Social Proximity, Influence, and the Dynamics of Language Death*

Adam C. Baker
University of Chicago
adamc@uchicago.edu

June 4, 2007

1 Model description

Abrams and Strogatz (2003) present a model of language death that they fitted to historical and census data they collected for Scottish Gaelic, Quechua, and Welsh. Their model idealizes many aspects of language competition by assuming fully connected, infinite populations and assigning status to the two languages in competition. Their model is an analytic model of language competition, rather than a simulated model.

In this paper I explore the consequences of relaxing some of these assumptions using a simulated model. I change their assumptions in two important ways. First, I assume that the populations are not fully connected, so an agent may base his decision about which language to use based on which language certain agents are using while ignoring other agents entirely. Second, I assume status is tied to agents rather than languages. I will report some preliminary results obtained by unsystematic experimentation with this modified model.

In the Abrams and Strogatz (2003) model, all speakers are assumed to only speak one of two languages, but each agent has a chance of changing to the other language that is

*Many thanks to Max Bane for his help implementing this model.

determined by the proportion of the population that speaks the other language and the status of that language. The dynamics of the model are described by equations 1 and 2. In these equations x is the proportion of the population that speaks language 0, y is the proportion of the population that speaks language 1, P_{xy} (P_{yx}) is the probability that a speaker of language 0 will change to language 1 (language 1 to language 0), s is the status people accord language 0 (and $1 - s$ is the status afforded to language 1), and a and c are parameters of the model that Abrams and Strogatz (2003) fitted to their data.

$$\frac{dx}{dt} = yP_{yx}(x, s) - xP_{xy}(x, s) \tag{1}$$

$$P_{yx} = cx^a s \qquad\qquad P_{xy} = c(1-x)^a(1-s) \tag{2}$$

After fitting their data, Abrams and Strogatz (2003) arrived at an average value for $a = 1.31$ with a standard deviation of 0.25.

Max Bane implemented this simple model as a simulation with a finite number of speakers who randomly changed languages each timestep according to equations 2. I altered his implementation by assigning each agent a real valued position and a real valued radius. These two aspects of an agent determine both who pays attention to that agent and determines that agent's status. A particular agent A is only affected by another agent B if the distance between the positions of A and B is less than B's radius value.

For every timestep in the simulation, an agent a has a probability of changing to language X (where X is not the agent's language), called P_{aX} given in eq 3, where p_x, r_x, and l_x are agent x's position, radius, and current language values, respectively, S is the set of all speakers in the model, and I is the indicator function, which returns 1 if it's argument is true and 0 if it is false.

$$P_{aX} = \left(\frac{\sum_{s \in S} I(|p_a - p_s| \leq r_s) I(l_s = X)}{\sum_{s \in S} I(|p_a - p_s| \leq r_s)} \right)^{1.31} \left(\frac{\sum_{s \in S} I(|p_a - p_s| \leq r_s) I(l_s = X) r_s}{\sum_{s \in S} I(|p_a - p_s| \leq r_s) r_s} \right) \tag{3}$$

The first factor in eq 3 is this model's version of the x factor in eq 2. It's just the number of speakers influencing a who speak X divided by the total number of speakers influencing a. The exponent on this factor is just the average a value found by Abrams and Strogatz

(2003), which I never varied. Likewise, the c parameter in their model is fixed to 1 in my model. The second factor is analogous to the status factor in eq 2. It is the total status held by agents influencing a who speak X divided by the total status held by agents influencing a.

This change in the model greatly affects how agents in the model affect other agents. In the Abrams and Strogatz (2003) model, all agents influence each other. In this model, agents only influence other agents positioned near them. Influence is not symmetric or transitive (though it is always reflexive).

This setup is best explained with a concrete example. Consider a model with three agents, A, B, and C. Let A's position be 25 and his radius be 10. Let B's position be 33. A influences B ($|25 - 33| \leq 10$), and if B's radius is less than 8, then A influences B asymmetrically. Let B's radius be 5 and C's position be 37. C is influenced by B ($|33 - 37| \leq 5$) but not by A ($|25 - 37| > 10$). A may be able to get B to change to A's language, and in the next timestep B might be able to get C to change to that language, and thus A may indirectly influence C through B, but A does not directly influence C.

