Using evolutionary models to infer leadership structure from linguistic data. Preliminary report.

Sociolinguistics suggests that language change is driven by a leading subset of a community. I consider the problem of inferring the size of this leadership core from historical linguistics data with a Markov chain model from population genetics. The model is a modified Moran process in which each individual has a factor $b$ less influence than the next most influential person on children as they learn language. Using a Monte Carlo approach, it is possible to estimate $b \approx 0.85$. This is accomplished through a maximum-likelihood calculation on data for a change in English syntax collected by Ellegård. That value suggests that about 10% of the population accounts for 95% of the total influence. In a more general form of the model, the new language variant is allowed to have a selection advantage over the old. In this case, the maximum-likelihood calculation indicates that the influence is spread out more evenly over the population. (Received September 20, 2007)