EVOLUTIONARY RATES AND DIRECTIONS OF WORD ORDER CHANGE ACROSS MAIN AND SUBORDINATE CLAUSES IN INDO-EUROPEAN

YINGQI JING[*1,2], PAUL WIDMER[1,2], and BALTHASAR BICKEL[1,2]

[*]Corresponding Author: yingqi.jing@uzh.ch
[1]Department of Comparative Language Science, Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland
[2]Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland

The evolution of major word order patterns remains an open issue. One question concerns potential differences between clause types. Some researchers suggest that subordinate clauses are more conservative in word order than main clauses (Givón, 1979; Bybee, 2002), whereas others claim that subordinate clauses are more innovative (Stockwell & Minkova, 1991), or that there is no difference at all (Kroch, 1989). A second issue is the direction of word order change. Some theories posit a preference for early head recognition, hence early V positions (Hawkins, 2014). Theories of dependency length minimization tend to predict a V-medial placement (Liu, 2008; Futrell, Mahowald, & Gibson, 2015; Ferrer-i-Cancho, 2015). Predictability maximization theory supports a V-final ordering, since this configuration can maximize the predictability of the head (McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011; Ferrer-i-Cancho, 2017).

Aiming at further resolution in these debates, we use Bayesian phylogenetic inference to test hypotheses on word order evolution in the history of Indo-European. We go beyond earlier approaches and include not only basic orders (Dryer, 2013), but all observed VA/AV, VO/OV, and AO/OA orders in main and subordinate clauses in a sample of Indo-European languages, with orders ranked by rough frequency estimates as reported in descriptive grammars. We assign probabilities for each word order in specific contexts according to grammars, e.g. a 2/3 probability for VS in all-new (thetic) main clause context, a 1/2 probability in subordinate clauses etc. With these context-dependent probabilities, we can approximate the probability distribution of word orders in real utterances by re-sampling word orders from each context (500 times), so as to incorporate uncertainty into our model.

We then model the evolution of word orders as discrete states (VA/AV, VO/OV, and AO/OA) on an explicit phylogeny (Chang, Cathcart, Hall, & Garrett, 2015), assuming a Continuous-time reversible Markov Chain model of evolution and es-

timating transition rates with a Bayesian MCMC approach. We choose a modified implementation in R's `phytools` package (Revell, 2012) which allows probability statements on states and does not enforce categorical decision. For example, the model allows a clause type in a language to be coded as having a .8 probability of OV rather than categorically as OV or VO.

Our results reveal a preferred direction of change towards AV and AO orders, but no preference in the change of the order of V and O (Figure 1). Evolutionary rates do not differ much between main and subordinate clauses, although the order of V and A tends to be slightly more stable in main clauses (mean rate of change: 0.0026) than in subordinate clauses (mean rate of change: 0.0046).

The evolutionary bias towards AV and AO is likely due to a general subject or agent first principle (Greenberg, 1963; Napoli & Sutton-Spence, 2014). The evidence for equal rates models in the placement of V and O challenges claims on the universality of trends from V-final to V-medial (Gell-Mann & Ruhlen, 2011; Maurits & Griffiths, 2014) and suggests that this placement might be less constrained than commonly assumed. Further research is needed to assess this with a higher-resolution corpus-based approach, and in other families, before we can draw firm conclusions.

Our findings also contradict the hypothesis that subordinate clauses are particularly conservative or particularly innovative. Instead, it is likely that the factors determining word order alternations might be more similar to each other in main and subordinate clauses than is commonly assumed. This also casts doubts on the notion that main clauses allow more variation than dependent clauses synchronically (Ross, 1973) and invites a less constrained view of the evolutionary dynamics of word order.
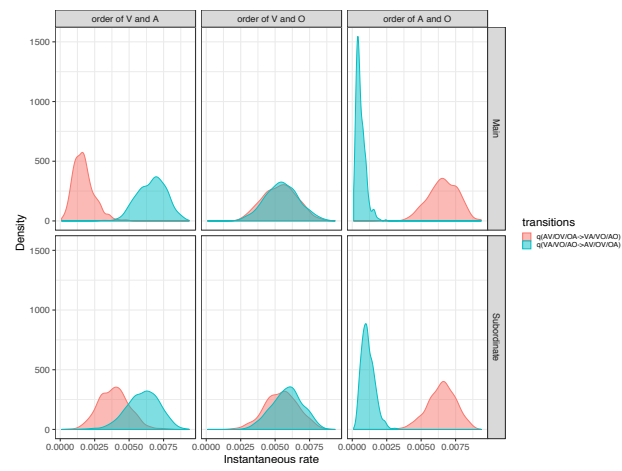
Figure 1. Posterior rate estimates

References

Bybee, J. (2002). Main clauses are innovative, subordinate clauses are conservative: Consequences for the nature of constructions. In M. N. Joan Bybee & S. Thompson (Eds.), *Complex sentences in grammar and discourse: essays in honor of Sandra A. Thompson* (pp. 1–17). Amsterdam: John Benjamins Publishing Company.

Chang, W., Cathcart, C., Hall, D., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, *91*(1), 194–244.

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology [`http://wals.info/`].

Ferrer-i-Cancho, R. (2015). The placement of the head that minimizes online memory: a complex systems approach. *Language Dynamics and Change*, *5*(1), 114–137.

Ferrer-i-Cancho, R. (2017). The placement of the head that maximizes predictability. an information theoretic approach. *Glottometrics*(39), 38–71.

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.

Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, *108*(42), 17290–17295.

Givón, T. (1979). *On understanding grammar* (Vol. 379). New York: Academic Press.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (p. 73-113). Cambridge, Mass.: MIT Press.

Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency.* OUP Oxford.

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, *1*(3), 199–244.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191.

Maurits, L., & Griffiths, T. L. (2014). Tracing the roots of syntax with bayesian phylogenetics. *Proceedings of the National Academy of Sciences*, *111*(37), 13576–13581.

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science*, *14*(2), 181–189.

Napoli, D. J., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: implications for all language. *Frontiers in Psychology*, *5*, 376.

Revell, L. J. (2012). phytools: an r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223.

Ross, J. R. (1973). The penthouse principle and the order of constituents. In T. C. S.-S. Claudia Corum & A. Weiser (Eds.), *You take the high node and I'll take the low node* (pp. 397–422). Chicago Linguistic Society.

Stockwell, R., & Minkova, D. (1991). Subordination and word order change in the history of English. In D. Kastovsky (Ed.), *Historical English syntax* (pp. 367–408). Walter de Gruyter.