Frequency distributions of dependency length differentials: The alternating word orders of object and adverbial

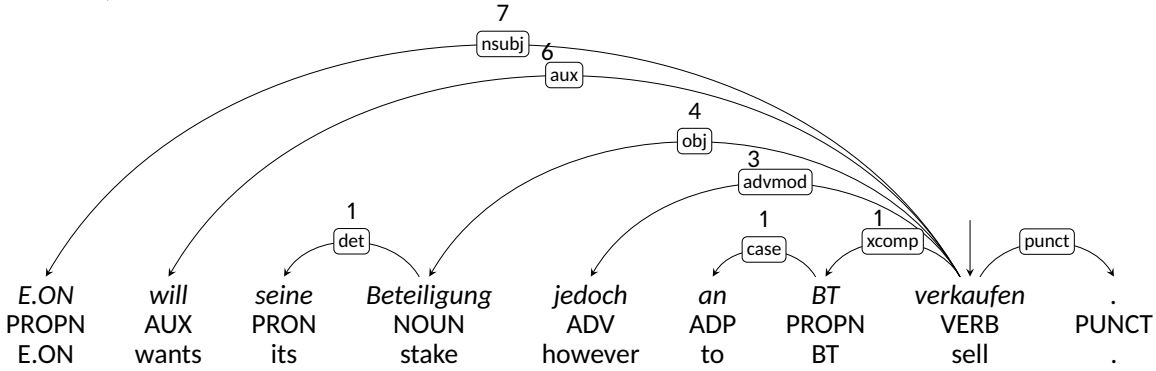
Hing-Yuet Fung
University of Hong Kong

Keywords: word order alternations, adverbial positions, frequency distributions, dependency length minimization, preverbal and postverbal domains

This study uses data similar to the large-scale study of dependency length minimization (Futrell, Mahowald, & Gibson, 2015), and suggests that dependency length analysis should be studied by word orders, which produce distributions of different shapes. We choose to study the alternating orders of a (direct) object (annotated by the dependency relation OBJ) and a (phrasal) adverbial modifier (ADVMOD), which have different tendencies to be placed first due to topicality and other factors. (1) shows a German sentence and its dependency analysis. Numbers above the arcs indicate dependency lengths.

(1) German

'However, E.ON wants to sell its stake to BT.'



In Futrell et al. (2015), the sum of dependency lengths in a sentence is observed to lie probabilistically within the boundary between a theoretical optimal value and a free random baseline. This distribution can be further analyzed by considering the relative dependency lengths between the observed order (2a) and a hypothetical alternative order (2b), in the spirit of length differentials in Hawkins (1994).

(2) a. E.ON will [seine Beteiligung]$_{\text{OBJ}}$ [jedoch]$_{\text{ADVMOD}}$ an BT verkaufen. (observed; sum=23)

b. E.ON will [jedoch]$_{\text{ADVMOD}}$ [seine Beteiligung]$_{\text{OBJ}}$ an BT verkaufen. (alternative; sum=24)

Results of 109k matched clauses from 68 eligible languages in the Universal Dependencies project (UD 2.12; Zeman et al., 2023) are presented in Figure 1. To study the interaction between object/adverb linearization and the position of the verb, data are separated into preverbal and postverbal domains, with reference to the dominant word order of the language as categorized in Dryer and Haspelmath (2013). Frequency counts of the adverb-first order is plotted upward, and the object-first order downward. Goodness-of-fit with the normal distribution was not meaningful due to the underdispersion of the data but each distribution is clearly unimodal. The dashed line of residuals between the upward and downward distributions illustrates the unaligned peaks and different skew directions. One implication of the results is that languages categorized as having no dominant order (such as German) behave similarly to OV languages in the preverbal domain (Figure 1a & e), and similarly to VO languages in the postverbal domain (Figure 1d & f).

