sinnum fyrir í prófunum á hugbúnaðinum. Við þessar hugleiðingar má bæta að oft er engin augljós leið til að ákvarða hvort tiltekið brottfall sé skyldubundið eða ekki. Þá er einfaldara að láta hljóðin standa frekar en að eltast við framburð sem jafnvel er á mörkum þess að teljast óskýrmæli.

Á atkvæðaskiptingunni er sú takmörkun að í fáeinum tilvikum getur verið vafi um að rétt sé skipt á mörkum orðhluta í samsettum orðum. Þetta á helst við þegar tveir samhljóðaklasar mætast en þá kann að hljóma eðlilegra að skipta samkvæmt orðhlutagerð en að láta öll samhljóð nema það síðasta fara í hala fyrra atkvæðis. Öfgafyllsta dæmi sem hægt er að taka um þetta í íslensku er orð eins *vatnsskrjóður* sem er raunar tilbúningur en gæti þó verið löglegt orð yfir ‘gamlan bát’ eða eitthvað slíkt. Þarna kemur fram klasi átta samhljóða [htnsskrj] og forritið myndi skipta því:

v(ah) t n s s k r - j (ou) - ð(ɣr.)

Líklega er þetta fráleit atkvæðaskipting, ef við gefum okkur að öll samhljóðin séu í raun borin fram, en það þýðir ekki að atkvæðaskipting fylgi alltaf orðhlutagerð. Þvert á móti virðist oft sem hljóð færast yfir í atkvæði annars orðhluta ef hljóðkerfislegar aðstæður eru fyrir hendi. Kristján Árnason tekur ekki afstöðu til þessa en heldur báðum möguleikum opnum (2005:219). Þetta vandamál fellur

raunar í svipaðan flokk og hitt sem áður var nefnt að því leyti að tveir stórir samliggjandi samhljóðaklasar, sem saman mynda annan enn stærri, eru sjaldgæft fyrirbæri í íslensku. Við prófanir á forritinu fór lítið sem ekkert fyrir vanda af þessu tagi.

Þar sem hvert orð er afgreitt sem stök eining óháð öðrum var meðal annars gert ráð fyrir að afröddun í lok segðar geti átt sér stað í hvaða orði sem er. Þetta er ekki til samræmis við samfelld talmál en það hvenær slík afröddun verður og hvenær ekki í talmáli lýtur duttlungum tilviljana og einfaldast þótti að samræma hljóðritunina með þessum hætti. Þetta ætti ekki að koma að verulegri sök ef tíðni raddaðra og óraddaðra hljóða er skoðuð með vitneskju um þetta í huga.

Eins og fram kemur í kafla um vélræna hljóðritun (2.1) eru einhver brögð að því að forritið rugli saman hljóðum sem ‘g’ getur staðið fyrir og eins kemur fyrir að giskað sé á rangt hljóð þegar ‘f’ eða ‘v’ kemur fyrir í innstöðu. Niðurstöður verður einnig að meta með þetta í huga.

Á heildina lítið er hægt að halda því fram að íslensk atkvæði henti ágætlega til tölfræðilegra rannsóka því ekki er flókið að skrifa hugbúnað sem nær sæmilega nákvæmum upplýsingum um atkvæði úr texta. Best væri þá líklega að vera með texta sem er skrifaður upp eftir talmáli en þó er líklegt að helstu tilhneigingar málsins komi fram þó unnið sé með ritmál. Þegar rætt er um atkvæði ritmáls hér á eftir er átt við atkvæði sem kæmu fram í upplestri sem endurspeglar hljóðritun hljóðaldsins.

