High verb frequency as an accessibility parameter promoting early verb placement in main clauses of OHG, OS, and OE

Karin Harbusch (Computer Science Faculty, University of Koblenz-Landau, Koblenz)
Gerard Kempen (Max Planck Institute for Psycholinguistics, Nijmegen)
Ans van Kemenade (Department of English Language, Radboud University, Nijmegen)

A substantial body of psycholinguistic research since Bock & Warren (1985) supports the importance of conceptual and lexical accessibility as a determinant of the order of constituents, particularly nominal constituents, within clauses: Speakers tend to assign early positions to NPs headed by highly accessible nouns. Among the factors promoting accessibility are animacy, frequency, and topicality. The role of these and other information-structure variables in language change is increasingly recognized in the historical-linguistic community (e.g., Struik & van Kemenade, 2018). Conspicuously absent are data on effects of accessibility of verbal clause constituents—presumably due to the immovable position of verbs, especially finites, in many present-day Germanic languages. However, the position of finite verbs is less fixed in Old High German (OHG), Old Saxon (OS), and Old English (OE). A practical reason to study effects of verb accessibility in these languages is the availability of substantial syntactically analyzed text corpora (“treebanks”): (B4-)(TATIAN for OHG (Petrova et al. 2009), HELIPAD for OS (Walkden 2016), and YCOE for OE (Taylor 2007). The accessibility factor we focus on is verb frequency: the total frequency of a verb (a “lemma”), calculated as the sum of the frequencies of all its (un)inflected forms in all spelling variants. We relied on lemmatizations provided in the published trees (OS), reconstructed from published materials (OHG), or done by ourselves (OE).

We report verb position data for finite verbs in main clauses only. The positions were labeled “P1” (first clause constituent, ignoring leading coordinating conjunctions and interjections); “P2” (second constituent, ignoring immediately preceding negation elements, verb particles, etc.); or “P>2” (verb in any later position). In order to obtain a dichotomous variable for linear position, we combined—somewhat arbitrarily—the P1 and P2 cases into a group called “early”, and renamed the P>2 group “late”. Likewise, in each treebank, we dichotomized the frequency spectrum into a High-frequency group comprising the 50 lemmas with highest total frequencies, and a Low-frequency group containing the remaining verb lemmas.

As shown in the figure, high verb frequency indeed promotes early verb placement (yellow bars higher than red bars). The mean placement difference between high- and low-frequent verbs is significant in each treebank (χ² tests with df=1, all yielding p < .01). (Discussion of the between-language differences is beyond the scope of this presentation.)

This result confirms the hypothesis that verb frequency functions as an accessibility parameter. High-frequent verbs are more likely to land in anterior clause positions allowed by the grammar than infrequent verbs because the speaker is able, or prefers, to retrieve them from the mental lexicon more easily, hence earlier, than infrequent verbs.

We suggest that the observed accessibility–anteriority link may have been a factor in the diachronic development of (S)OV–to–(S)VO main-clause word order in Germanic languages.
References