# The Structure, Computational Segmentation, and Translation to English of German Nominal Compounds

**Nori Heikkinen**
Computer Science Department
Swarthmore College
Swarthmore, Pennsylvania, USA
nori@sccs.swarthmore.edu

## Abstract

Among the most daunting words in German to a non-native speaker are the longest ones – those that near forty characters in length, and are comprised of many smaller words. A problem posed by these words is that of automatic segmentation into pieces in order to glean the meaning of the entire word. It is vital that these words be correctly parsed, interpreted, and translated as part of the ongoing work in multilingual computation linguistics, such as incorporation into German language spell-checkers.

Here, I describe the problem posed to morphological disambiguation in compounding languages, concentrating on the structure of German nominal compounds. Using regularities about their structure, I then devise a computational model for both determining linking elements of morphemes, and segmenting the words correctly. I discuss the inherent problems with such a model, and speculate about alternative approaches.

## 1 Note

This thesis was submitted as part of the requirements for the Bachelor of Arts in Linguistics and Computer Science. A full copy of this thesis can be obtained from the Linguistics Department at Swarthmore College.

## References

Becker, Thomas. Compounding in German. *Rivista di Linguistica*, 4(1):5-36, 1992.

Brill, Eric. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging. *Computational Linguistics*, 21(4):543-565, 1995.

Fleischer, Wolfgang. *Wortbildung der deutschen Gegenwartssprache.* VEB Bibliographisches Institut, Leipzig, 1969.

Jurafsky, Daniel and Martin, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, New Jersey, 2000.

Ngai, Grace and Florian, Radu. fnTBL. `http://nlp.cs.jhu.edu/~rflorian/fntbl/`.

Ngai, Grace and Florian, Radu. Transformation-Based Learning in the Fast Lane. In *Proceedings of NAACL-2001*, pages 40–47, 2001.

Ortner, Hanspeter and Ortner, Lorelies. *Zur Theorie und Praxis der Kompositaforschung*. Gunter Narr Verlag, Tübingen, 1984.



Appeared in: *Proceedings of the Class of 2005 Senior Conference*, page 15
Computer Science Department, Swarthmore College