

Usage-based evolutionary models reveal context-specific word order change in Indo-European

In the early days of Indo-European studies, Schleicher (1868) published his famous reconstruction of a fable in Proto-Indo-European to demonstrate that is possible to reconstruct all aspects of a language, from lexicon to morphology, domains where the comparative method has been applied widely and successfully, but also syntax. After the foundational work of Delbrück (1893-1900) and others (Wackernagel 1892; Brugmann 1925), the interest in syntactic reconstruction ceased for a couple of decades. In the 1970s, both Lehmann (1974) and Friedrich (1975) proposed reconstructions of basic word order patterns in Proto-Indo-European following the Greenbergian framework of conditional word order universals. Recently, computational phylogenetic methods have been applied to model the diachronic dynamics and reconstruct syntactic traits and other grammatical features (Greenhill et al. 2010, 2017; Dunn et al. 2011; Carling and Cathcart 2021, and more).

Proto-Indo-European word order was very likely flexible to some extent and allowed non-basic word order for emphasis and to mark information-structural properties (Viti 2014; Lühr 2015). Studies of documented word order changes highlight the importance of synchronic variation as a precondition for change (England 1991; Harris and Campbell 1995; Ross 2007; Heine 2008). Therefore, we propose a new approach to infer the evolutionary dynamics of word order under different pragmatic conditions. Instead of coding word order as an abstract type, we take observed instances in specific pragmatic contexts as a starting point.

To control for pragmatics and information-structure, we extracted a set of 46 sentences in 36 modern Indo-European languages from a parallel corpus (Levshina 2016, with additional data collected by the authors to enhance the coverage of Indo-European languages). Our sentence sample includes different types of subjects and objects (pronouns, nouns, and object clauses) to cover a wide range of constructions that are common in naturalistic speech. We use Bayesian phylogenetic comparative methods to infer transition rates between the states of a binary feature that encodes the order of object and verb. The follow-up analysis investigates whether some of these contexts are more prone to change than others.

The long-term probability of being in one state or the other varies between sentences, with some sentences having a higher probability for OV, while others have a higher probability for VO. This suggests the co-existence of different word order patterns in Proto-Indo-European.

By applying k-means clustering on the mean posterior rates, we identified sets of sentences that evolve in a similar way: one cluster contains verbs of speech and mental verbs with complement clauses which are almost exclusively VO in all modern Indo-European languages, even in those with basic OV order. A second cluster encompasses verbs with object pronouns which tend to precede the verb in many languages of the Romance and Slavic branch that otherwise prefer post-verbal nominal objects. The last cluster contains mostly nominal objects.

Our study did not provide evidence for the initial hypothesis that pragmatic factors lead to more variation and therefore faster rates of change. It is likely that our sentence sample was too small to have sufficient data for word order variation conditioned by pragmatic factors. Specific constructions that share semantic and structural properties can still be identified based on their distinct rates of change. This suggest that these factors play a major role in the evolution of word order.

References

- Brugmann, Karl. 1925. *Die Syntax Des Einfachen Satzes Im Indogermanischen*. Berlin; Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783111586304>.
- Carling, Gerd, and Chundra Cathcart. 2021. “Reconstructing the Evolution of Indo-European Grammar.” *Language* 97 (3): lan.0.0253. <https://doi.org/10.1353/lan.0.0253>.
- Delbrück, Berthold. 1893-1900. *Vergleichende Syntax Der Indogermanischen Sprachen*. Grundriss Der Vergleichenden Grammatik Der Indogermanischen Sprachen 3–5. Strassburg: Karl J. Trübner.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russel D. Gray. 2011. “Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals.” *Nature* 473: 79–82.
- England, Nora C. 1991. “Changes in Basic Word Order in Mayan Languages.” *International Journal of American Linguistics* 57 (4): 446–86.
- Friedrich, Paul. 1975. *Proto-Indo-European Syntax: The Order of Meaningful Elements*. Journal of Indo-European Studies. Butte: Montana College of Mineral Sciences.
- Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade, and Russell D. Gray. 2010. “The Shape and Tempo of Language Evolution.” *Proceedings of the Royal Society B: Biological Sciences* 277 (1693): 2443–50.
- Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson, and Russell D. Gray. 2017. “Evolutionary Dynamics of Language Systems.” *Proceedings of the National Academy of Sciences of the United States of America* 114: E8822–29.
- Harris, Alice C., and Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge Studies in Linguistics 74. Cambridge: Cambridge University Press.
- Heine, Bernd. 2008. “Contact-Induced Word Order Change Without Word Order Change.” In *Language Contact and Contact Languages*, edited by Peter Siemund and Noemi Kintana, 33–60. Amsterdam; Philadelphia: Benjamins.
- Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Austin; London: University of Texas Press.
- Levshina, Nathalia. 2016. “Verbs of Letting in Germanic and Romance: A Quantitative Investigation Based on a Parallel Corpus of Film Subtitles.” *Languages in Contrast* 16 (1): 84–117. <https://doi.org/10.1075/li.c.16.1.04lev>.
- Lühr, Rosemarie. 2015. “Traces of Discourse Configurability in Older Indo-European Languages?” In *Perspectives on Historical Syntax*, edited by Carlotta Viti, 203–32. Studies in Language Companion Series 169. Amsterdam & Philadelphia: Benjamins.
- Ross, Malcolm. 2007. “Calquing and Metatypy.” *Journal of Language Contact* 1: 116–43. <https://doi.org/10.1163/000000007792548341>.
- Schleicher, August. 1868. “Eine Fabel in Indogermanischer Ursprache.” *Beiträge Zur Vergleichenden Sprachforschung Auf Dem Gebiete Der Arischen, Celtischen Und Slawischen Sprachen* 5 (2): 206–8. <http://www.jstor.org/stable/23458804>.
- Viti, Carlotta. 2014. “Reconstructing Syntactic Variation in Proto-Indo-European.” *Indo-European Linguistics* 2 (1): 73–111. <https://doi.org/10.1163/22125892-00201004>.
- Wackernagel, Jakob. 1892. “Über Ein Gesetz Der Indogermanischen Wortstellung.” *Indogermanische Forschungen* 1: 333–436.