



Máltækni og málstefna – íslenska innan upplýsingatækninnar

Tilfni þessarar greinar eru þrjú:¹ 10 ára afmæli íslenskrar máltækni,² öndvegissstyrkur sem Rannsóknasjóður hefur nýlega veitt til máltækni-verkefna, og samþykkt íslenskrar málstefnu þar sem m.a. er fjallað um stöðu íslensku í tölvuheiminum. Gerð er grein fyrir uppbyggingu íslenskrar máltækni undanfarinn áratug og helstu verkefnum sem unnið hefur verið að, svo og þeim sem eru á döfinni. Í lokin er fjallað um meginmarkmið og aðgerðir íslenskrar málstefnu á sviði máltækni og hugbúnaðarþýðinga, og rætt um forsendur þess að íslenska eigi sér framtíð innan upplýsingatækninnar.

1. Inngangur

Í stefnuskrá Íslenskrar málnefndar 2006–2010 segir: „Upplýsingatækni verður æ mikilvægari þáttur í daglegu lífi Íslendinga og er afar brýnt að þeir geti notað móðurmálið á þeim vettvangi. Til að mynda er áriðandi að ýmiss konar grundvallarhugbúnaður (til dæmis stýrikerfi, ritvinnsluforrit, póstforrit og vafrar) verði fáanlegur með íslensku notendaviðmóti og að í algengum ritvinnsluforritum verði hjálparbúnaður fyrir íslensku, svo sem forrit sem leiðrétta stafsetningu, málfar og skipta orðum rétt á milli lína. Þá hafa einnig orðið örar framfarir í vélrænum þýðingum á undanförunum árum og er því nauðsynlegt að íslenska verði hluti af þeirri þróun. Til þess að staða íslenskrar tungu í upplýsingatækni sé tryggð þarf að eiga sér stað kröftugt rannsóknar-, þróunar- og uppbyggingarstarf í íslenskri tungutækni“ (http://www.islenskan.is/Stefnuskra_2006-2010.htm).

Þarna er talað bæði um máltækni (tungutækni) og hugbúnaðarþýðingar. Þetta tvennt er oft tengt saman og mörkin þarna á milli eru óskýr og í hugum margra jafnvel ekki til, en í raun og veru er hér þó um tvö sjálfstæð viðfangsefni að ræða. Það er alveg hægt að hugsa sér íslenskan hugbúnað án nokkurrar máltækni – þ.e. án íslenskra leiðréttingarforrita, án vélræna þýðinga milli íslensku og annarra mála, án íslenskra talgervla og talgreina, o.s.frv. Á sama hátt má vel hugsa sér íslenska máltækni þótt allt viðmót hugbúnaðarins sé á ensku, og það er reyndar staðan hjá mörgum notendum

– algengt er að menn séu með villuleitarforrit eins og *Púkann* en séu svo með Windows á ensku.

En það er eðlilegt að almennir notendur dragi ekki skil milli máltækni og hugbúnaðarþýðinga. Hvortveggja varðar tölvur og notkun tungumáls í tengslum við þær. Þetta snýst sem sé um hugtak sem hefur verið áberandi í málræktarumræðu undanfarin ár – *um dæmi* tungumálsins, sem í þessu tilviki er þá tölvuheimurinn og upplýsingatæknin. Verður íslenska ríkjandi í þessu umdæmi í framtíðinni – og hvað þarf til að svo verði? Í þessari grein ætla ég fyrst að rekja í stuttu máli þær breytingar sem hafa orðið á þessu sviði undanfarinn áratug, og velta síðan fyrir mér möguleikum íslenskrar tungu innan upplýsingatækninnar á næstu árum.

