

Introduction to Computational Linguistics I

Detmar Meurers, 684.01, Winter 2003

This introduction for graduates and advanced undergraduates provides:

- an introduction to theory-driven CL topics (“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: Monday and Wednesday 9³⁰–11¹⁸, 060 Derby Hall

Course web page (overheads, etc.): <http://ling.osu.edu/~dm/03/winter/684.01/>

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

Detmar's office hours and office location:

- Monday after class (11³⁰–12³⁰), or by appointment
- 201a Oxley Hall (tel. 292-0461)
- Email: dm@ling.osu.edu

Organization (2)

Course prerequisites:

- Introduction to 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 weekly homework assignments, some paper and pencil, some programming in Prolog (handed out Wednesday, returned Monday, discussed Wednesday).
- Final project implementing a grammar fragment for a short (10 sentences) text of your choice, to be handed in Friday, March 14.

Course outline

1. *Mon, 6. Jan.:* Organization/Introduction
2. *Wed, 8. Jan.:* Finite state machines and regular languages
3. *Mon, 13. Jan.:* Implementing finite state machines in Prolog
4. *Wed, 15. Jan.:* More on Prolog (recursion, negation) and implementing
5. *Mon, 20. Jan.:* *Martin Luther King Day*
6. *Wed, 22. Jan.:* Towards more complex grammar formalisms:
Basic formal language theory
7. *Mon, 27. Jan.:* From context free grammars to definite clause grammars
8. *Wed, 29. Jan.:* What to encode in a grammar: A DCG for English
9. *Mon, 3. Feb.:* How to process with a grammar: Intro to Parsing
10. *Wed, 5. Feb.:* Basic parsing strategies
11. *Mon, 10. Feb.:* More efficient parsing strategies
12. *Wed, 12. Feb.:* Remembering sub-results: Well-formed substring tables

13. *Mon, 17. Feb.:* Remembering subcomputations: The active chart
14. *Wed, 19. Feb.:* More complex data structures:
From atomic symbols to first order terms to feature structures
15. *Mon, 24. Feb.:* Term and feature structure unification
16. *Wed, 26. Feb.:* Parsing with complex categories
17. *Mon, 3. Feb.:* Implementing a grammar in a typed feature structure based parsing system
18. *Wed, 5. Mar.:*
19. *Mon, 10. Mar.:*
20. *Wed, 12. Mar.:*

Three aspects:

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

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)