Abstract for 'Causality in the Language Sciences' (for oral presentation).

Jeremy Collins, Radboud University, Nijmegen

Controlling for Language Contact when Inferring Causality in Linguistics

Language contact is a problem for studying causality, because a correlation between two traits may emerge due to geographical non-independence of languages. For example, tone has spread through language families in Southeast Asia and Africa, causing tone to correlate with other traits found in those two regions, such as SVO word order (Yiu and Matthews 2013) or, facetiously, acacia trees (Roberts and Winters 2012). This paper makes the following arguments about how to adequately distinguish causal links from accidental geographical similarity, using tone as an example throughout:

i) Current controls for geographical distance are inadequate, for example the Mantel test used by Dediu and Ladd (2007) when they claim that the ancestral alleles of ASPM and microcephalin are causally linked with tone; I show that mitochondrial DNA haplogroups also predict the distribution of tonal languages after controlling for language family and geographical distance, showing that genes and linguistic properties can travel tightly together, between language families and potentially across large distances.

ii) Phylogenetic methods so far control for genealogical relatedness, but not areal non-independence; a recent example is Dunn et al.(2011)'s finding of lineage-specific word order correlations, which might reflect causal linkages but also the acquisition of word orders by language contact. A possible extension of this method is to use phylogenies based on structural features: unrelated Southeast Asian languages for example can influence each other in their structures, which may mean that a 'phylogeny' of Southeast Asian languages inferred from their structures may be a better way of controlling for their non-independence.

iii) A causal hypothesis based on data in one part of the world should ideally be tested on another part of the world. Dediu and Ladd (2007) formulated their tone-gene hypothesis based on the visual similarity of the distribution of tonal languages and the distribution of those two genes in Africa and Eurasia, making the strong correlation that they find unsurprising. By contrast, I show that a correlation between simple coda structure and tone holds up in Africa/Eurasia, which holds up independently in the rest of the world.

iv) Variation within a single language could be exploited to show a causal link; for example, tone is more likely to be omitted in longer words in Ket (Vajda 2004:15) suggesting an inverse relationship between lexeme length and the use of tone, a correlation which is borne out cross-linguistically using ASJP data.

References

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. PNAS, 104, 10944-10949. doi:10.1073/pnas.0610848104.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. Nature, 473, 79-82. doi:10.1038/nature09923.

Roberts, S. G., & Winters, J. (2012). Social structure and language structure: The new nomothetic approach. Psychology of Language and Communication, 16, 89-112. doi:10.2478/v10057-012-0008-6.

Vajda, E. (2004). Ket. Munich: Lincom Europa. Languages of the World vol. 204.

Yiu, S., & Matthews, S. (2013). Correlations between tonality and word order types. Poster presentation at the Association for Linguistic Typology Tenth Biennial Conference, Leipzig, August 13-15 2013.