3-28-2014

Mergers, Migration, and Signaling

Christopher A. Ahern
University of Pennsylvania, christopher.ahern@gmail.com

Mergers, Migration, and Signaling

Abstract

The interaction between learning at the individual level and the trajectory of a population over time is fundamental to our understanding of linguistic change. Here we use game theory as a mathematical framework for formulating and testing hypotheses about this interaction. We formalize two hypotheses regarding the spread of vowel mergers and use them to derive the proportion of merged in-migrants that would precipitate a merger in a previously non-merged speech community. We test these predictions against the documented spread of the low-back merger. In light of these results, we consider the impact of social network structures on both models.

Mergers, Migration, and Signaling

Christopher Ahern[*]

1 Introduction

The interaction between learning at the individual level and the trajectory of a population over time is fundamental to our understanding of linguistic change. Here we use *game theory* as a mathematical framework for formulating and testing hypotheses about this interaction. We formalize two hypotheses regarding the spread of vowel mergers and use them to derive the proportion of merged in-migrants that would precipitate a merger in a previously non-merged speech community.

The first hypothesis stems from Herold's (1990) conjecture that mergers spread due to contact between merged speakers and non-merged hearers. Merged speakers do not use a phonemic distinction to differentiate meanings, but non-merged hearers rely on the distinction as a semantic cue. The misunderstandings caused by this mismatch lead non-merged learners to abandon the distinction. We will refer to this hypothesis as *merger-by-misunderstanding*.

The second hypothesis stems from Trudgill's (1986) supposition that accommodation plays a central role in contact-induced changes, and the observation that contact systematically favors mergers. If learners abide by the simple maxim "talk like others talk" (Keller, 1994), then this reveals an asymmetry in identifying how others actually talk. A tendency to misidentify non-merged speakers leads non-merged learners to abandon the distinction. We will refer to this hypothesis as *merger-by-misidentification*.

In the next section we provide a brief overview of game theory and offer a general characterization of the conditions for merger. With this more abstract criterion in place, we spell out the details of two models based on the hypotheses and derive concrete predictions. We test these predictions against the spread of the low-back merger along the Rhode Island-Massachusetts border (Johnson, 2010). In light of these results, we consider the impact of social network structures on both models. Finally, we conclude with a brief discussion of our results and avenues for future research.

2 Game Theory

A *game* is a mathematical structure that represents the decisions made by individual agents when the outcome depends on the decisions of others. As an intuitive example, consider two drivers approaching each other on the road. Both prefer avoiding a collision, so they do best when they drive on the complementary sides of the road. Crucially, each driver's decision depends on the other's. We can also think of the interaction in a community. If everyone else drives on one side, an individual does best by also driving on the right.

In game-theoretic terminology, each player must choose between the *strategy* of driving on the left or the right. The drivers' choices yield the *outcome* of the game. Each player has a *utility function* which expresses her preferences over these outcomes. Presumably, people prefer avoiding car accidents over not. In a larger community we can think of how well an individual does by driving on one side or the other given how everyone else behaves. An individual's *expected utility* expresses how well she does on average. Uncontroversially, individuals act to bring about their preferred outcomes. That is, they act so as to maximize their expected utilities. However, each individual's efforts to maximize her own expected utility are influenced by others' efforts, and vice versa.

This same interdependence can be found in language given how it is learned and used. That is, while children *acquire* language from the linguistic input of caretakers, the process of language *learning* continues as children come into contact with and reorient towards the patterns of the wider speech community (Payne, 1976). For most children, this occurs through introduction into a peer

group at school. In this case, each child is influenced by the linguistic input from her peers. But, in turn, she provides linguistic input for her peers. In this sense, language learning can be conceived of as a game. The actions of each individual learner affect everyone else, and vice versa.

Given our goal of understanding when mergers occur, we can be more specific about the components of the game. In doing so, we make the following assumptions. First, there are strategies corresponding to the behaviors of either maintaining or abandoning a phonemic distinction, which we will call D and M, Individuals either maintain the distinction in both perception and production, or abandon it in both perception and production.[1] Second, as in acquisition, learners only receive feedback as hearers. In those rare cases where learners actually receive feedback as speakers, it is ignored or irrelevant (Brown and Hanlon, 1970; Hirsh-Pasek et al., 1984). Third, acting to maximize expected utility does not involve conscious deliberation. Indeed, a variety of implicit learning dynamics are appropriate for the case of language (Fudenberg and Levine, 1998).

