Signaling Games and the Actuation Problem

Christopher Ahern
cahern@ling.upenn.edu

Abstract

The actuation problem is at once the most fundamental issue and recalcitrant problem in the study of language change. Here we present a game-theoretic framework that allows to explore the impact of cognitive and social factors in the actuation and spread of linguistic change. Signaling games are used to derive particular conditions for mergers to begin in and spread through a speech community. The predictions made by the two models are tested against empirical data and conceptual expectations.

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. The why and when of language change are intimately tied to how language is learned and used. Here we use signaling games (Lewis, 1969) as a framework for formulating and testing hypotheses regarding the factors that influence change. In particular, we focus on the role of cognitive and social factors in the actuation and spread of mergers.

The rationale for doing so is twofold. First, the search for the conditions that precipitate or preclude mergers is not a new one. The idea that there are functional pressures against the loss of distinctions has both an intuitive appeal and a contentious history (Gilliéron, 1917; Martinet, 1952; Hockett, 1967; King, 1967). Avoiding the pitfalls of previous accounts forces us to ground our analysis in our knowledge of how language is acquired and used. Second, mergers are notably free from social evaluation (Labov, 1994, 344). This simplifies our task by allowing us to leave aside questions of individual agency and concern stemming from notions of prestige in the course of change (Trudgill, 1986). Instead, we can focus on social factors that are, in the cases we consider here, arguably below the level of consciousness.

We outline signaling games and give a general characterization of the question of when mergers occur before spelling out two models and the predictions they make. Finally, we note the effect of network structure on the models and suggest ways of modeling cases where individuals are aware of and intentionally use phonological variants to certain ends.

Signaling Games

A Signaling Game is a dynamic game of incomplete information consisting of a tuple \( G = \langle S, R, T, \delta, M, A, U_S, U_R \rangle \). This tuple describes a sender, \( S \), and receiver, \( R \), engaged in a strategic interaction where the strategy and actions of the one affect the other. The game proceeds as follows. The sender has a state \( \tau \in T \), which is drawn based on a probability distribution \( \delta \). This state can be thought of as the sender’s intended meaning. The sender also has a strategy \( s \in [T \rightarrow M] \), which maps types to the set of messages, \( M \). Based on this strategy, the sender chooses a message \( m \in M \) and sends it to the receiver. In turn, the receiver has a strategy \( r \in [M \rightarrow A] \), which maps messages to a set of actions, \( A \). These actions can be thought of as the different ways in which the receiver can interpret the message. Based on this strategy, and the message sent, the receiver takes an action.

The outcome of the game is determined by the type of the sender, and the strategies of the two agents. The sender and receiver have preferences over these outcomes, which are specified by the utility relations \( U_S \) and \( U_R \), respectively. For example, these preferences might be for the successful communication of the intended meaning, or for the use of a particular signal to convey the meaning. From these utility relations and the probability distribution over types, we can calculate the expected utilities of particular sender and receiver strategies as the sum of the utility of the outcomes weighted by the probability of the states that result in those outcomes. Borrowing the notion of differential reproduction from evolutionary biology, expected utilities can be used to predict population dynamics. Strategies with the highest expected utility are the most successful in dealing with the linguistic environment and will thus proliferate. We present two models that define the available strategies and yield conditions under which mergers will be actuated.

Models

General Characterization

Asking when a merger takes place boils down to asking when a merged strategy does better than a non-merged (or distinct) strategy in a population. This is true when the expected utility of a merged strategy, \( EU(M) \), is greater than that of a non-merged strategy, \( EU(D) \).

\[
EU(M) > EU(D)
\]

(1)

Assume that individuals who use a merged (distinct) sender strategy also use a merged (distinct) receiver strategy, and that individuals only ever change behavior given feedback as listeners. We can specify the expected utilities of these strategies and the conditions where the inequality holds.

Let \( EU_{XY} \) be the expected utility of a receiver where \( X \) is the strategy of the sender and \( Y \) is the strategy of the receiver. For example, \( EU_{MD} \) represents the expected utility of a non-merged receiver when playing with a merged sender. Letting \( p \) be the proportion of merged individuals in the population, the expected utilities of both strategies can be given as:

\[
EU(M) = pEU_{MM} + (1-p)EU_{DM}
\]

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

(2)
In turn, this allows us to specify the conditions where a merger will occur according to the proportion of merged individuals in the population:

\[ p > \frac{EU_{DD} - EU_{DM}}{EU_{MM} + EU_{DD} - EU_{MD} - EU_{DM}} \]  

(3)

We can draw two general conditions from this inequality, even in the absence of any particular interpretation of what the strategies are. First, in cases of dialect contact between merged and non-merged speakers, we can determine the proportion of merged speakers that would cause a merger. Second, we can determine the conditions that suffice for actuation of a merger \((p \leq 0)\). This holds when \(EU_{DM} > EU_{DD}\).

