

ARI ARNALDS

Tæknimenn, tungan og tæknin

Eftirfarandi erindi var flutt á málræktarþingi í Hafnarfirði 17. nóvember 2001. Auk Ara töluðu þar Kristján Árnason, Auður Hauksdóttir, Karl Blöndal og Egill Helgason. –Ritstj.

Því heyrir oft fleygt að tæknimenn og aðrir þeir sem fjalla um sérgreinar af ýmsu tagi séu gjarnir á að nota slettur í máli sínu, jafnvel svo að íslenskri tungu stafi hætta af. Tæknimenn hafa oft lært fræði sín að verulegu leyti erlendis og er því tamt að nota erlend orð eftir heimkomuna, einkum yfir hugtök sem þeir hafa ekki þekkt áður en þeir fóru til náms. Ég var um rúmlega tuttugu ára skeið framkvæmdastjóri fyrir hugbúnaðarfyrirtæki og ætla að fjalla um þá reynslu sem ég hef af málnotkun starfsmanna í hátæknifyrirtæki.

Frá því á síðastliðnu ári hef ég verið formaður verkefnisstjórnar um tungutækni en það er verkefni sem menntamálaráðuneytið hefur hrundið af stað. Meginmarkmið verkefnisins er að

- tryggja að íslenska verði áfram lifandi tungumál íslensku þjóðarinnar í upplýsinga- og þekkingarsamfélagi 21. aldarinnar
- stuðla að því að til verði nauðsynleg tungutækniverkfæri fyrir þann tæknibúnað sem notaður verði í daglegu lífi og starfi
- skapa grundvöll fyrir tungutækniiðnað á Íslandi
- skapa grundvöll fyrir útflutning á iðnaðarframleiðslu og þekkingu í tungutækni

Mun ég nota mestan hluta tíma míns til að tala um það verkefni. En áður en ég vík að meginmálunum ætla ég að tala lítillega um hvernig ég skynjaði tungumálið áður en ég varð fræðingur á einhverju sviði.

Ég er nú kominn á þann aldur að ég get leyft mér að tala um það sem gerðist í gamla daga. Mitt minni nær svona um það bil til 1950 þannig að ég get miðlað af sýn minni á þróun íslenskrar tungu í hálfu öld. Ég nýt þeirra forréttinda að vera ekki fræðimaður á þessu sviði og get því leyft mér að láta út úr mér ýmislegt sem ef til vill stenst ekki gagnrýna skoðun frá fræðilegu sjónarmiði.

Móðir mín átti ættir að rekja til Danmerkur og bæði hún og pabbi dvöldu um tíma í Danmörku. Af þessum ástæðum var talsvert um dönskuslettur á heimili mínu. Þannig var oftast talað um *paalæg* í stað *áleggs* og pabbi, sem oft fór í bíó, talaði gjarnan um að kaupa *bílæti*. Eitt sinn, þegar pabbi hafði verið að sjá sakamálamynd, komst hann svo að orði: „Ég var svo mikill detektív að ég sá strax að klóninn var bandíttinn.“ Ég sá að sjálfsögðu ekkert athugavert við þessa orðanotkun. Þess skal getið að smám saman rjátlaðist þessi slettutilhneiging af foreldrum mínum og ég held að þau hafi alls ekki talað verra mál en gengur og gerist.

Frá þessum tíma minnst ég þess einnig að við strákarnir gerðum okkur gjarnan leik að því, á leiðinni úr skólanum þegar snjór var, að kasta snjóboltum í glugga sem á leið okkar urðu.

Eitt sinn köstuðum við í glugga í húsi við Mjóuhlíðina og þá kom strax út í gluggann kona, sem við kölluðum kellingu enda þótt þetta hafi ugglauast verið bráðhugguleg ung kona, sem kallaði til okkar: „Ef þið hættið þessu ekki strax kalla ég

á pólitíið.“ Þetta fannst okkur strákunum alveg bráðfyndið enda þekktum við ekki orðið, ekki einu sinni ég. Eftir þetta var föst regla að kasta í þennan glugga og hætta ekki fyrr en konan kom út og hótaði okkur pólitíinu.

