

Eiríkur Rögnvaldsson

Framtíð íslensku innan upplýsingatækninnar

Á íslenska sér framtíð innan upplýsingatækninnar?

Málþing Íslenskrar málnefndar og Tungutækniseturs, 7. mars 2008

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.“

Á þessu málþingi á að fjalla um 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 tungutækni – þ.e. án íslenskra leiðréttingarforrita, án vélrænna þýðinga milli íslensku og annarra mála, án íslenskra talgervla og talgreina, o.s.frv. Á sama hátt má vel hugsa sér íslenska tungutækni 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 tungutækni og hugbúnaðarþýðinga. Hvorttveggja varðar tölvur og notkun tungumáls í tengslum við þær. Þetta snýst sem sé um hugtak sem hefur verið áberandi í málræktarumræðu undanfarin ár – umdæmi 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 þessu erindi ætla ég fyrst að rekja í stuttu máli þær breytingar sem hafa orðið á þessu sviði undanfarin áratug, og velta síðan fyrir mér möguleikum íslenskrar tungu innan upplýsingatækninnar á næstu árum.

2. Sagan

Fyrir tíu árum, vorið 1998, var staða mála á þessu sviði ekki álitleg, hvort sem litið var til hugbúnaðarþýðinga eða tungutækni. 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. 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 tungutækni var síst betri – íslensk tungutækni 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 tungutækni, 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. Sama haust, 1998, fékk Björn Rögnvald Ólafsson dósent í eðlisfræði til að gera skýrslu um ástand og horfur í íslenskri tungutækni. Rögnvaldur fékk tvo menn með sér í skýrslugerðina og þessi starfshópur skilaði af sér vorið 1999. Í skýrslu hans 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.

Í framhaldi af þessu lagði starfshópurinn til að stjórnvöld beittu sér fyrir átaki á fjórum sviðum til eflingar íslenskri tungutækni næstu fimm árin:

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 tungutækni.
3. Fyrirtæki verði styrkt til þess að þróa afurðir tungutækni.
4. Menntun á sviði tungutækni og málvísinda verði efld.

Í tillögum starfshópsins var gert ráð fyrir að tungutækniátakið stæði í a.m.k. fjögur ár og heildarkostnaður á ári yrði 225-250 milljónir króna. Eftir það vonuðust menn til að íslensk tungutækni yrði orðin sjálfbær og þarfnaðist ekki meiri opinberra framlaga.

3. Staðan

Á fjárlögum árána 2000-2004 voru alls veittar 133 milljónir króna til tungutækni. 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 tungutækniáætlunin hafi skilað heilmiklu. Kennsla í tungutækni og samstarf við erlenda háskóla á þessu sviði er hafið, mikilvæg gagnasöfn hafa verið byggð upp, og ýmsum rannsóknar- og þróunarverkefnum hefur verið ýtt af stað. Ég skal nefna nokkur þau helstu.

- Íslenskur talgreinir (stakorðagreinir)
- Íslenskur talgervill (Ragga – vefpulan)
- Endurbætt ritvilluleit (Púki)
- Beygingarlýsing íslensks nútímamáls (258 þús. orð – 5,6 milljón beygingarmyndir)

- Leitarvél (Embla)
- Málfræðilegir markarar (forrit sem greina texta málfræðilega)
- Hlutapáttari (forrit sem greinir helstu setningarliði og setningarleg hlutverk)
- Mörkuð málheild (25 milljón orða safn fjölbreyttra texta, málfræðilega greint)
- Samhengisháð ritvilluleit

Því fer þó fjarri að íslensk tungutækni sé orðin sjálfbær, eins og stefnt var að, og eftir að tungutækniáætluninni lauk fyrir þremur árum hefur dregið mjög úr uppbyggingu, þróun og nýsköpun á þessu sviði. Og það er engin furða. Það kostar jafnmikið að byggja upp málleg gagnasöfn og tungutæknibúnað fyrir tungumál sem 300 þúsund manns tala og fyrir tungumál milljónaþjóða. Það er ekki von að fyrirtæki sjái sér hag í því að leggja í mikinn kostnað við að þróa og aðlaga tungutæknibúnað fyrir íslensku – sú fjárfesting skilar tæpast hagnaði. Rannsóknasjóður hefur undanfarin ár veitt nokkra styrki til tungutækniverkefna, og ber síst að vanþakka það; en það hrekkur skammt ef við ætlum að halda í við þróunina á þessu sviði erlendis. Á mörgum sviðum tungutækninnar 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.

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 á næstunni mun Windows Vista væntanlegt á íslensku, en það heyrum við meira um á eftir. 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, 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, rannsóknum sem stjórnisýslu.“ 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, a.m.k. ekki að fyrra bragði.

Þótt ég hafi nefnt Windows og Office hér sérstaklega, vegna yfirburðastöðu þessara kerfa á markaðnum, má ekki gleyma því að mikilsvert þýðingarstarf er unnið víða annars staðar. Þar má t.d. nefna öflugt starf áhugamanna að þýðingum á Linux-stýrikerfinu; Opera-vafrinn hefur verið íslenskaður; og fleira mætti nefna. Hins vegar skilst mér að ekki verði framhald á íslenskun Macintosh-stýrikerfa og hugbúnaðar.

