Rafræn bókaútgáfa
9. febrúar 2007

Rafrænir textar í rannsóknum og kennslu

Eiríkur Rögnvaldsson
Háskóla Íslands

Rafræn textasöfn (málheildir)

- Notkun rafrænna texta hefur valdið byltingu
  - á mörgum sviðum málrannsókna og málakennslu
- Víða hefur verið komið upp stórum söfnum
  - til að nýta í þessum tilgangi
- Þekktast er British National Corpus (BNC)
  - „a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written“

What sort of corpus is the BNC?

- **Monolingual:** It deals with modern British English, not other languges used in Britain. However non-British English and foreign language words do occur in the corpus.
- **Synchronic:** It covers British English of the late twentieth century, rather than the historical development which produced it.
- **General:** It includes many different styles and varieties, and is not limited to any particular subject field, genre or register. In particu-lar, it contains examples of both spoken and written language.
- **Sample:** For written sources, samples of 45,000 words are taken from various parts of single-author texts. Shorter texts up to a maximum of 45,000 words, or multi-author texts such as magazines and newspapers, are included in full. Sampling allows for a wider coverage of texts within the 100 million limit, and avoids over-representing idiosyncratic texts.

The main uses of the corpus

- Reference Book Publishing
  - Dictionaries, grammar books, teaching materials, usage guides, thesauri. Increasingly, publishers are referring to the use they make of corpus facilities: it's important to know how well their corpora are planned and constructed.
- Linguistic Research
  - Raw data for studying lexis, syntax, morphology, semantics, discourse analysis, stylistics, sociolinguistics...
- Artificial Intelligence
  - Extensive data test bed for program development.
- Natural language processing
  - Taggers, parsers, natural language understanding programs, spell checking word lists...
- English Language Teaching
  - Syllabus and materials design, classroom reference, independent learner research.

Nýting í málfræðirannsóknum

- Athugun á beygingu tiltekinna orða
  - hvernig er ef.ft. af *skutla*? *skutla* eða *skutlna*?
- Athugun á orðmyndun
  - hvaða viðskeyti eru mest notuð í íslensku?
- Athugun á orðasamböndum
  - hvort er algengara *oft og tíðum* eða *oft á tíðum*?
- Athugun á setningagerð
  - er „nýja þolmyndin“ *það var hrint mér* algeng?

Nýting í orðabókagerð

alldjarfliga, adv. med stor Dristighed eller megen Iver for at trænge sig frem ...

og berjast **alldjarflega.** Fann Þórður það brátt
og barðist sjálfur **alldjarflega.** Hann gekk mjög út á virkið er hann
og barðist þá **alldjarflega.** Litlu síðar heyrðu þeir mælt í
Þórir barðist **alldjarflega** og féll á skipi sínu með mikilli
Steingrímur barðist **alldjarflega** og varð fjögurra manna bani.
Víkingar lögðu að **alldjarflega** og þóttu hinir komnir í stilli.
og barðist **alldjarflega.** Þórir bað sína menn hlífa sér
og börðust **alldjarflega** því að Rauður var frækn maður

alldrengiliga, adv. paa saadan Maade, at man derved viser *mikinn drengskap*.

drengskapr, m. Tænkemaade, Opførsel, der gjør en til et saadant Menneske, som han bør være.

Þorgils varðist **alldrengilega** en féll þó fyrir þeim Gunnari og Grími.
en þeir verjast **alldrengilega.** En þó kom þar sem mælt er að ekki má
og segir **alldrengilega** frá för þeirra Þóris.
og sótti **alldrengilega.** Hjóst þá allmjög skjöldur Búa.
Varðist Þorbjörn þaðan **alldrengilega** með stokkinum
þeir Óspakur vörðust **alldrengilega.** Varð þeim þó handfátt
en hann varðist **alldrengilega.** Þar kom um síðir að þeir gátu drepið
Illugi varði þá báða **alldrenglega.** En Grettir var með öllu óvígur

Nýting í tungutækni

og annað er	lýtur	að fjórhjóladrifsbifreiðum
verknaði sem	lýtur	að framkvæmdinni sjálfri
um allt, sem	lýtur	að fundum húsfélaga
hvaðeina er	lýtur	að náttúruvernd
þess þáttar sem	lýtur	að sjúkdómsvörnum
Næsti hlekkur	lýtur	að því að ákveða
ef samþykki	lýtur	að því að ráðstafa
Þetta	lítur	alls ekki illa út
Þetta	lítur	auðvitað ekki vel út
þann sem	lítur	á heildina
Helga Kress	lítur	á íslenskar fornbókmenntir
kirkjan	lítur	á manninn sem eina heild
í stórum dráttum	lítur	á sig sem Íra
Vélin	lítur	á stöðuna
Íslenska óperan	lítur	á það sem sjálfsagðan hlut
þáttur umræðunnar	lýtur	án efa að þeirri samlíkingu
þannig að dæmið	lítur	betur út en áður

Nýting í kennslu

- Athuganir nemenda í verkefnum og ritgerðum:
  - athugun á beygingu
  - athugun á orðmyndun
  - athugun á orðasamböndum
  - athugun á setningagerð
  - athugun á merkingu orða
  - athugun á orðanotkun
  - athugun á myndmáli
  - athugun á stíleinkennum
    - o.s.frv.

Niðurstaða

- Rafrænir textar af ýmsu tagi eru mjög gagnlegir í rannsóknum og kennslu
- Nauðsynlegt er að kennarar og nemendur hafi gott aðgengi að fjölbreyttum textum
- Mikið skortir á slíkt aðgengi nú og það er hamlandi í kennslu og rannsóknum
- Bætt aðgengi að rafrænum textum myndi gerbreyta aðstöðu kennara og rannsakenda