2. Sagan

Fyrir tíu árum var staða mála íslensku innan upplýsingatækninnar ekki álitleg, hvort sem litið var til hugbúnaðarþýðinga eða máltækni. Staðan í hugbúnaðarþýðingum hafði þá raunar farið versnandi frá níunda áratugnum. Þegar frumstæð ritvinnslukerfi fóru að koma á markaðinn hér á landi upp úr 1980 voru þau undantekningarlítið á íslensku. Stýrikerfi og notendahugbúnaður Macintosh-tölva hefur verið á íslensku fram undir þetta, og einnig stóð IBM á Íslandi fyrir umfangsmiklum hugbúnaðarþýðingum í samstarfi við Orðabók Háskólans. En fyrsta alvöru ritvinnslukerfið sem var þýtt fyrir PC-tölvur var WordPerfect (*Word Perfect handbók* 1986). Sú þýðing var gerð árið 1986 og

náði um skeið yfirburðastöðu á markaðnum – án efa að verulegu leyti vegna tungumálsins.

En með útbreiðslu Windows upp úr 1990 náði Word yfirhöndinni og var ekki íslenskað. Um sama leyti fóru svo að koma tölvupóstkerfi eins og Outlook og Eudora, vafrar eins og Netscape og Mosaic, og síðar Internet Explorer. Þessi forrit voru öll á ensku, og með sprengingu í netnotkun á síðasta áratug vandist almenningur því að tölvur og allt sem þeim tengdist væri á ensku. Staðan í máltækni var síst betri – íslensk máltækni var varla til fyrir áratug. Við áttum að vísu ágætt ritvilluleitarforrit, *Púka* Friðriks Skúlasonar, og talgervil sem tæpast var boðlegur til almennrar notkunar þótt hann nýttist blindum og sjónskertum vel. En þar með var það upp talið.

Fyrir tíu árum komst svo hreyfing á málín, bæði á sviði hugbúnaðarþýðinga og máltækni, að frumkvæði Björns Bjarnasonar, þáverandi menntamálaráðherra. Sumarið 1998 hóf hann viðræður við Microsoft sem leiddu til þess að gerður var samningur menntamálaráðuneytisins og Microsoft um íslenskun á Windows 98 stýrikerfinu (sjá <http://www.bjorn.is/greinar/1999/01/20>). Sama haust, 1998, fékk Björn Rögnvald Ólafsson dósent í eðlisfræði til að gera skýrslu um ástand og horfur í íslenskri máltækni. Rögnvaldur fékk tvo menn með sér í skýrslugerðina og þessi starfshópur skilaði af sér fyrir tíu árum, vorið 1999. Í skýrslu hópsins var eftirfarandi slegið föstu:

Meginmarkmið Íslendinga hlýtur að verða að unnt verði að nota íslenska tungu, ritaða með réttum táknum, sem víðast innan tölvu- og fjarskiptatækninnar. Þar verður þó að sjálfsgöðu að sníða sér stakk eftir vexti. Það er mikið verkefni að gera íslensku gjaldgenga á öllum sviðum, við allar aðstæður. Því verður að leggja megináherslu á þá þætti sem varða daglegt líf og starf alls almennings, eða munu gera það á næstu árum (Rögnvaldur Ólafsson, Eiríkur Rögnvaldsson og Þorgeir Sigurðsson 1999:34).

Í framhaldi af þessu lagði starfshópurinn til að stjórnvöld beittu sér fyrir átaki á fjórum sviðum til eflingar íslenskri máltækni næstu fimm árin (Rögnvaldur Ólafsson, Eiríkur Rögnvaldsson og Þorgeir Sigurðsson 1999:28):

1. Byggð verði upp sameiginleg gagnasöfn, málsöfn, sem geti nýst fyrirtækjum sem hráefni í afurðir.
2. Fé verði veitt til að styrkja hagnýtar rannsóknir á sviði máltækni.
3. Fyrirtæki verði styrkt til þess að þróa afurðir máltækni.
4. Menntun á sviði máltækni og málvísinda verði efl.

