A Computational and Quantitative Revisitation of some Greenberg's Universals and Linguistic Types

Antoni Brosa Rodríguez
(Universitat Rovira i Virgili)

Keywords: Universal Dependencies, Typometrics, Token-based Typology, Greenberg, Fuzzy Logic

Linguistic typology evolved thanks to the proposal of Greenberg (1963). He proposed a series of language universals in word order or morphosyntax that show invariance between groups of different and distant languages united by the same shared linguistic type.

In subsequent years his claims have been revised by extending the original 30-language sampling to include more languages (Dryer 1992, 2009, 2019). However, the research methodology did not change.

Instead, nowadays, thanks to the availability of other methodologies such as the analysis of large language corpora, we can achieve a different and more precise type of knowledge, based on occurrences in real texts and obtain other metrics (Futrell et al. 2020, Gerdes et al. 2021, Guzmán Naranjo & Becker 2018, Levshina 2022, Liu 2010).

Therefore, in this research we ask whether it is possible to test and validate universals computationally through texts and whether we can formulate new universals from those proposed by Greenberg. In other words, we propose a quantitative (and not categorical and grammar-based, as in Greenberg) approach that allows us to observe in greater detail the precise behavior of the different languages of the world and to obtain non-prototypical structures.

The methodology used for the study is based on the computational analysis of data hosted in Universal Dependencies (2.11), relating to 143 different languages in 243 corpora of varying typology. For the processing of syntactic or morphosyntactic data, the Grew-Match tool is used (and we add graphical representation using Typometrics). In addition to the presentation of the data in quantitative format and observing the graphical distribution of the languages, we also propose to follow the guidelines of Gries (2013) and Levshina et al. (2021) to subsequently allow comparability with the Greenberg labels and to assess the differences found.

We offer a method for the determination of the linguistic types of the dominant word order, as Greenberg wanted, based on objective, falsifiable and quantificational criteria of languages not available in WALS, for example. We offer a formalization for universals 1, 4, 7 and 13. We analyze the (satisfactory) precise degree of fulfilment of these universals in UD. We propose modifications or extensions of these universals based on the alternative data we have analyzed. We propose a reading of the universals with reference to the clusters that can be formed in the distribution of languages.

Acknowledgments: Research Group on Mathematical Linguistics (GRLMC), 2021-SGR-00032.

References

Dryer, M. S. (1992). The Greenbergian Word Order Correlations. Language, 68(1), 81–138. http://www.jstor.com/stable/416370

Dryer, M. S. (2009). The Branching Direction Theory of Word Order Correlations Revisited. In S. Scalise, E. Magni, & A. Bisetto (Eds.), Universals of Language Today (pp. 185–208). Springer.

Dryer, M. S. (2019). On the order of Demonstrative, Numeral, Adjective, and Noun. Language, 94(4), 798–833.

Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. Language, 96(2), 371–412. https://doi.org/10.1353/lan.2020.0024

Gerdes, K., Kahane, S., & Chen, X. (2021). Typometrics: From implicational to quantitative universals in word order typology. Glossa, 6(1). https://doi.org/10.5334/gjgl.764

Greenberg, J. H. (1963). Universals of Language. The M.I.T. Press.

Gries, S. T. (2013). Statistics for Linguistics with R : a Practical Introduction. De Gruyer Mouton.

Guzmán Naranjo, M., & Becker, L. (2018). Quantitative word order typology with UD. 17th International Workshop on Treebanks and Linguistic, 91–104.

Levshina, N. (2022). Corpus-based typology: Applications, challenges and some solutions. Linguistic Typology, 26(1), 129–160. https://doi.org/10.1515/lingty-2020-0118

Levshina, N., Namboodiripad, S., Allassonnière-Tang, M., Kramer, M. A., Talamo, L., Verkerk, A., Wilmoth, S., Rodriguez, G. G., Gupton, T., Kidd, E., & Liu, Z. (2021). Why we need a gradient approach to word order. Linguistics, 61(4). https://doi.org/https://doi.org/10.31234/osf.io/yg9bf

Liu, H. (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. Lingua, 120(6), 1567–1578. https://doi.org/10.1016/j.lingua.2009.10.001