Sara Kazemi
Ling 795
Malouf

Simulating linguistic similarity as a function of geographical distance and population

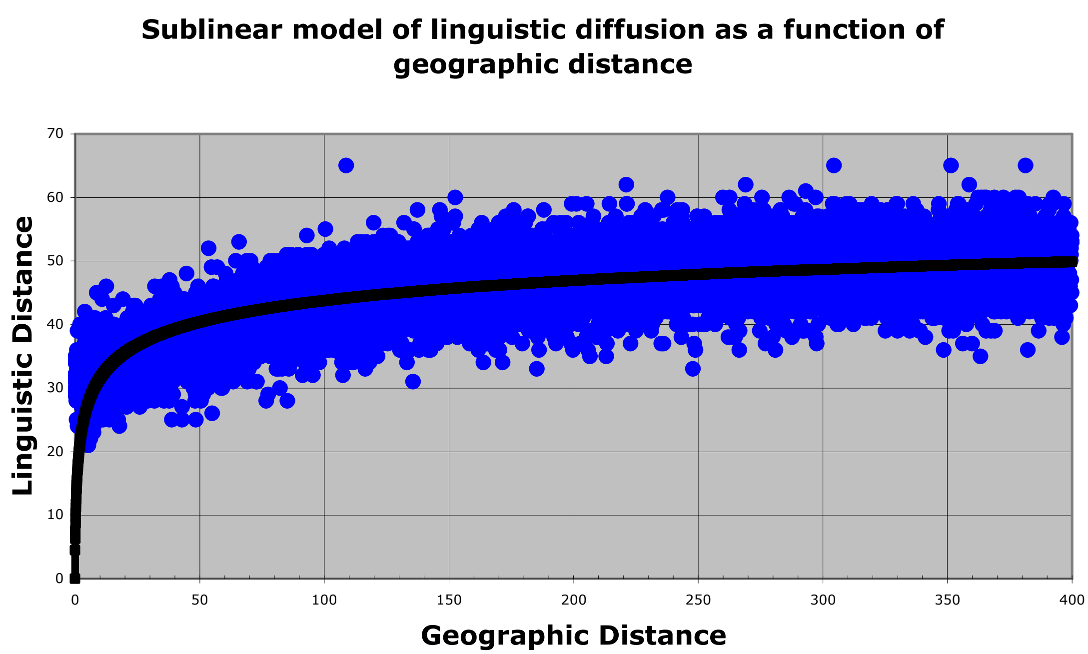
A prime focus of dialectometry is the geographical distribution of dialect similarity. Jean Seguy sparked this field of study when he found that linguistic distance was a sublinear function of the geographical distance between two sites (Nerbonne, 2008). Linguistic diffusion postulates that linguistic influences emanate from speakers of any given language. Speakers who co-communicate have a greater chance of sharing linguistic features and are more likely to share a common dialect of the language. On a wider scale, the closer settlements are to one another, the more likely they are to be linguistically similar.

