

Linguistic reproducibility and cognitive entrenchment: corpus-based evidence and applications for language teaching and translation

Jean-Pierre Colson
Université catholique de Louvain (Louvain-la-Neuve)

In this paper we will argue that the notion of linguistic reproducibility, as it has been used in the phraseological tradition, comes very close to the notion of cognitive entrenchment, as defined by cognitive grammar and by construction grammar.

In cognitive linguistics, entrenchment is related to one of the four general cognitive processes that (also) play a role in language, namely automatization (the other processes are: association, schematization and categorization). A number of experimental studies have already been carried out around the notion of entrenchment, following two different perspectives. On the one hand, psycholinguistic studies have made it possible to corroborate its theoretical underpinnings; on the other, recent research on *collostructions*, a statistical approach of constructions at the intersection of corpus linguistics, has made it possible to validate entrenchment as an observable feature on large linguistic corpora.

Elaborating on the *collostructional* approach, I will try to show that computational phraseology may shed some fresh light on the traceability of linguistic reproducibility and entrenchment in corpora. In particular, taking the perspective of language structure as a set of constructions (as claimed by construction grammar), a large part of which is idiomatic (as claimed by phraseology), makes it possible to trace down how those constructions behave in a probabilistic network.

From a practical point of view, such an investigation requires a new system for annotating corpora, which takes into account all the morpho-phonetic, syntactic and pragmatic aspects of constructions. As a first step in that direction, I will show that experiments with parts-of-speech tagged corpora already yield intriguing results as to the attraction between words, constructions and phraseological units. Those results might be used for a gradual selection of the most useful constructions in language learning, and for a number of practical tasks in translation, such as the identification of units of meaning and the search for equivalent constructions in the target language.