

Semantic distance and semantic transparency as determinants of compound stress assignment in English

Sabine Arndt-Lappe, Katja Böer & Ingo Plag
Universität Siegen

It has long been known that English compound stress assignment is subject to variation. Whereas the majority of binary noun-noun compounds is left-prominent, there is a substantial number of compounds which are right-stressed (cf. for example, *ópera glasses*, *wáтч-maker* and *Óxford Street* versus *steel bridge*, *morning páper*, and *Madison Avenue*). Among other factors, semantics has traditionally been assumed to play a role in explaining variation in compound stress assignment (e.g. Sampson 1980, Fudge 1984, Ladd 1984, Liberman & Sproat 1992, Sproat 1994, Olsen 2000, 2001, Spencer 2003, Giegerich 2006, Plag et al. 2007, 2008). However, the nature of these semantic factors is unclear. Why should certain semantic categories or relations favor rightward stress, while other categories or relations prefer leftward stress? In this paper we present the results of two studies that investigate potential semantic effects on compound stress assignment.

In order to better understand the semantic effects proposed in the literature, we first test an alternative to existing semantically based approaches. Whereas in the literature a large set of quite heterogeneous categories has been proposed to be relevant for compound stress assignment, we investigate the idea that semantic closeness between constituents is the underlying variable. For example, a locative relation between the constituents is less close than an argument-head relation, with pertinent consequences for stress assignment. Thus, a larger semantic distance would be expected to correlate with a higher probability of rightward stress (in line with predictions, for example, by Giegerich 2004, 2006). To study this, we employ the differentiation of conceptual relations as formalised in frame semantics as represented in the Berkeley FrameNet database (Fillmore & Baker, <http://framenet.icsi.berkeley.edu/>) to categorise the semantic relations between compound constituents. Our analysis shows that semantic closeness is indeed relevant in compound stress assignment – but not in the way expected by much of the literature. Thus, what matters is how close the second noun is to the first noun, but not how close the first noun is to the second.

Another kind of influence on stress assignment in compounds is lexicalization (e.g. Giegerich 2004). It has been shown by Plag and colleagues (2007, 2008) that the frequency and spelling of compounds (as one or two words) can be taken as a proxy of lexicalization and that more lexicalized compounds tend more towards leftward stress assignment. It is unclear, however, why such a correlation should exist in the first place. A semantic explanation suggests itself, to the effect that it is semantic opacity that goes together with leftward stress. In this paper we investigate both types of semantic effect.

To test semantic opacity, we use Latent Semantic Analysis (LSA, Landauer et al. 1998 et seq.) to measure the semantic transparency of some 10,000 compounds from different corpora. LSA is a computational model which constructs meaning relationships between words on the basis of large text corpora. It turns out that there is a significant relationship between the LSA scores and stress assignment, with the semantics of the left constituent having the greatest effect. Introducing this new and gradual measure of transparency, our analysis is the first to empirically validate the correlation between semantic transparency and compound stress assignment that was proposed in the literature.

In sum, the present paper in part confirms and in part challenges earlier assumptions about the role of semantic factors in explaining variation in English compound stress assignment. On the one hand we for the first time find strong empirical evidence for the two

factors semantic transparency and semantic closeness. On the other hand, these effects do not easily submit to an explanation in terms of traditional morphological theory. Thus, none of the effects is categorial, and it is the non-head that plays a more important role than the head.

References

- Fillmore, Charles & Collin F. Baker (2001 et seq.), The Berkeley FrameNet project. Berkeley: ICSI, <http://framenet.icsi.berkeley.edu>.
- Fudge, Erik C. (1984), *English Word-Stress*. London: George Allen and Unwin.
- Giegerich, Heinz J. (2004), Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8: 1–24.
- Giegerich, Heinz J. (2006), Attribution in English and the distinction between phrases and compounds. In Rösel, Petr (ed.), *Englisch in Zeit und Raum - English in Time and Space: Forschungsbericht für Klaus Faiss*. Wissenschaftlicher Verlag Trier.
- Ladd, D. Robert (1984), English compound stress. In Gibbon, Dafydd & Helmut Richter (eds.) *Intonation, Accent and Rhythm*. Berlin: Mouton de Gruyter: 253-266.
- Landauer, Thomas K., Peter W. Foltz & Darrell Laham (1998), Introduction to Latent Semantic Analysis. *Discourse Processes* 25: 259-284.
- Lieberman, Mark Y. & Richard Sproat (1992), The stress and structure of modified noun phrases in English. In: Sag, Ivan & Anna Szabolcsi (eds.), *Lexical Matters*, 131-181. Stanford: Center for the Study of Language and Information (Lecture Notes Number 24).
- Sampson, Geoffrey Richard (1980), Stress in English N+N phrases: a further complicating factor. *English Studies* 61: 264-270.
- Sproat, Richard (1994), English noun-phrase accent prediction for text-to-speech. *Computer Speech and Language* 8: 79–94.
- Olsen, Susan (2000), Compounding and stress in English: A closer look at the boundary between morphology and syntax. *Linguistische Berichte* 181: 55-69.
- Olsen, Susan (2001), Copulative compounds: A closer look at the interface between syntax and morphology. In Booij, Geert E. & Jaap van Marle (eds.), *Yearbook of Morphology 2000*, 279-320. Dordrecht/Boston/London: Kluwer.
- Plag, Ingo, Gero Kunter & Sabine Lappe (2007), Testing hypotheses about compound stress assignment in English: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory* 3.2, 199–233.
- Plag, Ingo, Gero Kunter, Sabine Lappe & Maria Braun (2008), The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language* 84.4: 760-794.
- Spencer, Andrew (2003), Does English have productive compounding? In Geert E. Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting* (Barcelona, September 20—22, 2001). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra: 329-341.