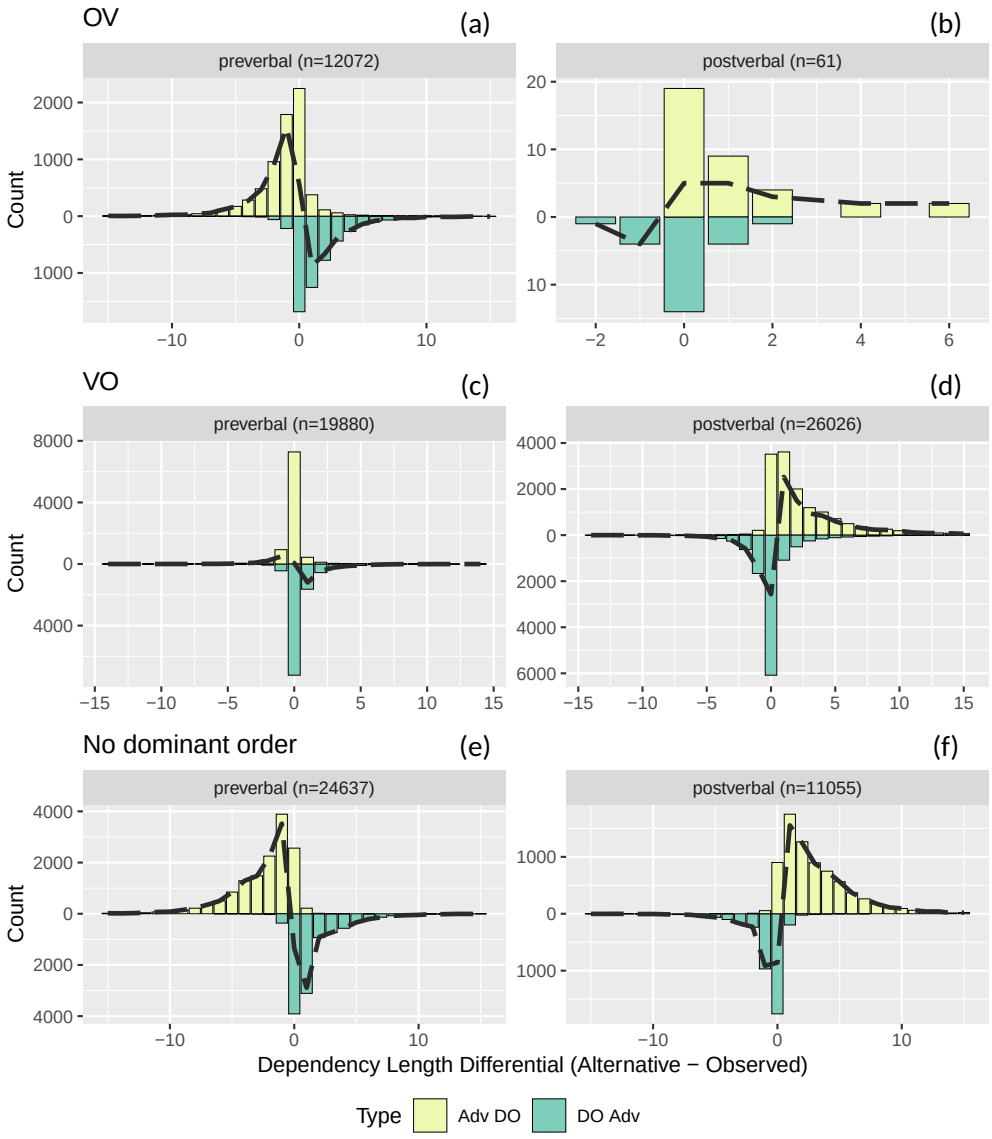


Figure 1: Distributions of dependency length differentials with alternating direct object (DO) and adverbial

References

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *WALS Online (v2020.3)*. Zenodo.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge University Press.

Zeman, D., Nivre, J., Abrams, M., Ackermann, E., Aepli, N., Aghaei, H., … Ziane, R. (2023). *Universal Dependencies 2.12*. Retrieved from `http://hdl.handle.net/11234/1-5150` (LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University)