Verb frequency as a sentence-production factor modulating historical Object-Verb to Verb-Object changes in West-Germanic languages

Gerard Kempen
MPI for Psycholinguistics, Nijmegen, The Netherlands

Karin Harbusch
Computer Science Department, University of Koblenz-Landau, Koblenz, Germany

In Germanic languages, the historical evolution from OV to VO took place earlier in Main clauses than in Subordinate clauses (e.g., see review by Kiparsky 1996). In this poster, we present corpus data suggesting that a psycholinguistic factor may have been causally involved in this time difference: usage frequencies of the finite verbs heading those clauses.

From treebanks of present-day German, Dutch and English (VERBMOBIL, CGN2.0, and Switchboard, respectively), we extracted all verbform tokens and categorized these as head of a Main clause (Main-FNT), of a finite Subordinate clause (Sub-FNT), or of a non-finite Subordinate clause (NONFNT). (We assume a 1-to-1 relation between verbs and clauses.) After lemmatizing these verbforms and calculating the total frequencies of the verb lemmas, we determined, for each lemma, its proportions of Main-FNT, Sub-FNT, and NONFNT tokens. The resulting numbers are plotted as percentages (GRAPH1 for Dutch; the Main-FNT gradient for German and Dutch are similar, both steeper than the one for English; see Kempen & Harbusch 2017a). We expected these percentages to be insensitive to the lemma’s total frequency; i.e., we predicted similar percentages of Main-FNT, Sub-FNT and NONFNT forms for high- and low-frequency verb lemmas. However, this prediction was confirmed only for the Sub-FNT forms (black curve in GRAPH1): In the set of Main-FNT forms, we observed an OVERREPRESENTATION of HIGH-FREQUENT VERB LEMMAS. Cross-language comparison of the overrepresented finite main verbs revealed considerable similarity: They include all auxiliaries and modals, and many light verbs and verbs of communication/cognition (for lists, see Kempen & Harbusch 2017b).

THEORY. In Germanic languages, the incremental planning of a HIERARCHY OF CLAUSES typically proceeds TOP-DOWN, starting with the Main clause, usually yielding a right-branching hierarchy of clauses (Main preceding Subordinate). High frequency (hence rapid accessibility and retrievability) of verbs facilitates the rapid and fluent onset of overt sentence production by enabling early placement of these verbs within clauses. Clauses embodying these properties—Main, sentence-initial and VO instead of Subordinate, sentence-final and OV—are more likely to be produced successfully, thereby promoting the survival of these properties in the form of sentence production rules or preferences.

GRAPH 1: Present-day spoken Dutch (CGN2.0)

GRAPH 2: Old High German (Tatián, “deviant” cases)

GRAPH 3: Old Saxon, Helaand (Helipad)

GRAPH 4: Percentage of VO orders

TESTING THE THEORY WITH PARSED CORPORA OF OLD HIGH GERMAN (OHG; B4.TATIAN) AND OLD SAXON (OS; HELIPAD) (work in progress). Order of verb and object in historical instantiations of Germanic languages tends to be more flexible than in modern versions. This creates an opportunity to test whether verb lemma frequency indeed affects the order of these constituents. In the poster, we report the result of a new study seeking to answer to two questions: whether (1) OHG and OS manifest the above overrepresentation pattern, and (2) whether verb frequency indeed can influence verb position. As shown in GRAPH2 and GRAPH3, the OHG as well as the OS corpora exhibit overrepresentation of high-frequent verbs in Main clauses, although only in embryonal form. As regards the second question, GRAPH4 confirms that high-frequency of a verb is capable of consigning that verb to an earlier position in the clause (VO instead of OV)—again, to a modest degree.

In conclusion, taking usage frequencies of verbs into account in historical-linguistic word-order studies could become a fruitful approach, supplementing the extant work on properties of the NPs governed by verbs.