Í tillögum starfshópsins var gert ráð fyrir að máltækniátakið stæði í a.m.k. fjögur ár og heildarkostnaður á ári yrði 225–250 milljónir króna (Rögnvaldur Ólafsson, Eiríkur Rögnvaldsson og Þorgeir Sigurðsson 1999:30). Eftir það

vonuðust menn til að íslensk máltækni yrði orðin sjálfbær og þarfnaðist ekki meiri opinberra framlaga.

3. Staðan

Á fjárlögum árunum 2000–2004 voru alls veittar 133 milljónir króna til máltækni (sjá <http://hamar.stjr.is/>). Það er verulegt fé, en þó aðeins u.þ.b. 1/8 þess sem starfshópurinn taldi að þyrfti til að ná tilætluðum árangri. Það er því ekki von að öllum verkefnum sem talin voru brýn í skýrslunni hafi verið gerð skil. Þó er óhætt að segja að furðu mikið hafi áunnist miðað við tilkostnað, og máltækniáætlunin hafi skilað heilmiklu. Þverfaglegt meistaranám í máltækni var í boði við Háskóla Íslands á árunum 2002–2004, en 2007 var námið endurskipulagt og er nú rekið í samvinnu Háskóla Íslands og Háskólans í Reykjavík. Nemendur í þessu námi hafa átt þess kost að taka námskeið í norræna máltækniháskólanum, Nordic Graduate School of Language Technology (NGSLT, <http://www.ngslt.org>) og hefur það verið ómetanlegt.

Einnig hafa mikilvæg gagnasöfn verið byggð upp, og ýmsum rannsóknar- og þróunarverkefnum hefur verið ýtt af stað. Ég skal nefna nokkur þau helstu – sex þau fyrsttöldu voru að talsverðu eða öllu leyti kostuð af máltækniáætlun menntamálaráðuneytisins.

- Íslenskur talgreinir (stakorðagreinir; sjá Eiríkur Rögnvaldsson 2004; Helga Waage 2004)
- Íslenskur talgervill (Ragga – vefpulan; sjá Eiríkur Rögnvaldsson, Björn Kristinsson og Sæmundur Þorsteinsson 2006)
- Endurbætt ritvilluleit (Púki; sjá Friðrik Skúlason 2004)
- Beygingarlýsing íslensks nútímamáls (hátt í 300 þúsund orð og hátt á sjöttu milljón beygingarmynda; sjá Kristín Bjarnadóttir 2004, 2005)
- Mörkuð málheild (25 milljón orða safn fjölbreyttra texta, málfræðilega greint; sjá Sigrún Helgadóttir 2004)
- Málfræðilegir markarar (forrit sem greina texta málfræðilega; sjá Sigrún Helgadóttir 2005, 2007; Hrafn Loftsson 2006, 2007)
- Leitarvél (Embla og leitarvél Já; sjá Hjálmar Gíslason 2006)
- Hlutabáttari (forrit sem greinir helstu setningarliði og setningarleg hlutverk; sjá Hrafn Loftsson og Eiríkur Rögnvaldsson 2007)
- Samhengisháð ritvilluleit (sjá Anton Karl Ingason o.fl. 2009)

Margt af þessum gagnasöfnum og búnaði nýttist nemendum og skólakerfinu beint. *Beygingarlýsingin* (<http://bin.arna stofnun.is/>) er t.d. ómetanlegt hjálpargagn í kennslu og verkefnavinnu. Talgervillinn *Ragga* nýttist blindum og sjónskertum nemendum mjög vel, og fjölmargir, m.a. lesblindir, nýta sér einnig þjónustu *Vefpulunnar* ([SKÍMA 1 • 2009](http://www.vefthulan.is/lesa-</p></div><div data-bbox=)

