ISCL Pro-/Hauptseminar
Summer Semester 2010, Prof. Meurers & Dr. Wunsch

Current Approaches to Dependency Parsing

Last update: July 8, 2010

Abstract:

Dependency grammar is an increasingly important grammar representation in computational linguistics, with a firm foundation in theoretical and descriptive linguistics. It is particularly well-suited for languages with relatively free word order and supports a transparent mapping from the string to a functor-argument representation needed for many computational applications. Dependency representations also are the new emerging standard for comparing the result of syntactic analysis across different grammar formalisms and parsing approaches.

This seminar will introduce the approaches to dependency parsing, existing implementations, and current research issues. Student projects will explore the training and evaluation of different parsing strategies, including the question of mapping existing treebank annotation schemes to dependency representations for training and testing of parsers.

Sessions

Note that the following session plan is subject to change; it only constitutes the current state of our planning as the semester unfolds.

1. Wednesday, April 14

   - Syllabus and course motivation

2. Monday, April 19

   - Introduction [DETMAR and HOLGER]
     - slides.pdf
   - Reading: Chapter 1 of Kübler et al. (2009), Chapter 1 of Mel'čuk (1988)

3. Wednesday, April 21

   - Introduction (continued)
   - Reading: Chapter 2 of Kübler et al. (2009)

4. Monday, April 26

   - Example dependency analyses and issues of converting constituency to dependency representation, using TüBa-DZ converted into a dependency bank (Versley 2005, originally created for) [HOLGER, DETMAR]

5. Wednesday, April 28

- Detecting errors in dependency annotation [DETMAR]
- Reading: Boyd et al. (2008)

6. Monday, May 3

   - First steps in Dependency Parsing [HOLGER]
   - Reading: Covington (2001)

7. Wednesday, May 5

   - Transition-Based Parsing [NIKO]
   - Reading: Chapter 3 from Kübler et al. (2009)

8. Monday, May 10

   - Transition-Based Parsing and the MaltParser [NIKO]
   - Reading: Nivre et al. (2007)

9. Wednesday, May 12

   - Transition-Based Parsing and the MaltParser (cont'd) [NIKO]
   - Reading: Nivre (2005b)
   - Hands-on-session: MaltParser [HOLGER, NIKO]
   - Reading: MaltParser user's manual, Hall et al. (2009)
     http://www.sfs.uni-tuebingen.de/~wunsch/lehre/ss-10/dep/ex1.pdf

10. Monday, May 17

    - Arc-Factored Parsing [JULIA, JOHANNES]
    - Reading: Chapter 4 from Kübler et al. (2009)

11. Wednesday, May 19

    - Arc-Factored Parsing (continued)

12. *Monday, May 24: Holiday*

13. *Wednesday, May 26: Holiday*

14. Monday, May 31

    - Learning arc-factored models [JULIA, JOHANNES]
    - MaltParser hands-on session: discussion of results

15. Wednesday, June 2

    - Linguistic features for dependency parsing [ANAS]
    - Reading: Nivre (2005a)

16. Monday, June 7

    - Grammar-based and constraint-based dependency parsing [JANINA]

- Reading: Chapter 5 from Kübler et al. (2009)

17. Wednesday, June 9

    - Grammar- and constraint-based dependency parsing (continued)
    - Discussion of exercise 2 (Implementation of Eisner's algorithm)
      http://www.sfs.uni-tuebingen.de/~wunsch/lehre/ss-10/dep/ex2.pdf

18. Monday, June 14

    - Grammar- and constraint-based dependency parsing (continued)
    - Reading: Weighted Constraint Dependency Grammar parser (WCDG, Foth et al. 2004)

19. Wednesday, June 16

    - Hands-on-session: WCDG parser
    - Reading: Menzel (2010)
    - **Location: Computer Pool, 2.26**

20. Monday, June 21

    - Evaluation [MARIA]
    - Reading: Chapter 6 from Kübler et al. (2009), Carroll et al. (2003)

21. Wednesday, June 23

    - Evaluation (continued) [MARIA]