With these two conditions in mind, we now proceed to give particular interpretations to the strategies involved based on cognitive and social factors.

### Cognitive Factors

Cognitive factors are taken to be those that “influence the acquisition of the linguistic system that conveys information on states of affairs” (Labov, 2010, 2). Mergers lead to homophony, homophony to ambiguities, and ambiguities to an increase in the probability of misunderstandings. That is, mergers have a detrimental impact on the transmission of information about states of affairs via linguistic signals. However, as noted above, there are conditions that make adopting a merger a better strategy for dealing with the linguistic environment.

For example, Herold (1990) conjectures 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 can increase the successful transmission of information. The first part of this conjecture is borne out in a corpus of naturally occurring misunderstandings (Labov, 1994, 2010); misunderstandings almost always occur between merged speakers and non-merged hearers. The second part of this conjecture depends on what it is that merged speakers actually do. Yang (2009) develops a model that spells out the particular strategies employed by speakers and hearers. We reframe this model as a signaling game and use it to further examine the properties of (3).

For the minimal pairs affected by the merger, the sender strategies of concern can be stated as follows. Non-merged senders use different phonemes (messages) conditioned on an intended meaning (state), whereas merged senders only use one phoneme regardless of the intended meaning. For receiver strategies, non-merged receivers rely on the phonemic difference as a semantic cue, whereas merged receivers ignore the distinction. Given that mergers render minimal pairs homophonous, we can be more specific about the strategy employed by merged receivers. Namely, evidence from homophone processing suggests that merged hearers assume the more frequent interpretation of homophonous pairs (Boland and Blodgett, 2001; Bonin and Fayol, 2002).

We adopt the following notation, which addresses the low-back merger, to express the expected utilities of the combinations of strategies. /o/ represents the phoneme in the word cot and /oh/ represents the phoneme in the word caught.

Let \(h'_{oh}\) and \(h_o\) be the frequencies of the words containing /oh/ and /o/ in the ith minimal pair. Let \(H_{oh} = \sum h'_{oh}\) and \(H_o = \sum h_o\), the sum of the frequencies of the words containing the different phonemes. Let \(H = H_{oh} + H_o\), the sum of the frequencies of all the words concerned. Let \(e_{oh}\) be the probability of mistaking /oh/ for /o/ and \(e_o\) be the probability of mistaking /o/ for /oh/ (Peterson and Barney, 1952; Hillenbrand, Getty, Clark, and Wheeler, 1995).

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 or additional processing costs, when the message is perceived correctly. Whether the word is in \(H_o\) or \(H_{oh}\), errors will only be made when the message is misperceived. When paired with a merged speaker, a non-merged listener will not do so well. Without loss of generality, assume that merged senders only ever use /oh/. This means that whenever words from \(H_o\) are intended, the listener will make an error of interpretation, except in case of an error in perception.

\[ EU_{DD} = H_o(1 - e_o) + H_{oh}(1 - e_{oh}) \]
\[ EU_{MD} = H_o(e_{oh}) + H_{oh}(1 - e_{oh}) \]

(4)

The expected utility of non-merged hearers can be calculated as follows. As noted, evidence from homophone processing suggests that the more frequent of the pair is assumed. Let \(max(h'_{oh}, h_{oh})\) give the maximum of a given minimal pair. If merged hearers treat non-merged speakers alike, then the expected utility of a merged receiver can be expressed as follows.

\[ EU_{DM} = EU_{MM} = \frac{\sum max(h'_{oh}, h_{oh})}{H} \]

(5)

If we make the simplifying assumption that error rates are roughly equal \((e = e_o \approx e_{oh})\), then the proportion of merged speakers can be represented fairly compactly as a linear function of \(EU_{DM}\), where \(\gamma = \frac{1}{2(1 - e)}\).

\[ p > \gamma(1 - e) - \gamma(EU_{DM}) \]

(6)

This allows us to revisit the conditions we specified above in more detail. As the asymmetry between maximum and minimum in each pair grows, so does \(EU_{DM}\). 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. Absent contact, we can also calculate conditions for actuation. These obtain exactly when \(EU_{DM} > 1 - e\). A low error rate will require a particularly
lopsided set of minimal pairs for a merger to actuate. For a high error rate, even a largely balanced set of minimal pairs can result in a merger.