textu) til að lesa texta upphátt. Markarinn *IceTagger* er aðgengilegur á netinu (http://nlp.ru.is:8080/IceNLPWeb/icenlp_isl.html) og geta nemendur og aðrir nýtt sér það við málfræðilega greiningu orða – og að hluta til einnig við setningagreiningu. *Púkinn* er mjög naskur við að vinna stafsetningar- og innsláttarvillur, en hann ræður þó ekki við ýmsar algengar villur þar sem leyfileg orðmynd er rituð á röngum stað (t.d. *himin* í stað *himinn*, *leiti* í stað *leyti* o.s.frv.). Til að finna slíkar villur þarf *samhengisháða ritvilluleit*, þ.e. búnað sem skoðar samhengi orðanna í stað þess að líta bara á einstakar orðmyndir. Frumgerð slíks búnaðar hefur nú verið þróuð fyrir íslensku og verður gerð aðgengileg á netinu innan skamms (á <http://www.tungutaekni.is>).

Þrátt fyrir þetta fer því fjárfari að íslensk máltækni sé orðin sjálfbær, eins og stefnt var að – og það er engin furða. Það kostar jafnmikið að byggja upp málleg gagnasöfn og máltækniþúnað fyrir tungumál sem 300 þúsund manns tala og fyrir tungumál milljónþjóða. Það er ekki von að fyrirtæki sjái sér hag í því að leggja í mikinn kostnað við að þróa og aðlaga máltækniþúnað fyrir íslensku – sú fjárfesting skilar tæpast hagnaði. Á mörgum sviðum máltækninnar er nánast allt óunnið. Það hefur t.d. sáralítið verið gert á sviði vélrænna þýðinga, svo að nefnt sé eitt verkefni sem tilgreint er í stefnuskrá Málnefndarinnar sem vitnað var til í upphafi.

Þótt illa hafi gengið að auka áhuga fyrirtækja á gerð máltækniþúnaðar hafa rannsóknir á sviði máltækni eflst verulega undanfarin ár. Rannsóknasjóður hefur veitt nokkra styrki til máltækniverkefna, m.a. til gerðar hlutabáttara og samhengisháðrar ritvilluleitar sem nefnd eru hér að framan. Nú í ársbyrjun veitti Rannsóknasjóður svo öndvegissstyrk, u.þ.b. 14,5 milljónir á ári í þrjú ár, til nýs máltækniverkefnis sem fræðimenn og framhaldsnemar við Háskóla Íslands, Háskólann í Reykjavík og Stofnun Árna Magnússonar í íslenskum fræðum standa að. Það verkefni skiptist í þrjá þætti. Sá fyrsti er vélræn merkingargreining texta; annar verkþáttur er vélræn grófþýðing milli ensku og íslensku; og þriðji verkþátturinn felst í vélrænni setningagreiningu og gerð íslensks trjábanka, en trjábanki er texti sem hefur verið greindur setningafræðilega. Afurðir allra þessara verkþátta munu nýtast í margvíslegum búnaði, s.s. við þýðingar, leit á netinu og í gagnabönkum, vélræna málfarsleiðréttingu o.m.fl.

Á sviði hugbúnaðarþýðinga eigum við því láni að fagna að langútbreiddasta stýrikerfið og notendahugbúnaðurinn, Windows og Office, er til á íslensku. Windows XP og Office 2003 komu í íslenskri þýðingu árið 2004, og í fyrra komu Windows Vista og Office 2007 á íslensku. Þarna strandar meira á notendunum sjálfum – þýðingarnar hafa ekki fengið þá útbreiðslu sem vænta mætti. Það skiptir miklu máli að notendur venjist íslenskuðum hugbúnaði frá byrjun, og þess

vegna er höfuðatriði að hugbúnaður í skólakerfinu sé á íslensku.