So, if agents act to maximize expected utility, then mergers occur just when the expected utility of the merged strategy, $EU(M)$, exceeds that of the non-merged strategy, $EU(D)$. These expected utilities, in turn, depend on the composition of the speech community as a whole: hearers will receive some input from merged and non-merged speakers. Let p be the proportion of merged individuals in the population, and $(1 - p)$ the proportion of non-merged individuals. Where EU_{MD} represents the expected utility of a non-merged hearer when interacting with a merged speaker, and EU_{DD} the expected utility when interacting with a non-merged speaker, we can express $EU(D)$ as the following.

$$EU(D) = pEU_{MD} + (1 - p)EU_{DD} \tag{1}$$

We can do the same for $EU(M)$, where EU_{MM} represents the expected utility of a merged hearer when interacting with a merged speaker, and EU_{DM} the expected utility when interacting with a non-merged speaker.

$$EU(M) = pEU_{MM} + (1 - p)EU_{DM} \tag{2}$$

We are interested when one exceeds the other.

$$EU(M) > EU(D) \tag{3}$$

Which allows us to specify the conditions where a merger will occur according to the proportion of merged individuals in the population:

$$p > \frac{EU_{DD} - EU_{DM}}{EU_{MM} + EU_{DD} - EU_{MD} - EU_{DM}} \tag{4}$$

This is just a general expression of the threshold for merger. The value of the threshold depends entirely on how we define strategies and preferences over outcomes. Once we provide the definitions we will have a prediction. We now turn our attention to providing two such sets of definitions.

3 Models

3.1 Merger-by-misunderstanding

Mergers lead to homophony, homophony to ambiguities, and ambiguities to an increase in the probability of misunderstandings. This line of reasoning suggests that maintaining a distinction will be the better strategy. However, as noted above, there are conditions that render the opposite true. For example, Herold (1990) suggests that mergers spread due to errors in interpretation made by non-merged hearers in contact with merged speakers, and that ceasing to rely on a phonemic distinction as a cue for a semantic distinction might actually decrease the rate of misunderstandings.

---

[1] The phenomenon of *near-mergers* (Labov et al., 1991) demonstrates that perception and production are indeed dissociable. This is a simplifying, but useful assumption.

The first part of this conjecture is overwhelmingly borne out in a corpus of naturally occurring misunderstandings (Labov, 1994). Misunderstandings almost always occur between merged speakers and non-merged hearers. Yet, it seems implausible that such misunderstandings occur frequently enough to prompt learners to adopt a new strategy. The corpus consists of hundreds of misunderstandings, but these were gathered over several decades. The second part of this conjecture depends entirely on what it is that merged speakers actually do.

Drawing from psycholinguistic evidence, Yang (2009) develops a compelling model that addresses both these points. First, misunderstandings need not rise to the level of conscious awareness to affect learning. Small delays in processing, akin to *garden paths*, can be taken as a proxy for the utility of a strategy. The more delays a strategy incurs, the lower its expected utility. Second, mergers render minimal pairs homophonous. Results from homophone processing suggests that merged hearers will assume the more frequent interpretation of homophonous pairs (Bonin and Fayol, 2002). Moreover, higher level information does not allow for disambiguation. Boland and Blodgett (2001) show that this frequency effect overwhelms syntactic information, and Swinney (1979) demonstrates that contextual information is only available much later on in processing.

These results can be thought of straightforwardly as a *signaling game* (Lewis, 1969).[2] Signaling games take place between a sender and a receiver. The sender observes some state of the world, sends a signal to a receiver, and the receiver then takes some action given this signal. In the case of language, a speaker has some intended meaning that he wants to convey, he chooses a word to say, and the hearer interprets it. The outcome of the game depends on how speakers signal their meaning, and how hearers interpret those signals. To define the strategies employed by merged and non-merged individuals, we adopt the following notation. In what follows we will be concerned with the use and interpretation of minimal pairs affected by the low-back merger. Let $t_\text{ɔ}^i$ and $t_\text{ɑ}^i$ be the frequencies of the words containing /ɔ/ and /ɑ/ in the ith minimal pair. Let $T_\text{ɔ} = \sum_i t_\text{ɔ}^i$ and $T_\text{ɑ} = \sum_i t_\text{ɑ}^i$ be the sum of the frequencies of the words containing the different phonemes. Let $T = T_\text{ɔ} + T_\text{ɑ}$ be the sum of the frequencies of all the words concerned. Let $m_\text{ɔ}^i$ and $m_\text{ɑ}^i$ be the phonetic forms for non-merged speakers of the ith minimal pair. More generally, let $m_\text{ɔ}$ and $m_\text{ɑ}$ represent the phonetic forms of the two vowels for non-merged speakers. Finally, let $\varepsilon_\text{ɑ}$ be the probability of mistaking $m_\text{ɑ}$ for $m_\text{ɔ}$, and $\varepsilon_\text{ɔ}$ be the probability of mistaking $m_\text{ɔ}$ for $m_\text{ɑ}$ (e.g., Peterson and Barney (1952)).