En nú ætla ég að víkja að málnotkun forritara og annarra starfsmanna í hugbúnaðariðnaði. Skemmst er frá því að segja að reynsla mín er að þeir leggja flestir metnað sinn í að tala og skrifa góða íslensku. Ég tel mig geta fullyrt að í þessari atvinnugrein sé alls ekki meira um erlendar slettur en dæmin sem ég nefndi að ofan gefa til kynna að verið hafi um miðja síðustu öld.

Það er að vísu rétt að þegar nýjungar koma fram í tækninni á erlendum vettvangi er að sjálfsögðu ekki til neitt orð á íslensku til að lýsa nýjunginni og þá geta menn lítið annað gert en að nota erlenda orðið eða einhverja íslenskaða mynd af því. En það líður sjaldnast langur tími þar til íslensk orð fyrir fyrirbærið skjóta upp kollinum. Í því efni eiga orðanefndir hinna ýmsu fagfélaga mikinn heiður skilið en einnig er áhugi starfsmanna á viðkomandi sviði oftast mikill. Ég tel ekki leika neinn vafa á því að langflestir tæknimenn, eins og aðrir Íslendingar, vilja tala góða íslensku og nota íslensku í daglegum störfum sínum.

Ég er því þeirrar skoðunar að það að menn sækja þekkingu til annarra landa og upplýsingar í erlend tungumál muni alls ekki verða hættulegt íslenskunni. Svo lengi sem almenningur, þar með talið tæknifólk, á þess kost að nota íslensku í daglegu lífi muni það gera það.

Starfsfólk við hugbúnaðargerð hefur allt frá upphafi þeirrar greinar hér á landi þurft að notast við hugbúnaðarverkfæri sem oftast eru skrifuð á ensku. Hér á ég við hugbúnað sem notaður er við að búa annan hugbúnað til. Eins og ég sagði áðan sýnist mér að þetta hafi ekki komið að sök.

Nær allur hugbúnaður, sem framleiddur hefur verið hér á landi, er hins vegar á íslensku. T.d. hefur það ekki hvarflað að mönnum að skrifa bókhaldshugbúnað á öðru máli en íslensku og öll erlend bókhaldskerfi, sem náð hafa umtalsverðri útbreiðslu hér á landi, hafa verið þýdd á íslensku, að minnsta kosti sá hluti þeirra sem almennir notendur hafa aðgang að. Sömu sögu er að segja um flestan þann hugbúnað sem er í almennri notkun.

Það er kostnaðarsamt að þýða erlendan hugbúnað á íslensku og hugbúnaður er mjög misjafnlega vel fallinn til þýðinga. Stundum er allur texti geymdur í sérstökum töflum og þá er aðgengilegt að þýða enda þótt um talsverða vinnu geti verið að ræða. Í öðrum tilvikum er texta ekki haldið þannig til haga og þá er að öllu leyti erfiðara um vik.

Á síðustu misserum hafa komið upp raddir um að réttast væri að Íslendingar hætti að standa í þýðingum og tileinki sér frekar að vinna á ensku. Sem betur fer held ég að þetta sjónarmið eigi ekki miklu fylgi að fagna og ég trúi því að það sé sett fram af aðilum sem eru að selja erlendan hugbúnað sem er illa skrifaður með tilliti til þýðingar. Engu að síður er þetta sjónarmið komið fram og ég held það sé full ástæða til að vera vel á varðbergi.

Þá hefur það einnig heyrst að íslenskur hugbúnaðariðnaður byggji nú meir og meir á útflutningi og sé því að verða svo alþjóðlegur að ekki taki því að skrifa íslenska útgáfu hugbúnaðar og að réttast sé að skrifa hann á ensku. Ég hef ekki mikla trú á þessari röksemdafærslu.