22. Monday, June 28

    - Constraint Grammar, CG3 [RAMON]
    - Reading: Tapanainen & Järvinen (1997), Bick (2005)

23. Wednesday, June 30

    - Hands-on-session: Constraint Grammar (CG 3) [HOLGER, RAMON]
    - **Location: Computer Pool, 2.26**
      http://www.sfs.uni-tuebingen.de/~wunsch/lehre/ss-10/dep/ex4.pdf

24. Monday, July 5

    - Dependency Parsing of Long-Distance Dependencies [PAOLINA]
    - Reading: Schneider (2008)

25. Wednesday, July 7

    - Dependency Parsing of Long-Distance Dependencies (cont.)

26. Monday, July 12

    - Extensible Dependency Grammar (XDG) [ADRIANE]

- Reading: Debusmann (2006)
- Adriane Boyd's thesis research

27. Wednesday, July 14

- Dependency annotation for learner data [HOLGER, DETMAR]
- Reading: Dickinson & Ragheb (2009)
- wrapping up

Instructors:

- Detmar Meurers
  - *Office:* Room 1.28, Blochbau (Wilhelmstr. 19)
  - *Email:* dm@sfs.uni-tuebingen.de
  - *Office hours:* Mondays 11:30–12:30

- Holger Wunsch
  - *Office:* Room 1.26, Blochbau (Wilhelmstr. 19)
  - *Email:* wunsch@sfs.uni-tuebingen.de
  - *Office hours:* Tuesdays, 14:00–15:00

Course meets: in Seminarraum 1.13, Blochbau (Wilhelmstr. 19)
- Mondays, 16ct-18
- Wednesdays, 16ct-18

Credits and Campus:
- Register in Campus between June 1 and 30
- Credit Points: 6 (BA ISCL), 10 (MA ISCL)

Syllabus (this file):

- html-Version (http://purl.org/dm/10/ss/dep)

- pdf-Version (http://purl.org/dm/10/ss/dep/syllabus.pdf)

Moodle page: http://moodle01.zdv.uni-tuebingen.de/course/view.php?id=347

Nature of course and our expectations: This is a Pro-/Hauptseminar intended to provide an overview of the current approaches in this active research area. Each participant is expected to

1. regularly and actively participate in class, read the papers assigned by any of the presenters and post a question on Moodle to the "Reading Discussion Forum" on each reading *at the latest on the day before it is discussed* in class. (30% of grade for Hauptseminar, 50% for Proseminar)

    Note: Following the rules of the Neuphilologische Fakultät, missing more than two meetings unexcused, automatically results in failing the class.

2. explore and present a topic (30% of grade for Hauptseminar, 50% for Proseminar):

   - select one of the sub-topics during the first week of the semester
   - thoroughly research the topic, taking our literature pointers *as a starting point*
   - prepare the presentation with slides and discuss the presentation with Holger or Detmar during the office hours *a week before the presentation*
   - start a new Moodle thread on the "Reading Discussion Forum" specifying what every course participant should read to prepare for your presentation *a week before your presentation*
   - present the topic in class

3. for a Hauptseminar Schein, work out a term paper (40% of grade for Hauptseminar)

   - select a topic and submit a one-page abstract *by the end of Friday, July 16.*
     - The term paper topic must be related to the seminar topic.
   - email the term paper in pdf format to both instructors *before the beginning of the next semester, i.e., by September 30, 2010.*
     - Note for ISCL students: The term paper must be produced in LaTeX, and BibTex must be used for the bibliography.

Academic conduct and misconduct: Research is driven by discussion and free exchange of ideas, motivations, and perspectives. So you are encouraged to work in groups, discuss, and exchange ideas. At the same time, the foundation of the free exchange of ideas is that everyone is open about where they obtained which information. Concretely, this means you are expected to always make explicit when you've worked on something as a team – and keep in mind that being part of a team always means sharing the work.

For text you write, you always have to provide explicit references for any ideas or passages you reuse from somewhere else. Note that this includes text "found" on the web, where you should cite the url of the web site in case no more official publication is available.