We can now define the strategies for merged and non-merged individuals. For each minimal pair, a non-merged speaker will use distinct forms for the two different meanings, whereas a merged speaker will only use one. That is, a non-merged speaker will condition his message on the state, sending $m_\text{ɔ}^i$ if intending meaning $t_\text{ɔ}^i$, and $m_\text{ɑ}^i$ if intending meaning $t_\text{ɑ}^i$. A merged speaker will send a single message in both states. Without loss of generality, we will assume that merged speakers always use $m_\text{ɔ}^i$. A non-merged hearer will condition her action on the message received. If she hears $m_\text{ɑ}^i$ she will interpret as the speaker meaning $t_\text{ɑ}^i$. If she hears $m_\text{ɔ}^i$ she will interpret it as the speaker meaning $t_\text{ɔ}^i$. This process is not without the possibility of error. Occasionally one signal will be mistaken for another. In contrast, merged hearers ignore any differences in the message. Regardless of whether they receive $m_\text{ɑ}^i$ or $m_\text{ɔ}^i$, they assume the speaker meant the more frequent of the two possible meanings.

We are now in a position to calculate how well these strategies do when paired together. The expected utility of a non-merged hearer can be calculated as follows. When paired with a non-merged speaker, the hearer will be able to recover the intended meaning of the speaker, without any delays and additional processing costs, when the message is perceived correctly. Whether the word is in $T_\text{ɑ}$ or $T_\text{ɔ}$, errors will only be made when the message is misperceived. These values are normalized by the total frequency of the words involved.

$$EU_{DD} = \frac{T_\text{ɑ}(1 - \varepsilon_\text{ɑ}) + T_\text{ɔ}(1 - \varepsilon_\text{ɔ})}{T} \tag{5}$$

When paired with a merged speaker, a non-merged hearer will not do so well. Again, assuming

---

[2]This is a reformulation of Yang's model modulo the assumption that learning under a Bush and Mosteller (1951) linear reward penalty scheme a *winner-take-all* dynamic applies. At some point the learner simply assumes the more likely of the two options. It is not clear why such a process should occur, nor when it should take place. Here we obtain the same results without this stipulation and the resulting conceptual difficulty.

that merged speakers only ever use $m_{\mathfrak{d}}$. This means that whenever words from T_{a} are intended, the hearer will make an error of interpretation, except in case of an error in perception.

$$EU_{MD} = \frac{T_{\mathrm{a}}(\varepsilon_{\mathfrak{d}}) + T_{\mathfrak{d}}(1 - \varepsilon_{\mathfrak{d}})}{T} \tag{6}$$

Without knowing anything at all about the various frequencies we can note an important asymmetry between how non-merged hearers fare. If error rates are fairly low, as we would expect them to be, then non-merged hearers always do worse against merged speakers than against non-merged speakers. This captures Herold's intuition that it is the interaction with merged speakers that causes problems for non-merged hearers.

The expected utility of merged hearers can be calculated as follows. As noted above, evidence from homophone processing suggests that the more frequent of the pair is assumed. Let $max(t_{\mathrm{a}}^i, t_{\mathfrak{d}}^i)$ yield the meaning in the minimal pair with the maximum frequency. If merged hearers treat both merged and non-merged speakers alike, then the expected utility of a merged receiver can be expressed as follows.

$$EU_{DM} = EU_{MM} = \frac{\sum_i max(t_{\mathrm{a}}^i, t_{\mathfrak{d}}^i)}{T} \tag{7}$$

We have now specified all the components of the model, which we will use for the actual calculations of the threshold for merger in Section 4.

But, before moving on, we pause to note several general properties of this model of *merger-by-misunderstanding*. To do so, we make the simplifying assumption that error rates are roughly equal ($\varepsilon : \varepsilon_{\mathrm{a}} \approx \varepsilon_{\mathfrak{d}}$), which allows us to represent the threshold as a linear function of a merged hearers expected utility, $EU_{DM} = EU_{MM}$, where $\gamma = \frac{T_{\mathrm{a}}}{T}(1 - 2\varepsilon)$.

