

Correlations between linguistic features are reflected in their geospatial patterning: Introducing the geo-typological Sandwich Conjecture

A wide variety of work in both the Greenbergian typological tradition (Greenberg 1963, 1978; Hawkins 1979; Dryer 1989; Nichols 1992) and in the comparative-syntax paradigm (Baker 2001, 2008; Longobardi and Guardiano 2009; Roberts 2019) has established that different surface/typological features are frequently *correlated*; individual feature values can favour or disfavour others. Such correlations arise over historical time in a dynamical process in which both “vertical” genetic descent and “horizontal” contact processes play a role, and it has been suggested that the present-day global distribution of language types is reflective of the stationary state of this stochastic process (Greenberg 1978; Maslova 2000; Kauhanen et al. 2021; Jäger and Wahle 2021). This makes possible, among other things, the estimation of linguistic rates of change using a variety of techniques (Dediu and Cysouw 2013; Murawaki and Yamauchi 2018; Kauhanen et al. 2021).

What has received far less attention so far is the question of how *geographical* patterns of linguistic features relate to the purely linguistic properties of those features. Although Kauhanen et al. (2021) put forward a method for inferring rates of change of features from their present-day geospatial distributions, their method is restricted to statistically independent features. In this paper, using a set of word order features as our data, we show that feature correlations and geospatial distributions in fact stand in a systematic relationship. We show that typologically dispreferred types (such as the disharmonic combination OV & prepositions) tend to be surrounded by a greater variety of types than typologically preferred types (such as the harmonic combination OV & postpositions), which favour more uniform geographical environments. In particular, languages of dispreferred type are often found geographically “sandwiched” between languages of preferred type.

This *Sandwich Conjecture* is operationalized in terms of a notion of *neighbourhood entropy*, an information-theoretic measure of the extent of typological variability in a language’s immediate geographical neighbourhood. We apply this method to 28 word order feature pairs harvested from WALS (Dryer and Haspelmath 2013), and use an unrelated (and demonstrably uncorrelated) feature (presence/absence of voicing contrast) as a control feature. Measuring the typological correlation of two features with the usual ϕ coefficient (cf. Jäger and Wahle 2021), measuring “sandwichness” with neighbourhood entropy, and using permutation tests to factor out random noise, we then show that a statistically significant linear relationship exists between feature correlations and geospatial patterning: the more two features are correlated purely typologically (i.e. simply by virtue of the number of languages exhibiting the various possible feature combinations), the more sandwiched the corresponding geospatial distribution of types is. Hence, feature correlations are reflected in the geospatial patterning of said features.

This finding suggests a possible mechanism for the diachronic stability of dispreferred types, an otherwise surprising fact given that disharmonic feature combinations should be expected to resolve one way or the other into harmony over sufficient historical time. The Sandwich Conjecture explains this paradoxical stability by proposing that dispreferred types are preferentially found in typologically rich geographical environments, meaning that ample opportunities for horizontal transfer (contact) exist to sustain disharmony.

In addition to the above quantitative results, we present converging evidence from proof-of-concept computer simulations in which the same finding emerges.

References

- Baker, Mark. 2001. *The Atoms of Language*. New York: Basic Books.
- . 2008. “The Macroparameter in a Microparametric World.” In *The Limits of Linguistic Variation*, edited by Theresa Biberauer, 351–73. John Benjamins.
- Dediu, Dan, and Michael Cysouw. 2013. “Some Structural Aspects of Language Are More Stable Than Others: A Comparison of Seven Methods.” *PLOS ONE* 8: e55009.
- Dryer, Matthew S. 1989. “Large Linguistic Areas and Language Sampling.” *Studies in Language* 13: 257–92.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/>.
- Greenberg, Joseph H. 1963. “Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements.” In *Universals of Human Language*, edited by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
- . 1978. “Diachrony, Synchrony, and Language Universals.” In *Universals of Human Language*, edited by Joseph H. Greenberg, 1:61–91. Stanford, CA: Stanford University Press.
- Hawkins, John A. 1979. “Implicational Universals as Predictors of Word Order Change.” *Language* 55 (3): 618–48.
- Jäger, Gerhard, and Johannes Wahle. 2021. “Phylogenetic Typology.” *Frontiers in Psychology* 12: 682132.
- Kauhanen, Henri, Deepthi Gopal, Tobias Galla, and Ricardo Bermúdez-Otero. 2021. “Geospatial Distributions Reflect Temperatures of Linguistic Features.” *Science Advances* 7: eabe6540.
- Longobardi, Giuseppe, and Cristina Guardiano. 2009. “Evidence for Syntax as a Signal of Historical Relatedness.” *Lingua* 119 (11): 1679–1706.
- Maslova, Elena. 2000. “A Dynamic Approach to the Verification of Distributional Universals.” *Linguistic Typology* 4: 307–33.
- Murawaki, Yugo, and Kenji Yamauchi. 2018. “A Statistical Model for the Joint Inference of Vertical Stability and Horizontal Diffusibility of Typological Features.” *Journal of Language Evolution* 3: 13–25.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Roberts, Ian. 2019. *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.