Markaðsmenn segja mér að það sé næsta fátítt að fyrirtæki nái góðum árangri á erlendum markaði hafi það ekki góða fótfestu á heimamarkaði sínum enda tel ég að nær allur sá hugbúnaður, sem nú er fluttur út, hafi upphaflega verið þróaður fyrir innlendan markað. Þá er þess líka að gæta að enska er alls ekki það alþjóðlega tungumál sem menn vilja vera láta. Ég er hræddur um að menn komist ekki langt með

að markaðssetja hugbúnað, sem er ætlaður almenningi, á ensku, til dæmis í Danmörku.

Ég held að svar íslenskra hugbúnaðarframleiðenda verði að vera að skrifa hugbúnað með þeim hætti að auðvelt sé að þýða hann á hin ýmsu tungumál.

Þessi misserin eru að verða miklar breytingar á sambúð tungumála og tækni. Það er að verða æ algengara að ýmis tæki, svo sem heimilistæki, bílar og hljómflytningstæki, gefi notendum upplýsingar á rituðu máli í stað þess að nota til dæmis vísi á mæli.

Þess er líklega ekki langt að bíða að ýmis tæki og búnaður fari að taka við skipunum á mæltu máli. Mér er ekki grunlaust um að ef og þegar að því kemur að menn þurfi að tala við heimilistækin, bílinn, símakerfið og margt fleira á erlendu tungumáli þá fari íslenskunni að stafa veruleg hætta af.

Menntamálaráðuneytið hefur, með afgerandi hætti, markað stefnu um það hvernig brugðist skuli við þessari hættu með því að setja tungutækniverkefninu þau tvö markmið sem ég gat um í upphafi; að

- tryggja að íslenska verði áfram lifandi tungumál íslensku þjóðarinnar í upplýsinga- og þekkingarsamfélagi 21. aldarinnar
- stuðla að því að til verði nauðsynleg tungutækniverkfæri fyrir þann tæknibúnað sem notaður verði í daglegu lífi og starfi

Verkefnisstjórn tungutækniverkefnisins fær einnig drjúgt veganesti til að ná þessum markmiðum því á fjárlögum þessa árs eru 104 milljónir króna ætlaðar til verksins.

Verkefnisstjórnin vinnur eftir verkefnaáætlun sem menntamálaráðherra hefur samþykkt en hér á eftir mun ég gera nokkra grein fyrir þeim aðgerðum sem þar er fjallað um og verkefnisstjórnin hyggst gangast fyrir til þess að stuðla að því að þessum markmiðum verði náð.

Ég geri ráð fyrir að margir hér inni viti hvað átt er við með tungutækni. Ég ætla samt sem áður að rifja það upp og vona að þeir, sem ekki þurfa á upprifjun að halda, sýni biðlund.

Mín skilgreining á tungutækni er að hún fjalli um tæknibúnað af ýmsu tagi sem auðveldi mönnum að nota tungumálið til tjáningar. Bæði getur verið um að ræða búnað sem auðveldar mönnum að tjá hugsun sína og búnað sem „skilur“ tungumál. Slíkur búnaður þarf þá að „skilja“ íslensku til að markmiðum tungutækniverkefnisins verði náð.

Sem dæmi um tungutæknitól má nefna ýmis tæki sem auðvelda notkun og samningu ritaðs máls svo sem forrit til að leiðbeina um og leiðrétta stafsetningu, málfræði og setningaskipan, forrit til að skipta orðum milli lína og fleira. Enda þótt nokkuð sé til af slíkum búnaði fyrir íslensku eru Íslendingar talsvert á eftir mörgum öðrum þjóðum hvað þetta varðar. Framfarir á þessu sviði hafa verið örar víða erlendis þannig að þetta bil er sífellt að stækka.