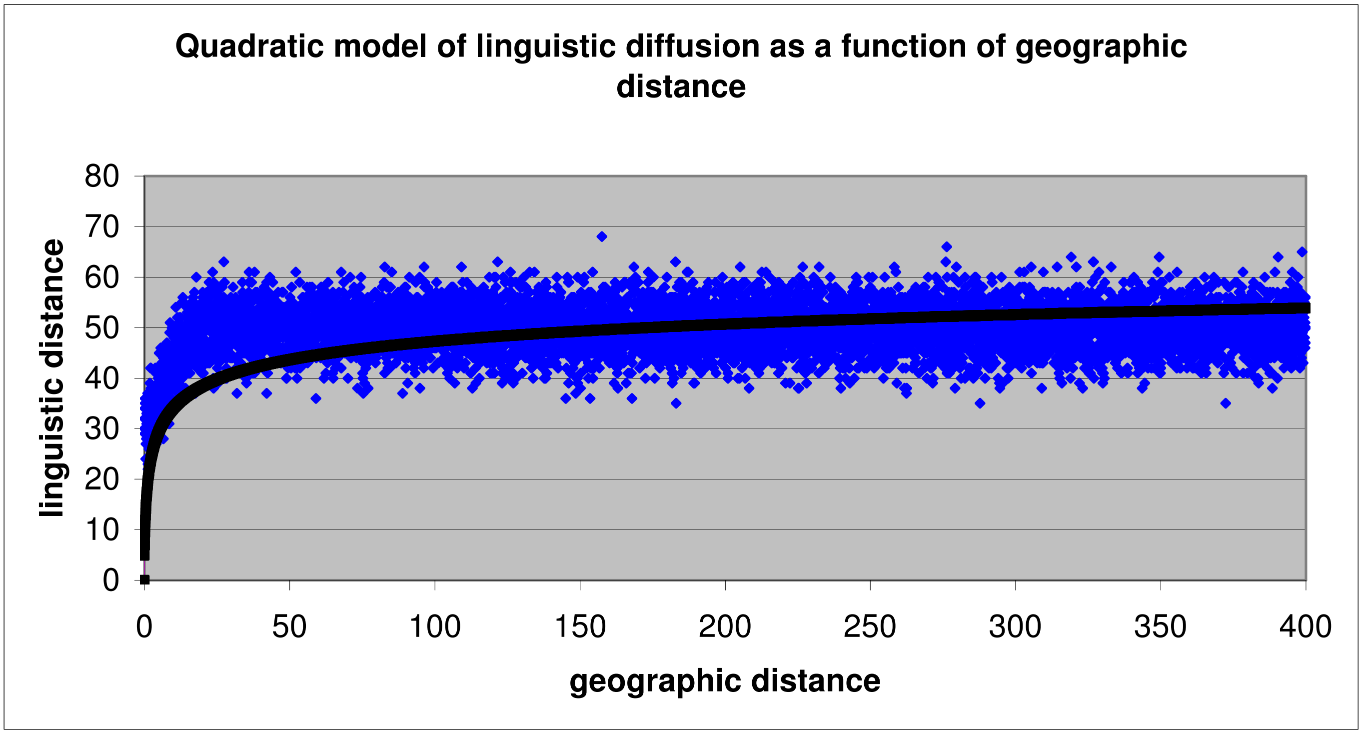
Trudgill's gravity model (1974) of linguistic innovation assumes that population as well as geographical distance are predictors of communication, reasoning that linguistic innovations may emanate from a reference point to a larger (more populous) site before actually reaching a site that is geographically nearer. Nerbonne and Heeringa (2007) derived a quadratic equation from Trudgill's formulation of the gravity model such that the linguistic influence of a reference site upon another site is equal to the product of the both sites' populations divided by the squared distance between them. In their study, Nerbonne and Heeringa concluded that population is not strongly determinant of linguistic similarity and, further, that the relation between linguistic similarity and geographical distance is sublinear, not quadratic.

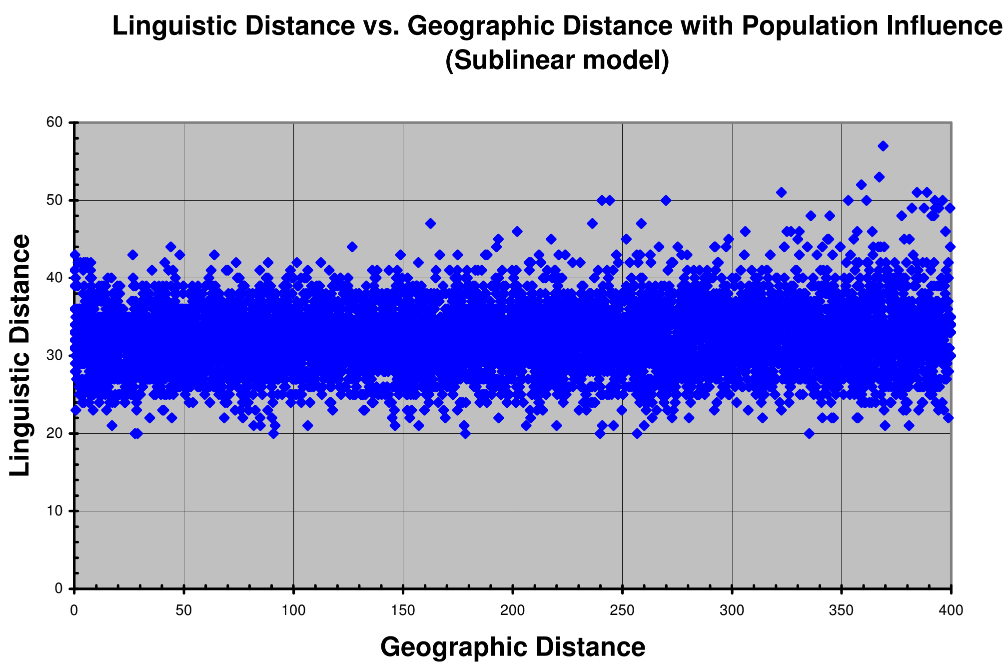
Nerbonne (2008) extends this assertion by constructing a model that ignores population. He implements the model in two versions: quadratically and sublinearly.