These results fit nicely with those reported by Wedel, Jackson, and Kaplan (to appear). 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 of the pair is close to 1 — yields the best model for predicting whether a merger has occurred or not. A large number of balanced minimal pairs does not entail a small value for $EU_{DM}$. For example, there may be multiple low frequency pairs that are quite balanced, but one or two extremely unbalanced pairs. The model presented here could easily be used to make predictions given any particular phonemic contrast and frequencies.

**Social Factors**

Social factors are taken to be those that influence the actions by and interactions between individuals with particular characteristics (age, gender, class, etc.). Given that mergers are not conditioned by any characteristics in particular, we will focus on the effects of face-to-face interaction between individuals. In particular, we will examine the impact of accommodation, which is evident at various levels of linguistic structure (Natale, 1975; Gregory, 1990; Pardo, 2006; Branigan, Pickering, and Cleland, 2000; Brennan and Clark, 1996; Garrod and Doherty, 1994).

Accommodation to fine phonetic detail has been shown across both laboratory (Goldinger, 1997) and more natural conversational settings (Pardo, 2006). If people act as if to obey the simple maxim “talk like others talk” (Keller), whether by some automatic process or in order to reduce social distance (Giles, Coupland, and Coupland, 1991), the resulting change can be substantial. In the case of mergers, the preference for coordinating on using the same phoneme given a particular lexical item may drive individuals to adopt the merger. We can spell out the strategies and preferences that describe the effect of accommodation.

Sender strategies are the same as above. Non-merged senders condition on the intended meaning, merged senders do not. For receiver strategies, suppose that receivers are able to recover the intended meaning of the sender, but have a particular preference for using the same signal. That is, suppose that receivers compare the signal to what they would have sent if in the role of the sender. For example, a non-merged hearer receiving the word *cot* from a merged speaker will take it as a failure to coordinate. The same holds for a merged hearer receiving *cot* from a non-merged speaker. The process of identifying what is the ‘same’ or ‘different’ is not without error. We use $e_{oh}$ and $e_o$ to represent these probabilities, but not that they bear a different interpretation than in the last model. In principle, these error rates are open to empirical investigation in a largely analogous way.

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 that successful coordination has occurred except in those cases of misperception.

$$EU_{DD} = \frac{H_o(1-e_o) + H_{oh}(1-e_{oh})}{H}$$

When paired with a merged speaker, a merged hearer does not do as well. When words from $H_o$ are intended the listener only thinks the same phoneme was used when the signal is misperceived. When words from $H_{oh}$ are intended they are more successful, less errors in perception.

$$EU_{MD} = \frac{H_o(e_{oh}) + H_{oh}(1-e_{oh})}{H}$$

The expected utility of a merged listener can be calculated as follows. When paired with a non-merged speaker, a merged hearer is in a similar situation. When words from are intended the listener only thinks the same phoneme was used when the signal is misperceived. When words from $H_{oh}$ are intended they are more successful, less errors in perception.

$$EU_{DM} = \frac{H_o(e_o) + H_{oh}(1-e_{oh})}{H}$$

When paired with a merged speaker, a merged hearer always believes to have coordinate on the use of the same phoneme regardless of whether the word came from $H_o$ or $H_{oh}$.

$$EU_{MM} = \frac{H_o + H_{oh}}{H}$$

Again, if we make the simplifying assumption that error rates are equal, we can represent the corresponding threshold in a fairly compact manner, where $\gamma$ is defined as above.

$$p > \frac{\gamma^{-1}}{2\gamma^{-1} + e}$$

In the limit, the threshold is $\frac{1}{2}$. However, we know that $\gamma^{-1} = \frac{H_o}{H}(1 - e) < 1$, so we will only be concerned with a smaller interval. Namely, we know that when $\gamma^{-1}$ is very close to zero, the threshold for merger is as well. As it increases, so does the threshold. The rate at which it does so is determined by the error rate. When the error rate is low the threshold quickly approximates $\frac{1}{2}$. When the error rate is high, the threshold approximates $\frac{1}{2}$ much more slowly.