The model is initialized by assigning each agent a position and radius which do not change during the simulation, and a language, which of course does. The position and radius values for each agent are drawn randomly from a normal distribution with a mean and standard deviation that are associated with the language that the agent is initially assigned. The distance between the two population means, the standard deviation of each population, and the mean radius for each population are all parameters of the model that are varied in my experiments. The number of speakers assigned to each population is also varied in the experiments.

To understand the predictions made by this model, it helps to interpret the parameters and values used in the model and the questions it attempts to address. Abrams and Strogatz interpret their model as idealizing speakers as all being monolingual. Another interpretation which may be more plausible for my model is to idealize all the speakers in the model as being bilingual. At each timestep a speaker decides which language to use based on which language the agents he's paying attention to are using. On this interpretation, a language that no one is using is not going to be used in the future, and so will die very soon without outside incentive to use the language again.

The absolute position of two agents does not have any meaningful interpretation on this model. What is important is the distance between the positions of two agents. This is an abstract “social distance”, which might correlate to real world factors like physical proximity, kinship status, or occupational status (where two people who work at the same place are socially closer). The smaller the distance between two agents, the more likely they are to interact. If we switch from thinking about agents to thinking about populations, where a population is the set of all agents who were assigned a particular initial language, then the distance between the two population centers represents the degree of integration of the two populations. When the two populations are centered on the same point, there will be a great deal of interaction. When the two populations are far away, interaction will happen only on the fringes.

The standard deviation assigned to a population’s position measures how tightly knit a language community is. A tightly knit community would be a population of speakers clustered around a particular point, so that all of them (or most of them) influence each other. This is accomplished by assigning that population a small standard deviation for their position values.

The radius value associated with agents should be interpreted as that agents status within society. A high status agent in this simulation will affect more of the population and will have a greater affect on those who he influences than an agent with lower status. When thinking about the status assigned to a population, the mean radius value for a population is somewhat like the status assigned to the language in the Abrams and Strogatz model. The standard deviation of the radius value for a population was kept at 1/10th of the mean radius for that population.

The two important questions I explore with this model are: Are there any settings of the models parameters that result in stably bilingual societies? Given the most plausible assumption I can come up with for modeling certain social situations, what does the model predict about the survivability of both languages in those situations?

This model assumes that language death happens mostly without speakers either actively trying to preserve or discard their languages. It assumes that they merely use whichever language those around them are using. As such, it does not try to model the attempts at language preservation described by Fishman (2001). It also assumes peaceful language shift, so it does not model cases of language death through the violent extermination of indigenous populations described in Trask (1996).

2 Tests of the model

I started by trying to simulate the Abrams and Strogatz model with identical status and starting population sizes for the two languages. In our simulation version of their model, starting with population sizes of 300 and status values of 0.5 for each language resulted in one language or the other dying after about 50 timesteps. Which language died and which survived was essentially random.

By piling all the speakers on one point and giving them all the same radius value, I attempted to achieve this behavior in my model. But language death occurs even faster in my model model, usually before 20 timesteps, since a speaker changing languages not only adds another speaker to the new language but also adds more status to that language.

I altered this simple scenario by increasing the standard deviation of the position value of one of the populations so that it was more spread out. This scenario predictably resulted in the death of the more spread out language in every run of the simulation.

My first real set of tests involved separating the two populations and assigning one a much higher average status and the other a much lower standard deviation for the agents position. I held the distance between the two groups constant at 30. The first group always had a position standard deviation of 15 and an average radius value of 15, which ensured that the first population would exert a strong influence on the second population. I assigned the second population an average radius value of 5, so the second population had roughly one third the status of the first. I then manipulated the standard deviation of the position values of the second population in an effort to find a stable configuration.

For my first simulation runs, I assigned the second population a standard deviation of 2. The first run produced the results in Fig 1 (in this and all subsequent graphs, the red curve is the population with the higher average status).

This first run was not representative, however. I ran the simulation many more times with these settings (with fewer speakers to speed up the simulation. This had no effect on the results other than slightly amplifying noise). Most of the time the lower status language (henceforth: language B, with the higher status language called language A) died pretty quickly in the simulation.

