

## Questioning the lemma in usage-based linguistics

John Newman  
University of Alberta, Monash University  
john.newman@ualberta.ca

Keywords: corpus methods, lemma, inflection, collocates, linguistic unit

Usage-based approaches to language often make use of a “lemma”, i.e., a unit of representation and analysis that encompasses all the inflected forms of a word. As such, the lemma is an intermediate unit of analysis, more abstract than the orthographic or phonological words of raw text/speech, and utilized in studies that purport to discover usage-based knowledge of speakers. Cognitive linguistic and corpus-based studies of argument structure, syntactic patterning, and collocational patterning typically express results in terms of patterning around lemmas rather than the inflected word forms. In one of the few corpus-based studies that systematically investigate the value of working with lemmas vs. inflected forms, Gries (2011) concluded that for the most part, a lemma-based analysis was sufficiently informative in a study of a specific syntactic alternation of English. Notwithstanding Gries’ valuable, but circumscribed study, I question the practice of positing the lemma as a default unit of analysis. I argue in the spirit of Sinclair (1991), Rice & Newman (2005), and Newman & Rice (2006), that the lemma can obscure much of the interestingly differential behaviour of individual inflected forms in corpus-based studies.

In this corpus-based paper I begin in a very targeted way with a comparison of the adjectival collocates occurring immediately before the noun *child* (singular) vs. *children* (plural), *children* having already been identified in the literature as an inflected noun form of particular interest in English and some other languages (Tiersma 1982). The data is drawn from the SPOKEN component of the Corpus of Contemporary American English (COCA, Davies 2008-) and the adjectival collocates were ranked according to the measures of Attraction and Reliance (cf Schmid 2010). Interesting and unexpected differences between the adjectives preferred by the singular and plural nouns emerge from this comparison. To take just one unexpected result, adjectives of the form *x-year-old* (*six-year-old* etc.) are clearly preferred collocates of *child*, but not *children*, in the corpus measured by both Attraction and Reliance scores. Following on from this particular observation, I then broaden the focus to adjectives occurring before approximate synonyms of *child/children* (*daughter(s)*, *son(s)* etc.) and show that the same contrast between adjective + singular patterns vs. adjective + plural patterns holds for this larger set of synonyms as well. These findings suggest that there is indeed good reason to study the behaviour of adjectival collocates of nouns at the inflected word form level. The study considers differential effects of patterning around inflected forms of verbs vs. nouns, as well as different consequences associated with patterns involving relatively abstract syntactic structure vs. lexical collocates.

### References

- Davies, Mark (2008-). *The Corpus of Contemporary American English: 450 million words, 1990-present*. <http://corpus.byu.edu/coca/>.
- Gries, Stefan Th. (2011). Corpus data in usage-based linguistics: What’s the right degree of granularity for the analysis of argument structure constructions? In Mario Brdar, Stefan Th. Gries, & Milena Žic Fuchs (eds.), *Cognitive linguistics: Convergence and Expansion*, 237-256. Amsterdam & Philadelphia: John Benjamins.
- Newman, John & Sally Rice (2006). Transitivity schemas of English EAT and DRINK in the BNC. In S. Th. Gries, & A. Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*, 225–260. Berlin & New York: Mouton de Gruyter.
- Rice, Sally & John Newman. (2005). Inflectional islands. Paper presented at the International Cognitive Linguistics Conference, Seoul, South Korea.
- Schmid, H.-J. (2010). Does frequency in text instantiate entrenchment in the cognitive system? In D. Glynn & K. Fischer (eds.), *Quantitative Methods in Cognitive Semantics*, 101-133. Berlin & New York: Mouton de Gruyter.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Tiersma, Peter Meijes (1982). Local and general markedness. *Language* 58 (4): 832-849.