Easy accessibility and retrievability of high-frequent verbs as a production-based factor modulating the speed of historical SOV-to-SVO word order changes

Gerard Kempen, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands (Gerard.Kempen@mpi.nl)
Karin Harbusch, Department of Computer Science, University of Koblenz-Landau, Germany

COMPARATIVE-LINGUISTIC BACKGROUND. Historical- and typological-linguistic studies over the past few decades have corroborated the hypothesis that SOV (Subject–Object–Verb) is the basic and original word order of languages in most language families (cf. Dryer 1989, 1991; Newmeyer 2000), and have shown that, if this order changes, the most likely next historical step is to SVO (Gell-Mann & Ruhlen 2011). The number of SVO languages currently approaches the number of SOV languages (Dryer 2013). The present study focuses on a special case of the SOV-to-SVO change: the fact that, in all Germanic languages, the historical evolution of VO in main clauses is faster than in subordinate clauses (e.g., see the survey by Kiparsky 1996).

CORPUS-LINGUISTIC ANALYSES. As part of a corpus study into SVO and SOV word orders in extemporaneously spoken present-day Dutch and German (Kempen & Harbusch 2017), we encountered an unexpected effect of verb frequency. After extracting and lemmatizing all verbform tokens in two treebanks, we categorized the forms of each verb (lemma) into three morpho-syntactic groups: (1) finite forms functioning as head of a main clause (MAIN-FNT), (2) finite forms heading a subordinate clause (SUBOR-FNT), and (3) nonfinite forms heading a subordinate clause (infinitives and participles; NONFNT). Verbs of any type were included (lexical/full, auxiliary, modal, copular). Rationale: In Dutch and German main clauses, the finite head verb occupies an earlier position ("verb-second", V2) than in virtually all types of subordinate clauses ("verb-final"; V-FNL). We calculated the MAIN-FNT, SUBOR-FNT and NONFNT frequencies of the inflected forms of each verb lemma, as well as the total frequency (TotFreq) of each lemma (i.e., the sum of its MAIN-FNT, SUBOR-FNT and NONFNT frequencies), and expressed these subfrequencies as percentages of the verb's TotFreq. We expected these percentages to be insensitive to the lemma's TotFreq: The percentages of MAIN-FNT, SUBOR-FNT and NONFNT forms of a lemma should be similar, irrespective of whether the lemma is high- or low-frequent.

However, this prediction was confirmed only for the SUBOR-FNT forms (red lines in Fig. 1): The percentages of MAIN-FNT forms turned out to INCREASE with increasing lemma TotFreq. (The high-frequent verbs include not only auxiliaries, copulas and modals but also lexical verbs.) As a sentence may have any number of (non)finite subordinate clauses but at most a single (coordination of) main clause(s), we characterize the observed data pattern as the MAIN-CLAUSE BIAS SHIFT (MCBS) OF HIGH-FREQUENT VERBS. An additional treebank study revealed a McBS also for English. Furthermore, high-frequent verbs in SENTENCE-INITIAL finite subordinate clauses did not reveal a McBS although these clauses occupy the same early sentential position as most main clauses (in Fig. 1, compare the left and right columns). Finally, the McBSs were considerably weaker in WRITTEN than in spoken corpora (Kempen & Harbusch 2017).

THEORY. Our explanation proceeds from assumptions about (A) planning, (B) utility, and (C) learning: (A) 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. (B) High frequency and hence easy accessibility of verbs facilitate the rapid and fluent onset of overt sentence production by enabling leftward placement of these verbs within clauses—VO instead of OV—, and of these clauses within the sentence (main before subordinate). (C) As A and B cause VO to combine with main clauses more readily than with subordinate clauses, learning processes promote the acquisition of VO order primarily in main clauses. Taking the present as a source of information about the past, we argue that (A) through (C) provide a possible cognitive explanation for the more rapid evolution of VO in main than in subordinate clauses of languages where the hierarchy of clauses tends to be planned top-down.

