Ready-made chunks or schematic generalizations? Using agent-based simulation and corpora to discover how constructions are processed

Dirk Pijpops[1]
[1]RU Lilith, University of Liège, dirk.pijpops@uliege.be

Cognitive linguists generally agree that language users can cognitively store ready-made multi-word phrases (Arnon and Snider 2010) and employ more schematic generalizations (Perek and Goldberg 2017). Still, the question remains when exactly they use either. This talk will present a way of tackling that question with agent-based simulation and corpora. The key is to use contamination phenomena such as constructional and lectal contamination (Pijpops and Van de Velde 2016; Pijpops 2022), which rely on ready-made phrases to induce lexical biases in language variation. Put concretely, agent-based simulations are used to show that ready-made storage is indeed crucial to create the biases, and corpus research can show when these biases are present in the real world.

As an example, I first present two agent-based simulations of lectal contamination. This is an effect whereby language users of Variety A are more likely to use a variant typical of Variety B when using words that are more often used in Variety B. For example, Belgians are more likely to produce the 'Netherlandic' partitive genitive variant with -s ending (e.g. *iets bijzonders* 'something interesting-s', *iets interessants* 'something interesting-s') in phrases containing words that are more often used by people from the Netherlands, such as *bijzonder* 'special' or *boeiend* 'fascinating'. The simulations show that this effect indeed emerges if the simulated agents employ ready-made language chunks, but does not emerge when schematic constructions are used.

Next, we attempt to observe this effect in the real world through two corpus studies. The first looks at the Dutch partitive genitive introduced above, while the second investigates the variation between the determiners *zulke* 'such' and *zo'n* 'such', as in *zulke mensen* vs. *zo'n mensen* (both 'such people'). These case studies are alike in that they both present lectally stratified variation in Dutch NP's: the partitive genitive variant with -s and the determiner *zulke* 'such' are more popular in the Netherlands, while the partitive genitive without -s and the determiner *zo'n* 'such' are more common in Belgium. They are crucially different, however, in that ready-made chunks are likely employed when processing the partitive genitive. There are three reasons for this: (i) the variant is a bound morpheme; (ii) the partitive genitive appears with a comparatively limited amount of highly frequent phrases, although it is a productive construction; and (iii) it has an internal structure that is highly atypical for Dutch grammar, viz. an adjectival phrase postmodifying a pronoun. By contrast, NP's with *zulke* and *zo'n* are more likely candidates to be processed through schematic generalizations, because (i) the variant is a unbound determiner; (ii) its phrases seem lexically much more diverse; and (iii) its phrases adhere to the typical structure of Dutch NP's, with determiners followed by adjectives followed by nouns. As a result, it may be expected that the effect of lectal contamination does emerge with the partitive genitive, while it does not emerge with the *zulke-zo'n* variation. The results show that the expected lexical biases do indeed appear with the partitive genitive, while they cannot be observed for the *zulke-zo'n* variation.

References

Arnon, Inbal and Neal Snider. 2010. More Than Words: Frequency Effects for Multi-Word Phrases. *Journal of Memory and Language* 62(1). 67–82.

Perek, Florent and Adele Eva Goldberg. 2017. Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition* 168(C). 276–293.

Pijpops, Dirk. 2022. Lectal contamination: Evidence from corpora and from agent-based simulation. *International Journal of Corpus Linguistics* 27(3). 259–290.

Pijpops, Dirk and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50(2). 543–581.