OSU Department of Linguistics
Spring 2001
W. Detmar Meurers

LING 795K
Tue and Thu 15^{30}–17^{18}
211 Botany & Zoology

Constraint-based processing and advanced parsing techniques

Course web page:
- http://www.ling.ohio-state.edu/~dm/2001/spring/795K/

Course participants email list:
- 795k@ling.ohio-state.edu

Office hours:
- Tuesdays, 14^{00}–15^{00} or by appointment
- 201a Oxley Hall
- Tel.: 292-0461
- Email: dm@ling.ohio-state.edu

Course prerequisites:
- 684.01 (or other background on syntactic processing)

Successful course participation requires:
- Regular attendance and active participation (20% of grade)
- Presentation of a topic in class (30% of grade)
- Writing of a research paper (50% of grade)
  due date?

Dates
1. *Tue, 27. March*: Introduction
2. *Thu, 29. March*: Chunk parsing (Erhard)
   read before seminar: Abney (1995, 1996)
3. *Tue, 3. April*: Processing with HPSG linguistics (Detmar)
   read before seminar: Meurers (1994)
4. *Thu, 5. April*: Processing with HPSG linguistics (Detmar)
5. *Tue, 10. April*:
6. *Thu, 12. April*: AT UPENN, RESCHEDULE?
7. *Tue, 17. April*:
8. *Thu, 19. April*:
9. *Tue, 24. April*:
10. *Thu, 26. April*:
11. *Tue, 1. May*:
12. *Thu, 3. May*:
13. *Tue, 8. May*:
14. *Thu, 10. May*:
15. *Tue, 15. May*:
16. *Thu, 17. May*:
17. *Tue, 22. May*:
18. *Thu, 24. May*:
19. *Tue, 29. May*:
20. *Thu, 31. May*:

Topics

1. Processing HPSG theories as close as possible to linguistics:
   - **Detmar:** Implementing an HPSG theory in a phrase structure backbone system: Meurers (1994)
   - Compiling HPSG theories into logic programs:
     Götz and Meurers (1995, 1997, 1999); Götz (1999)
   - **Wesley:** Lexical Rules: Meurers and Minnen (1997)
2. Advanced parsing techniques:
   - **Vanessa:** Head-Driven Parsing:
     - primary: van Noord (1997)
     - background: Kay (1990); Sikkel and op den Akker (1992); Bouma and van Noord (1993); Nederhof and Satta (1994)
   - GLR Parsing
     - primary: Caroll (1993, chs. 4 and 5)
     - background: Aho and Ullman (1977, ch. 6), Tomita (1986, 1987); Lang (1991)
     - chart parsing alternative to Tomita: Nederhof and Satta (1996)
   - Improved Earley Parsing: Nerhof and Satta (1997)
   - Indexing, etc.: Kiefer et al. (1999), Malouf et al. (2000), Götz et al. (2001)
3. **Mike:** Encoding of Typed Feature Structures: Penn (2000, ch. 8)
4. **Anton:** Packing ambiguity: Griffith (1996), Oepen and Caroll (2000), Lavie and Penstein Rosé (2000)
5. Parser evaluation:
   - starting points: Caroll and Oepen (2000); Caroll et al. (1998)
   - others: Caroll (1994); Lin (1995); Caroll et al. (2000)
   - general considerations and relation to linguistic theory: Berwick and Weinberg (1982)
6. Preprocessing and combined systems:
   - **Anton:** Basili et al. (2000); Kay (2000); Neumann et al. (2000)
   - **Erhard:** Chunk parsing: Abney (1995), Abney (1996) (later appeared as Abney, 1997)
   - **Markus:** Approximations:
     - CFG approximation of HPSG (Kiefer and Krieger, 2000)
     - Regular approximation of CFLs (Nederhof, 2000)
7. **Anna and Arantxa:** Parsing with linearization grammars:
   - LSL: Suhre (1999); Götz and Penn (1997); Fouvry and Meurers (2000)
   - DPSG: van der Sloot (1999), RDPSG Vogel and Erjavec (1994)
   - various: Bunt and van Horck (1996), Johnson (1985); Reape (1991); van Noord (1991); Müller (1996)

References

Abney, Steven (1995). Chunks and Dependencies: Bringing Processing Evidence to Bear on Syntax. In J. Cole, G. M. Green, and J. L. Morgan (Eds.), *Linguistics and Computation*, pp. 145–164. Stanford, CA: CSLI Publications.

Abney, Steven (1996). Partial Parsing via Finite-State Cascades. In *Proceedings of the ESSLLI '96 Robust Parsing Workshop*.

Abney, Steven (1997). Partial Parsing via Finite-State Cascades. *Natural Language Engineering 2*(4), 337–344.

Aho, Alfred V. and Ullman, Jeffrey D. (1977). *Principles of Compiler Design*. Reading, MA: Addison Wesley.

Basili, R., Pazienza, M.T., and Zanzotto, F.M. (2000). Customizable Modular Lexicalized Parsing. In *Proceedings of the Sixth International Workshop on Parsing Technologies (IWPT 2000)*, pp. 41–52.

