
Corpus driven analysis of obsolescence of multi-word expressions in Late Modern English

Ondřej Tichý
Faculty of Arts, Charles University
ondrej.tichy@ff.cuni.cz

This paper explores a new methodology for extracting multi-word constructions, that were once common but has since become obsolete (Tichý, 2018), from large corpora (esp. from the Google n-grams dataset of the Google Books project).

It proceeds from a novel method for a bottom up multi-word construction extraction (Wahl & Gries, forthcoming), to the formulation of a semi-automatic procedure for identifying constructions that may have become lost or obsolete in the course of the last three centuries, from the Late Modern era to Present-day English (1700-2000). The procedure makes use of both relative frequencies to establish currency, as well collocational measures such as mutual information.

In the analytical part, the paper focuses on analysing select constructions of the type *there needs no* (as in “there needs no priest”) whose obsolescence is indicative of wider structural and typological changes, e.g. as in this case a decline of impersonal constructions.

Conditions, circumstances and consequences of the loss of such constructions are considered with a focus on the competing forms expressing similar functions that may be recognized as supplanting the old forms

References: Michel, Jean-Baptiste, Shen, Yuan Kui, Aiden, Aviva Presser, Veres, Adrian, Gray, Matthew K., Pickett, Joseph P., Hoiberg, Dale et al. (2011): Quantitative analysis of culture using millions of digitized books. *Science* 331(6014): 176–182. • Tichý, Ondřej. (2018): Lexical obsolescence 1700–2000. In: *Applications of Pattern-driven Methods in Corpus Linguistics*. Kopaczkyk, J & Tyrkkö, J (eds). *Studies in Corpus Linguistics*, John Benjamins • Wahl, Alexander & Stefan Th. Gries. (forthcoming): Multi-word expressions: A novel computational approach to their bottom-up statistical extraction. In Pascual Cantos (ed.), *Lexical collocation analysis: advances and applications*. Berlin & New York. Springer.