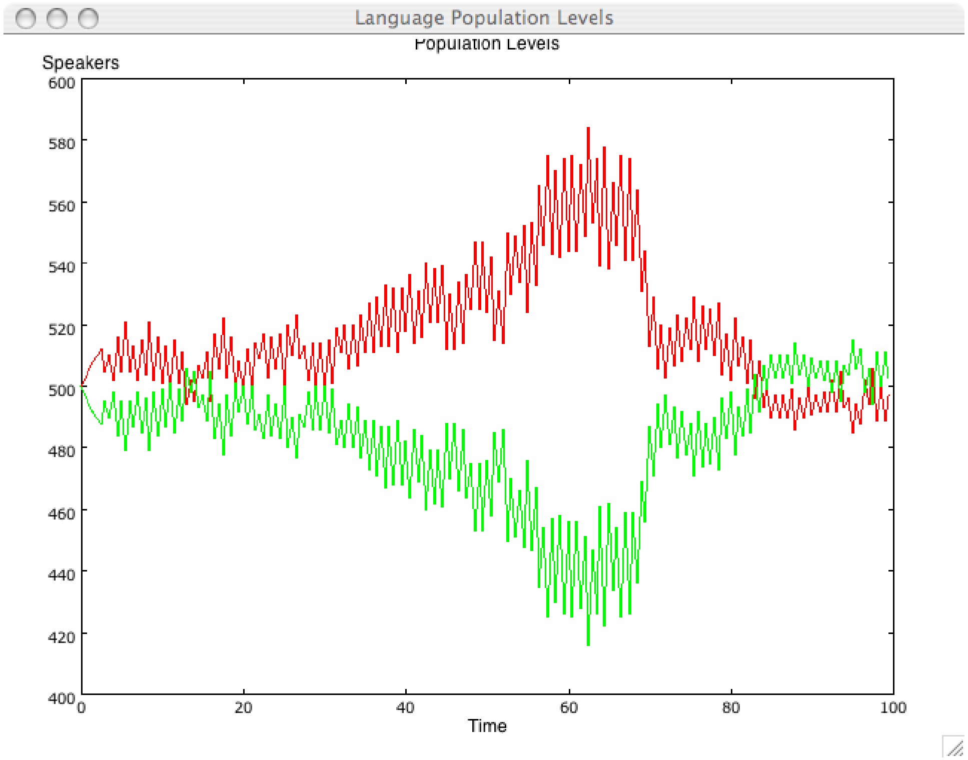

Fig 1: Distance: 30: Pop A: 500 speakers, 15 SD, 15 status; Pop B: 500 speakers, 2 SD, 5 status

Since I was trying to produce a stably bilingual society I tried tightening the standard deviation of the second population to 0.7. The results were often stable oscillation of a few speakers throughout the entire run of the simulation, shown in Fig 2. But at least as often, language B died quite quickly.

