Diachronic pathways of definite articles distribution

Definite articles are commonly suggested as an example of an unstable grammatical feature (Croft 1996, Wichmann & Holman 2009, Dediu and Levinson 2012, Greenhill et al. 2017). Frajzyngier (2008: 18) describes grammatical instability as a system "[i]n which given functional distinctions are neutralized in significantly extensive environments", a situation which may cause articles to emerge and subsequently erode from a system. However, the properties promoting article emergence, loss, and persistence on a global scale remain unclear. Different linguistic features have been posited as the predictors of their distribution – gender marking and nominal classifiers, verb typology, topic and focus marking, verb-final word order, flexible word order, switch reference (Givón 1978, 1983), ergative alignment (Du Bois 1987, 2017), case marking (Comrie 1989), and perfective aspect (Leiss 2000). Empirical studies identified case, verb-final word order, flexible subject order, and ergative alignment as potentially robust predictors of the absence of definite articles (Evers 2020). These relationships have been explained in the light of 1) efficiency (case and flexible argument order can serve to mark definiteness and languages with these features might be less likely to gain another grammatical means with the similar function) and 2) the information status in languages with ergative alignment and verb-final languages (A arguments are prototypically definite, and clause-final arguments are prototypically indefinite). However, the relationships between the four predictors are complex: case marking is typically found in verb-final and flexible word order languages (e.g. Levshina 2021), and many ergative languages have case marking. An explicit causal approach is thus required to reveal the diachronic processes behind the cross-linguistic distribution of definite articles.

Here we disentangle the causal relationships underlying the synchronic distribution of definite articles on a global sample of 1232 languages matched for typological information from Grambank (The Grambank Consortium 2022) and the global EDGE tree (Bouckaert et al. 2022). We use logistic regression with phylogenetic correction within phylogenetic path analysis in *phylopath* package (van der Bijl 2018) to first establish the existing causal relationships between the predictors themselves and then to evaluate how and whether case, word order, and ergativity explain the distribution of definite articles.

We find that verb-final languages are more likely to maintain and gain case marking, and languages with case marking are likely to develop flexible ordering of core arguments. We use this causal model as the basis for several competing models predicting definite articles distribution. Out of the tested predictors of definite articles, only verb-final word order proves robust on a global sample: articles are more likely to be absent or disappear in verb-final word order languages. This suggests that word order dependent preferences for negotiating information status primarily shape the evolution of definite articles. The previously observed patterns of complementary distribution between 1) definite articles and 2) case, ergative alignment, and flexible word order are not supported by causal inference. Instead, these suggested relationships might be products of complex interactions between case, word order, and ergativity.

References

Bijl, Wouter van der. 2018. *phylopath*: Easy phylogenetic path analysis in R. *PeerJ*. PeerJ Inc. 6. e4718.

Bouckaert, Remco, David Redding, Oliver Sheehan, Thanos Kyritsis, Russell Gray, Kate E Jones & Quentin Atkinson. 2022. Global language diversification is linked to socio-ecology and threat status [Preprint]. https://osf.io/preprints/socarxiv/f8tr6/.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80(1). 1–28.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.

Croft, William. 2002. *Typology and universals*. Cambridge University Press.

Dediu, Dan & Stephen C Levinson. 2012. Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLoS ONE* 7. e45198.

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.

Du Bois, John W. 2017. Ergativity in discourse and grammar. In Jessica Coon, Diane Massam & Lisa D. Travis (eds.), *The Oxford Handbook of Ergativity*, 23–58. Oxford: Oxford University Press.

Evers, Stephanie. 2020. *To 'the' or not to 'the': Cross-linguistic correlations between existing morphosyntax and the emergence of definite articles*. Buffalo, NY: The State University of New York at Buffalo PhD Thesis.

Frajzyngier, Zygmunt. 2008. Grammaticalization, typology and semantics. In María José López-Couso & Elena Seoane (ed.), 61–102. Amsterdam: John Benjamins.

Givón, Talmy. 1978. Definiteness and referentiality. *Universals of Human Language* 4. 291–330.

Givón, Talmy. 1983. Topic continuity in discourse: the functional domain of switch reference. In John Haiman & Pamela Munro (eds.), *Switch Reference and Universal Grammar*, vol. 2, 51–82. Amsterdam: Benjamins.

Greenhill, Simon J, Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C Levinson & Russell D Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*. National Acad Sciences 114(42). E8822–E8829.

Leiss, Elisabeth. 2000. Artikel und Aspekt. Die grammatischen Muster von Deftnitheit. In *Studia Linguistica Germanica 55*. Berlin: Walter de Gruyter.

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. 2791.

The Grambank Consortium (eds.). 2022. Grambank 1.0. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://grambank.clld.org.

Wichmann, Søren & Eric W Holman. 2009. *Assessing temporal stability for linguistic typological features*. Munich: LINCOM Europa.