

## Informationsteknologien og små sprogsamfund

- (1) Smá sprogsamfunds bødler:
  - a. Satellite-tv
  - b. CD-ROM
  - c. Internet/WWW
  
- (2) “Þær breytingar sem í hönd fara verða því að styrkja íslenska menningu og trú þjóðarinnar á eigin sérstöðu og hlutverk. Svo vill til að íslenskur menningararfur er að langmestu leyti fólgin í upplýsingum. Upplýsinga- og fjarskiptatækni líkur því ekki einungis upp nýjum möguleikum er varða framtíðarþróun íslensks samfélags, heldur mun þessi tækni valda straumhvörfum í kynningu og skilningi á þeim verðmætum sem þjóðin hefur skapað á liðnum öldum.”  
(Davíð Oddsson: *Framtíðarsýn ríkisstjórnar Íslands um upplýsingasamfélagið.*)
  
- (3) **Markmið**
  - Standa vörð um íslenska tungu í upplýsingasamfélaginu og tryggja að tungumálið fái að þróast í samræmi við íslenska málstefnu.
  - Hvatt verði til að íslenskt efni verði í auknum mæli á boðstólum á tölvutæku formi.
  - Efla þarf nýyrðasmíð sem tengist upplýsingatækni og tölvuvæð-ingu.
  - Stefnt skal að því að íslenska algengustu gerðir hugbúnaðar.  
(Í krafti upplýsinga. Tillögur menntamálaráðuneytisins um menntun, menningu og upplýsingatækni 1996-1999.)
  
- (4) “Með upplýsingatækni hefur framboð á erlendu efni á geisladiskum og Interneti aukist verulega. Íslendingar verða að leggja áherslu á að styrkja íslenska framleiðslu í samkeppni á þessu sviði sem og öðrum. Styrkja þarf notkun íslenskrar tungu í upplýsingatækni og stuðla að nægu framboði efnis þannig að hægt sé að nálgast sem fjölbreyttast efni á íslensku. Framleiðendur hér á landi verða að geta nýtt sér nýja tækni og stuðlað að góðu framboði á íslensku efni á geisladiskum og Interneti á komandi árum.”

*(Í krafti upplýsinga. Tillögur menntamálaráðuneytisins um menntun, menningu og upplýsingatækni 1996-1999.)*

(5) Reaktióner:

- a. Modersmálsundervisning.
- b. Produktion af íslensk materiale.
- c. Oversættelser.

(6) “Nú þegar er mikið framboð á margmiðlunardiskum með hvers kyns fræðslu- og skemmtiefni, og ýmsir slíkir diskar fylgja oft nýjum tölvum. Þetta efni er að sjálfsögðu allt á erlendum tungumálum, einkum ensku. Fyrir lítið málsamfélag eins og það íslenska er mikilvægt að íslensk tunga sé nothæf og notuð við öll tækifæri, við nám, vinnu og leik. Það er hnignunarmerki á tungumáli þegar málnotendur treysta sér ekki lengur til að nota málið á öllum sviðum daglegs lífs, en telja sig þurfa að grípa til erlends máls í tilteknu samhengi.

Ef ungir tölvunotendur venjast því að hugsa um margmiðlun sem erlent efni - oftast engilsaxneskt - gæti íslensk tunga verið í hættu stödd.”

*(Ritstjórnarstefna Verkefnisstjórnar Lýðveldissjóðs.)*

(7) “In an ideal world, every user would be able to access the information they required using their own language, and be able to receive the information in their own language. ‘Information’ in this context includes business and commercial information (including advertisements, orders, bills, payments, etc.), educational materials, entertainment, and so on.”

*(Language and Technology.)*

(8) Date: Mon, 4 Aug 97 12:35:55 EST  
From: Jacques Guy <j.guy@trl.telstra.com.au>  
Subject: Computational linguistics: The AltaVista "refine" option

A colleague of mine told me about the "refine" option offered by the AltaVista search engine and how good it was. In a nutshell, the "refine" options returns a list of synonyms of, and notions related to the words in your query. Indeed, it gave extremely sensible responses.

