In spoken German and Dutch, high-frequency finite verbs populate main clauses more densely than subordinate clauses, but much less so in spoken English: A corpus-linguistic study into VO versus OV word order

Gerard Kempen (MPI for Psycholinguistics, Nijmegen)  
Karin Harbusch (Computer Science Department, University of Koblenz-Landau)

As part of a study into the frequency of VO and OV orders in spoken German and Dutch, we extracted all occurrences of finite verbs in a German (Verbmobil; Stegmann et al., 2000) and a Dutch (CGN2.0; van Eerten, 2007) treebank. (We excluded forms that can only occur in main clauses, e.g. imperatives, and added the frequencies of all finites belonging to the same citation form.) For both languages, we found a correlation between the total corpus frequency of a finite verb on the one hand, and the distribution of its exemplars across main (VO) versus subordinate (OV) clauses on the other: High-frequency finites populate main clauses more densely than subordinate clauses; lower-frequency finites tend toward a balanced VO/OV distribution when frequency drops. Specifically, the proportion of finites functioning as head of a main clause (expressed as a percentage of their total frequency) decreases when going from highest- to lowest-frequency verbs: 92%-to-58% in German, 83%-to-56% in Dutch. To check whether the effect is due to within-clause linear order (VO/OV) or to hierarchical position of the clause (main/subordinate), we conducted the same study with spoken English, where all clauses are VO (SWITCHBOARD; Godfrey et al., 1992). The German/Dutch pattern returned in essence, but in much weakened form: 65%-to-54%.

A possible explanation departs from two assumptions. First, high-frequency verbs have a higher level of cognitive availability (accessibility) than low-frequency ones, and thus can enter the grammatical encoding process earlier. Second, hierarchically higher clauses tend to be encoded prior to hierarchically lower clauses. In German and Dutch, high-frequency finite verbs facilitate the grammatical encoding of main clauses in two ways: Available early, they are more likely than low-frequency verbs (1) to end up in hierarchically higher clauses, e.g. main clauses, and (2) to be ready for an early position within those clauses, i.e. VO rather than OV. In English, the second factor cannot work.

The observation that the verb frequency effect is stronger in German than in Dutch, could reflect the more posterior position of finite verbs in German than in Dutch subordinate clauses, due to contrasting rules for placement of finite verbs in verb clusters, and of other constituents in the Endfield/Nachfeld.

The sketched availability-based mechanism invites speculations on whether it could have been a driving force behind historical (S)OV-to-(S)VO developments.

References