Talgervlar eru verkfæri sem koma í stað mannsraddar og geta lesið upp tölvutækan texta. Talgervlar eru mjög gagnlegir blindu fólki en þeir geta verið gagnlegir á fleiri sviðum. Með þeim má láta tölvuna lesa upp fyrir sig tölvupóst í gegnum síma eða lesa texta á meðan ekið er. Talgervlar geta einnig svarað sjálfvirkir í síma og eru nú að verða hluti af símkerfum.

Á ensku og fleiri tungumálum eru til forrit sem nema talað mál og rita það sem texta inn í tölvu. Það getur verið mikið hagræði í að geta lesið tölvunni fyrir í stað þess að þurfa að slá inn allan texta með lyklaborðinu. Mörg tæki eru nú svo smá að

þar er ekki pláss fyrir lyklaborð og því er mjög gagnlegt að geta notað röddina til samskipta við þau.

Loks skal getið um vélrænar þýðingar sem felast í því að hugbúnaður tekur við texta á einu tungumáli og skilar honum á öðru.

Þannig fjallar tungutækni um sambúð tungumálsins og tölvutækni í víðum skilningi. Hún snýst um smíði ýmiss konar forrita sem fást við tungumál og tengir því saman málfræði, upplýsingatækni og tölvutækni.

Eitt stærsta vandamál, sem Íslendingar standa frammi fyrir ef þeir ætla að hefja öflugt og markvisst starf á sviði tungutækni, er skortur á fólki með menntun, reynslu og þekkingu á því sviði. Mikilvægt er að hafa í huga að enda þótt skammt sé síðan tungutækni varð að iðngrein í grannlöndum okkar styðst það starf við margra ára rannsóknir og kennslu á háskólastigi. Erlendis er máltölvun (computational linguistics) víða kennd í háskólum, ýmist sem sérstök grein eða innan málvísindadeilda eða tölvunarfræðideilda.

Mikill vöxtur hefur verið í þessum greinum á undanförunum árum. Langflest störf, sem málfræðingum standa til boða við háskóla í ensku- og þýskumælandi löndum um þessar mundir, eru t.d. á þessu sviði. Vitaskuld er beint samband milli þenslunnar á þessi sviði í háskólum og hinnar öru þróunar sem hefur verið í tungutækni sem iðngrein en hún veltir nú háum fjárhæðum. Þörf atvinnulífsins fyrir fólk með menntun á þessu sviði hefur stórauðist og þar með vilji til að efla slíka kennslu á háskólastigi.

Háskóli Íslands hefur nú hafið undirbúning að meistaranámi í tungutækni. Menntamálaráðuneytið gerði samning við skólann um þetta nám og um þjónustu við verkefnisstjórn í tungutækni í tengslum við það. Markmiðið með náminu er að mennta nemendur til þess að vinna að þróun tungutækni fyrir íslensku, taka að sér störf á þessu sviði í íslenskum upplýsingatækni fyrirtækjum og miðla þekkingu sinni til annarra. Unnið verður að rannsóknum á tungutækni fyrir íslenska málsamfélagið með sérstakri áherslu á lausnir sem gætu nýst öðrum fámönnum málsamfélögum.

Til þess að örva þróun og framleiðslu tungutækniþóla fyrir íslensku mun menntamálaráðherra, að fengnum tillögum verkefnisstjórnar um tungutækni, veita styrki til tvenns konar verkefna. Annars vegar styrki til þess að koma upp málsöfnum og þekkingargrunnum sem nauðsynlegir eru taldir við þróun tungutækniþóla. Hins vegar er ætlunin að styrkja fyrirtæki til þess að þróa, framleiða og setja á markað tiltekið tungutækniþól.

Til þess að unnt sé að þróa og framleiða tungutækniþóla er nauðsynlegt að miklar upplýsingar um íslenskt mál liggi fyrir í formi sem er aðgengilegt fyrir hin ýmsu verkfæri. Hér skortir talsvert á að ástandið hér á landi sé sambærilegt við það sem gerist meðal nágranna okkar. Það eru þekkingargrunnar af þessu tagi sem gert er ráð fyrir að styrkja á vegum verkefnisstjórnarinnar.

