Morphology and Corpora:
The Case of Quantitative Productivity

Marco Baroni (`baroni@sslmit.unibo.it`)

Granada Corpus Linguistics Seminars series
May 22-25 2006

General information

This seminar provides an overview of past and current work on quantitative productivity, one of the main areas of application of corpora in morphology.

After a theoretical introduction to the field, the seminar will feature a "hands-on" part, in which we will analyze the productivity of various classes of words in sections of the British National Corpus. Through this task, the participants will familiarize themselves with two powerful and free tools for corpus analysis: the IMS Corpus WorkBench[1] and the R statistical computing software.[2]

Handouts and papers on relevant topics will be made available during the meetings.

Detailed program

Corpora and their use in morphology

- Corpora: General overview

- Data in morphology (dictionaries and corpora)

- The "importance of low frequency events" dilemma

Productivity: theory and examples

- Introduction and motivation

- Lexical statistics and quantitative productivity

- Applications to English derivational morphology

---

[1] `http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/`
[2] `http://www.r-project.org/`

- Methodological issues: pre-processing, the "variable constant" problem, randomness

- The linguistic interpretation of quantitative productivity: parsing and productivity, productivity and semantic transparency, need and competition

- Conclusion: the past and future of quantitative productivity studies

Laboratory practice

- Introduction to the IMS Corpus WorkBench (CWB)

- Exploring a corpus with CWB/cqp

- Extraction of frequency lists

- Introduction to the basic functionalities of R

- Studying lexical richness and productivity with CWB and R