Perversely perhaps, I tried it in French, using "vin" (what else!) as the keyword. Bingo. This is what it returned:

72% Etait, etre, annees, meme, apres, etaient, derniers  
59% Egalement, particulierement, differentes, possibilite  
52% Qualite, vins, vin, vignoble, vigne, crus, vignes, vigneron, vigneron  
[etc...]

Far from satisfactory. "Eau" and "pain" returned similar nonsense, featuring "etaient", "etre", "egalement" et alia in prominent positions. In fact, AltaVista "refine" seems decidedly adverse to foodstuff in French, "fruit", "poisson" and "sandwich" failing equally miserably (so did "sable", "mer", "lac").

So I was quite surprised when Italian queries about "wine" returned sensible synonyms:

60% vino, vini, vigneti, uve  
40% quantita, ettari, vitigni  
39% sapore, profumo, invecchiamento  
[etc.]

"Acqua" and "pane" fared equally well. So I turned my attention to Spanish. Spanish did quite as badly as French. This is quite puzzling for the size of the Spanish data is quite large.

[...]

Do take a break and experiment a bit with AltaVista "refine" option in your favourite languages (Polish was as nonsensical as French). It is quite amusing. [...]

(9) a. **vín (islandsk)**

8% Thad, thar, thess, ekki, fyrir, thau, eru, hafa, meira  
8% Hja, segja, okkur, betur  
8% Hefur, eftir, starfa  
8% Nokkur, verda, ekkert, okkar, hins, einn, vegar, sinni  
8% Heldur, hafi, setja, kemur, vid  
8% Sem, getur, einnig

b. **vin (dansk)**

99% Vaere, forste, sadan  
67% Kobenhavn, hoteller, underholdning  
64% Havde, gik, fik, sagde, hende, svarede, lagde, stod, spurgte  
63% Laenge, slet, kvinde, sidste  
61% Nar, doren, gaden  
59% Noget, jeg, sige, mig  
58% Lidt, meget, kender  
57% Bade, hjaelp, oplysninger, dansk, danske, samarbejde, vigtige, mulighed, sprog  
55% Hvad, siger, bange  
52% Drikke, kalder, kalde, mindste  
52% Nogle, selvom, bruge, derefter, hvis, bruger  
50% Maned, udstyr, gange, hver  
49% Lige, svarer, pige, vejret  
49% Hoj, udenfor, smuk, gores, frokost, eftermiddagen  
48% Nogen, altid, elsker  
46% Hore, herre, smukke  
44% Fandt, vejen, forbi  
43% Rigtig, ligesom, specielt  
42% Masser, frankrig, forskellige, kob, salg, hjemmeside  
42% Begyndte, engang, vidste, holdt

- (10) Tre faktorer:
- a. Vigtige områder.
  - b. Hele befolkningen.
  - c. Dagligt liv.
- (11) “We should see Europe’s linguistic diversity not as a weakness, however, but as one of its strengths. National and regional differences - which reflect and are reflected by language - lead to a rich diversity of attitudes and approaches to solving problems and creating solutions. This leads to the creation of a wider variety of products, in many languages.”  
*(Language and Technology.)*
- (12) PAROLE:  
14 vesteuropæiske sprog.  
Korpus: 20 million tekstord.  
Leksikon: 20 tusind leksemer.
- (13) “We have seen already how the concentration of information in a small number of languages (notably English) has the potential for discriminating against speakers of other languages, and particularly minority languages and languages used in less well-developed regions. The efforts of the language industry must be directed towards lessening this imbalance, with the ultimate aim of making information available in the language of the person who needs it, and easily accessed by them in their own language. The realization of this objective is some considerable time away - some would say it is unrealistic.”  
*(Language and Technology.)*