Class etiquette: Please do not read or work on materials for other classes in our seminar. Come to class on time and do not pack up early. When our seminar meets in the computer lab, only use the computers when you are asked to do a specific activity – do not read email or browse the web. All portable electronic devices such as cell phones should be switched off for the entire length of the flight, oops, class. If for some reason, you must leave early or you have an important call coming in, or you have to miss class for an important reason, please let Detmar or Holger know *before* class.

References

Bick, E. (2005). Turning Constraint Grammar Data into Running Dependency Treebanks. In M. Civit, S. Kübler & M. A. Martí (eds.), *Proceedings of the 4th Workshop on Treebanks and Linguistic Theory*. Barcelona, pp. 19–27.

Boyd, A., M. Dickinson & D. Meurers (2008). On Detecting Errors in Dependency Treebanks. *Research on Language and Computation* 6(2), 113–137. URL http://purl.org/dm/papers/boyd-et-al-08.html.

Carroll, J., G. Minnen & T. Briscoe (2003). Parser evaluation: using a grammatical relation annotation scheme. In A. Abeillé (ed.), *Treebanks: Building and Using Parsed Corpora*, Dordrecht: Kluwer.

Covington, M. A. (2001). A Fundamental Algorithm for Dependency Parsing. In J. A. Miller & J. W. Smith (eds.), *Proceedings of the 39th Annual ACM Southeast Conference*. pp. 95—-102. URL http://www.stanford.edu/~mjkay/covington.pdf.

Debusmann, R. (2006). Extensible Dependency Grammar: A Modular Grammar Formalism Based on Multigraph Dimension. Ph.D. thesis, Universität des Saarlandes.

Dickinson, M. & M. Ragheb (2009). Dependency Annotation for Learner Corpora. In *Proceedings of the Eighth Workshop on Treebanks and Linguistic Theories (TLT-8)*. Milan, Italy. URL http://jones.ling.indiana.edu/~mdickinson/papers/dickinson-ragheb09.html.

Foth, K. A., M. Daum & W. Menzel (2004). A broad-coverage parser for German based on defeasible constraints. In *In KONVENS 2004, Beiträge zur 7. Konferenz zur Verarbeitung natürlicher Sprache*. pp. 45–52. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.6728&rep=rep1&type=pdf.

Hall, J., J. Nilsson & J. Nivre (2009). *Homepage of MaltParser*. URL http://maltparser.org.

Kübler, S., R. McDonald & J. Nivre (2009). Dependency Parsing. In G. Hirst (ed.), *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. State University of New York Press.

Menzel, W. (2010). CDG homepage. URL http://nats-www.informatik.uni-hamburg.de/view/CDG/WebHome.

Nivre, J. (2005a). *Dependency Grammar and Dependency Parsing*. MSI report 05133, Växjö University: School of Mathematics and Systems Engineering. URL http://stp.lingfil.uu.se/~nivre/docs/05133.pdf.

Nivre, J. (2005b). Pseudo-Projective Dependency Parsing. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. URL http://w3.msi.vxu.se/~nivre/papers/acl05.pdf.

Nivre, J., J. Nilsson, J. Hall, A. Chanev, G. Eryigit, S. Kübler, S. Marinov & E. Marsi (2007). MaltParser: A Language-Independent System for Data-Driven Dependency Parsing. *Natural Language Engineering* 13(1), 1–41. URL http://w3.msi.vxu.se/~nivre/papers/nle07.pdf.

Schneider, G. (2008). Hybrid long-distance functional dependency parsing. Ph.D. thesis, University of Zurich. URL https://www.zora.uzh.ch/7188/.

Tapanainen, P. & T. Järvinen (1997). A non-projective dependency parser. In *Proceedings of the fifth conference on Applied natural language processing*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 64–71. URL www.ldc.upenn.edu/acl/A/A97/A97-1011.pdf.

Versley, Y. (2005). Parser Evaluation across Text Types Fourth Workshop on Treebanks and Linguistic Theories. In *Proceedings of the Fourth Workshop on Treebanks and Lin-*

guistic Theories (TLT-05). Barcelona, Spain. URL http://www.versley.de/versley_tlt05.pdf.