Því miður virðist meginhluti grunn- og framhaldsskóla nota Windows á ensku (sjá Hulda Hreiðarsdóttir 2008), þótt ágæt íslensk útgáfa stýrikerfisins hafi verið til undanfarin ár. Þarna þarf einhverja hugarfarsbreytingu hjá forráðamönnum skólanna. Ekki er ástandið betra í Háskóla Íslands, sem þó stærir sig af því að hafa sett sér málstefnu þar sem segir m.a.: „Málstefnan hefur að leiðarljósi að talmál og ritmál Háskólans er íslenska, jafnt í kennslu, rannsóknnum sem stjórnarsýslu“ (<http://www.hi.is/page/malstefna>). Samt sem áður býður Reiknistofnun Háskólans eingöngu upp á Windows og Office á ensku í tölvuverum, og starfsmönnum er ekki boðið upp á íslenska útgáfu hugbúnaðarins að fyrra bragði.

4. Framtíðin

En hverjar eru þá framtíðarhorfur íslenskrar máltækni, og íslensku innan upplýsingatækninnar? Ég er sannfærður um að þær velta algerlega á íslenskum almenningi. Ef almennir málnotendur vilja hafa upplýsingatækni á íslensku og sætta sig ekki við annað þá verður hún á íslensku. Til þess þarf að halda áfram uppbyggingu íslenskrar máltækni, en á því eru engar tæknilegar hindranir. Hindranirnar eru fyrst og fremst fjárhagslegar, en það er samt smámál miðað við jarðgöng, virkjanir, mislæg gatnamót og annað sem við leggjum fé í (eða lögðum fé í til skamms tíma, þótt e.t.v. verði eitthvað minna um það á næstunni).

Það er ekki hægt að búast við því að stjórnálamenn eða fyrirtæki vilji leggja fé í búnað sem enginn hefur áhuga á eða vill nota. Ég held því að við þurfum á vitundarvakningu meðal almennings að halda. Við þurfum að sannfæra fólk um að það er engin ástæða til lítillþægni – upplýsingatæknin getur verið á íslensku og á að vera það. Það gerist ekkert nema almennir málnotendur vilji geta notað íslensku innan upplýsingatækninnar og sýni þann vilja í verki. Og þetta er ekki bara okkar mál – það er beinlínis skylda okkar við komandi kynslóðir að gera tungumálið gjaldgengt innan upplýsingatækninnar. Ef við missum það svið algerlega til enskunnar náum við því aldrei til baka.

Í nýrri íslenskri málstefnu, sem Alþingi samþykkti 12. mars sl., er sett fram eftirfarandi meginmarkmið um notkun íslensku innan upplýsingatækninnar (*Íslenska til alls* 2008: 49):

- Að íslensk tunga verði nothæf – og notuð – á öllum þeim sviðum innan tölvu- og upplýsingatækninnar sem varða daglegt líf alls almennings.

Þetta merkir í fyrsta lagi að viðmót algengs hugbúnaðar (valmyndir, hjálpartextar o.s.frv.) þarf að vera íslenskt; í öðru lagi að til þarf að vera ýmiss konar hugbúnaður sem liðsinnir og leiðbeinir notendum við notkun íslensks

máls (leiðréttingarforrit, þýðingarforrit, hjálparforrit fyrir fatlaða); og í þriðja lagi að unnt á að vera að nota íslensku sem samskiptamál við ýmiss konar tölvu- og tækniþúnað (upplýsingakerfi, þjónustuver, tölvustýrð tæki af ýmsu tagi) (*Íslenska til alls* 2008:49).

Í framhaldi af þessu er svo sett fram aðgerðaáætlun í niú liðum, m.a. eftirfarandi (*Íslenska til alls* 2008:50):

- Að allur almennur notendahugbúnaður í íslensku skóla-kerfi, frá leikskólum til háskóla, verði á íslensku innan þriggja ára.
- Að stöðugt verði unnið að uppbyggingu og eflingu mállegra gagnasafna sem eru forsenda fyrir þróun og smíði margs kyns máltækniþúnaðar.
- Að hugbúnaður til að lagfæra og leiðrétta íslenskt málfar verði gerður og kominn í notkun innan þriggja ára.
- Að nothæf þýðingarforrit milli íslensku og valinna erlendra mála, a.m.k. ensku, verði gerð innan fimm ára.
- Að íslenskur talgervill og talgreinir sem gerðir voru á vegum tungutækniátaks menntamálaráðuneytisins verði endurbættir og lagaðir að nýjustu tækni.