3. Athuganir

Til að svara því hvort vélræn nálgun er gagnleg við rannsóknir á íslenskum atkvæðum liggur beinast við að prófa að keyra niðurstöður út úr forritinu sem annar kafli lýsir og skoða útkomuna. Notast var við þrjá skáldaða texta, sem hver er rúmlega 20 þúsund lesmálsorð. Í heild voru því um 70 þúsund lesmálsorð hljóðrituð og þeim skipt í atkvæði. Höfundur lagði sjálfur til einn frumsaminn texta en hinir voru fengnir hjá Netútgáfunni, *Borgir* Jóns Trausta og *Völsunga saga*. Þrátt fyrir að vera allir skáldaðir eru þessir textar ólíkir að uppruna og ætla mætti að ef efnistösk texta hefðu mikil áhrif á tölfræði atkvæða hans kæmu fram einhverjar vísbendingar um slíkt.

3.1. Tíðni mismunandi atkvæðagerða í íslensku

Fyrri athugunin af tveimur sem hér verða kynntar er samantekt á tölfræði um atkvæðagerð íslensku, það er hve alengar atkvæðagerðir eins og CV, CVC og CCV eru í málinu. Sams konar athuganir hafa verið gerðar á öðrum tungumálum og er áhugavert að bera íslenskuna saman við slíkar

niðurstöður. Vegna þeirrar hugmyndar að lokasérhljóð einkvæðra orða standi út af í íslenskum atkvæðum verða tölurnar fyrst birtar með tvennu mismunandi móti. Annars vegar þar sem öll samhljóð eru tekin með og hins vegar þar sem lokasérhljóðum einkvæðra orða er sleppt.

Í málum heimsins eru opin atkvæði algengari en lokuð og yfirleitt er CV algengasta atkvæðagerð hvers máls. Áður hefur verið nefnt að til eru tungumál sem hafa nánast aðeins CV atkvæði en einnig má nefna að í frönsku eru um 80% atkvæða opin, þar af 70% á forminu CV (Adda-Decker o.fl. 2002) og í sænsku eru um 60% atkvæða opin, þar af 40% af taginu CV (Bannert 1998).

Í *töflu 1* má sjá tíðni 10 algengustu atkvæðagerða í þeim þremur textum sem teknir voru til athugunar. Aðrar atkvæðagerðir komu sjaldnar en 100 sinnum fyrir og var sleppt en það skýrir hvers vegna samanlagðar prósentur eru lægri en hundrað. Hér er rétt að benda á að engu breytir fyrir þessar tölur hvort forritið velur rangt hljóð fyrir bókstafinn ‘g’ eða ruglast á ‘f’ og ‘v’. Að því leyti er um að ræða marktæka talningu þó að spurningum um ólíka atkvæðagerð talmáls og ritmáls sé ósvarað.

Athygli vekur að fylgni virðist milli aldurs textanna og hlutfalls milli opinna og lokaðra atkvæða. Þetta gæti verið tilviljun þar sem textarnir eru fáir en þó vakna spurningar um hvort hluti af þróun málsins yfir lengri tímabil sé að fjölga opnum atkvæðum á kostnað lokaðra. Það virðist ekki fráleitt því eins og minnst hefur verið á eru dæmi um herra hlutfall opinna atkvæða í öðrum germönskum málum. Hugsanlega er þetta þó ekki rétt sýn á atkvæðagerð íslensku eins og næst verður vikið að.

Anton		
Opin atkvæði		
CV	11063	32,7%
V	1747	5,2%
CCV	1246	3,7%
Samtals:	41,5%	
Lokuð atkvæði		
CVC	10453	30,9%
VC	3602	10,6%
CVCC	2806	8,3%
CCVC	1206	3,6%
VCC	657	1,9%
CCVCC	426	1,3%
CVCCC	310	0,9%
Samtals:	57,4%	

Borgir		
Opin atkvæði		
CV	11499	31,2%
V	1718	4,7%
CCV	1392	3,8%
Samtals:	39,6%	
Lokuð atkvæði		
CVC	11425	31,0%
VC	4321	11,7%
CVCC	3448	9,3%
CCVC	1468	4,0%
VCC	585	1,6%
CCVCC	430	1,2%
CVCCC	329	0,9%
Samtals:	59,7%	

