

USING CORPORA FOR LANGUAGE RESEARCH (COGS523 – 1)

Course Objectives:

Suitable for graduate students interested in doing theoretical or applied (computational, ELT, etc.) linguistic research using corpora.

The study of language using corpora. Usage of corpora within linguistics and cognitive science. Definition and varieties of corpora. Building a corpus: sampling, representativeness, encoding and annotation. Characteristics of major available corpora. Necessary statistics to interpret corpus data. Using corpora: corpora in psycholinguistics, corpora and syntax, semantics, and discourse; statistical natural language processing. Using (or writing tools) for corpus-based language studies. Conducting a research project on available corpora.

References:

1. McEnery, T. and A. Wilson. *Corpus Linguistics*. Edinburgh University Press, 2001 (2nd Edition, P98 .M32 2001, in 3-day reserve)
2. McEnery, T, R. Xiao, and Y. Tono. *Corpus-Based Language Studies: An Advanced Resource Book*, 2006 (P128 .C68 M37 2006, in 3-day reserve)
3. Sampson, G. and D. McCarthy (eds). *Corpus Linguistics: Readings in a Widening Discipline*, 2005.
4. Kennedy, G. *An Introduction to Corpus Linguistics*, Longman, 1998.
5. Farghaly, *Handbook for Language Engineers*, 2003.
6. Garside, R., G. Leech and T. McEnery (eds.). *Corpus Annotation: Linguistic Information from Computer Text Corpora*, Longman, 1997.
7. Biber, D. Conrad, S. and Reppen, R. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press, 1998
8. Oakes, M. P. *Statistics for Corpus Linguistics*, Edinburgh University Press, 1998.
9. Manning, C. D. and H. Schutze. *Foundations of Statistical Natural Language Processing*, MIT Press, 1999.
10. Mason, O. *Programming for Corpus Linguistics: How to do text analysis with Java*, Edinburgh University Press, 2000.

11. Meyer, C. English Corpus Linguistics: An Introduction. Cambridge University Press, 2002.

12. Various Research Articles.

Grading Policy:

Term Project: (progress report 10%, presentation and/or demo 15%, final project report 15%) 40%

Evaluation of a Corpus-Related Tool: (Presentation 10 %, Report 15%) 25%

Take Home Final Exam 25%

Participation (including online - presentation to the forum an exam question for 60% of the papers and inclass activities): 10%