Til þess að tryggja að opinbert fé, sem þannig er varið til þróunar þekkingargrunna, nýtist sem best, verður einungis stutt við gerð þekkingargrunna sem atvinnufyrirtæki á sviði tungutækni hyggst nýta til þróunar tiltekins tungutækniþóls. Gert er ráð fyrir að fyrirtækið, sem óskar eftir að þekkingargrunnur sé búinn til, taki einhverja áhættu um að niðurstöður verði í raun nýttar. Þá er það skilyrði að niðurstöður rannsókna af þessu tagi verði öllum opnar til áframhaldandi rannsókna eða til að þróa önnur tungutækniþóla.

Sé ekki hægt að veita fullan aðgang að þekkingargrunni mun tungutækniþólið einungis fjármagna þann hluta verksins sem er opinn öllum. Þannig öðlast styrkþegi ekki eignarrétt á grunninum.

Markmiðið með þessum styrkjum er að framleiðendur tungutæknitóla hafi aðgang að þeirri þekkingu og þeim málsöfnum sem þeir þurfa á að halda og að þróun slíkra tóla fyrir íslensku verði þar með ódýrari og auðveldari.

Af þeim upplýsingum, sem nú liggja fyrir, virðist sem það fyrsta, sem þarf að koma upp vegna tungutækniverkefna, sé stór textagrunnur sem greindur verði eftir málfræðilegum atriðum. Þannig verði hvert orð greint í orðflokk og gerð grein fyrir beygingum og öðrum málfræðiatríðum eftir því sem við á.

Textagrunnur af þessu tagi mun ekki einungis nýtast framleiðendum tungutæknitóla. Hann mun einnig nýtast við margháttaðar rannsóknir, bæði í tungutækni og við almennar rannsóknir á íslenskri tungu.

Mikið er til af texta í tölvutæku formi, til dæmis hjá Orðabók Háskólans, Morgunblaðinu, bókaútgefendum og fleirum, sem nýst getur til að mynda slíkan grunn. Það er hins vegar talsverð vinna að greina orðaforða grunnsins eins og lýst er að framan. Því verki ætti þó að vera hægt að ljúka á einu til tveimur árum.

Ljóst er að það er mikil vinna sem eftir á að vinna við hagnýtar rannsóknir á tungumálinu til þess að unnt verði að framleiða mörg þau verkfæri sem vonandi líta dagsins ljós á næstu árum.

Það er hins vegar eitt af markmiðum verkefnisstjórnarinnar að stuðla að því að innan skamms tíma sjáist árangur af þessu átaki menntamálaráðuneytisins. Við í verkefnisstjórninni erum líka þeirrar skoðunar að unnt sé að þróa gagnleg tungutæknitól með því að nota nær eingöngu þá þekkingu sem þegar er til staðar.

Af þessum ástæðum verða veittir beinir styrkir til fyrirtækja sem óska eftir að framleiða einstök verkfæri til að markaðssetja og selja. Þegar um er að ræða styrk af þessu tagi eru skilmálar nokkuð á annan veg en þegar um þróun upplýsingagrunns er að ræða. Afurðin verður algerlega eign þess fyrirtækis sem þróar hana. Þó getur verið um einhverja endurgreiðslu að ræða, t.d. í því formi að fyrirtækið leyfi takmörkuð afnot tólsins í skólum án endurgjalds. Ekki er gert ráð fyrir að styrkja mörg slík verkefni og verkefni, sem skila fljótt árangri, munu fá forgang.

Markmiðið með þessum styrkjum er að efla áhuga á tungutækni, bæði hjá notendum og framleiðendum, og að auka notkun tungutæknitóla hér á landi.

Að lokum vil ég hvetja þá sem vilja fræðast meir um tungutækniverkefni menntamálaráðuneytisins til að kynna sér vef verkefnisins, www.tungutaekni.is.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.