Völsunga saga		
Opin atkvæði		
CV	12291	30,3%
V	1556	3,8%
CCV	1476	3,6%
Samtals:	37,8%	
Lokuð atkvæði		
CVC	13021	32,1%
VC	5445	13,4%
CVCC	3422	8,4%
CCVC	1836	4,5%
VCC	638	1,6%
CCVCC	376	0,9%
CVCCC	215	0,5%
Samtals:	61,5%	

Tafla 1: Tíðni atkvæðagerða í þremur textum

Í töflu 2 hefur forsendum talningar verið breytt með því móti að lokasamhljóð einkvæðra orða eru látin standa út af þegar skipt er í atkvæði, þ.e. þeim er sleppt, samkvæmt lýsingu Kristjáns Árnasonar (2005:180-222). Áhrif þessa á lengdarreglu íslenskra sérhljóða hafa áður verið rædd en nú kemur í ljós að önnur afleiðing er sú að hlutfall opinna og lokaðra atkvæða verður nú sambærilegt við til dæmis sænsku. Hér verður ekki lagt mat á hvort hér er um tilviljun að ræða eða hvort þetta rennir stoðum undir að útafstæð samhljóð eigi rétt á sér á þeim forsendum að eðlilegt sé að náskyld tungumál eins og íslenska og sænska hafi svipað hlutfall opinna og lokaðra atkvæða. Hér væri frekari samanburður íslensku við önnur tungumál áhugaverður.

Anton				Borgir				Völsunga saga			
Opin atkvæði				Opin atkvæði				Opin atkvæði			
CV	13587	40,1%		CV	14350	38,9%		CV	15634	38,5%	
V	4149	12,2%		V	4579	12,4%		V	6004	14,8%	
CCV	1653	4,9%		CCV	1899	5,1%		CCV	2346	5,8%	
Samtals:			57,2%	Samtals:			56,5%	Samtals:			59,1%
Lokuð atkvæði				Lokuð atkvæði				Lokuð atkvæði			
CVC	8814	26,0%		CVC	10051	27,3%		CVC	11688	28,8%	
CVCC	2037	6,0%		CVCC	2071	5,6%		CVCC	1562	3,8%	
VC	1520	4,5%		VC	1803	4,9%		VC	1467	3,6%	
CCVC	1002	3,0%		CCVC	1204	3,3%		CCVC	1280	3,2%	
VCC	386	1,1%		VCC	256	0,7%		VCC	233	0,6%	
CCVCC	250	0,7%		CVCCC	229	0,6%		CCVCC	123	0,3%	
CVCCC	207	0,6%		CCVCC	211	0,6%		CVCCC	75	0,2%	
Samtals:			42,0%	Samtals:			42,9%	Samtals:			40,5%

Tafla 2: Tíðni atkvæðagerða í þremur textum miðað við útafstætt lokasamhljóð einkvæðra orða

Ekki er hægt að yfirgefa þessa umræðu án þess að víkja að samanburði textanna sem skoðaðir voru. Meginniðurstaða úr þeim samanburði hlýtur að vera sú að afar lítill munur er á textunum. Þetta má túlka sem vísbendingu um að tíðnitölur af þessu tagi séu að miklu leyti óháðar efnistöfum texta. Jafnvel óvanaleg frávik í textunum sjálfum, eins og það að tíðni samtengingarinnar *og* er ótrúlega há í *Völsunga sögu* (er rúmlega 7% allra lesmálsorða), virðast ekki hafa veruleg áhrif á heildarmyndina. Lítil tilbrigði í niðurstöðum milli textanna þriggja renna stoðum undir að rannsóknaraðferðin gefi marktækar niðurstöður. Jafnvel má leiða að því líkur að ef textar með jafnólík efnistöfum og orðaval og þessir gefa sömu niðurstöðu þá geti svipað átt við um talmál. Slíkt er þó ekki hægt að fullyrða án samanburðar við tölfræði um atkvæði í íslensku talmáli.

