

Language and Computation foundational course: Case, Scrambling and Default Word Order

Teachers

- Miriam Butt (miriam.butt@uni-konstanz.de)
- Heike Zinsmeister (heike.zinsmeister@uni-konstanz.de)

Abstract

Many of the world's languages are so-called "free word order" languages, whereby the major arguments of a clause can be scrambled quite freely. This scrambling generally goes hand-in-hand with a robust case marking system and some means of verb-argument agreement (usually verb-subject agreement, but not always), which allows the identification of the various arguments of the clause (i.e., which is the agent, the patient, the goal, etc.). Sometimes, however, the correct identification of which syntactic argument encodes which of the semantic participants of a verb/clause can only be achieved by world or contextual knowledge. Additionally, effects of so-called word order freezing can be observed, whereby suddenly the word order is not free, but is fixed if one wants a certain mapping of semantic participants to syntactic arguments. Finally, one generally also refers to a "default word order" exhibited by languages which in principle allow for the (more-or-less) free scrambling of syntactic arguments. The theoretical status of this default word order is not clear and this course will examine the topic of argument scrambling, word order freezing and default word order with respect to two main perspectives: 1) theoretical linguistics; 2) corpus linguistics.

With respect to the theoretical perspective, students will be introduced to current theories of case and word order so that we can examine what (if anything) these theories have to say about default word order and word order freezing in particular. With respect to the computational perspective, we will examine to what degree information from corpora can help guide the analysis and help us understand why things are scrambled when they are scrambled and what status the "default word order" actually has in terms of frequency and distribution. As part of the course (one day), we will also present psycholinguistic studies that identify (combinations of) features that determine word order preferences.

For additional material and references see:

<http://ling.uni-konstanz.de/pages/home/zinsmeister/essli2009/>

Monday July, 27: Introduction: Word Order in German and Urdu

Background reading:

- Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford, CA: CSLI Publications. Chapter 2.

Electronically available in the ESSLLI reader:

- Müller, Gereon. 1999. Optimality, Markedness, and Word Order in German. *Linguistics* 37, 777–818. (pre-print version: 02-Mueller1999.pdf)
- Müller, Stefan. 2002. Multiple Frontings in German. In Gerhard Jäger, Paola Monachesi, Gerald Penn, and Shuly Winter (eds.): *Proceedings of Formal Grammar*, Trento, Italy. (03-Mueller2002.pdf)

Tuesday July, 28: Case Theory and Linking

Background reading:

- Butt, Miriam. 2006. *Theories of Case*. Textbook. Cambridge University Press.

Wednesday, July 29: Corpora

Electronically available in the ESSLLI reader:

- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. Predicting the Dative Alternation. In G. Boume, I. Kraemer, and J. Zwarts (eds.): *Cognitive Foundations of Interpretation*, Amsterdam: Royal Netherlands Academy of Science, pp. 69–94. (pre-print version: 04-BresnanEtAl2007.pdf)
- Evert, Stefan. 2006. How random is a corpus? The library metaphor. *Zeitschrift für Anglistik und Amerikanistik*, 54 (2), 177–190. (pre-print version: 05-Evert2006.pdf)
- Lüdeling, Anke, Stefan Evert, and Marco Baroni. 2007. Using Web data for linguistic purposes. In M. Hundt, N. Nesselhauf, and C. Biewer (eds.): *Corpus Linguistics and the Web*, volume 59 of Language and Computers - Studies in Practical Linguistics. Amsterdam, New York: Rodopi, pp. 7–24. (pre-print version: 06-LuedelingEvertBaroni2007.pdf)
- Meurers, Detmar W. 2005. On the use of electronic corpora for theoretical linguistics. Case studies from the syntax of German. *Lingua*. 115 (11), 1619–1639. (pre-print version: 07-Meurers2005.pdf)

Thursday, July 30, and Friday, July 31: Results of Corpus Studies and Psycholinguistic Evidence

Electronically available in the ESLLI reader:

- Bader, Markus and Jana Häussler (in press) Word Order in German: A Corpus Study. To appear in *Lingua*. (pre-print version: 08-BaderHaeussler.pdf)
- Cahill, Aoife, Martin Forst, and Christian Rohrer. 2007. Stochastic Realisation Ranking for a Free Word Order Language. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*. Saarbrücken, Germany, pp. 17–24. (09-CahillForstRohrer2007.pdf)
- Filippova, Katja and Strube, Michael. 2007. Generating Constituent Order in German Clauses. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague, Czech Republic, pp. 320–327. (10-FilippovaStrube2007.pdf)
- Forst, Martin. 2007. Filling Statistics with Linguistics - Property Design for the Disambiguation of German LFG Parses. In *Proceedings of the ACL Workshop on Deep Linguistic Processing*, Prague, Czech Republic. (11-Forst2007.pdf)
- Keller, Frank, 2000. Evaluating competition-based models of word order. In: Gleitman, L. R., Joshi, A. K. (eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 747–752. (12-Keller2000.pdf)
- Patil, Umesh, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry, and Shravan Vasishth. 2008. Focus, Word Order and Intonation in Hindi. *Journal of South Asian Linguistics* 1 (1), 55–72. (13-PatilEtAl2008.pdf)
- Schulte im Walde, Sabine. 2002. A Subcategorisation Lexicon for German Verbs induced from a Lexicalised PCFG. *Proceedings of the 3rd Conference on Language Resources and Evaluation (LREC)*. Las Palmas de Gran Canaria, Spain. (14-SchulteImWalde2002.pdf)