$$p > \frac{(1 - \varepsilon) - EU_{MM}}{\gamma} \tag{8}$$

First, mergers are always possible, even in the special case where error rates are negligible ($\varepsilon \approx 0$). In this case the threshold is then given by $p > \frac{T - \sum_i max(t_{\mathrm{a}}^i, t_{\mathfrak{d}}^i)}{T_{\mathrm{a}}}$, which is always less than 1. Second, as the asymmetry between maximum and minimum in each pair grows, so does EU_{MM}. The more lopsided the pairs become, the lower the threshold for merger becomes. For situations of contact between merged and non-merged dialects, this means that asymmetric contrasts are more susceptible to merger. Third, taken even further, this also gives the conditions for when a contrast could be sufficiently lopsided for the *actuation* of a merger in the absence of dialect contact ($1 - \varepsilon = EU_{MM}$). This is suggestive in light of the work reported in Wedel et al. (2012), where, in a survey of 482 phonemic contrasts, they find that the number of balanced minimal pairs — minimal pairs where the ratio between the minimum and the maximum frequency of the pair is close to 1 — is a fairly good predictor of whether a merger has occurred or not.

3.2 Merger-by-misidentification

Mergers lead to homophony, homophony to ambiguities, and ambiguities might not matter for much. That is, misunderstandings might be irrelevant in light of more fundamental pressures to "talk like others talk" (Keller, 1994). Trudgill posits that, in cases of dialect contact, speakers accommodate to each other "by reducing the dissimilarities between their speech patterns and adopting features from each other's speech" (1986, 39).

While this supposition is, in a certain sense, undeniable, it misses a crucial aspect of change. That is, accommodation undoubtedly permeates linguistic interactions at various levels of structure (e.g., Natale (1975); Gregory (1990); Pardo (2006); Brennan and Clark (1996); Branigan et al. (2000)). But, accommodation presupposes the ability to identify when our own speech differs from that of others. If mergers are favored in cases of dialect contact, then there must be some asymmetry in identifying how it is that others actually talk.

Accommodation can be taken as a *coordination game*. That is, agents prefer to coordinate their behavior with others. As in the case of driving conventions, an individual does best by coordinating

her behavior with that of the population as a whole. However, in the case of language conventions, everything depends on the evidence that a learner receives that her interlocutors are using one strategy or another. To define the strategies employed by merged and non-merged individuals, we use the same notation as in the previous model.

Speaker behavior is as in the previous model: non-merged speakers condition their message on the intended meaning, whereas merged speakers do not. However, here we assume that hearers can recover the speaker's intended meaning. Crucially, what is at issue is not *what* is said, but rather, *how* it is said. That is, coordination is a matter of having the same meaning-form mappings. For example, the word *cot* from an ɔ-merged speaker will be taken as evidence as a failure to coordinate by a non-merged hearer. The same would hold if the roles of speaker and hearer were reversed. The important thing is whether the hearer would have used the same form given the intended meaning. Again, this process is not without error. There is some probability of mistakes between the two forms. The important thing is how often a hearer has evidence that a speaker shares the same meaning-form mappings.

We are now in a position to calculate how well these strategies do when paired together. The expected utility of a non-merged hearer can be calculated as follows. When paired with a non-merged speaker, the hearer always has evidence of successful coordination except in those cases of misperception.

$$EU_{DD} = \frac{T_{\mathrm{a}}(1-\varepsilon_{\mathrm{a}}) + T_{\mathrm{ɔ}}(1-\varepsilon_{\mathrm{ɔ}})}{T} \tag{9}$$

When paired with a merged speaker, a merged hearer does not do as well. When words from T_{a} are intended the hearer only thinks the speaker shares the same meaning-form mapping when the signal is misperceived. When words from $T_{\mathrm{ɔ}}$ are intended they have evidence of successful coordination, less errors in perception.

$$EU_{MD} = \frac{T_{\mathrm{a}}(\varepsilon_{\mathrm{ɔ}}) + T_{\mathrm{ɔ}}(1-\varepsilon_{\mathrm{ɔ}})}{T} \tag{10}$$

The expected utility of a merged hearer can be calculated as follows. When paired with a non-merged speaker, a merged hearer is in a parallel situation to when the roles are reversed. When words from T_{a} are intended the hearer only has evidence of successful coordination when the signal is misperceived. When words from $T_{\mathrm{ɔ}}$ are intended they have evidence of successful coordination, less errors in perception.

$$EU_{DM} = \frac{T_{\mathrm{a}}(\varepsilon_{\mathrm{a}}) + T_{\mathrm{ɔ}}(1-\varepsilon_{\mathrm{ɔ}})}{T} \tag{11}$$

When paired with a merged speaker, a merged hearer always has evidence of successful coordination regardless of whether the word came from T_{a} or $T_{\mathrm{ɔ}}$.

$$EU_{MM} = \frac{T_{\mathrm{a}} + T_{\mathrm{ɔ}}}{T} \tag{12}$$

We now have all the components of the model, which will be used for calculations of the threshold for merger in Section 4.