Þessi stefnumörkun er mikilsverð þótt óljóst sé um efndir og ekki sé líklegt að mikið fé verði lagt í aðgerðir á þessu sviði á næstunni. En forsenda þess að stefna af þessu tagi skili árangri er sú að hún njóti almenns stuðnings. Fátt væri verra fyrir íslenska tungu en opinber málstefna sem væri aðeins fögur innantóm orð en hefði ekki tengsl við almenna málnotendur og stuðning þeirra.

Þegar mikilvægi íslenskrar máltækni er metið verður að líta til þess að upplýsingatæknin er orðin mikilvægur þáttur í daglegu lífi alls almennings í landinu. Ef ekki verður hægt að nota íslensku innan hennar kemur upp splunkuný staða, sem ekki á sér hliðstæðu fyrr í málsögunni. Þá verður orðið til mikilvægt svið í daglegu lífi venjulegs fólks, þar sem móðurmálið er gagnslítið eða ónothæft. Hvaða áhrif hefði slíkt umdæmistap á málnotendur og málsamfélagið? Hvað gerist ef móðurmálið er ekki lengur nothæft í nýrri tækni og öðru sem er nýtt og spennandi; á sviðum þar sem nýsköpun af ýmsu tagi á sér stað; og á sviðum þar sem ný atvinnutækifæri bjóðast? Menn þurfa varla að velja þessu lengi fyrir sér til að sjá hættumerkin.

En það er ástæðulaust að meta þörf á íslenskri máltækni eingöngu út frá sjónarmiði tungumálsins og varðveislu þess. Við eigum einnig og ekki síður að líta á þetta út frá þörfum og hagsmunum okkar, almennra málnotenda. Við eigum kröfu á því að geta notað móðurmálið hvar sem er í íslensku málsamfélagi – líka innan upplýsingatækninnar. Við eigum að krefjast þess að hugbúnaðurinn sem við notum sé á íslensku, að við fáum leiðréttingarhugbúnað fyrir íslenskan texta, að við getum talað við ýmis tölvustýrð tæki á íslensku,

að við fáum þýðingarforrit sem geti þýtt milli íslensku og annarra mála, að við getum unnið flóknar upplýsingar úr íslenskum texta- og gagnasöfnum og leitað í þeim á margvíslegan hátt, o.s.frv. Við eigum þetta skilið – og íslenskan á það skilið.

Tilvísanir

1 Greinin er að hluta byggð á erindinu *Framtíð íslensku innan upplýsingatækninnar* sem haldið var á málþinginu *Á íslenska sér framtíð innan upplýsingatækninnar?* á vegum Tungutækniáseturs og Íslenskrar málnefndar í Háskólanum í Reykjavík 7. mars 2008.

2 *Máltækni* snýst um margs konar notkun tölvu við meðferð tungumálsins. Þetta svið nefnist á ensku *language technology* og hefur verið kallað *tungutækni* á íslensku undanfarin ár en það orð er á ýmsan hátt óheppilegt og margir þeirra sem vinna mest á þessu sviði hafa því tekið upp orðið *máltækni* í staðinn.

Ritaskrá

Anton Karl Ingason, Skúli Bernhard Jóhannsson, Eiríkur Rögnvaldsson, Sigrún Helgadóttir og Hrafn Loftsson. 2009. Context-Sensitive Spelling Correction and Rich Morphology. *Proceedings of NODALIDA 17*.

