Introduction to Symbolic Computational Linguistics

Detmar Meurers, 684.01, Winter 2001

This introduction for graduates and advanced undergraduates provides:

- an introduction to symbolic CL topics, mainly syntax related
- formal background
- practical experience implementing algorithms and small grammars, based on PROLOG

The course is part of the two course introduction to CL, together with the Introduction to Statistical Approaches in Computational Linguistics (684.02, spring quarter, Chris Brew, Mike Brow)

Organization (1)

Class meets:

- Tuesday and Thursday 8^{30}–10^{18}
- 166 Ramseyer Hall

Course web page (slides, exercise sheets, . . .):

- http://www.ling.ohio-state.edu/~dm/2001/winter/684.01/

Course participants email list (installed by 8. Jan.):

- 684.01@ling.ohio-state.edu

Detmar's office hours and office location:

- Tuesday after class (10^{30}–12^{00}), or by appointment
- 201a Oxley Hall
- Tel.: 292-0461
- Email: dm@ling.ohio-state.edu

Organization (2)

Course prerequisites:

- Introductions to linguistics and syntax required.
- A course on formal methods in linguistics and previous experience with Prolog helpful.

Successful course participation requires:

- Regular attendance and active participation
  (20% of grade)
- Weekly homework assignments, some paper and pencil, some programming in Prolog
  (40 % of grade)
- Final project implementing a grammar fragment for a short text of your choice.
  (40 % of grade)

Course outline

1. *Tue, 2. Jan.*: Introduction
2. *Thu, 4. Jan.*: Finite state techniques I
3. *Tue, 9. Jan.*: Finite state techniques II
4. *Thu, 11. Jan.*: Finite state techniques III
5. *Tue, 16. Jan.*: Formal language theory I
6. *Thu, 18. Jan.*: Formal language theory II
7. *Tue, 23. Jan.*: DCGs as grammar formalism I
8. *Thu, 25. Jan.*: DCGs as grammar formalism II
9. *Tue, 30. Jan.*: Computability and complexity
10. *Thu, 1. Feb.*: Parsing strategies I
11. *Tue, 6. Feb.*: Parsing strategies II
12. *Thu, 8. Feb.*: Well-formed substring tables
13. *Tue, 13. Feb.*: The active chart
14. *Thu, 15. Feb.*: Unification
15. *Tue, 20. Feb.*: PATR-II
16. *Thu, 22. Feb.*: Typed feature structures I
17. *Tue, 27. Feb.*: Typed feature structures II

18. *Thu, 1. Mar.*: Parsing with unification grammars
19. *Tue, 6. Mar.*: Parsing as deduction
20. *Thu, 8. Mar.*: Wrap-up

Two aspects:

- declarative: ways of formulating grammars
- procedural: algorithms for parsing (or generating) with those grammars

Reading material

A basic script based on Chris Mellish, Pete Whitelock and Graeme Ritchie (1994): *Techniques in NLP 1*. Module Workbook, University of Edinburgh.

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.

The weekly reading material is generally made available in 201 Oxley Hall, in the shelf next to my office door.

Two perspectives on computational linguistics

From a practical perspective, computational linguistics provides the computational means to deal with spoken and written natural language:

- information extraction
- automatic translation
- natural language frontends to databases
- message generation
- spelling correction
- . . .

From a linguistic perspective, computational linguistics provides a possibility to

- formalize and computationally test linguistic theories
- obtain example data relevant to linguistic theorizing

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