A phylogenetic model of trade-offs in strategies for determining 'who did what to whom'

Annemarie Verkerk - Universität des Saarlandes (Germany)
Luigi Talamo – Universität des Saarlandes (Germany)
Natalia Levshina – Max Planck Institut for Psycholinguistics (Netherlands)

Keywords: phylogenetics, modeling diachronic change, token-based typology, case marking, word order

As syntax textbooks teach us, the languages of the world have three major strategies to distinguish the two argument roles in a transitive sentence from each other: fixed or rigid word order, case marking, and indexing of argument roles on the verb (verb agreement). The relation between word order type and various types of morphological marking has been investigated for decades (*verb-initial word order implies head-marking*, Nichols 1992: Sect 3.3; *free word order implies case-marking*, Sinnemäki 2008). The negative correlation between case marking and rigid word order has been interpreted as a efficiency trade-off: language users do not need all possible cues for determining 'who did what to whom', hence the expectation is that most languages dominantly use only word order or case marking (as demonstrated by Fedzechkina et al. 2017). As for verb agreement, no evidence supporting trade-offs has been found (Sinnemäki 2008).
However, while the explanation in terms of an efficiency trade-off may make sense from a synchronic, language-in-use perspective (although more research of users' choices in particular contexts is needed), we must not forget that head- and dependent-marking as well as clausal word order patterns are diachronically stable features of languages that nevertheless show great cross-linguistic variability (Helmbrecht 2001, Dunn et al. 2011). Strong crosslinguistic evidence for the efficiency trade-off provided by Nichols (1992), Sinnemäki (2008), and others must be rooted in diachronic explanations. These have remained underexplored. A recent causal scenario for the interdependency of a concrete constellation of these variables (case marking, rigid word order, verb-medial, and so-called tight semantics, which represents strong associations between syntactic roles and lexemes that express them (Levshina 2020, based on Hawkins 1986, 1995)) was sketched by Levshina (2021). She argues for a tradeoff between rigid word order and case marking as a manifestation of efficient behavior mediated by specific sociolinguistic context, especially the impact of L2 speakers and their preference for rule-based learning. This calls for an explicit diachronic investigation, taking into account linguistic data on the trade-off for argument resolution as well as extra-linguistic variables, most importantly sociolinguistic ones.
The current study takes an explicit diachronic approach by modeling the trade-off between case marking, rigid word order, verb-medial, and tight semantics and their interaction in an explicit phylogenetic context, including sociolinguistic variables on the number of L2 speakers and language use and history. Our data consist of (parallel) corpora, parsed with Universal Dependencies parsers (Marneffe et al. 2021), in 34 Indo-European languages (Eastern Armenian, Hellenic (2 languages), Indo-Aryan (4), Celtic (2), Slavic (12), Romance (7), and Germanic (6)), including 7 historical languages. We model the four variables in terms of mutual information and entropy (Levshina 2020, 2021) and sociolinguistic variables on the phylogeny by Bouckaert et al. (2012) using mixed effect models in brms (Bürkner 2017). Additionally, we use SPELT (Nunn and Cooper 2015) to determine the direction of change. Preliminary results suggest a strong negative relation between case marking and rigid word order, confirming earlier work, with change in case marking lagging behind change in rigidity of word order. We have no results on including sociolinguistic variables yet, but hypothesize that the number of L2 speakers may be a too coarse/uninformative variable (see Dobrushina & Moroz 2021). We propose alternatives in line with the sociolinguistics of the Indo-European language families.

References

Bouckaert, R, P Lemey, M Dunn, S J Greenhill, A V Alekseyenko, A J Drummond, R D Gray, M A Suchard, and Q D Atkinson. 2012. 'Mapping the Origins and Expansion of the IndoEuropean Language Family'. *Science* 337 (6097): 957–60.

Bürkner, Paul-Christian. 2017. 'Brms: An R Package for Bayesian Multilevel Models Using Stan'. *Journal of Statistical Software* 80 (1): 1–28.

Dobrushina, Nina, and George Moroz. 2021. 'The Speakers of Minority Languages Are More Multilingual'. *International Journal of Bilingualism*, June, 136700692110231.

Dunn, Michael, Simon J Greenhill, Stephen C Levinson, and Russell D Gray. 2011. 'Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals'. *Nature* 473 (7345): 79–82.

Fedzechkina, Maryia, Elissa L. Newport, and T. Florian Jaeger. 2017. 'Balancing Effort and Information Transmission During Language Acquisition: Evidence From Word Order and Case Marking'. *Cognitive Science* 41 (2): 416–46.

Hawkins, John A. 1986. *A Comparative Typology of English and German: Unifying the Contrasts*. London: Croom Helm.

Hawkins, John A. 1995. Argument-predicate structure in grammar and performance: A comparison of English and German. In Irmengard Rauch & Gerald F. Carr (eds.), *Insights in Germanic Linguistics, Vol. 1 Methodology in Transition*, 127-44. Berlin: Mouton de Gruyter.

Helmbrecht, Johannes. 2001. 'Head-Marking vs. Dependent-Marking Languages'. In *Language Typology and Language Universals: An International Handbook. Volum 2*, edited by Martin Haspelmath et al., 1424–1432. Berlin: Walter de Gruyter.

Levshina, Natalia. 2020. How tight is your language? A semantic typology based on Mutual Information. In *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, 70-78. Düsseldorf: ACL. https://www.aclweb.org/anthology/2020.tlt 1.7.pdf

Levshina, Natalia. 2021. Cross linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology* 12: 648200.

Marneffe, Marie-Catherine de, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. 'Universal Dependencies'. *Computational Linguistics*, May, 1–54.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Nunn, Charles L., and Natalie Cooper. 2015. 'Investigating Evolutionary Lag Using the Species-Pairs Evolutionary Lag Test (SPELT)'. *Evolution* 69 (1): 245–53.

Sinnemäki, Kaius. 2008. 'Complexity Trade-Offs in Core Argument Marking'. In *Language Complexity: Typology, Contact, Change*, edited by Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson, 67–88. Amsterdam: Benjamins.