We pause to note several properties of this model of *merger-by-misidentification*. To do so, we make the simplifying assumption that error rates are equal, which allows to represent the corresponding threshold in a compact manner, where $\gamma = \frac{T_{\mathrm{a}}}{T}(1-2\varepsilon)$, as above.

$$p > \frac{\gamma}{2\gamma+\varepsilon} \tag{13}$$

First, in the case where error rates are negligible ($\varepsilon \approx 0$), a majority is required for merger to occur, exactly as we would expect. Thus, any tendency for merger requires some non-negligible error rate. Second, when the error rate is non-negligible ($\varepsilon > 0$), the impact of the error rate is determined by the frequency distribution. That is, the more lopsided the distribution, the greater the impact the error rate has. The smaller T_{a} is, the more likely that a merger will occur resulting in /ɔ/. The larger T_{a} is,

the less likely that a merger is to occur resulting in $/\mathrm{ɔ}/$. In fact, the larger $T_\mathrm{ɑ}$ is, the more likely that a merger is to occur resulting in $/\mathrm{ɑ}/$. For situations of contact, we expect that contrasts that are more lopsided, in terms of the total frequency of the words containing the two phonemes, are more likely to undergo merger. Third, conditions for *actuation* only occur vacuously in this model ($\gamma = 0$). That is, these conditions obtain just when there is no semantic distinction to be had, or when the forms are perfectly confusable. However, the actuation of mergers may be attributable to other causes. For example, based economic and demographic evidence, Herold (1990) argues that the actuation of the low-back merger in Eastern Pennsylvania was brought about by an influx of foreign-born anthracite coal miners. It is safe to assume that an adult would not be able to learn the distinction if it were not already present in his or her native language, at least, given that merged children older than fourteen without a native distinction categorically fail to learn it (Chambers, 1992). These foreign-born adults would have provided the linguistic input for their own children, who would have eventually come into contact with peers. While demographic data is not conclusive as to native dialects, foreign-born or first-generation individuals constituted the majority in the towns where the merger occurred. Thus, *contact*, in some broader sense, might always be the source of actuation.

4 Evaluation

Up to this point, we have done two things. First, we stated a general formula for calculating the threshold for merger. Second, we formalized two hypotheses about the details of this general formula. Now, we are in a position to make the calcuations for each model that yield the proportion of merged individuals sufficient to bring about a merger. We should note that the predictions made by the two models are generally distinct. Thus, the two models can be taken as competing explanations. These predictions, in turn, can then be compared to the documented spread of the merger.

The thresholds for merger were computed for both models over a range error rates, from no errors to the case where the two messages are perfectly confusable. We used minimal pair frequencies from a tokenized version of the Wortschatz Corpus (Biemann et al., 2007).[3] Minimal pairs where one member did not appear at least ten times in the corpus were excluded from the calculations. Given that this eliminates low frequency pairs it increases the threshold for the model of *merger-by-misunderstanding*, but leaves the model of *merger-by-misidentification* largely untouched. The results of all the calculations can be seen in Figure 1 where the horizontal and the vertical axes represent the error rates $\varepsilon_\mathrm{ɑ}$ and $\varepsilon_\mathrm{ɔ}$, respectively. The threshold for each of these combinations is represented as a color. The darker the color the lower the threshold, the lighter the color the higher the threshold. Both models yield their maximum thresholds when there are no errors: the maxima are $p = .27$ for the model of *merger-by-misunderstanding*, and $p = .5$ for the model of *merger-by-misidentification*.

Johnson (2010) documents the spread of the low-back merger along the border between Massachussets and Rhode Island. The dialect boundary between merged Massachussets and non-merged Rhode Island has been historically stable, despite contact between adults from both sides. Johnson argues that the spread of the merger is due to an influx of merged children into previously distinct areas. In South Attleboro, MA the merger went to completion within a few years starting in roughly 1990. In Seekonk, MA the merger went to completion within a few years starting in roughly 2000. Johnson estimates the proportion of merged children at the inception of the merger to have been roughly 50% in South Attleboro and 20% in Seekonk. Assuming that error rates are fairly low ($\varepsilon < .1$), as suggested by work on vowel confusability (Peterson and Barney, 1952), both models find support in these empirical thresholds. That is, the model of *merger-by-misunderstanding* predicts a threshold of 20% and the model of emphmerger-by-misidentification predicts a threshold of

