

A formal model of lexicogrammatical individuality

Andrea Nini

University of Manchester, andrea.nini@manchester.ac.uk

Keywords: idiolect, linguistic individuality, linguistic unit, Cognitive Grammar, chunking, forensic linguistics, authorship analysis

Individuality in language is an under-researched area of linguistics. Although frameworks of different kinds presuppose the existence of idiolects, the study of why and how individuals are different and/or unique in their lexicogrammatical system is an area that is not well developed. This is a problem for forensic linguists who provide evidence of authorship of questioned documents in a court of law. Despite the significant advances in this field made by computational linguists, the reason for the success of these techniques is unknown. One of the most fascinating and yet unexplained facts is that the frequency of function words can differentiate individuals. In this talk I will present a formal theory of linguistic individuality fundamentally based on cognitive linguistics that proposes an explanation for this phenomenon within a formal architecture based on cognitive linguistic principles. The central argument of the theory is that each individual possesses a unique repertoire of linguistic units, defined following Langacker (1987) as structures that a person can produce automatically. A linguistic unit defined in this way largely corresponds to the cognitive psychology notion of a chunk (Miller, 1956; Gobet et al., 2001), which in turn are traces of memory, predominantly procedural. Usage-based theories predict that these repertoires are different because entrenchment, the process of unit creation, is mostly idiosyncratic (Schmid, 2015; Langacker, 1987; Dąbrowska, 2014). The argument proposed is that the frequency of function words is a statistical representation reflecting these unique repertoires. Initial empirical evidence in support of this hypothesis will also be presented in the form of a comparison between two authorship analysis systems, one based on function word frequencies and one on comparison of repertoires of units. This evidence seems to support Langacker's claim that "the set of units mastered by all members of a speech community might be a rather small proportion of the units constituting the linguistic ability of any given speaker" (Langacker, 1987, p. 62).

References

- Dąbrowska, Ewa. 2014. Words that go together. Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon* 9(3). 401–418. doi:10.1075/ml.9.3.02dab. <http://www.jbe-platform.com/content/journals/10.1075/ml.9.3.02dab>.
- Gobet, Fernand, Peter C.R. Lane, Steve Croker, Peter C-H. Cheng, Gary Jones, Iain Oliver & Julian M Pine. 2001. Chunking mechanisms in human learning. *Trends in Cognitive Sciences* 5(6). 236–243. doi:10.1016/S1364-6613(00)01662-4. <http://tics.trends.com1364>.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar*, vol. 1. Stanford, CA: Stanford University Press.
- Miller, George A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2). 81–97. http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0000U6&topic_id=1.
- Schmid, Hans-Jörg. 2015. A blueprint of the Entrenchment-and-Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association* 3(1). 3–25. doi:10.1515/gcla-2015-0002.