We can specify the conditions for merger in this model. Note that there are two possible sources for a low threshold. $\gamma^{-1}$ might be very small if $\frac{H_o}{H}$ is very small, if $e$ is rather large, or some combination of the two. 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. Unlike in the previous model, we need not restrict ourselves to only considering minimal pairs. That is, $H_o$ could be the sum of all words across the lexicon containing /o/. Regardless of the set of words used to calculate the threshold, there is an asymmetry in the predictions of the model. The smaller $H_o$ is, the more likely that a merger will occur resulting in /oh/. The larger $H_o$ is, the less likely that a merger is to occur resulting in /oh/. The rate
of error determines how asymmetric the comparison must be, in a particular direction, for the merger to take place at a given threshold.

There are two conditions that suffice for actuation under this model. The first we can discard as tautological. That is, if $\frac{H_0}{H_f} = 0$, then $p = 0$. If a contrast has disappeared, then a contrast will disappear. The second occurs when $\frac{H_0}{H_f}(2e - 1) > 0$. This is true just when $e > \frac{1}{2}$. When error rates are high enough, a merger will be actuated. The plausibility of this situation hinges on how we interpret the error rates and empirical verification of reasonable values.

**Evaluation**

Both models make predictions we can test against empirical data. We consider the spread of the low-back merger along the Rhode Island-Massachusetts border documented by Johnson (2010). The merger went to completion within a few years starting in roughly 1990 in the town Seekonk, MA and 2000 in South Attleboro, MA. The dialect boundary between merged and non-merged regions was historically stable, despite contact between adults from both sides. Johnson takes this to indicate the role of children in the merger.

In particular, he suggests that the merger is due to an influx of merged children into previously distinct areas. He estimates the proportion of merged children at the inception of the merger to have been roughly 50% in South Attleboro and 20% in Seekonk. The thresholds for merger were computed for both models over various error rates using minimal pair frequencies from a tokenized version of the Wortschatz Corpus. The horizontal and the vertical axes represent the error rates $e_o$ and $e_{oh}$, respectively, and 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.

The cognitive model predicts the 20% threshold for a range of error rates, however. However, it fails to predict the 50% threshold for any combination of error rates. In contrast, in the social model there are combinations of error rates that yield both thresholds. However, the plausibility of these predictions rests on our interpretation of error rates. In terms of actuation, the cognitive model offers an intuitive prediction. If error rates exceed roughly 15%, then a merger will be actuated. This seems reasonable given that the merger is so widespread (Labov, Ash, and Boberg, 2006). The conditions for actuation on the social model are much more restricted and, again, hinge on the value of the error rates. In total, the cognitive model seems to yield predictions that match with both some of the data as well as conceptual expectations regarding actuation. To see if more of the data can be predicted by the cognitive model, we consider the role of network structure.

**Network Structure**

Implicit in the models is the assumption that the likelihood of encountering an agent with a particular strategy was the same as the proportion of the population playing that strategy. In terms of network structure this amounts to a fully-connected network. For any individual in a network of size $n$ the number of other individuals it is connected to — the degree of that individual— is $n-1$. If individuals interact by randomly choosing, then the proportion of individuals is equivalent to the likelihood of interacting with one. However, if the network is not complete then the *effective* proportion of individuals may differ from the actual proportion of individuals. A few highly-connected individuals may exert an outsized influence on the population. They will play a greater than average part in the linguistic environment of others.

Given a particular distribution of degrees of connectedness,
we wish to specify the way in which location influences interactions. That is, we wish to determine the influence that constraints on placement in the network have on the actual proportion of agents using a given strategy. We suggest preliminary steps towards determining the extent to which these sorts of constraints might affect the predictions made by the two models.

**Conclusion**

We have used signaling games as a framework for formulating hypotheses regarding the factors that affect the spread of mergers. Conditions for merger in cases of contact and actualization are predicted and can be compared to empirical data. The approach we have taken here can be extended to cases where the use of variants is subject to conscious control. Sender and receiver strategies can be made contingent upon characteristics of both interlocutors and we might model the use of a prestige variant as some sort of increase in utility. The utility derived from the use of a particular variant might change depending on the sender and the receiver. Moreover, the strategies available to an agent might be conditioned on familiarity with or knowledge of another. While these extensions are tentative, the ease of incorporating a wide range of phenomena recommends the framework presented here.

**References**