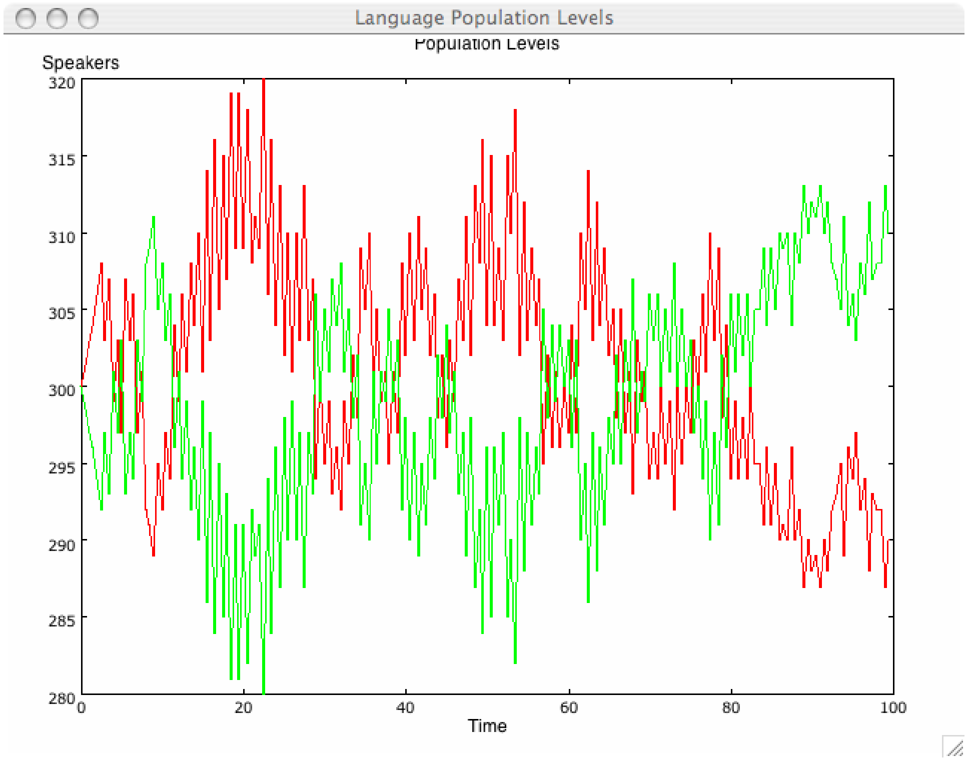

Fig 2: Distance: 30: Pop A: 300 speakers, 15 SD, 15 status; Pop B: 300 speakers, 0.7 SD, 5 status

I wasn't sure if these results occurred because of the randomness in the initial setup of the two populations (ie sometimes the agents in population B would be configured in a way that allowed language B to survive) or whether it was cause by the randomness in the agents' decision making. To test this I ran the simulation for 500 timesteps and tried to see if the simulation would ever oscillate for a long time (more than 100 timesteps) and then language B would die off. This did occur, and the plot is shown in Fig 3.

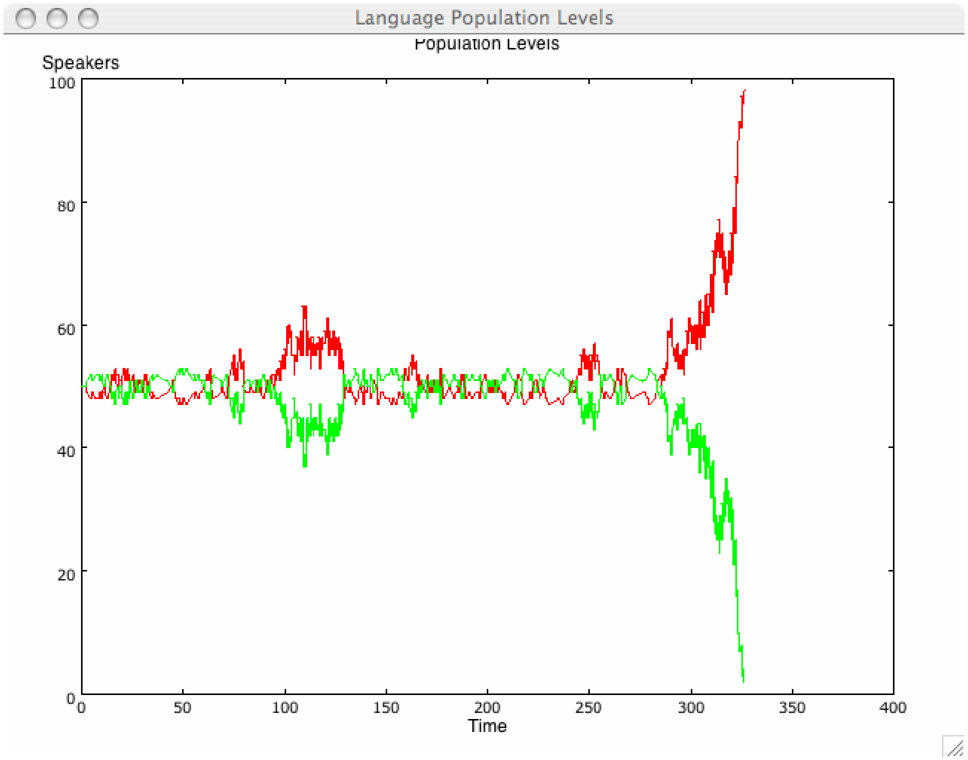

Fig 3: Distance: 30: Pop A: 300 speakers, 15 SD, 15 status; Pop B: 300 speakers, 0.7 SD, 5 status

This led me to believe that the apparently stable bilingual situations that occasionally occurred at these parameter settings were actually unstable. A chance shifting of enough speakers of language B towards language A could happen at any point and would quickly kill off language B.

