



Eiríkur Rögnvaldsson
Tungutækni

Tungutækni og hugbúnaðarþýðingar

Málþing um mótun íslenskrar málstefnu
í safnaðarheimili Neskirkju
19. janúar 2008

Hvað er tungutækni?

- Samvinna tungumáls og tölvutækni
 - í einhverjum hagnýtum tilgangi
- Tvær hliðar samvinnunnar:
 - notkun tækninnar í þágu tungumálsins
 - leiðréttingarforrit, þýðingarforrit, talgervlar, hjálpartæki fatlaðra, tölvuorðabækur, kennsluforrit ...
 - notkun tungumálsins í þágu tækninnar
 - upplýsingaleit í gagnabönkum og á neti, stjórn ýmiss konar tækja, samskiptakerfi ...

Starfshópur um tungutækni, 1999

- Meginmarkmið Íslendinga
 - hlýtur að vera að unnt verði að nota íslenska tungu, ritaða með réttum táknum, sem víðast innan tölvu- og fjarskiptatækninnar
- 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
- Eftirtalin forgangsverkefni voru skilgreind:

1. Þýðing tölvuforríta

- *Helstu tölvuforrit á almennum markaði verði á íslensku (Windows, Word, Excel; Netscape, Internet Explorer; Eudora; ...)*
- Windows XP og Microsoft Office er komið á íslensku
 - og einnig ýmis önnur forrit
 - en útbreiðsla þýðinganna er takmörkuð
 - hætt að þýða stýrikerfi Macintosh á íslensku

2. Íslenskir bókstafir

- *Unnt verði að nota íslenska bókstafi (áéíóúýðþæö ÁÉÍÓÚÝÐÞÆÖ) við allar aðstæður; í tölvum, GSM-símum, textavarpi og öðrum tækjum sem almenningu notar.*
- Hér hefur staðan batnað
 - m.a. með aukinni útbreiðslu Unicode
- Nú er hægt að nota íslenska stafi í GSM
 - með einhverjum takmörkunum þó

3. Málgreining

- *Unnið verði að þróun málgreiningar fyrir íslensku, með það að markmiði að geta greint íslenskan texta í orðflokka og setningarliði.*
- Nokkur slík verkefni hafa verið unnin
 - beygingarlýsing íslensks nútímamáls
 - tveir málfræðilegir markarar
 - vélræn íslensk setningagreining
 - hlutapáttun íslensku

3.1 Textaheild – 3.2 Orðasafn

- *Koma þarf upp stórri tölvutækri textaheild með íslenskum textum af sem fjölbreyttustum toga til að byggja áframhaldandi vinnu á.*
- Verið er að byggja upp slíka málheild
- *Koma þarf upp fullgreindu orðasafni (með málfræðilegri og merkingarlegri greiningu) til nota í áframhaldandi vinnu.*
- Ekkert slíkt orðasafn er til
 - þótt til sé hráefni sem vinna mætti út frá

4. Hjálparforrit við ritun

- *Til verði góð hjálparforrit við ritun texta á íslensku, s.s. orðskiptiforrit, stafsetningarleiðréttingarforrit, málfarsleiðréttingarforrit o.fl.*
- Nýtt forrit til stafsetningarleiðréttingar hefur verið unnið á vegum Microsoft
 - en stendur *Púkanum* að baki
- Unnið er að samhengisháðri ritvilluleit
 - sem einnig nýtist við málfarsleiðréttingu
- Málfarsleiðréttingaforrit eru ekki til enn

5. Íslenskur talgervill

- *Til verði góður íslenskur talgervill sem geti lesið upp íslenskan texta með skýrum og auðskiljanlegum framburði og eðlilegu tónfalli og sem sé skiljanlegur án þjálfunar.*
- Nýr talgervill, *Ragga*, er kominn á markaðinn
 - byggist á fullkomnari tækni en fyrri talgervlar
 - fær yfirleitt góða dóma
 - virðist ætla að ná verulegri útbreiðslu

6. Talgreining

- *Unnið verði að þróun talgreiningar fyrir íslensku, með það að markmiði að til verði forrit sem geti túlkað eðlilegt íslenskt tal.*
- Háskólinn og fjögur fyrirtæki stóðu að *Hjali*
 - íslenskri stakorðagreiningu
 - íslenskur talgreinir er nú til og virkar vel
 - en hefur þó ekki verið nýttur að ráði
- Langt er í greiningu samfellds máls