---

[3]The minimal pairs used, with frequencies from lowest to highest, are: pod (24), pawed (11); fond (1318), fawned (12); bock (20), balk (16); sot (20), sought (1703); collar (403), caller (21); clod (23), clawed (238); hock (25), hawk (127); knotty (25), naughty (195); sod (30), sawed (37); pond (258), pawned (31); yon (67), yawn (36); cot (39), caught (2444); bot (52), bought (995); nod (180), gnawed (53); Otto (67), auto (260); odd (830), awed (80); stock (1456), stalk (105); rot (109), wrought (336); moll (188), maul (121); knot (226), naught (156); Don (1052), Dawn (736).
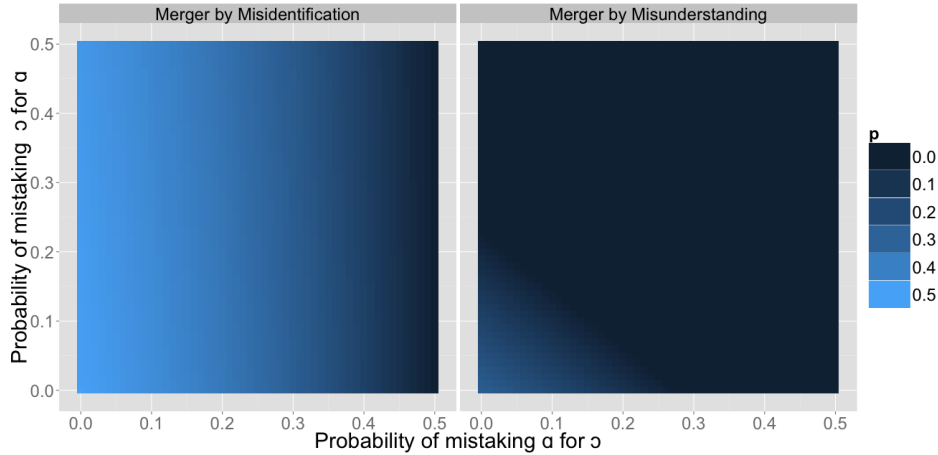
Figure 1: Threshold for merger for model of *merger-by-misidentification* (left) and *merger-by-misunderstanding.* (right)

roughly 50%.

What are we to make of these results? We certainly cannot posit that learners behave differently in different locations. So, the fact that there are two distinct thresholds is problematic for both models. This follows from the fact that both models are deterministic. That is, given a set of lexical frequencies and error rates, each model predicts a unique threshold. If either model has any possibility of accounting for the data, we must posit some difference in the components of the model. That is, we must consider whether the frequencies or error rates can vary from location to location. We consider these two options in turn.

Lexical frequencies are drawn from a corpus, which is taken as representative of the general linguistic environment of speakers. Obviously, individual speakers have unique histories of linguistic input. Thus, we might think of different distributions over experiences in a population. However, it is not clear why any such experiences should vary so widely across communities. Accounting for the differences across the two communities in this manner would essentially be an appeal to chance factors. That is, the children in South Attleboro would somehow have had to have heard more of the less frequent of all the minimal pairs. Or, in the other direction, the children of Seekonk would have had to have heard many more tokens of words with meanings from the set $T_{\mathfrak{o}}$. At best, these considerations push the difference between the two communities further into the mysterious.

Error rates seem a more promising route in explaining the difference between the two communities. It is uncontroversial that size of the distinction between two categories may vary with location. Further, it is reasonable to assume that the size of the distinction factors into error rates; the smaller the distinction, the larger the error rates. Note that this approach adjudicates in favor of the model of *merger-by-misidentification*. That is, there are error rates that supply both the South Attleboro ($\varepsilon \approx .1$) and Seekonk ($\varepsilon \approx .3$) thresholds. The maximum threshold value of the model of *merger-by-misunderstanding* is less than the South Attleboro threshold, regardless of the error rate. However, we would need convincing evidence that the error rate would have been higher in Seekonk, thus lowering the threshold for merger. That is, we would need evidence that the vowels approximated each other much more closely in Seekonk than in South Attleboro.