Berwick, Robert C. and Weinberg, Amy S. (1982). Parsing Efficiency, Computational Complexity, and the Evaluation of Grammatical Theories. *Linguistic Inquiry 13*(2), 165–191.

Bouma, Gosse and Noord, Gertjan van (1993). Head-Driven Parsing for Lexicalist Grammars: Experimental Results. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics (EACL 93)*.

Bunt, Harry and Horck, Arthur van (Eds.) (1996). *Discontinuous Constituency*, Volume 6 of *Natural Language Processing*. Berlin and New York, NY: Mouton de Gruyter.

Caroll, John, Briscoe, Ted, and Sanfilippo, Antonio (1998). Parser evaluation: a survey and a new proposal. In *Proceedings of the 1st International Conference on Language Resources and Evaluation*, pp. 447–454. Granada, Spain.

Caroll, John, Minnen, Guido, and Briscoe, Ted (2000). Parser evaluation: using a grammatical relation annotation scheme. In *Building and using syntactically annoted corpora*. Kluwer Academic Publishers.

Caroll, John and Oepen, Stephan (2000). Efficient Large Scale Parsing: A survey. In J. Caroll, R. C. Moore, and S. Oepen (Eds.), *Proceedings of the workshop on Efficiency in Large-Scale Parsing Systems held at the 18th International Conference on Computational Linguistics (COLING-2000)*, pp. 7–12.

Caroll, John Andrew (1993). *Practical Unification-based Parsing of Natural Language*. Ph. D. thesis, Computer Laboratory, University of Cambridge.

Caroll, John Andrew (1994). Relating complexity to practical performance in parsing with wide-coverage unification grammars. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL 94)*, pp. 287–294. Las Cruces, NM.

Fouvry, Frederik and Meurers, Detmar (2000). Towards a platform for linearization grammars. In E. W. Hinrichs, D. Meurers, and S. Wintner (Eds.), *Proceedings of the Workshop on Linguistic Theory and Grammar Implementation*, pp. 153–168. Birmingham, UK: ESSLLI 2000.

Götz, Thilo (1999). *Feature Constraint Grammars*. Ph. D. thesis, Seminar für Sprachwissenschaft, Universität Tübingen, Tübingen. Published 2000 as Volume 148 in Arbeitspapiere des SFB 340.

Götz, Thilo, Lougee-Heimer, Robin, and Nicolov, Nicolas (2001). Efficient Indexing for Typed Feature Structures. In *Ms. IBM Watson Research Center*.

Götz, Thilo and Meurers, Walt Detmar (1995). Compiling HPSG Type Constraints into Definite Clause Programs. In *Proceedings of the 33rd Annual Meeting of the ACL*, pp. 85–91. MIT, Cambridge, MA.

Götz, Thilo and Meurers, Walt Detmar (1997). Interleaving Universal Principles and Relational Constraints over Typed Feature Logic. In *Proceedings of the 35th Annual Meeting of the ACL and 8th Conference of the EACL*, pp. 1–8. Universidad Nacional de Educación a Distancia, Madrid.

Götz, Thilo and Meurers, Walt Detmar (1999). The importance of being lazy – Using lazy evaluation to process queries to HPSG grammars. In G. Webelhuth, J.-P. Koenig, and A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, Studies in Constraint-Based Lexicalism, pp. 249–264. Stanford, CA: CSLI Publications. Published version of a paper presented at the 3rd Int. Conference on HPSG, May 1996, Marseille.

Götz, Thilo and Penn, Gerald (1997). A Proposed Linear Specification Language. Technical report, SFB 340, Universität Tübingen.

Griffith, John (1996). Modularizing Contexted Constraints. In *Proceedings of the 16th Conference on Computational Linguistics (COLING-96)*.

Johnson, Mark (1985). Parsing with discontinuous constituents. In *Proccedings of the 23rd Annual Meeting of the ACL*, pp. 127–132. Chicago, IL.

Kay, Martin (1990). Head-Driven Parsing. In M. Tomita (Ed.), *Current Issues in Parsing Technology*. Dordrecht: Kluwer Academic Publishers. Previously published in the proceedings of the International Workshop on Parsing Technologies, 1989.

Kay, Martin (2000). Guides and Oracles for Linear-Time Parsing. In *Proceedings of the Sixth International Workshop on Parsing Technologies (IWPT 2000)*, pp. 6–9.

Kiefer, Bernd and Krieger, Hans-Ulrich (2000). A Context-Free Approximation of Head-Driven Phrase Structure Grammar. In *Proceedings of the Sixth International Workshop on Parsing Technologies (IWPT 2000)*, pp. 135–146.

Kiefer, Bernd, Krieger, Hans-Ulrich, Caroll, John, and Malouf, Rob (1999). A bag of useful techniques for efficient and robust parsing. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pp. 473–480. University of Maryland, College Park.