3.2. Dreifing samhljóða á stuðul, kjarna og hala

Áður hefur verið minnst á það viðhorf að lýsing á hljóðskipun sem einblínir á algildar hömlur kunni að horfa fram hjá markverðum upplýsingum um það tungumál sem fengist er við (Kessler 1997). Í *töflu 3* eru upplýsingar um dreifingu íslenskra samhljóða milli stuðuls, kjarna og hala samkvæmt vélrænni talningu. Allir textarnir úr fyrri tilraun eru með í þessum tölum en hverfandi munur var á niðurstöðum fyrir einstaka texta.

Dæmi um lýsingu á hljóðskipun í íslensku er sú vel þekkta staðreynd að fráblásin lokhljóð koma aðeins fyrir í framstöðu hjá meirihluta málhafa. Eins liggur fyrir að ófráblásin lokhljóð geta komið fyrir í framstöðu, innstöðu og bakstöðu og að í stórum dráttum sé dreifing fráblásinna lokhljóða annars vegar og ófráblásinna hins vegar sú sama (Eiríkur Rögnvaldsson 1989:28-29). Í umfjöllun um að sterk hljóð hafi meiri tilhneigingu en veik til að standa fjarri sérhljóðum í atkvæðum eru íslensk lokhljóð einnig sett undir sameiginlegan hatt (Kristján Árnason 2005:215).

Af ofangreindu virðist svo vera að enginn sérstakur munur sé til dæmis á dreifingu ófráblásnu hljóðanna [p] og [t]. Sú er líka raunin ef aðeins er horft á hvað sé leyfilegt og hvað óleyfilegt í málinu. Sýnin er þó önnur þegar *tafla 3* er skoðuð því þar virðist talsverð tilhneiging til að [p] standi í stuðli atkvæðis (76% tilvika) og sjaldgæft að hljóðið standi í hala (12% tilvika), en á sama tíma er eins og [t] dreifist mun jafnar á milli stuðla og hala (55% á móti 35%). Ekki virðist þó vera mikill munur á tilhneigingu hljóðanna til að mynda samhljóðakjarna. Af þessu að dæma lítur út fyrir að í raun sé nokkur munur á dreifingu [p] og [t] þó að hljóðin geti komið fyrir í sams konar umhverfi. Þá vakna spurningar um hvort upplýsingar af þessu tagi eigi heima í umfjöllun um íslenska tungu eða hvort rétt sé að sniðganga þær á þeim forsendum að málfræðireglur eigi fást við hið algilda? Slíkum spurningum verður hver málfræðingur að svara fyrir sig.

Nokkur dæmi eru um óeðlilega stöðu hljóða í töflunni, svo sem að ‘h’ og ‘j’ komi fyrir í hala. Þetta skýrist af því að í gögnunum koma fyrir fáeinir óvanalegar orðmyndir sem ýmist eru af erlendum uppruna eða eru ekki formlegar orðabókarflettur. Þannig ber persónan Natasha úr yngsta textanum ábyrgð á flestum eða öllum dæmum þess að ‘h’ sé talið til hala en hún er iðulega nefnd *Tash* af vinum sínum. Upphrópunin *nöhj* skýrir það að ‘j’ lendi einu sinni í hala. Eðlilegra þótti að hafa þessi frávik með en sleppa þeim þar sem segðir af þessu tagi eru raunverulegur hluti málsins. Tilvist þeirra er í raun frekari rökstuðningur fyrir því að tíðnitölur og hlutföll eigi erindi inn í mállýsingu þegar fyrir liggur að tungumálið þróast og ryður fyrri hljóðskipunarhömlum úr vegi.

