Statistical physics of language change inferred from time evolving maps

James Burridge[1], Bert Vaux
[1]University of Portsmouth, Portsmouth, United Kingdom

Language change may be studied by tracking linguistic variant frequencies through space and time. Observed frequencies may be viewed as realizations of a statistical field. The statistical physics of language change should explain the dynamics of this field beginning from microscopic (speaker level) interactions and motion. Here we construct a statistical field model and infer its parameters from high resolution survey data. We first develop an efficient, cross validated method for inferring the most probable evolution of linguistic state fields from data. The method assumes a Gaussian process prior with parameters which control the length and time scales of field fluctuations. These parameters are adjusted to maximize the probability of out-of-sample data according to the inferred field. We apply our inference procedure to a range of linguistic variables surveyed in the USA by the Cambridge Online Survey of World Englishes, which contains approximately 100,000 responses. Image 1 shows the inferred spatial-temporal evolution of the fractions of three linguistic variants used to describe a sweetened carbonated beverage in the USA between 1950 and 2000. We then present our statistical field model, derived from microscopic assumptions, including an empirically fitted migration model, real population distributions, diffusive daily motion, and linguistic variant selection based on a form of stochastic replicator dynamics which encodes both inherent biases and linguistic accommodation (conformity to the majority). We infer the parameters of this model by comparing changes in model fields with changes in the fields we inferred from survey data. Our inferences demonstrate that spatial linguistic patterns can be rapidly destroyed by migration but that conformity can create them, protect them, and in combination with diffusion, produce partially predictable pattern dynamics. These dynamics, forms of which have been proposed in previous theoretical models, are analogous to phase ordering in physics, but mediated by variations in population density. Our results provide new evidence for such surface tension driven phase ordering, and to our knowledge represent the first example where field parameters have been inferred from the combined spatial and temporal evolution of language features. We quantify the ability of our model to forecast the future by computing the Kullback-Leibler divergence between model forecasts and observations.