Introduction to Computational Linguistics I

Detmar Meurers, 684.01, Winter 2004

This introduction for graduates and advanced undergraduates provides:

- an introduction to theory-driven computational linguistics (sometimes referred to as "symbolic CL"), focusing on syntax/parsing
- some formal background
- practical experience implementing algorithms and small grammars, based on PROLOG

The course is part of the two course introduction to CL. The second half, 684.02, focuses on data-intensive, statistical CL and is offered by Chris Brew in Spring.

Organization (1)

Class meets: Tuesdays and Thursdays 3^{30}–5^{18}, 340 Central Classrooms

Course web page (overheads, etc.): http://ling.osu.edu/~dm/04/winter/684.01/

Course participants email list: 684.01@ling.osu.edu

Detmar's office hours and office location:

- Tuesdays, 12^{00}–13^{00} (or by appointment, best arranged by email)
- 201a Oxley Hall (tel. 292-0461)
- Email: dm@ling.osu.edu

Organization (2)

Course prerequisites: An understanding of

- why linguistic analysis?
- what is syntax? (LING 602.01 or equiv.)
- formal foundations (LING 680 or equiv.)

Successful course participation requires:

- Regular attendance and active participation
- Taking reading assignments serious and completing about six homework assignments, some paper and pencil, some programming in Prolog (handed out Thursday, completed Tuesday's class, discussed Thursday).
- Final project implementing a grammar fragment for a short (10 sentences) text of your choice, to be handed in Friday, March 12.

Course contents

Three aspects:

- data structures
- formalisms for expressing grammars using these data structures
- parsing algorithms for processing with those grammars

Outline

1. *Tue, 6. Jan.*: Organization/Introduction
2. *Thu, 8. Jan.*: Finite state machines and regular languages
3. *Tue, 13. Jan.*: Implementing finite state machines in Prolog
4. *Thu, 15. Jan.*: More on Prolog (recursion, negation) and implementing
5. *Tue, 20. Jan.*: Towards more complex grammar formalisms: Basic formal language theory
6. *Thu, 22. Jan.*: From context free grammars to definite clause grammars
7. *Tue, 27. Jan.*: What to encode in a grammar: A DCG for English
8. *Thu, 29. Jan.*: How to process with a grammar: Intro to Parsing
9. *Tue, 3. Feb.*: Basic parsing strategies

10. *Thu, 5. Feb.*: More efficient parsing strategies
11. *Tue, 10. Feb.*: Remembering sub-results: Well-formed substring tables
12. *Thu, 12. Feb.*: Remembering subcomputations: The active chart
13. *Tue, 17. Feb.*: More complex data structures: From atomic symbols to first order terms to feature structures
14. *Thu, 19. Feb.*: Term and feature structure unification
15. *Tue, 24. Feb.*: PATR-II Parsing with complex categories
16. *Thu, 26. Feb.*: Chart-Parsing with complex categories
17. *Tue, 2. Feb.*: Implementing a grammar in a typed feature structure based parsing system
18. *Thu, 4. Mar.*:
19. *Tue, 9. Mar.*:
20. *Thu, 11. Mar.*:

Reading material

A basic script as backbone to the material is on the course web page.

General background reading material:

- Gerald Gazdar and Chris Mellish (1989): *Natural Language Processing in Prolog*. Wokingham, England et al.: Addison-Wesley.
- Fernando Pereira and Stuart Shieber (1987): *Prolog and Natural-Language Analysis*. Stanford: CSLI Publications.
- Daniel Jurafsky and James H. Martin (2000): *Speech and Language Processing*. Upper Saddle River, NJ: : Prentice Hall.

These books and other assigned reading material can be found in 201 Oxley.

Reading assignment No. 1: Chapter 1 of Jurafsky & Martin (2000)