hljóð	stuðull	kjarni	hali	samtals	st %	kj%	ha%
Lokhljóð							
c	4494			4494	100,0%		
c ^h	360			360	100,0%		
k	4719	937	1389	7045	67,0%	13,3%	19,7%
k ^h	2088			2088	100,0%		
p	3244	520	503	4267	76,0%	12,2%	11,8%
p ^h	249			249	100,0%		
t	9796	1759	6251	17806	55,0%	9,9%	35,1%
t ^h	1918			1918	100,0%		
Önghljóð							
ç	415			415	100,0%		
ð	5280	945	98	6323	83,5%	14,9%	1,5%
f	3329	849	980	5158	64,5%	16,5%	19,0%
γ	1451	348	7	1806	80,3%	19,3%	0,4%
h	6424	4561	33	11018	58,3%	41,4%	0,3%
j	4318	4	1	4323	99,9%	0,1%	0,0%
s	9463	3343	2708	15514	61,0%	21,5%	17,5%
v	9431	663	102	10196	92,5%	6,5%	1,0%
x		634	4450	5084		12,5%	87,5%
θ	5718	1721	5770	13209	43,3%	13,0%	43,7%
Ne fhljóð							
m	5529	811	38	6378	86,7%	12,7%	0,6%
m̥		2766	1888	4654		59,4%	40,6%
n	8244	8164	379	16787	49,1%	48,6%	2,3%
n̥	54	1593	5186	6833	0,8%	23,3%	75,9%
ɲ		490		490		100,0%	
ɲ̥		7		7		100,0%	
ŋ		1381		1381		100,0%	
ŋ̥		120		120		100,0%	
Hliðarhljóð							
l	6157	2103	230	8490	72,5%	24,8%	2,7%
l̥	272	762	1432	2466	11,0%	30,9%	58,1%
Sveifluhljóð							
r	8667	4172	33	12872	67,3%	32,4%	0,3%
r̥	240	8237	4424	12901	1,9%	63,8%	34,3%
Samtals	101860	46890	35902	184652	55,2%	25,4%	19,4%

Tafla 3: Dreifing íslenskra samhljóða á stuðul, kjarna og hala

Ekki verður lagst yfir frekari túlkanir á dreifingu íslenskra samhljóða en *tafla 3* þess í stað látin tala fyrir sig sjálf. Fyrirvari er settur um að eins og áður var nefnt er mest hættá á að forritið mistúlki ‘g’, ‘f’ og ‘v’ en þó án þess að niðurstöður eigi að bera þess mikinn keim. Tölur um tannvaramælt önghljóð eru líklega bjagaðar með því móti að tíðni ‘v’ er lítillega ofmetin á kostnað ‘f’.

Lokaorð

Markmiðið með þróun og prófunum á hugbúnaði sem hljóðritar og skiptir í atkvæði var annars vegar að velta upp tæknilegum hliðum slíks forrits og hins vegar að meta möguleika og takmarkanir aðferða af þessu tagi til málvísindalegra rannsókna á íslenskum atkvæðum.

Prýðilega gekk að útfæra lausnina tæknilega og má segja að forritið, sem smíðað var á tiltölulega stuttum tíma, sinni sínum verkefnum með nægilegri nákvæmni til að teljast hagnýtt verkfæri. Minniháttar villur sem koma stöku sinnum upp í tengslum við einstök máhljóð eru galli en þó ekki svo stór að hann spilli notagildi forritsins að neinu marki.

Rök hafa verið færð fyrir því að tölvumálvísindi af þessu tagi geti lagt markverðar upplýsingar til lýsingar tungumálsins á sviði eins og atkvæðarannsóknum og þá ekki síst hvað varðar atriði sem kalla á tíðnitölur og hlutföll sem ekki er gott að lýsa með algildum lögmálum.

Niðurstöður úr athugunum sem gerðar voru (3.1 og 3.2) eru einar sér ekki ýkja merkilegar en þjóna sem dæmi um möguleika þessarar rannsóknaraðferðar. Sambærilega aðferð mætti nota á ýmsum sviðum þar sem atkvæði koma við sögu (sbr. 1.4). Nefna má dæmi eins og samanburð íslensku við önnur germönsk mál í nútíð og fortíð og eins mætti nota svipaða aðferðafræði og Kessler (1997) til að varpa ljósi á hvort stuðull-rím líkanið á við um íslenska atkvæðið. Einnig má sjá fyrir sér að almennar tíðnirannsóknir á íslenskum atkvæðum kæmu að góðum notum til að stýra áherslum við þróun íslensks formendatalgervils.

