

On Domain Minimization: PP ordering revisited

Research into the serialization of syntactic constituents has accumulated increasing evidence for the idea that language users tend to prefer constituent orders that impose fewer demands on (verbal) working memory (Hawkins, 1994, 2004; Gibson, 1998; Wasow, 2002; *inter alia*). One influential proposal in this context is John Hawkins' theory of linguistic efficiency and the principles of domain minimization formulated therein. Crucially, these principles predict that speakers will prefer a structure S over a possible alternative S' in proportion to the overall difference in efficiency between S and S', where efficiency depends on the number of linguistic units that need to be perceived to recognize domains of dependent elements. The latest version of the theory considers both syntactic and semantic dependency domains, whose minimization contributes to an increase in efficiency of a structure. The relative ordering of multiple PPs as shown in the examples in (i) – (ii) may serve as an illustration (example taken from Hawkins 2004:114):

- (i) They VP[v counted PP1[on his son] PP2[in his old age].

- (ii) They VP[v counted PP2[in his old age] PP1[on his son]

The solid lines indicate the sequence of words that have to be parsed so as to recognize the internal structure of the respective verb phrases (the *phrasal combination domain*), which stretches from the verb to the head of the second PP. The ordering in (i) is a little more efficient as one word less has to be processed to recognize the immediate constituents of the VP. In addition to this syntactic domain, the sentence also exhibits a lexical-semantic dependency (the *lexical domain*), which ranges from *counted* to *on* as indicated by the dashed line. Such domains are characterized by the fact that certain semantic properties of the predicate can only be assigned once both of these elements have been processed. Again the theory predicts that speakers prefer (i) over (ii) as the lexical-semantic dependency domain is much shorter in (i), thereby imposing fewer demands on working memory. Hawkins (2000: 253) argues that it is syntactic domain minimization that is the strongest ordering constraint relegating semantic dependencies to a secondary role.

The present study sets out to (re-)assess the role of both syntactic and semantic domain minimization through multifactorial analysis. To this end, we queried the British component of the International Corpus of English for all V PP PP sequences and manually weeded out all those instances that exhibit a hierarchical ordering of the PPs. Following Hawkins (2000), we also excluded all instances in which any other material appeared in the relevant clause apart from the two PPs ($N_{\text{remaining instances}} = 1256$). We then annotated the data with respect to syntactic and lexical-semantic dependencies. The magnitude of domain minimization was expressed as the difference in domain-length of the observed order and the (not actualized) alternative order. Subsequently, we pitted syntactic and semantic dependency domain minimizations against each other and estimated their relative strength and importance using logistic regression models without an intercept (Levy, *in progress*: 123-128). Our results contrast with previous analyses (most notably Hawkins 2000, 2004) as they suggest that the lexical-semantic dependency constitutes a stronger constraint on serialization than the weight related syntactic one. More specifically, our results show that while syntactic minimization has much greater data coverage – it applies to a much larger proportion of the data (978/1257 data points ~ 78% of the data) –, the lexical-semantic factor has a much greater effect size, thus is much more seldomly violated (LEXICAL MINIMIZATION: Regression coefficient estimate = 0.67, SE = 0.06, $z = 11.43$, $\text{Pr}(>|z| < 2e-16)$; SYNTACTIC MINIMIZATION: Regression coefficient estimate = 0.39, SE = 0.03, $z = 11.37$, $\text{Pr}(>|z| < 2e-16)$). We interpret our findings as presenting evidence for a more semantically-driven theory of constituent serialization.

References

- Hawkins, John. 1994. *A performance theory of order and constituency*. Cambridge: CUP.
- 2000. The relative order of prepositional phrases in English: Going beyond manner-place-time. *Language Variation and Change* 11: 231-266.
- 2004. *Efficiency and Complexity in Grammars*. Oxford: OUP.
- Gibson, Edward. 1998. Linguistic complexity: Locality and syntactic dependencies. *Cognition* 68:1-76.
- Levy, Roger. in progress. *Probabilistic Models in the Study of Language*. Cambridge, MA: MIT Press. (June 22, 2011 - <http://idiom.ucsd.edu/~rlevy/textbook/text.html>)
- Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: CSLI Publications.