A third option for exploring the explanatory power of the two models lies in relaxing our assumptions about the structure of interactions between individuals. Thus far our models have carried the implicit assumption that the likelihood of encountering an individual with a particular strategy is the proportion of individuals using that strategy. In the next section, we relax this assumption and consider the consequences for the two models.
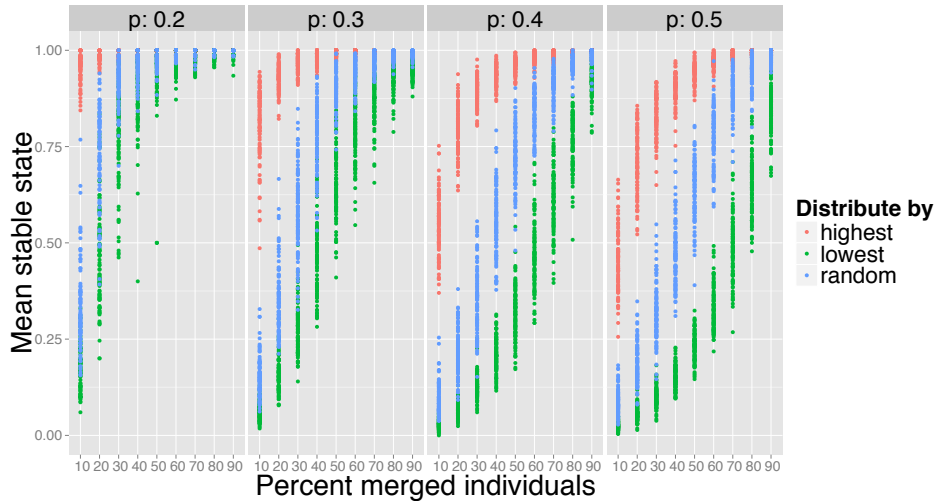
Figure 2: Mean stable state for various proportions of merged speakers.

5 Network Structure

Real networks are rarely uniform; individuals occupy unique social niches. Here we explore the effects of network structure by constructing networks via preferential attachment (Barabási and Albert, 1999). The resulting networks consist of a few highly connected nodes, whereas most nodes have only a few connections. This parallels a network where a few individuals are highly-connected, whereas most are more peripheral. A few highly-connected individuals may exert an outsized influence on the population. In this case, they will play a greater than average part in the linguistic environment of others. There are two questions we wish to answer. First, can we explain the threshold in South Attleboro by the model of *merger-by-misunderstanding*? That is, can 50% of the population be placed in a network to only have the effect of 20%? Second, can we explain the threshold in Seekonk by the model of *merger-by-misidentification*? That is, can 20% of the population be placed in a network to have the effect of 50%? As a first step towards answering these questions we performed the following simulations.

We constructed networks of 500 agents by preferential attachment and simulated the end state of the population. Each agent began with an initial state, either non-merged (0) or merged (1). Merged agents were placed into the network by three conditions. They were either placed in the most-connected positions, placed randomly, or placed in the least-connected positions. Agents began by calculating the proportion of their neighbors wiht the merger. If this proportion exceeded a threshold, they adopted the merger as well. This process was iterated until no agents changed their state. The average state of all agents was recorded. We considered multiple values for the threshold and various proportions of merged speakers. For each proportion of merged speakers we ran the simulation 100 times. The results can be seen in Figure 2.

We are now able to address our two questions. First, we address whether 50% can have the effect of 20% when the threshold is as predicted by the model of *merger-by-misunderstanding*. Looking at the panel in Figure 2 where $p = .2$, we want to establish several things. For example, it appears that the effect of 20% varies quite widely with placement in the network. When placed in the most connected positions, the merger easily spreads. When placed randomly or in the lowest positions, the merger is not guaranteed to spread. In contrast, if more than 20% of the population is merged, regardless of placement, the merger spreads. This would seem a good thing, insofar as 50% can have the same effect as 20%. That is, they can both result in merger. But, this fails to explain why the merger in South Attleboro occurred when it did. To see why, consider the following. If the proportion of merged speakers in South Attleboro exceeded 20% at any point prior to 1990,

then the merger would have begun earlier. The fact that it did not leaves us with two options. It could be the case that there was a dramatic influx of merged children just prior to 1990. That is, the proportion went from 20% to 50% in a short period of time. Citing census data, Johnson argues that such a dramatic demographic change in South Attleboro is unlikely. The only other option is that the threshold for merger is not as predicted by the model of *merger-by-misunderstanding*. Thus, the network structure we have tested suggests that the model of *merger-by-misunderstanding* cannot explain both thresholds.

Second, we address whether 20% can have the effect of 50% when the threshold is as predicted by the model of *merger-by-misidentification*. Again, consulting Figure 2 in the panel where $p = .4$, we want to establish what the effect of 50% is. In this case, it depends on placement. That is, 50% can bring about merger if placed in the most connected positions or randomly, but fail to do so if placed in the lowest positions. Similarly, 20% can bring about a merger if placed in the most connected positions, but fail to do so if placed randomly or in the least connected positions. We might argue that the two different thresholds just reflect different network structures. But, why should the merged children in South Attleboro occupy less connected positions, and the merged children in Seekonk more connected positions?