Lokaniðurstaðan er því sú að vélræn nálgun sé áhugaverð og gagnleg til rannsókna á íslenska atkvæðinu. Um leið má þó segja að þessi niðurstaða varpi fram fleiri spurningum en hún svarar því að þó ýmsir möguleikar séu fyrir hendi þarf reynslan að skera úr um hve vel aðferðin fellur að ólíkum viðfangsefnum – hvort sem er í ýmsum geirum málvísinda eða á hagnýtum vettvangi.

Viðauki: Dæmi um úttak forrits

Hér verða til viðbótar gefin nokkur dæmi um hvernig texti lítur út eftir að hafa farið í gegnum hljóðald (2.1) og atkveðald (2.2).

Inntak (úr Borgum Jóns Trausta):

Það er hart fyrir ungan aðstoðarprest að sitja yfir nýsaminni ræðu um réttlæti Fariseanna og fá engan til að hlusta á hana. Meðhjálparann taldi hann ekki. Hann var landseti Torfa gamla og bjó í Grundarseli, þar lengra inni í dalnum. Hann var stöðugur heimagangur á prestssetrinu, og nú var hann kominn inn í baðstofuna til vinnufólksins og sat þar með skyrask á hnjánum. Ekki einu sinni forsöngvarinn ómakaði sig til kirkjunnar þennan dag.

Úttak:

θ(a)θ	h(a)-n(a)	st(œ)-ð(ʏ)-ʏ(ʏ)
(ε)ɾ	m(εð)-ç(au)̥-p(a)-r(an)̥	h(ei)-m(a)-ʏ(au)̥-k(ʏ)
h(aɾ)t	t ^h (al)-t(i)	(au)
f(i)-r(ɾ)	h(an)̥	p ^h r(εs)ts-s(ε)-tr(i)-n(ʏ)
(u)̥-k(a)̥	(εh)-c(i)	(ɔ)x
(að)s-t(ɔ)-ð(aɾ)-pr(εs)t	h(an)̥	n(u)
(a)θ	v(a)ɾ	v(a)ɾ
s(i)-tj(a)	l(an)t-s(ε)-t(i)	h(an)̥
(i)-v(ɾ)	t ^h (ɔr)-v(a)	k ^h (ɔ)-m(m)̥
n(i)-s(a)-m(m)-n(i)	k(am)-l(a)	(m)̥
r(ai)-ð(ʏ)	(ɔ)x	(i)
(ʏ)̥	pj(ou)	p(að)s-t(ɔ)-v(ʏ)-n(a)
rj(εh)t-l(ai)-t(i)	(i)	t ^h (i)̥
f(a)-r(i)-s(ε)-(an)-n(a)	kr(ʏn)-t(aɾ)-s(ε)-l(i)	v(m)-n(ʏ)-v(ou)l-s(m)s
(ɔ)x	θ(a)ɾ	(ɔ)x
f(au)	l(ei)̥k-r(a)	s(a)t
(ei)̥-k(a)̥	(m)-n(i)	θ(a)ɾ
t ^h (i)̥	(i)	m(ε)θ
(a)θ	t(al)-n(ʏ)̥	sc(i)-r(as)k
l̥(ʏs)-t(a)	h(an)̥	(au)
(au)	v(a)ɾ	̥j(au)-n(ʏ)̥

(eh)-c(i)	(ou)-m(a)-k(a)-ð(i)	θ(εn)-n(a _g)
(ei)-n(γ)	s(i)x	t(a)x
s(m)-n(i)	t ^h (i) _l	
f(ɔ _r)-s(æi _η)k-v(a)-r(m) _η	c ^h (i _r)-c(vn)-n(a _r)	

Hljóðritunarverkefni

Eftirfarandi er dæmi um úttak þar sem inntak er hljóðritunarverkefni úr námskeiðinu Hljóðfræði og hljóðkerfisfræði við Háskóla Íslands. Hér reynir á afar fjölbreytileg atriði í hljóðritun.