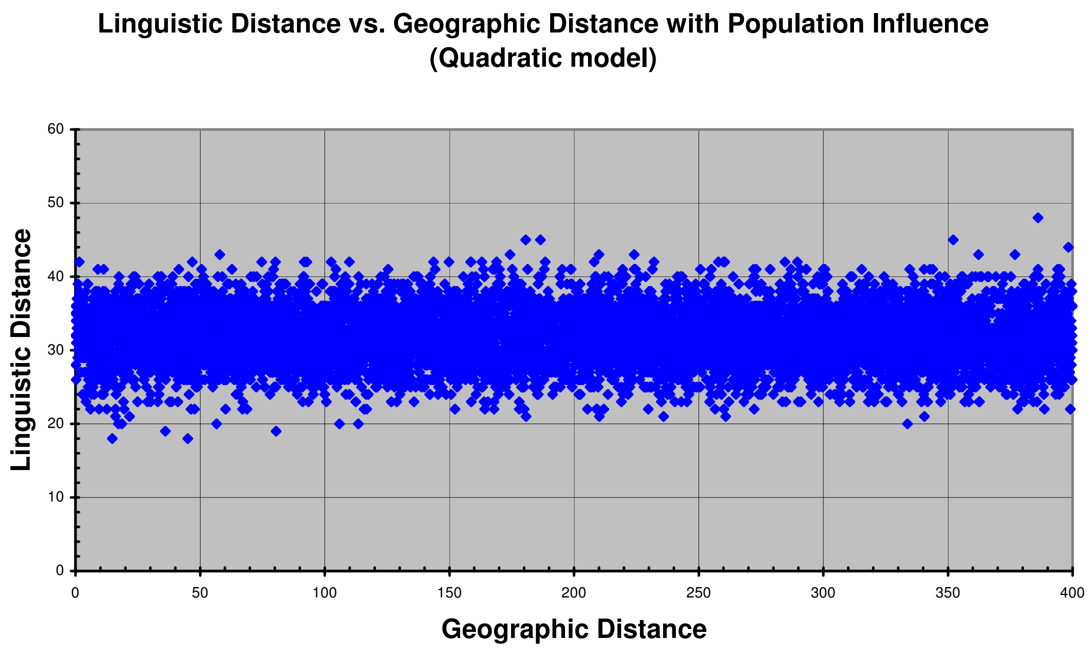
While his results do indicate that the sublinear curve does fit more closely with Seguy's categorically plotted curve, it would be interesting to compare Nerbonne's results to sublinear and quadratic models of diffusion, which also consider population and to measure just how little or how much the results may change.

Using Nerbonne's (2008) pseudocode, the language diffusion model was replicated in the Java programming language. In addition to running the simulations originally run by Nerbonne, sublinear and quadratic models that integrate population were also run. The graphs of these simulations illustrate how the results fit Seguy's curve, using different normalizing constants.




Quadratic model of linguistic diffusion as a function of geographic distance
linguistic distance
0
10
20
30
40
50
60
70
80
0
50
100
150
200
250
300
350
400
geographic distance


Linguistic Distance vs. Geographic Distance with Population Influence (Sublinear model)
Linguistic Distance
0
10
20
30
40
50
60
0
50
100
150
200
250
300
350
400
Geographic Distance



The first two graphs, which ignore population, closely resemble Nerbonne's (2008) results. As in Nerbonne's study, the sublinear model results fit Seguy's curve more closely than the results derived from the quadratic model. Proceeding to the last two graphs, it becomes clear why Nerbonne does not integrate population into his simplified model, which assumes distances at equal, linear intervals. While population may not significantly determine linguistic diffusion, population is probably affected by physical geography. Such an assumption is not modeled by the current study—population was randomly generated. Adding a randomly generated variable to already randomized data seems to have equalized the results; no real curve to which to compare to Seguy's curve emerges.

References

Nerbonne, John. (December, 2008). "Measuring the Diffusion of Linguistic Change." Manuscript read at the "Cultural and Linguistic Diversity" conference, AHRC Centre for Evolution of Cultural Diversity, London.

Nerbonne, John and Heeringa, Wilbert. 2007. Geographic Distributions of Linguistic Variation Reflect Dynamics of Differentiation. In eds, Featherston, Sam Featherston and Sternefeld, Wolfgang. *Roots: Linguistics in Search of its Evidential Base*, pp. 267–297. Berlin: Mouton De Gruyter.