MA in Computational Linguistics
Pre-Course Reading Suggestions

Doug Arnold

University of Essex

doug@essex.ac.uk

[You can also get this as a PS or pdf file]

The MA scheme in Computational Linguistics (CL) involves three different kinds of expertise, and getting some background in any of them will be a help, especially at the start of the scheme. You don't *need* to read any of the books mentioned here (still less try to look at all of them), but anything you do manage to do will probably help you.

Much of the course material is accessible via the World Wide Web, e.g. via my own web page: `http://privatewww.essex.ac.uk/~{}doug`.

Linguistics: Reading almost anything about linguistics will help. Time spent reading any reasonable introduction (and that means essentially any introduction) will be time well spent, as will time spent looking at descriptive grammars of English (such as Huddleston (1984), or Quirk et al. (1985), or Huddleston and Pullum (2002)). But perhaps the most useful basis for the course as a whole is good idea about phrase structure grammar, for which Borsley (1991), Borsley (1996) Sag and Wasow (1999), and Kim and Sells (2008) can be recommended.

Computational Aspects Most of the computational material on the course is approached via Prolog. There are many good introductions to Prolog. Among the best and easiest is Matthews (1998), which also has the advantage of being oriented towards linguistics. This is the recommended text book for the first part of the Prolog module. Another excellent book is Rogers (1986), which is sadly out of print, but will still exist in libraries. Dougherty (1994) is also a very gentle introduction, with a linguistic orientation.

Many books on Computational Linguistics give a broader perspective, looking at general issues, and avoiding implementation and other practical details. These can be useful, especially for giving background. Allen (1987), Arnold et al. (1993), and Smith (1991) are good examples.

A good, recent, and rather comprehensive text book is Jurafsky and Martin (2000). Especially, the new, second edition, is recommended.

On the course itself, apart from Matthews (1998), the Computational Linguistic books that will be most useful include Pereira and Shieber (1987), and Covington (1994), though Gazdar and Mellish (1989) contains a great deal of good material.

Computational Tools Following this degree scheme involves a good deal of practical work with computers in a Unix style environment using X-Windows. Any experience you can get with such an environment will be helpful. There are many good books providing introductions to Unix. You will soon become familiar with a collection of tools for writing programs, preparing documents, and so on. The most important of these are: Emacs, which is described very nicely in Cameron and Rosenblatt (1991), for editing (and very many other tasks), and LATEX for preparing documents Lamport (1994) (see `http://www.essex.ac.uk/linguistics/external/clmt/latex4ling/` for a longer list of readings on LATEX).

References

J. Allen. *Natural Language Understanding.* Benjamin Cummings Pub. Co., Inc., Menlo Park, Ca., 1987.

D.J. Arnold, Lorna Balkan, Siety Meijer, R.Lee Humphreys, and Louisa Sadler. *Machine Translation:*

an Introductory Guide. Blackwells-NCC, London, 1993.

Robert D. Borsley. *Syntactic Theory: a Unified Approach.* Edward Arnold, London, 1991.

Robert D. Borsley. *Modern Phrase Structure Grammar.* Number 11 in Blackwell textbooks in linguistics. Blackwell Publishers, 1996.

Deb Cameron and Bill Rosenblatt. *Learning GNU Emacs.* O'Reilly & Associates, Inc., Sebastopol, CA., 1991.

Michael A Covington. *Natural Language Processing for Prolog Programmers.* Prentice Hall, Englewood Cliffs, NJ, 1994.

Ray C. Dougherty. *Natural Language Computing.* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1994.

G. Gazdar and C. Mellish. *Natural Language Processing in Prolog.* Addison Wesley, Wokingham, 1989.

R. Huddleston. *Introduction to the Grammar of English.* CUP, Cambridge, 1984.

Rodney Huddleston and Geoffrey K Pullum, editors. *The Cambridge grammar of the English language.* Cambridge University Press, Cambridge, 2002.

Daniel Jurafsky and James H Martin. *Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall, New Jersey, 2000.

Jong-Bok Kim and Peter Sells. *English Syntax: An Introduction.* CSLI Pub., Stanford, Ca., 2008.

Leslie Lamport. *LaTeX — A Document Preparation System – User's Guide and Reference Manual.* Addison Wesley, New York, 1994.

Clive Matthews. *An Introduction to Natural Language Processing through Prolog.* Learning About Language Series. Longman, London, 1998.

Fernando C.N. Pereira and Stuart M. Shieber. *Prolog and Natural-Language Analysis*, volume 10 of *CSLI Lecture Notes.* Chicago University Press, Stanford, 1987. URL `http://www.mtome.com/Publications/pnla.html`.

R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. *A Comprehensive Grammar of the English Language.* Longman, London, 1985.

J.B. Rogers. *A Prolog Primer.* Addison-Wesley, Wokingham, 1986.

Ivan A Sag and Thomas Wasow. *Syntactic Theory: a Formal Introduction.* Number 92 in CSLI Lecture Notes. CSLI Publications, 1999.

George W Smith. *Computers and Human Language.* Oxford University Press, New York, 1991.