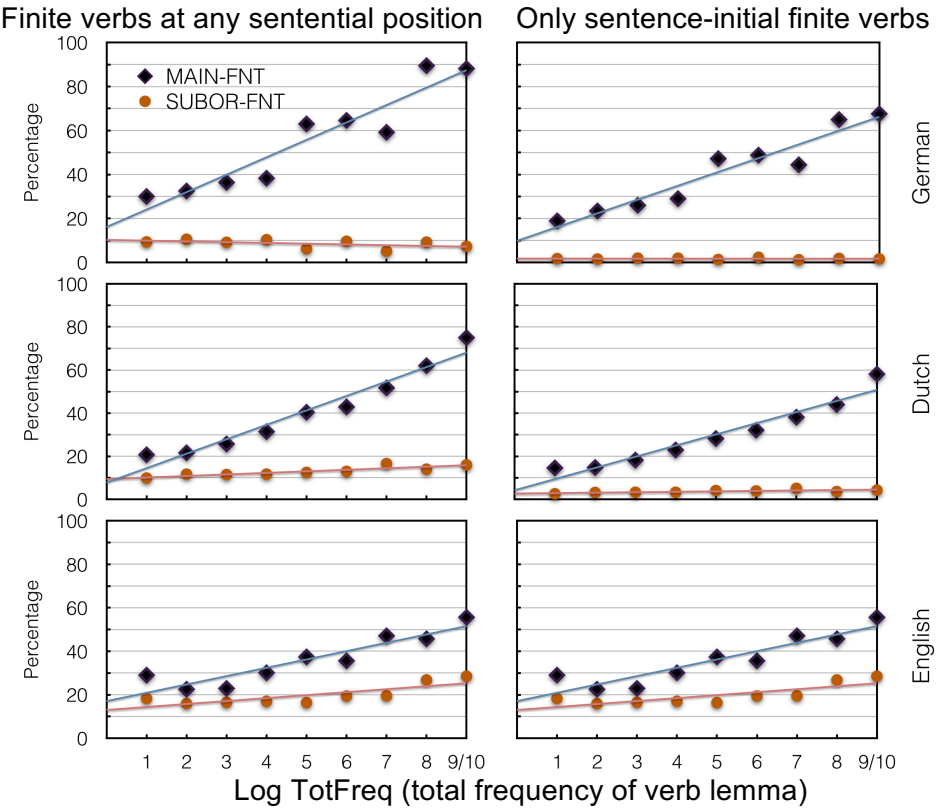

Figure 1. Increasing overrepresentation of high-frequent verbs in main clauses as a function of total verb frequency (TotFreq, plotted logarithmically: $\log_e$). This effect obtains in all three target languages and to similar degrees: In an ANOVA (ARTool; Kay & Wobbrock 2016) of the MAIN-FNT data underlying the left column of this figure, the 3*3 interaction of Language with TotFreq (3 TotFreq ranges) is not significant: $F(4, 63)=1.97$, $p=.11$ (only verbs with TotFreq≥8 were included due to a sparse data problem for verb lemmas with lower TotFreqs: many empty cells). Data source: the corpora VERBMOBIL (German), CGN2.0 (Dutch), and Switchboard (English).

References

Dryer, M. S. 1989. Large linguistic areas and language sampling. *Studies Lang, 13,* 257–292.

Dryer, M. S. 1991. SVO languages and the OV:VO typology. *J. Linguistics, 27,* 443–482.

Dryer, M. S. 2013. Order of Subject, Object and Verb. In Dryer, M. S. & Haspelmath, M. (Eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Online at http://wals.info/chapter/81, accessed on 2016-12-07.)

Gell-Mann, M., & Ruhlen, M. 2011. The origin and evolution of word order. *PNAS, 108*, 17290–5

Kay, M. & Wobbrock J. 2016. *ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs*. R package version 0.10.4, https://github.com/mjskay/ARTool.

Kempen, G. & Harbusch, K. 2017. Frequential test of (S)OV as unmarked word order in Dutch and German clauses: A serendipitous corpus-linguistic experiment. In Reckman, H.; Cheng, L.L.S.; Hijzelendoorn, M., & Sybesma, R. (Eds.). *Crossroads Semantics: Computation, experiment and grammar.* Amsterdam: Benjamins.

Kiparsky, P. (1996). The Shift to Head-Initial VP in Germanic. In Thráinsson, H., Epstein, S. D., & Peter, S. (Eds.). *Studies in comparative Germanic syntax, Volume II*. Dordrecht: Kluwer.

Newmeyer, F. J. (2000). On the Reconstruction of 'Proto-World' Word Order. In Knight, C., Studdert-Kennedy, M., & Hurford, J. R. (Eds.). T*he evolutionary emergence of language*: *Social Function and the Origins of Linguistic Form*. Cambridge UK: CUP.