

Towards an empirical study of sentence semantics: Examples from the domains of scalar operators and concessives in English and German

Volker Gast

The study of sentence semantics in a model-theoretic framework (formal semantics), which grew out of logic and mathematics (e.g. Frege 1879, Russell 1905, Montague 1973, Cresswell 1973) and which abstracts away from any extra-sentential or non-linguistic information, has increasingly taken the factor of ‘context’ into account, thus leading to the establishment of a ‘formal’ branch of pragmatics (e.g. Kadmon 2001, Potts 2010). The consideration of the context is of particular relevance to the study of elements that make reference to alternatives, such as focus particles (Rooth 1985), or that refer to contextually given background assumptions, such as concessives (König 1991). Given that contextual embedding is arguably much less intuitively accessible than truth conditions, and given that information relating to the context of speech is often implicit, the study of language in context requires a methodology that relies on naturalistic data, and that goes beyond categorical judgements of truth or falsity, or (in)felicitousness.

While in descriptive linguistics a canon of methods has been developed that is based on natural language data (as represented in linguistic corpora, e.g. Biber et al. 1998), and that allows for the integration of probabilistic (in addition to categorical) generalizations (cf. Baayen 2008), studies in formal semantics or pragmatics rarely rely on naturalistic, contextually embedded data. This is probably at least partly due to the lack of corpora with annotations as are required for the analysis of elements such as focus particles or concessives. For example, for a proper analysis of focus operators information about the scope and the focus of a particle is indispensable, and information about the focus alternatives (for any given example) is at least useful. What is required, thus, is corpora with ‘rich’ annotations, i.e. annotations at various levels of linguistic analysis (minimally syntax, semantics and pragmatics) that are theoretically informed and provide the type of information that is needed for a fine-tuning of the formal analyses.

In this talk I will aim to show how the formal study of sentence meaning can be enriched by using a corpus-based, quantitative methodology, and by creating *ad hoc* corpora annotated at various levels. Such annotations are highly specific and require a combination of automatic annotations – for instance, syntactic parsing (Klein & Manning 2003) and word sense disambiguation (Pedersen & Kolhatkar 2009) – with manual annotations (Gast et al. 2015) – such as indications of the scope and focus of a particle, and the (implicit or explicit) focus alternatives available in the discourse environment.

I will present two case studies, one on scalar additive operators (e.g. Engl. *even*) and one on concessives (e.g. Engl. *although*). I will compare the relevant English expressions to their German equivalents. Specifically, I will show what distinguishes German scalar operators such as *selbst* and *sogar* from each other – and how they relate to Engl. *even* – and what determines the distribution of English concessive subordinators such as *although* and *while*, in comparison to the corresponding German expressions.

References

- Baayen, Harald. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.
- Biber, D., S. Conrad & R. Reppen. 1998. *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Cresswell, Max. 1973. *Logics and languages*. London: Methuen.
- Frege, Gottlob. 1879. *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle.
- Gast, Volker, Lennart Bierkandt & Christoph Rzymiski. 2015. Annotating modals with GraphAnno, a configurable lightweight tool for multi-level annotation. In *Proceedings of the Workshop on Models for Modality Annotation, held in conjunction with IWCS 11, 2015*, 19–28. Stroudsburg, PA.
- Kadmon, Nirit. 2001. *Formal Pragmatics: Semantics, Pragmatics, Presupposition and Focus*. Oxford: Blackwell.
- Klein, Dan & Christopher Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st meeting of the association for computational linguistics*, 423–430.
- König, Ekkehard. 1991. Concessive relations as the dual of causal relations. In D. Zaefferer (ed.), *Semantic Universals and Universal Semantics*, 190–209. Berlin: de Gruyter.
- Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In P. Suppes, J. Moravcsik, & J. Hintikka (eds.), *Approaches to Natural Language*, 221–242. Dordrecht: Reidel.
- Pedersen, Ted & Varada Kolhatkar. 2009. WordNet::SenseRelate::AllWords - A Broad Coverage Word Sense Tagger that Maximizes Semantic Relatedness. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 17–20. Stroudsburg, PA: Association for Computational Linguistics.
- Potts, Christopher. 2010. Formal pragmatics. In L. Cummons (ed.), *The Routledge Pragmatics Encyclopaedia*, 167–170. London: Routledge.
- Rooth, Mats. 1985. Association with focus. Doctoral Dissertation, University of Massachusetts.
- Russell, Bertrand. 1905. On denoting. *Mind* 14. 479–493.