

The truth behind the numbers: corpus methods in the research on Old English syntax

Corpus methods in the form of computer-assisted analysis of large samples of linguistic material have become very popular over the recent years, showing that one of the basic advantages of corpus linguistics is the possibility of detecting patterns and regularities which otherwise may easily be missed. The growing number of increasingly advanced tools supporting linguistic research is a great opportunity for scholars working within various theoretical frameworks and in various areas of linguistics. Historical linguistics has also witnessed a digital revolution, and the great role of electronic language corpora in both synchronic and diachronic studies of contemporary and historical languages is now absolutely unquestioned.

In the case of English, the creation of Penn corpora, including the York-Toronto-Helsinki Parsed Corpus of Old English prose (YCOE, Taylor et al. 2003), has been a great breakthrough. The corpus is small from the point of view of contemporary corpus linguistics (only 1.5 million words), but it could not be bigger since the corpus compilers included at least one version of all the existing prose texts in Old English excluding glosses (on the basis of their closeness to the Latin sources and limited use for syntactic research). The syntactic annotation makes it a perfect tool for any kind of study on Old English syntax, but its specific composition is also a great danger for any linguist trying to work with YCOE.

In this paper I would like to present both the great advantages of YCOE and its limitations, which must be controlled for the results of any corpus investigation to be reliable. In order to illustrate the great potential of corpus data for the verification of hypotheses first formulated in the pre-corpus era of research on Old English syntax, I am going to present a case study of Old English conjunct clauses and their alleged preference for the V-final order signalled in Mitchell (1985: §1685), Traugott (1992: 277), Fischer et al. (2000: 53), Ringe & Taylor (2015: 419) and many others, later questioned by Bech (2017) and then once again supported in Zimmermann (2017). The threats related to working with YCOE as if it were a well-balanced representative collection of texts will be shown in another case study of the so-called narrative inversion, which is a specific construction with a very skewed distribution (Mitchell 1985: §3933; Ohkado 2004; Calle-Martín & Miranda-García 2010). Close inspection of the data retrieved from YCOE shows that the structure requires a great deal of qualitative analysis for the numbers to give us any valuable information at all.

References

- Bech, K. 2017. Old truths, new corpora: Revisiting the word order of conjunct clauses in Old English. *English Language and Linguistics* 21(1). 1–25.
- Calle-Martín Javier & Antonio Miranda-García. 2010. ‘Gehyrdon ge þæt gecweden wæs’ – a corpus-based approach to verb-initial constructions in Old English. *Studia Neophilologica* 82. 49-57.
- Fischer, O., A. van Kemenade, W. Koopman & W. van der Wurff. 2000. *The syntax of early English*. Cambridge: Cambridge University Press.
- Mitchell, B. 1985. *Old English syntax*. Oxford: Clarendon Press.
- Ohkado, Masayuki. 2004. On the structure and function of V1 constructions in Old English. *English Studies* 85. 2-16.
- Ringe, D. & A. Taylor. 2015. *The development of Old English*. Oxford: Oxford University Press.
- Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE).
- Traugott, E. 1992. Syntax. In R. M. Hogg (ed.), *The Cambridge history of the English language*, vol. 1: *The Beginnings to 1066*, 168-289. Cambridge: Cambridge University Press.
- Zimmermann, Richard. 2017. Formal and quantitative approaches to the study of syntactic change: Three case studies from the history of English. Unpublished Ph.D. dissertation: University of Geneva.