Title: Exploring the dynamics of linguistic typological features in time and space using computer simulations

Despite recent advances in comparative and quantitative methods, there remain many questions concerning the influence of geography and genealogy on the global distribution of linguistic typological features. The well-known problem of spatial autocorrelation limits the usefulness of a direct statistical approach, and as a result various other approaches have been innovated. There are a number of studies exploring the stability or biases of typological features within language families (Dediu & Cysouw 2013; Dediu & Levinson 2012; Maslova 2000; Wichmann & Holman 2009). However, among those that account for geography directly, it is limited to comparisons of languages within families or between families. Other approaches which account for both genealogy and geography are limited to a particular region (Cysouw 2013), use typological distance measures that do not allow the comparison of features separately (e.g. Comrie & Cysouw 2012), or apply a ratio of genealogical and geographic measures that reduces independently varying factors to their quotient (Parkvall 2008).

In the last two decades, computer simulations have emerged as a promising method by which to explore these and related questions (e.g., Holman et al. 2007; Oliveira et al. 2008). However, nearly all of these studies have been limited to abstract representations of space such as a lattice. A recent exception is Wichmann (2017), who uses 'populated places' from the GeoNames database to add geographic realism in his simulations of language family expansions.

In this paper, we set out to explore the dynamics of linguistic features in time and space using simulations of languages in a realistic geographic environment. The foundation of our simulations—and a prerequisite for exploring feature dynamics—is a realistic model of language migration, speciation, and extinction. Following Wichmann (2017), language locations are limited to 'populated places' in the GeoNames database. At each time step, parameter settings determine the probability that a language will migrate to a new location, split into daughter languages, or disappear. Migration (including the placement of a new language in speciation) is determined by a quadrilateral method described in Wichmann (2017). Additional mechanisms such as preferred migration into underpopulated areas and speciation probabilities based on surrounding language density are also tested.

To ensure that our simulations closely model real language family distributions, we perform several tests to determine the best parameter settings. For example, we investigate a range of speciation and extinction probabilities to determine which levels yield the power-law distribution of language family sizes described in Wichmann (2005), and in doing so, we discuss the role of phylogenetic 'momentum' in our simulations. Similarly, we explore the relationship among various migration parameters in achieving realistic language family densities.

Equipped with a realistic model of language phylogeny and movement, we turn our attention to the dynamics of linguistic features. We assign feature values to languages, and these values are subject to a probability of change at each time step. This change is either random (due to pressures internal to the language) or the result of diffusion (based on the values of surrounding languages). Without the interference of internal change or diffusion, daughter languages will inherit feature values from their parents. We compare the output of our simulations to real language data from WALS (Dryer & Haspelmath 2013) using metrics

established in the literature such as Parkvall's (2008) genealogical and areal homogeneity measures. We find clear evidence for spatial autocorrelation in our simulations, and this could be used in the future to better understand this phenomenon and its specific role in linguistic typology.

To demonstrate its utility, we put the simulation to work on additional research questions. For example, to what extent do structured (such as ordinal) features behave differently from unstructured features? And, can we apply differing transitional probabilities to values of a feature to produce desired frequency disparities among those values?

In this paper, we share various insights into the dynamics of language phylogeny and migration, as well as linguistic feature dynamics. Yet, perhaps the most important contribution of this research is the model itself, which can be adapted for a variety of different research questions. Allowing us to model interactions among both phylogenetic and feature-based characteristics of languages in a real geographic landscape, it opens the door for further discoveries in the evolution of language and patterns of human migration and interaction.

References:

Comrie, Bernard & Michael Cysouw. 2012. New Guinea through the eyes of WALS. *Language & Linguistics in Melanesia* 30.

Cysouw, Michael. 2013. Disentangling geography from genealogy. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szmrecsanyi (eds.), *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, 21–37. Berlin: De Gruyter Mouton.

Dediu, Dan & Michael Cysouw. 2013. Some structural aspects of language are more stable than others: A comparison of seven methods. *PLoS ONE* 8(1).

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(9).

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Holman, Eric W., Christian Schulze, Dietrich Stauffer, & Søren Wichmann. 2007. On the relation between structural diversity and geographical distance among languages: observations and computer simulations. *Linguistic Typology*, 11. 395-423.

Maslova, Elena. 2000. Stochastic models in typology: obstacle or prerequisite? *Linguistic Typology* 4. 357–364.

Oliveira, Paulo Murillo Castro de, Dietrich Stauffer, Søren Wichmann, & Susana Moss de Oliveira. 2008. A computer simulation of language families. *Journal of Linguistics*, 44. 659-675.

Parkvall, Mikael. 2008. Which parts of languages are the most stable? *STUF*, 3. 234-250.

Wichmann, Søren. 2005. On the power-law distribution of language family sizes. *Journal of Linguistics*, 41(1). 117-131.

Wichmann, Søren. 2017. Modeling language family expansions. *Diachronica*, 34(1). 79-101.

Wichmann, Søren & Eric W. Holman. 2009. *Temporal stability for linguistic typological features*. Munich: Lincom Europa.