For my first attempt to simulate an actual situation that could cause language death, I tried simulating a situation where a relatively small group of foreign conquerers were ruling over a native population. For this simulation, the "foreign conquerer" population (population A) were given 1/8th the number number of speakers as the "native" population (population B), but 3 times the status. Population A was placed 20 units away from population B, with a standard deviation in position of only 1 unit. Population B was given a standard deviation of 10 units. This creates a situation where Population A is relatively secluded and is only influenced directly by less than 5% of population B. It should be noted that these parameter settings are only my naive intuitions as to how to model this social situation, and are not based on studies of this situations.

My hypothesis was that Language A would dominate the population and kill off language B. But several runs have all produced results like those shown in Fig 4: a stably bilingual society with about half the population using each language. Most runs actually end with about 200 speakers of language A and 250 speakers of language B, though I occasionally got runs where language A had 300 speakers and language B had only 150 speaker. In all cases, this configuration was stable and language death never resulted in any run.

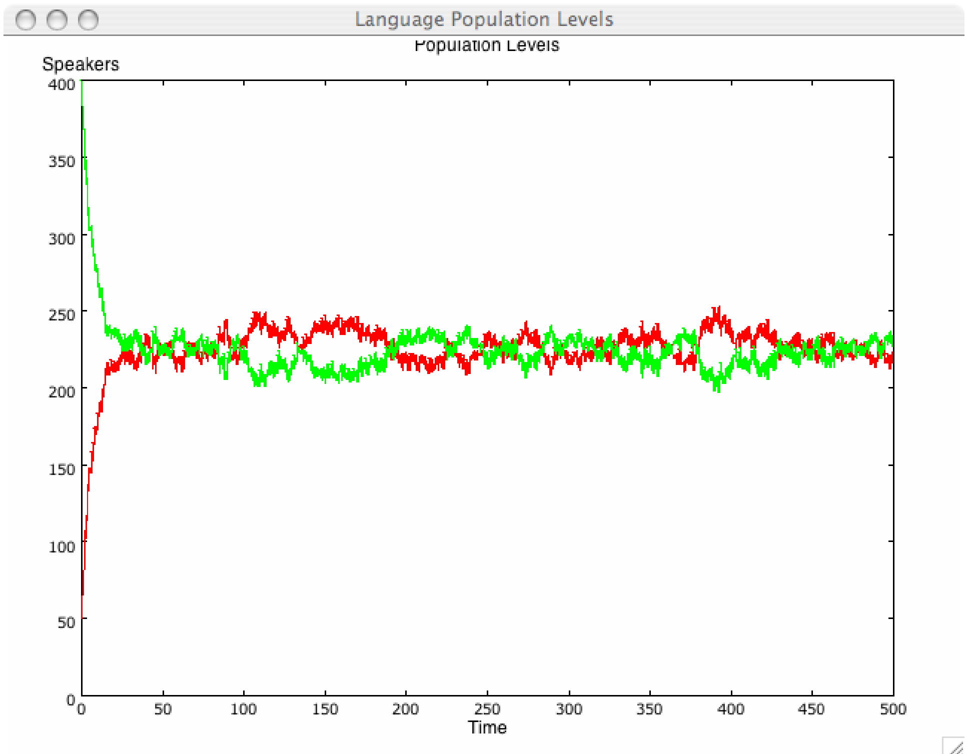

Fig 4: Distance: 20: Pop A: 50 speakers, 1 SD, 15 status; Pop B: 400 speakers, 10 SD, 5 status

For my next trial, I attempted to simulate and immigrant community embedded in a larger community, like the many immigrant neighborhoods we have in Chicago. It is not clear that this is possible in this model, but I tried. I gave the immigrant population B an average status of 5, an initial population of 100, and a standard deviation of 1. The surrounding population was centered on the same point with a status of 7, a standard deviation of 20, and 300 speakers. The results are shown in Fig 5.