Johnson suggests a possible explanation, albeit speculatively, when considering the social situation of merged children moving into Seekonk (2010, 203):

> The children of families from Greater Boston — especially those who live in new homes in expensive subdivisions — might have high prestige. A few such children, who happen to have the merger, might have more influence on their distinct peers.

Of course, we might also think of prestige in terms of connectedness. That is, the merged children entering Seekonk may be highly connected with their peers. It is an empirical question as to whether the economic status of one's parents translates to anything meaningful in terms of social network status. But, it does not seem unreasonable to think that a higher economic status might afford more opportunities for engagement with the peer group in both school and extracurricular activities. This line of reasoning also offers a plausible explanation for the placement of merged children in South Attleboro. For example, we might suppose that rising housing prices in Boston and the expansion of commuting would displace those of lower economic means first. Thus, the merged children moving into South Attleboro before 1990 would not have, on average, occupied more highly connected positions in their peer networks. This, in turn, would have allowed for the higher observed threshold. Again, this rests on an assumed correlation between the parents' economic and a child's social status.

The effect of social networks can be quite strong. But, we should note that the shape of children's actual social networks is an empirical matter. Here we have merely offered a proof of concept that such structures can matter. Much more work would need to be done in order to substantiate our conjectures. However, on the whole we have seen that even these conceptual considerations favor the model of *merger-by-misidentification*.

6 Conclusion

We have used game theory to formalize hypotheses regarding the factors that affect the spread of mergers. We derived conditions for *merger-by-misunderstanding* and *merger-by-misidentification*, and compared the predictions to empirical thresholds. Given the existence of distinct empirical thresholds, we considered what aspects of the two models could be reasonably varied across locations. Both error rates and network structure weighed in favor of *merger-by-misidentification*.

While game theory offers us a powerful framework for exploring how learning impacts language change, much work remains to be done. We have asked several questions regarding the shape of social networks and their impact on learning and change. Answering them will be an empirical matter. More broadly, if all of sound change can be thought of as (conditioned) mergers and splits, then we have a wealth of evidence at our disposal to continue testing and refining our hypotheses.

References

Barabási, Albert-László, and Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286:509–512.

Biemann, Chris, Gerhard Heyer, Uwe Quasthoff, and Matthias Richter. 2007. The Leipzig corpora collection-monolingual corpora of standard size. In *Proceedings of Corpus Linguistics 2007*.

Boland, Julie E., and Allison Blodgett. 2001. Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language* 45:391–411.

Bonin, Patrick, and Michel Fayol. 2002. Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology* 14:289–313.

Branigan, Holly, Martin Pickering, and Alexandra Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition* 75:B12–25.

Brennan, Susan, and Herb Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22:1482–1493.

Brown, Roger, and Camille Hanlon. 1970. *Derivational complexity and order of acquisition in child speech*, 11–53. Wiley Online Library.

Bush, Robert R., and Frederick Mosteller. 1951. A mathematical model for simple learning. *Psychological Review* 58:312–323.

Chambers, Jack K. 1992. Dialect acquisition. *Language* 673–705.

Fudenberg, Drew, and David K. Levine. 1998. *The theory of learning in games*. MIT press.

Gregory, Stanford. 1990. Analysis of fundamental frequency reveals covariation in interview partner's speech. *Journal of Nonverbal Behavior* 14:237–251.

Herold, Ruth. 1990. Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania. Doctoral dissertation, University of Pennsylvania.

Hirsh-Pasek, Kathy, Rebecca Treiman, and Maita Schneiderman. 1984. Brown and Hanlon revisited: Mothersrq sensitivity to ungrammatical forms. *Journal of child language* 11:81–88.

Johnson, Daniel Ezra. 2010. *Stability and Change Along a Dialect Boundary: The Low Vowels of Southeastern New England*. American Dialect Society.

Keller, R. 1994. *On language change: The invisible hand in language*. Routledge.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Labov, William, Mark Karen, and Corey Miller. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3:33–74.

Lewis, David. 1969. *Convention*. Cambridge: Harvard University Press.

Natale, Michael. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32:790–804.

Pardo, Jennifer. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119:2382–2393.

Payne, Arvilla. 1976. The Acquisition of the Phonological System of a Second Dialect. Doctoral dissertation, University of Pennsylvania.

Peterson, Gordon E., and Harold L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24:175–184.

Swinney, David A. 1979. Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of verbal learning and verbal behavior* 18:645–659.

Trudgill, P. 1986. *Dialects in contact*. Oxford: Blackwell.

Wedel, Andrew, Scott Jackson, and Abby Kaplan. 2012. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. Unpublished Manuscript.

Yang, Charles. 2009. Population Structure and Language Change.

Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104
cahern@ling.upenn.edu