7. Vélrænar þýðingar

- *Unnið verði að þróun forrita til vélrænna þýðinga milli íslensku og annarra tungumála, m.a. til að auðvelda leit í gagnabönkum.*
- Hér hefur lítið gerst
 - ýmsir hafa unnið með þýðingarminni
 - engin nothæf þýðingarforrit hafa birst
 - en *Alnet* vinnur að íslensk-ensku þýðingarforriti

8. Ábyrgðaraðilar

- *Ákveðnum aðilum (stofnunum eða fyrirtækjum) verði falin ábyrgð á einstökum verkefnum.*
- Sett var á fót verkefnisstjórn í tungutækni
 - til að ýta verkefnum af stað og samræma aðgerðir
- Þetta skilaði góðum árangri
 - en verkefnisstjórnin var lögð niður í árslok 2004
- Stofnað hefur verið *Tungutæknisetur*
 - en það hefur ekkert fé til umráða

Staðan að lokinni tungutækniáætlun

- Tungutækniáætlunin skilaði árangri
 - menntun á sviði tungutækni er hafin
 - Íslendingar eru farnir að fara í nám erlendis
 - gagnasöfn hafa verið byggð upp
 - ýmsum verkefnum hefur verið ýtt af stað
- En íslensk tungutækni er ekki orðin sjálfbær
 - nú þegar tungutækniáætluninni er lokið
 - og einmitt þyrfti meira fé í rannsóknir og þróun

Ógnar upplýsingatæknin tungunni?

- Þrjú einkenni upplýsingatækni skipta máli
 - þegar áhrif hennar á íslenska tungu eru metin
- Hún er orðin
 - mikilvægur þáttur
 - í daglegu lífi
 - alls almennings
- Þess vegna verður hún að vera á íslensku
 - að öðrum kosti er tungan feig

Þrengt notkunar svið móðurmálsins

- Hvað ef móðurmálið er ekki gjaldgengt á sviði
 - sem er mikilvægt
 - í daglegu lífi
 - alls almennings?
- Hvað ef það er ekki nothæft
 - í nýrri tækni og öðru sem er nýtt og spennandi
 - á sviðum þar sem nýsköpun á sér stað
 - og þar sem ný atvinnutækifæri bjóðast?

Tungumál í hættu

- Við þær aðstæður hefst dauðastríð tungunnar
 - móðurmálið verður víkjandi
 - aðeins hæft til heimabruks
 - en ekki til neinna alvarlegra hluta
- Ungt fólk sér þá ekki tilgang í að læra málið
 - heldur leggur áherslu á að tileinka sér enskuna sem best
- Hvað er þá til ráða?

Tveir kostir í stöðunni

- Að hafna tækninni en halda tungunni
 - látið eiga sig að tileinka okkur ýmsar nýjungar
 - ef móðurmálið er ekki gjaldgengt á þessu sviði
- Þessi kostur er ekki raunhæfur
- Að fórna tungunni en fylgjast með tækninni
 - nota ensku í upplýsinga- og tölvutækni
 - ef íslenska verður ekki nothæf á því sviði
- Þessi kostur er óviðunandi!

– og sá þriðji:

- Að gera eitthvað í málinu!
 - gera átak í tungutækni og hugbúnaðarþýðingum
 - gera íslensku gjaldgenga innan upplýsingatækni
- Það er eini valkostur okkar
 - ef við viljum halda áfram að nota íslensku
 - á öllum sviðum þjóðlífsins
- Annars verður málið fljótlega forngripur
 - dauðadæmt og gæti dáið út á fáum áratugum

Tungutækni fyrir málnotendur

- Tungutækni snýst ekki bara um málvernd
 - einnig um þjónustu og sjálfsvirðingu
- Eigum við að sitja við sama borð og aðrir
 - eða eigum við að sitja skör lægra?
- Við eigum kröfu á að geta notað móðurmálið
 - sem víðast, við sem fjölbreyttastar aðstæður
- Allt annað er uppgjöf!

Fordæmi Eista

- Eistar eru smáþjóð eins og við
 - aðeins um ein milljón talar eistnesku
- Þeir hafa gert áætlun um þróun tungutækni
 - til sjö ára, 2004-2011
- 2004 voru þeir á svipuðu stigi og við
 - en þeirra tungutækniáætlun var þá að byrja
 - okkar var að enda
- Ætlum við að láta hér við sitja?

Þakka ykkur fyrir áheyrnina

eirikur@hi.is