Lang, Bernard (1991). Towards a uniform formal framework for parsing. In M. Tomita (Ed.), *Current Issues in Parsing Technology*, pp. 153–172. Dordrecht: Kluwer Academic Publishers.

Lavie, Alon and Penstein Rosé, Carolyn (2000). Optimal Ambiguity Packing in Context-Free Parsers With Interleaved Unification. In *Proceedings of the Sixth International Workshop on Parsing Technologies (IWPT 2000)*, pp. 147–158.

Lin, Dekang (1995). A Dependency-based Method for Evaluating Broad-Coverage Parsers. In *Proeedings of the 14th International Joint Conference on AI (IJCAI-95).* Montreal, Quebec, Canada.

Malouf, Rob, Caroll, John, and Copestake, Ann (2000). Efficient feature structure operations without compilation. *Natural Language Engineering 6*(1), 29–46.

Meurers, Walt Detmar (1994). On Implementing an HPSG Theory – Aspects of the Logical Architecture, the Formalization, and the Implementation of Head-Driven Phrase Structure Grammars. In E. W. Hinrichs, W. D. Meurers, and T. Nakazawa (Eds.), *Partial-VP and Split-NP Topicalization in German – An HPSG Analysis and its Implementation*, Arbeitspapiere des SFB 340 Nr. 58, pp. 47–155. Tübingen: Universität Tübingen.

Meurers, Walt Detmar and Minnen, Guido (1997). A Computational Treatment of Lexical Rules in HPSG as Covariation in Lexical Entries. *Computational Linguistics 23*(4), 543–568.

Müller, Stefan (1996). The Babel-System–An HPSG Prolog Implementation. In *Proceedings of the Fourth International Conference on the Practical Application of Prolog*, pp. 263–277. London and New York, NY. http://www.dfki.de/˜stefan/Pub/babel.html.

Nederhof, Mark-Jan (2000). Regular approximation of CFLs: a grammatical view. In H. Bunt and A. Nijholt (Eds.), *Advances in Probabilistic and other Parsing Technologies*, pp. 221–241. Dordrecht: Kluwer Academic Publishers.

Nederhof, Mark-Jan and Satta, Giorgio (1994). An Extended Theory of Head-Driven Parsing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL 94)*, pp. 210–217. Las Cruces, NM.

Nederhof, Mark-Jan and Satta, Giorgio (1996). Efficient Tabular LR Parsing. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL 96)*, pp. 239–246. Santa Cruz, CA.

Nerhof, Mark-Jan and Satta, Giorgio (1997). A variant of Earley parsing. In *AI*IA 97: Advances in Artificial Intelligence, 5th Congress of the Italian Association for Artificial Intelligence*, Volume 1321 of *Lecture Notes in Artificial Intelligence*, pp. 84–95. Berlin: Springer-Verlag.

Neumann, Günter, Braun, Christian, and Piskorski, Jakub (2000). A Divide-and-Conquer Strategy for Shallow Parsing of German Free Texts. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 00)*, pp. 239–246. Seattle, Washington.

Oepen, Stephan and Caroll, John (2000). Ambiguity packing in constraint-based parsing—practical results. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 00)*, pp. 162–169. Seattle, WA.

Penn, Gerald (2000). *The Algebraic Structure of Attributed Type Signatures.* Ph. D. thesis, School of Computer Science, Carnegie Mellon University.

Reape, Mike (1991). Parsing Bounded Discontinuous Constituents: Generalisations of some common algorithms. In M. Reape (Ed.), *Word Order in Germanic and Parsing*, DYANA Deliverable R1.1.C, ESPRIT Basic Research Action BR 3175, pp. 41–70. Edinburgh: Centre for Cognitive Science, University of Edinburgh.

Sikkel, Klaas and Akker, Rieks op den (1992). Head-Corner Chart Parsing. In *Proceedings of Computing Science in the Netherlands (CSN'92).* Utrecht.

Suhre, Oliver (1999). Computational Aspects of a Grammar Formalism for Languages with Freer Word Order. Diplomarbeit, Department of Computer Science, University of Tübingen. Published 2000 as Volume 154 in Arbeitspapiere des SFB 340.

Tomita, Masaru (1986). *Efficient parsing for natural language: a fast algorithm for practical systems.* Boston: Kluwer Academic Publishers.

Tomita, Masaru (1987). An efficient augmented-context-free parsing algorithm. *Computational Linguistics 13*(1), 31–46.

Sloot, Ko van der (1999). The TENDUM 2.7 parsing algorithm for DPSG. Technical report, ITK, Tilburg, The Netherlands.

Noord, Gertjan van (1991). Head Corner Parsing for Discontinuous Constituency. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp. 114–121.

Noord, Gertjan van (1997). An Efficient Implementation of the Head-Corner Parser. *Computational Linguistics 23*(3), 425–456.

Vogel, Carl M. and Erjavec, Tomaz (1994). Restricted Discontinuous Phrse Structure Grammar and Its RAmifications. In C. Martín-Vide (Ed.), *Current Issues in Mathematical Linguistics*, pp. 131–140. Elsevier Science B.V.