

EVIDENCE FOR HIGH FREQUENCY OF LINGUISTIC RARA IN EUROPEAN LANGUAGES

ARTURS SEMENUKS^{*1}

^{*}Corresponding Author: asemenuk@ucsd.edu

¹ *Department of Cognitive Science, University of California, San Diego, San Diego, USA*

The concept of an absolute linguistic universal (ALU) – a feature that is shared or avoided by all languages – is important and useful in linguistic theory building. Indeed, presence of a truly absolute universal in a linguistic domain would impose an *absolute constraint* on the structures of explanatory theories for that domain, and thus highly likely have fundamental consequences for understanding cognitive and/or extracognitive pressures on linguistic structure, for the questions of how and why language evolved and evolves, and, ultimately, for specifying what are the defining features of language (Bickel, 2007; Piantadosi & Gibson, 2014).

However, although convincingly defending the existence of an ALU is a useful theoretical goal, some have put forward critical arguments for the position that currently there is no good evidence for the existence of many, if any, theoretically profound absolute universals (Dryer, 1998; Evans & Levinson, 2009). A single counterexample is enough to negate a proposed universal claim, which has been the fate of many of them, highlighting the variability in the potential structures that human language can obtain (Evans & Levinson, 2009). Furthermore, computational modeling suggests that due to sampling limitations rarely can typological investigations alone warrant considering a pattern an ALU, even for cases without exceptions among languages surveyed so far (Piantadosi & Gibson, 2014).

Because of this, it is crucial to better understand what leads to emergence of linguistic *rara* – features of languages that are extremely typologically uncommon and limited to only a handful of languages (Cysouw & Wohlgemuth, 2010). Such knowledge would provide assistance in finding counterevidence for ALUs by suggesting which languages a violation of a proposed ALU might be found in. The current work, in particular, aimed to investigate whether European

languages are overrepresented among the known languages with linguistic rara due to the researchers typically having more intimate knowledge of them.

To examine this, I web scraped the University of Konstanz electronic database of linguistic rara (Plank, 2006). The rara were matched with their corresponding area and genealogical stock using the AUTOTYP database (Nichols, Witzlack-Makarevich & Bickel, 2013). In order to safeguard against a stock being overrepresented due to retainment of a diachronically old and stable rare feature and simplify further analyses, only those rara that were reported to be found in a single language were left in the sample.

Statistical analyses of the data suggest that European languages are reported to contain a linguistic rarum that has been observed only in a single language more frequently than would be expected. Bootstrapping simulations also suggest that European languages contain unique rara more frequently than should be expected by chance (Figure 1).

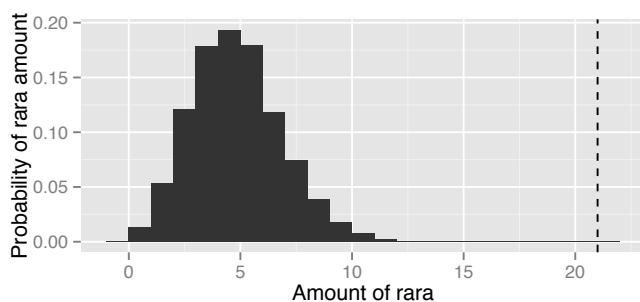


Figure 1. The probabilities of European languages having different amounts of rara based on 10000 bootstrapping simulations. The dotted line denotes the observed amount of rara.

The results suggest two primary interpretations of the data. The first interpretation is that having a unique linguistic feature is more widespread among the world's languages than is usually assumed, and European languages stand out in the data due to implicit ethnocentrism of the researchers who are more familiar with the features of languages spoken in Europe, and so are more likely to spot a rarum in them. This could mean that the amount of ALUs is overestimated. The second interpretation suggests that European languages are indeed more unusual than the “cross-linguistic mean”, which has implications for the sociological aspect of linguistic theory building. I will discuss the implications of both interpretations for the study of linguistic universals, and argue for the former interpretation.

References

- Bickel, B. (2007). Typology in the 21st century: major current developments. *Linguistic Typology*, 11(1), 239-251.
- Cysouw, M., & Wohlgemuth, J. (2010). The other end of universals: Theory and typology of rara. In: J. Wohlgemuth & M. Cysouw (Eds.), *Rethinking Universals. How Rarities Affect Linguistic Theories*. Berlin, New York: Mouton de Gruyter, 1–10.
- Dryer, M. S. (1998). Why statistical universals are better than absolute universals. *Chicago Linguistic Society 33: The Panels*, 123-145.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*, 32(5), 429-448.
- Nichols, J., Witzlack-Makarevich, A., & Bickel, B. (2013). *The AUTOTYP genealogy and geography database: 2013 release*. Zurich: University of Zurich.
- Piantadosi, S. T., & Gibson, E. (2014). Quantitative standards for absolute linguistic universals. *Cognitive Science*, 38(4), 736-756.
- Plank, F. (ed.) (2006). *Das grammatische Raritätenkabinett*. Electronic database, online at: <http://typo.uni-konstanz.de/rara/intro/>