The grammar of knowledge: Investigating the emergence of conventions in the evolution of Wikipedia

Peeter Tinits

School of Humanities, University of Tallinn
Tallinn, Estonia
peeter.tinits@gmail.com

Stefan Hartmann

German Department, University of Bamberg
Bamberg, Germany
stefan1.hartmann@uni-bamberg.de

The grammar of knowledge: Investigating the emergence of conventions in the evolution of Wikipedia

In recent years, approaches like cognitive linguistics, construction grammar, and other usage-based paradigms have called attention to the observation that communicative conventions exist at multiple levels of abstraction, from the level of the morpheme to syntax (e.g. Croft 2001). However, even beyond the sentence level, phenomena like genre and text type are governed by linguistic and cultural conventions. In this paper we investigate the directionality in the evolution of text writing conventions in the large-scale collaborative writing project of Wikipedia.

Wikipedia, the free encyclopedia that anyone can edit, exists since 2001, and has over time gathered an extremely large userbase around it. (English Wikipedia alone has over 31 million registered users, around 300,000 have made more than 10 edits to the site.) As Wikipedia maintains a public record of all the edits made to it, it provides an interesting test case to study the evolution of writing conventions.

There are two dimensions in which directionality in the evolution of linguistic conventions is to be expected: 1) In smaller and larger communities, simple accommodation to other community members will usually lead communicative norms to converge (e.g. Kerswill 1996). This happens in face-to-face interactions (e.g. Pickering & Garrod 2004) or can happen nowadays in long-distance coordination in large groups (Kooti et al. 2012). 2) Over repeated usage, linguistic conventions tend to become more efficient for its purposes (Zipf 1949). In fact, establishing norms itself can often allow communication to be more efficient and precise (e.g. Fusaroli et al. 2012).

Both accommodation and the drive towards efficiency can result in standardization. As such, we should expect increasing conventionalization in the textual practices within Wikipedia. We present a pilot study aiming to test the degree of conventionalization in Wikipedia writing practices. Particularly, we focus on the evolution of the structure of the Wikipedia articles as it relates to their content based on diachronic datasets of the English Wikipedia.

More specifically, we test:

a) whether specific structural patterns become increasingly likely to be used in new articles,

b) whether there are specific structural patterns that tend to emerge as most "stable" within categories, in being less likely to be replaced in particular categories than other patterns

The first development, which can, for example, be found in the diachronic evolution of headings and subheadings used in Wikipedia articles, indicates at least some increase in conventionalization in the structures of the texts. This provides an interesting view on the alignment of linguistic practices above the level of the sentence. A trend towards stable structural patterns in article composition could be interpreted in terms of a directional development of language use as a result of long-lasting coordination of communicative norms.

References

Croft, W. (2001). *Radical Construction Grammar: syntactic theory in typological perspective.* Oxford: Oxford University Press.

Fusaroli, R; Bahrami, B; Olsen, K; Roepstorff, A; Rees, G; Frith, C; Tylén, K; (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. Psychological Science, 23 (8) pp. 931–939. 10.1177/0956797612436816.

Kerswill, P. (1996). Children, adolescents, and language change. *Language variation and change*, *8*(2), 177-202.

Kooti, F., Yang, H., Cha, M., Gummadi, K. P., & Masion, W. A. (2012). The emergence of conventions in onlie social networks. In *6th International AAAI Conference on Weblogs and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4661

Pickering, M. J./Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27:2, 169–226. doi:10.1017/S0140525X04000056.

Zipf, G. K. (1949): *Human behavior and the principle of least effort.* Redding: Addison-Wesley.