



HÁSKÓLI ÍSLANDS

Eiríkur Rögnvaldsson
Háskóla Íslands

Tungutækni og íslenskan

Kynning á tungutækni
fyrir starfsmenn Símans
2. mars 2006

Áhrif tungutækni á íslenskt mál

- Hvað gerist ef móðurmálið er ekki gjaldgengt
 - á sviði sem er mikilvægt í daglegu lífi 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?
 - þar sem ný atvinnutækifæri bjóðast?
- Undir slíkum kringumstæðum er tungan feig
 - þess vegna verður tungutækni að vera á íslensku

Tungutækniátak

- Tungutækni getur verið á íslensku
 - en til þess þarf mikla uppbyggingu
 - á gagnasöfnum, hugbúnaði og tólum
- Menntamálaráðuneytið stóð fyrir átaki
 - til að efla íslenska tungutækni
 - í kjölfar skýrslu starfshóps vorið 1999
- Til þess var varið 133 milljónum króna
 - á árunum 2001-2004

Íslensk tungutækni verkefni

- Beygingarlýsing íslensks nútímamáls
 - nýtt m.a. í leitarvélinni *Emblu* og orðabókum
- Markari (grammatical tagger)
 - mun nýtast við leiðréttingar, vélrænar þýðingar
- Mörkuð málheild (tagged corpus)
 - mun nýtast í ýmsum verkefnum
- Hlutapáttari (shallow parser)
 - mun nýtast í leitarvélum, við leiðréttingar o.fl.

Greining og málfarsleiðrétting

- Málfarsleiðrétting er útilokuð án greiningar:
 - villur felast sjaldan í notkun óleyfilegra mynda
 - *föðurs* í stað *föður*
 - *keyptu* í stað *kauptu*
 - fremur í að nota réttar myndir á röngum stöðum
 - *Ég hitti systir þína* > *systur*
 - *vegna þeirrar tilhneigingu* > *tilhneigingar*
 - *fjöldi manna komu* > *kom*
 - *mér langar* > *mig langar*

Stafsetning og vélrænar þýðingar

- Sama gildir um stafsetningarleiðréttingu
 - margar villur finnast aðeins með málgreiningu
 - *það er kominn morgun* > *morgunn*
 - *ég hitti Kristinn* > *Kristin*
 - *hann er farin* > *farinn*
- Vélrænar þýðingar krefjast málgreiningar
 - annars eru þær bara uppfletting í orðasafni
 - *hot spring river this book* (hver á þessa bók)

Meðferð þular á tölum

- Greining skiptir líka máli fyrir þul
 - til að hann geti lesið rétt úr ýmsu í texta
 - skammstöfunum og ekki síst tölum
- Hér er Valdís Ólafsdóttir M.A.
 - hún var að ljúka meistaraþrófsritgerð
 - um reglur um það hvernig eigi að lesa úr tölum
- Valdís tekur nú við
 - og kynnir viðfangsefni sitt lauslega