4. Framtíðin

En hverjar eru þá framtíðarhorfur 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. 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. Ég kem meira að því síðar.

Fyrir 11 árum flutti ég erindi um tungutækni á norrænu málnefndarþingi í Færeyjum. Þetta var áður en nokkuð fór að gerast í íslenskri tungutækni, en ég var nýbúinn að átta mig á þeirri hættu sem tungumálum væri búin ef þau yrðu ekki gjaldgeng innan upplýsingatækninnar, og var dálítið uppveðraður yfir þessu öllu saman. Ýmsum þótti ég

vera með dómsdagsspár í þessu erindi, og vissulega hefur þróunin í notkun tungumáls í samskiptum manns og tölvu ekki orðið jafn hröð og margir töldu fyrir áratug. Tungutækniátakið og sú þróun sem það kom af stað hefur leitt til þess að við höfum ekki alveg misst af lestinni – við eigum möguleika á að taka þátt í því sem er að gerast á þessu sviði. En þá verðum við að spýta í lófana.

Ef við viljum nota íslensku innan upplýsingatækninnar á annað borð munum við þurfa að borga fyrir það, með einum eða öðrum hætti. Spurningin er bara sú hvort ríkið eigi að greiða þennan kostnað, eða hvort við eigum að gera ráð fyrir að einkafyrirtæki greiði þróunarkostnaðinn og velti honum út í verðlagið, og láta málnotendum það svo eftir að ákveða hvort þeir kaupa dýran tungutækniþúnað fyrir íslensku eða ódýrari þúnað sem gerður er fyrir ensku.

Ég er ekki viss um að sú leið sé vænleg til árangurs. Við getum kannski verndað – eða reynt að vernda – t.d. íslenskan landbúnað með tollum, innflutningshömlum o.s.frv., en við getum ekki verndað íslenska tungu á sama hátt. Fólk væri hugsanlega til í að greiða eitthvað meira fyrir hugbúnað á íslensku og tungutækniþól fyrir íslensku, en ef munurinn er verulegur eru allar líkur á að fólk halli sér fremur að erlendum búnaði – þ.e.a.s. enskum. Fyrir flesta er það svo sem ekkert stórmál. Við erum vön því að hafa ensku allt í kringum okkur allan liðlangan daginn hvort eð er. Ef einkafyrirtæki á þessu sviði fá ekki meiri opinberan stuðning eru litlar líkur á því að við eigum eftir að sjá mikla þróun í gerð tungutækniþúnaðar fyrir íslensku á næstu árum.

Það er ekki hægt að búast við því að stjórnámálemenn 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.

5. Að lokum

Þegar mikilvægi íslenskrar tungutækni 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 hefur það á málnotendur og mál-samfélagið? 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 velta þessu lengi fyrir sér til að sjá hættumerkin.

En það er ástæðulaust að meta þörf á íslenskri tungutækni 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 ekki að þurfa að sitja skör lægra en aðrir, og vera neydd til þess að nota erlend mál við hversdagslegar aðstæður. 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.

Auðvitað kostar þetta eitthvað. En við vitum að það er dýrt að vera fámenn þjóð í stóru landi, tala sérstakt tungumál og halda uppi sjálfstæðu menningarsamfélagi. Við höfum axlað ýmsar byrðar sem þessi staða leggur okkur á herðar. Við leggjum t.d. hátt á sjötta hundrað milljóna til Sinfóníuhljómsveitar Íslands á þessu ári, og svipaða upphæð til Kvikmyndasjóðs. Við gætum alveg sparað okkur þessi útgjöld, en við gerum það ekki vegna þess að við teljum það mikilvægan þátt í menningu þjóðarinnar að halda uppi sinfóníuhljómsveit og íslenskri kvikmyndagerð. Ég er sist af öllu að telja þetta eftir – en hvers virði er íslensk tunga? Verður hún metin til fjár? Er til of mikils mælst að samfélagið verji nokkrum tugum milljóna á ári í það að gera hana að fullgildum samskiptamiðli í upplýsingaþjóðfélaginu? Ef stjórnvöld finna fyrir þrýstingi almennings í þessa veru hef ég ekki trú á að þau hlaupist undan merkjum.

Tilfni þessa málþings er vinna Íslenskrar málnefndar að mótun íslenskrar málstefnu. Sú stefna getur haft mótandi áhrif á viðgang íslenskrar tungu á næstu árum og verið vegvísir fyrir stjórnvöld í lagasetningu, fjárveitingum og fleiru. Það er mikilvægt að í málstefnunni verði lögð rík áhersla á öflugt starf á sviði tungutækni og hugbúnaðarþýðinga. 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. Það er hlutverk okkar – Íslenskrar málnefndar, Tungutækniseturs, og allra sem vinna við og hafa áhuga á íslensku og upplýsingatækni – að sýna fram á að upplýsingatæknin getur verið á íslensku. Með því móti öflum við almenns stuðnings við málstefnuna og tryggjum framgang hennar.