Eiríkur Rögnvaldsson. 2004. The Icelandic Speech Recognition Project *Hjal*. Henrik Holmboe (ritstj.), *Nordisk Sprogteknologi. Árbog 2003*, s. 239–242. Museum Tusulanums Forlag, Kaupmannahöfn.

Eiríkur Rögnvaldsson, Björn Kristinsson og Sæmundur Þorsteinsson. 2006. Nýr íslenskur þulur að koma á markað. *Morgunblaðið* 20. janúar.

Fríðrik Skúlason. 2004. Endurbætt tillögugerðar- og orðskiptiforrit Púka. *Samspil tungu og tækni*, s. 29–31. Menntamálaráðuneytið, Reykjavík.

Helga Waage. 2004. Hjal – gerð íslensks stakorðagreinis. *Samspil tungu og tækni*, s. 49–53. Menntamálaráðuneytið, Reykjavík.

Hjálmar Gíslason. 2006. Resourcebehov i informationsøgning. Erindi flutt á málstefnunni *Språkteknologisk infrastruktur i Norden*, Gautaborg, 26. október 2006. <http://www.nordisk-sprakrad.no/Bilaga%20presentationer.pdf>

Hrafn Loftsson. 2006. Tagging Icelandic Text: A Linguistic Rule-Based Approach. *Technical Report CS-06-04*, Department of Computer Science, University of Sheffield.

Hrafn Loftsson. 2007. Tagging and Parsing Icelandic Text. Doktorsritgerð, Department of Computer Science, University of Sheffield.

Hrafn Loftsson og Eiríkur Rögnvaldsson. 2007. IceParser: An Incremental Finite-State Parser for Icelandic. Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek og Mare Koit (ritstj.): *Proceedings of the 16th Nordic Conference of Computational Linguistics NODALIDA-2007*, s. 128–135. Tartu.

Hulda Hreiðarsdóttir. 2008. *Getur tungan leyst hugvit úr læðingi? Umfjöllun um áhrif notkunar á stýrikerfi í íslenskri þýðingu á tölvulæsi og notkun tölvunýrrða*. BA-ritgerð í íslenskri málfræði við Háskóla Íslands.

Íslenska til alls. 2008. Tillögur Íslenskrar málnefndar að íslenskri málstefnu. Menntamálaráðuneytið, Reykjavík.

Kristín Bjarnadóttir. 2004. Beygingarlýsing íslensks nútímamáls. *Samspil tungu og tækni*, s. 23–25. Menntamálaráðuneytið, Reykjavík.

Kristín Bjarnadóttir. 2005. Modern Icelandic Inflections. Henrik Holmboe (ritstj.), *Nordisk Sprogteknologi. Árbog 2005*, s. 49–50. Museum Tusulanums Forlag, Kaupmannahöfn.

Rögnvaldur Ólafsson, Eiríkur Rögnvaldsson og Þorgeir Sigurðsson. 1999. *Tungutækni. Skýrsla starfshóps*. Menntamálaráðuneytið, Reykjavík.

Sigrún Helgadóttir. 2004. Mörkuð íslensk málheild. *Samspil tungu og tækni*, s. 67–71. Menntamálaráðuneytið, Reykjavík.

Sigrún Helgadóttir. 2005. Testing Data-Driven Learning Algorithms for PoS Tagging of Icelandic. Henrik Holmboe (ritstj.), *Nordisk Sprogteknologi. Árbog 2004*, s. 257–265. Museum Tusulanums Forlag, Kaupmannahöfn.

Sigrún Helgadóttir. 2007. Mörkun íslensks texta. *Orð og tunga* 9:75–107.

WordPerfect handbók. 1986. Íslensk þýðing Eiríkur Rögnvaldsson lektor og Vilhjálmur Sigurjónsson kerfisfræðingur. Rafreiknir hf., Reykjavík.

■ *Höfundur er prófessor í íslenskri málfræði við Háskóla Íslands.*