Collostructions revisited: making it simpler, making it better

Stefan Th, Gries, UC Santa Barbara & JLU Giessen, stgries@linguistics.ucsb.edu

One of the most widespread corpus-linguistic methods to study constructions is collostructional analysis, a family of three methods applying an association measure-based approach towards the co-occurrence of different kinds of constructions. The three methods differ in what the units are that are involved in these co-occurrence data:

1. **collexeme analysis** (Stefanowitsch & Gries 2003): one looks at how much constructions (usually words/lemmas) (dis)like to occur in a slot of one usually more schematic constructions (as opposed to elsewhere);

2. **distinctive collexeme analysis** (Gries & Stefanowitsch 2004): one looks at how much constructions (usually words/lemmas) (dis)like to occur in a slot of a usually more schematic construction as opposed to alternative, functionally similar constructions;

3. **covarying collexeme analysis**: one looks at how much constructions (usually words/lemmas) in one slot of a more abstract construction (dis)like to co-occur with constructions (usually words) in another slot of the same more abstract construction.

In spite of the fairly widespread success of these methods, corpus linguistics in general has evolved to a degree that updating/revisiting aspects of this family of methods is overdue. In this paper, we make two sets of suggestions of how collostructional analysis can be updated.

The first set of suggestions involves simplifying the analysis for descriptive/exploratory purposes while at the same time enriching it with confidence intervals. More specifically, we propose to use a different kind of statistic as the main association measure, which has several advantages: (i) it speeds up the relevant computations by multiple orders of magnitude and avoids computationally expensive calculations to avoid the notorious problems of infinite results values, which then (ii) allows something that no collostructional studies have done so far, namely compute bootstrapped confidence intervals for collexeme strengths. We will exemplify this by revisiting Stefanowitsch & Gries's original case studies of ditransitives and particle verbs.

The second set of suggestions involves the idea that we should move away from a single kind of association measure for theoretical/exploratory purposes and instead quantify collostructional attraction as a tuple of, minimally, three ideally orthogonal dimensions: frequency, association, and dispersion, because only this kind of analysis can address all dimensions that are relevant to cognitive/usage-based approaches -- entrenchment, contingency, and learning/exposure; this will be exemplified on the basis of the ditransitive and the *way*-construction.

Finally, we end with a (renewed) plea to take the notion of construction more seriously than most analyses have done. With some simplification, most existing work has actually not looked at the associations of (lexical) constructions to ('syntactic') constructions (i.e. true pairings of forms *and* meaning), but at associations of forms to syntactic constructions; using polysemous phrasal verbs, we exemplify the degree to which polysemous senses of lexical constructions exhibit wildly different constructional associations that most traditional analyses are likely to miss.

References

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97-129.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2). 209-243.