(ap)l	h(eim̥)-t(a)	p ^h (at)-l(i)
(af)-t(ʏr̥)	h(εm̥)-p(a)	p ^h (at)-l(ʏr̥)
p(au̯j̥)-c(i)	ç(au̯)-p(a)	r(au̯)-l(a)
p(ar)tn̥	ç(ou)-l(a)	s(a)̯t
p(as)t̥l̥	l̥(au)-t(ʏr̥)	s(en)-t(a)
p(ah)t-n(a)	l̥(εs)-s(a)	s(m)-n(a)
p(ɪr̥)-c(i)	l̥(u̯j̥)-k(ʏr̥)	sc(eij̥)-c(a)
pl(iθ̥)-k(a)	ŋ̥(ε)-v(i)	sc(ɪf)s
pl(ai)-j(a)	r̥(ei)-v(in̥)k	sc(ɪf)-t(a)
p(u̯j̥)-c(i)	r̥(eih)k-n(ɪr̥)	sk(ɔh)-k(a)
p(is)tn̥	r̥(uh)t-n(ʏm̥)	sk(ɔ)̯-t(ʏr̥)
t(ap)-n(a)	k ^h v(ɔf)-t(ʏr̥)	sk(ɔr̥)-p(a)
t(ʏ)-ʏ(a)	h(œp)ŋ̥	stl(œi)-v(a)
(εf)s-t(ʏr̥)	j(au̯j̥)-k(a)	stl(au)
(εh)-c(a)	j(au)-t(a)	stl(ɔh)-p(ʏr̥)
f(an̥)-t(ʏr̥)	k ^h (ap)-n(a)	sm(au)r̥
f(ar̥)-s(i)	k ^h (at)-l(a)	stn(ar)t̥l̥
f(ax)s	k ^h (a)̯t	stnj(ou)r̥
f(au̯l̥)-c(i)	c ^h (εl)-t(a)	stn(u)-(a)
f(ɪx)-t(a)	c ^h (as)-s(a)	stn(ɪr̥)-t(a)
f(it)ŋ̥	k ^h l(ɪf)-t(i)	s(ou)-v(i)
fl(ɔx)s	k ^h r(ai)-j(i)	sp(εk)-l(a)
k(au̯j̥)s	k ^h r(εf)-t(i)	sp(ɪt)-l(a)
k(ar)-k(a)	c ^h (ai)-v(a)	sp(ɔh)-t(i)
k(ar)tn̥	c ^h (ait)ŋ̥	spr(εh)-t(a)
c(ε)-v(a)	l(au̯j̥)t	st(am̥)-p(ʏr̥)
c(εk)ŋ̥	l(eijn̥)-c(i)	st(εr̥)t
c(εk)-n(a)	l(ou)-v(i)	st(εr̥)-k(ʏr̥)
k(ɔr̥)-t(a)	l(ɪx)-t(a)	st(eif)-t(i)
c(i)s	m(ar̥)t	st(ɪ)̯-k(ʏm̥)
c(ai)-j(as)t	m(ɪj̥)-k(a)	stj(our)t-n(a)
h(af)s	n(ei)-t(a)	st(ɔh)-p(a)
h(ar̥)-k(a)	n(ai)-j(a)	strj(u)-k(a)

str(œiŋ)k	t ^h (ɔ)-ʏ(a)	θj(out)ŋ
st(œx)-t(ʏ)	t ^h (ɔh)-p(ʏŋ)	θ(ɔh)-c(i)
sv(auŋ)-k(ʏŋ)	t ^h r(aθ)-k(a)	θv(eiŋ)s
sv(ar)̄t	v(ar)t-l(a)	θ(œk)̄ŋ
sv(eit)̄ŋ	v(ah)tns	(ais)-k(a)
sv(ai)-v(a)	v(eŋ)st-l(a)	(aih)t-l(a)
s(iŋ)-c(a)	v(i)-t(a)	(œ)-v(ʏn)t
s(i)-n(a)	v(aih)t-l(a)	(œ)-r(ʏx)t
t ^h (ah)-c(i)	(ih)t-n(i)	(œx)st̄ŋ
t ^h (auŋ̄)-k(ʏŋ)	θ(a)-ð(aŋ)	

Textar

Anton Karl Ingason. 2006. Óútgefin saga.

