Holger Keibel, Cyril Belica, Marc Kupietz, Rainer Perkuhn (Mannheim, Germany)

Approaching grammar: Detecting, conceptualizing, and generalizing paradigmatic variation

Abstract

In previous work we argued that any viable attempt to develop an *explanatory* theory of grammar is necessarily rooted in empirical research from the outset (Kupietz/Keibel 2009; Keibel/Kupietz 2009). The same papers further outlined an empirical-linguistic research program that seeks to uncover grammatical regularities from authentic language productions as recorded in corpora. The point of departure is a perspective on grammar according to which grammatical regularities are emergent by nature, emerging from language use in an ongoing process (cf., for instance, Hopper 1998; Hoey 2005). More specifically, such regularities are thought of as arising in the grammars of individual speakers, as the continuous result of aggregating experience with language. For a speaker, these regularities are therefore considered psychologically real, taking the form of cognitive routines (Hopper 1998:165). Because the language experience of different speakers is never identical, their individual grammars are likely to differ. Grammar as a supra-individual phenomenon, then, is characterized as what the individual grammars of most speakers of the respective language community have in common.

The causal relation between language use and grammar is in fact reciprocal: Language use constantly shapes the emergent grammars of individual speakers, and their grammars in turn shape their language productions. In consequence, the emergent grammatical regularities are likely to have correlates in sufficiently large and stratified collections of authentic language productions. The research program is interested in the grammatical regularities of the language community, but as these cannot be studied directly it attempts to uncover them through their correlates in corpora. Moreover, because authentic data of language use are in general lexically specific, the best, if not the only, choice to do so is to start from lexical items and incrementally derive hypotheses about increasingly complex and abstract regularities around these items. This inductive approach to grammatical regularities is the general strategy taken here, and it is likely that a large proportion of the more abstract regularities have originally become psychologically real for many speakers along very similar inductive paths. The research program further requires that induction proceeds in small steps, and that each step is motivated deductively by psychological premises that might play a role in the emergence of grammatical phenomena and that may be tested independently. This requirement minimizes the risk of a blind induction that any empirically driven approach is faced with.

How the first inductive steps may look like was explored elsewhere (Keibel/Kupietz/Belica 2008). While these first steps emphasized syntagmatic generalizations which are lexically specific, the current paper goes one step further by presenting explorations on how to uncover partly schematic syntagmatic patterns and how these patterns may be conceptualized in an appropriate way. The corpus correlate of such a schema can be described as the co-occurrence relation of a word or word sequence with an entire set of words. This paper follows the rationale that such a cooccurrence relation is likely to constitute an emergent schematic pattern in the language community if the latter words are sufficiently similar to each other, where similarity between words is understood in terms of similar use. The relevant aspects of similarity may differ across patterns. When they primarily concern grammatical or semantic similarity, the pattern may be described in terms of *colligation* or in terms of what Sinclair (1998) calls *semantic* preference, respectively. In many cases, however, the pattern will be a blend of both and also involve other types of similarity. Importantly, these interpretations and classifications are deferred until after the putative schema has been identified on the basis of usage-based similarities.

References

Hoey, M. (2005): Lexical Priming: A New Theory of Words and Language. London: Routledge.

Hopper, P. J. (1998): Emergent Grammar. In Tomasello, M. (Ed.): The New Psychology of Language: Cognitive and functional approaches to language structure (pp. 155-175). Mahwah, New Jersey: Erlbaum.

Keibel, H. / Kupietz, M. (2009): Approaching grammar: Towards an empirical linguistic research programme. In: Working Papers in Corpus-based Linguistics and Language Education, No. 3 (pp. 61-76). Tokyo: Tokyo University of Foreign Studies (TUFS). http://cblle.tufs.ac.jp/assets/files/publications/working_papers_03/section/061-076.pdf

Keibel, H. / Kupietz, M. / Belica, C. (2008): Approaching grammar: Inferring operational constituents of language use from large corpora. In: Šticha, F. / Fried, M. (Eds.): Grammar & Corpora 2007: Selected contributions from the conference Grammar and Corpora, Sept. 25-27, 2007, Liblice, Czech Republic (pp. 235-242). Prague: ACADEMIA.

Kupietz, M. / Keibel, H. (2009): Gebrauchsbasierte Grammatik: Statistische Regelhaftigkeit. In: Konopka, M. / Strecker, B. (Eds.): Deutsche Grammatik – Regeln, Normen, Sprachgebrauch (pp. 33-50). Berlin, New York: de Gruyter.

Sinclair, J. (1998): The lexical item. In: Weigand, E. (Ed.): Contrastive lexical semantics. Amsterdam, Philadelphia: Benjamins.