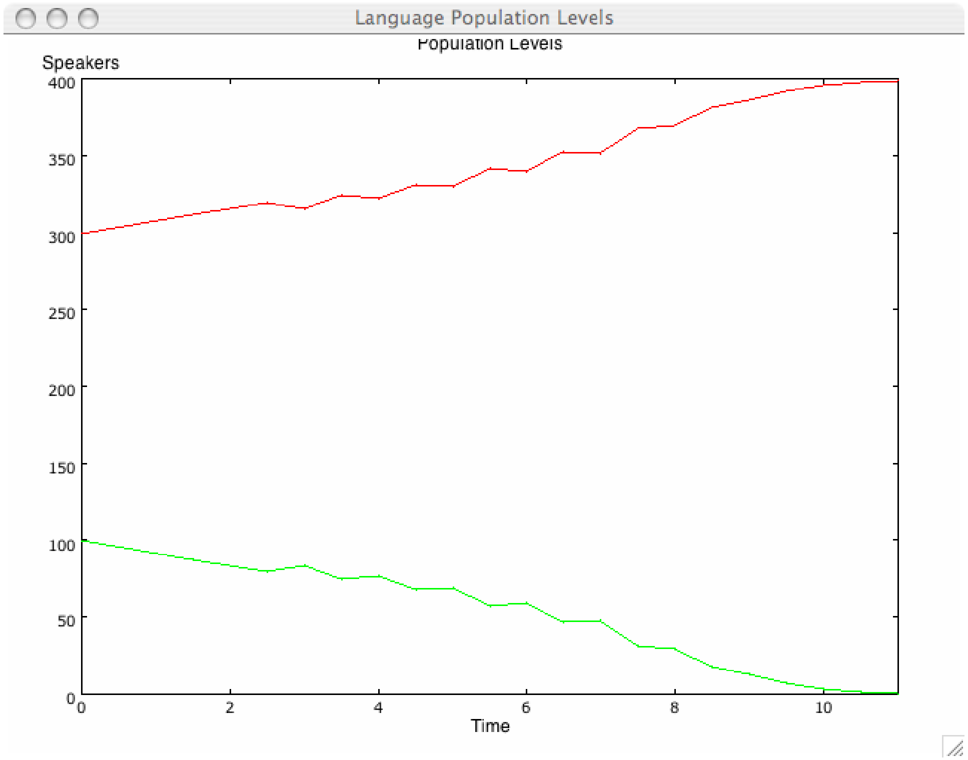

Fig 5: Distance: 0: Pop A: 300 speakers, 20 SD, 7 status; Pop B: 100 speakers, 1 SD, 5 status

All runs on this trial resulted in death in around 10 timesteps.

3 Shortcomings and possibilities for future models

There are clearly a large number of idealizations in this model that could be made more realistic. It could be augmented to handle birth and death or to allow agents to change position and status in the social network. A more realistic model of language could be used, and agents could be allowed to have multiple languages and could be given a more complicated mechanism for when they choose to abandon a language.

But these are all radical changes that might be better implemented by discarding the present model and starting with a more complex framework. Many small changes can be made to improve the model as it is and allow it to better simulate the situations I attempted to model in the last section.

One of the simplest and possibly one of the most important changes that can be made is to have the agents give more weight to other speakers influencing them who are closer. This is done by dividing the radius value of any speaker affecting the agent by the distance between the agent and that speaker. This changes eq 3 to eq 4.

$$P_{aX} = \left(\frac{\sum_{s \in S} I(|p_a - p_s| \leq r_s) I(l_s = X)}{\sum_{s \in S} I(|p_a - p_s| \leq r_s)} \right)^{1.31} \left(\frac{\sum_{s \in S} I(|p_a - p_s| \leq r_s) I(l_s = X) \frac{r_s}{|p_a - p_s|}}{\sum_{s \in S} I(|p_a - p_s| \leq r_s) \frac{r_s}{|p_a - p_s|}} \right) \tag{4}$$

This change would make tightly clustered populations more resistant to language change.

A second change that could be made would be to allow more ways to setup the initial population locations. For an example, the immigrant population simulation I attempted in the last section would be better modeled by having the immigrant population center surrounded on both sides by the other population with substantial overlap in the population. If the model setup were to use a bimodal distribution (a combination of Gaussians with two different mean values would be appropriate here) rather than a single gaussian distribution then this setup could be achieved.

A third, probably less important change that could be made to achieve better accuracy would be to assign radius values from an exponential distribution rather than a Gaussian distribution.

References

Abrams, Daniel M., and Steven H. Strogatz. 2003. Modelling the dynamics of language death. *Nature* 424:900.

Fishman, JA. 2001. *Reversing language shift*. Multilingual Matters.

Trask, RL. 1996. *Historical linguistics*. Arnold.