Jón Trausti. [Ár vantar] *Borgir*. Sótt 18. september 2006 af <<http://www.snerpa.is/net/roman/borgir.htm>> Netútgáfan.

Völsunga saga. [Ár vantar]. Sótt 18. september 2006 af <<http://www.snerpa.is/net/forn/volsung.htm>> Netútgáfan.

Heimildir

Bagemihl, Bruce. 1991. „Syllable structure in Bella Coola.“ *Linguistic Inquiry*, 22: 589–646. Tilvitnað eftir *Wikipedia*. Sótt 5. desember 2006 af <<http://en.wikipedia.org/wiki/Syllable>>

Bannert, Robert. 1998. „Two thousand and one syllables in spoken Standard Swedish: aspects of syllabification.“ Sótt 5. desember 2006 af <<http://www.ling.umu.se/~bannert/papers/Fon98.2001.syll.pdf>> [Áður birt í *PHONUM* 6, 51-81]

Eiríkur Rögnvaldsson. 1989. *Íslensk hljóðfræði*. Kennslukver handa nemendum á háskólastigi. Málvísindastofnun Háskóla Íslands, Reykjavík.

Elbert, Samuel H. og Mary Kawena Pukui. 1979. *Hawaiian Grammar*. The University Press of Hawaii, Honolulu. Tilvitnað eftir *Wikipedia*. Sótt 5. desember 2006 af <http://en.wikipedia.org/wiki/Hawaiian_language>

Kessler, Brett og Rebecca Treiman. 1997. „Syllable Structure and the Distribution of Phonemes in English Syllables.“ Sótt 5. desember 2006 af <<http://brettkessler.com/SyllStructDistPhon/CVC.html>> [Áður birt í *Journal of Memory and Language* 37, 295-311]

King, Stephenson o.fl. 1998. „Speech recognition via phonetically featured syllables.“ Sótt 5. desember 2006 af <http://www.cstr.ed.ac.uk/downloads/publications/1998/King_Stephenson_Isard_Taylor_Strachan_icslp1998.pdf>

Kristján Árnason. 2005. *Hljóð*. Ritröðin Íslensk tunga. Almenna bókafélagið, Reykjavík.

- Levelt, Clara C., Niels O. Schiller og Willem J. M. Levelt. 1999. „A Developmental Grammar for Syllable Structure in the Production of Child Language.“ Sótt 5. desember 2006 af <http://www.libc.leidenuniv.nl/schiller/content_docs/levelt_schiller_levelt_b_1_1999.pdf> [Áður birt í *Brain and Language* 68, 291-299]
- Lyovin, Anatole V. 1997. *An Introduction to the Languages of the World*. Oxford University Press, New York. Tilvitnað eftir *Wikipedia*. Sótt 5. desember 2006 af <http://en.wikipedia.org/wiki/Hawaiian_language>
- M. Adda-Decker, P. Boula de Mareüil, G. Adda, L. Lamel. 2002. „Investigating syllabic structure and its variation in speech from French radio interviews.“ Sótt 5. desember 2006 af <<http://www.clsp.jhu.edu/pmla2002/cd/papers/adda-decker.pdf>>
- Obleser, Jonas, Brigitte Rochstroh og Carsten Eulitz. 2004. „Gender differences in hemispheric asymmetry of syllable processing: Left-lateralized magnetic N100 varies with syllable categorization in females.“ Sótt 5. desember 2006 af <<http://www.blackwell-synergy.com/links/doi/10.1111/j.1469-8986.2004.00204.x/pdf>> [Áður birt í *Psychophysiology*, 41 (2004), 783–788]
- Sigríður Sigurjónsdóttir. 2002. „Hljóðþróun í máltöku barna.“ *Talfræðingurinn*, 16(1), 6-10.