Population Structure and Language Change

Charles Yang*
Department of Linguistics & Computer Science
University of Pennsylvania

October 28, 2009

The parallels between language change and biological changes were noted by none other than Darwin himself. However, the development of a mathematical foundation for the theory of evolution has not been matched by similar progress in the study of language change, even as tools from the quantitative genetics have found applications in the linguistic arena (Cavalli-Sforza 1997, Ringe, Warnow, & Taylor 2002, Gray & Atkinson 2003).

These notes take some initial steps toward a synthetic theory of language change. For the study of cultural transmission, which broadly includes language, Cavalli-Sforza & Feldman's classic treatment (1981) remains the benchmark. Nevertheless, empirically meaningful models of change must incorporate an accurate understanding of what is being transmitted (i.e., linguistic structures) and how transmission takes place (i.e., language learning by children). And it is in these aspects that the Cavalli-Sforza & Feldman approach is lacking, and that the present work aims to address. After all, the mathematical theory of biological evolution is established upon appropriate characterizations of genetic structures and inheritance.

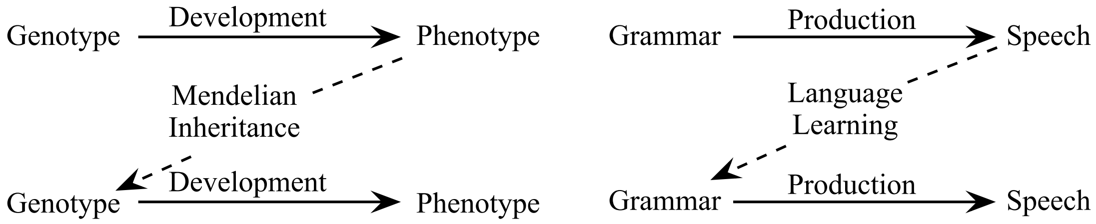


Figure 1. Biological and linguistic change via transmission over time.

We model change as a selectionist process among linguistic hypotheses that successive generations of language users may be associated with (Yang 2002). Its formal properties in the general case turn out to be identical to those of the theory of Natural Selection, a connection

---

*The development of this paper has benefited from numerous conversations with Bill Labov and Gillian Sankoff, for which I am grateful. Daniel Ezra Johnson's work and subsequent personal communications have also been instrumental.

that is facilitated by the general characteristics of language learning. In the present paper, we first summarize the essential features of the model of change via learning, and then consider a specific application to a case of phonological change induced by linguistic migration at the Massachusetts-Rhode Island border. We show that the model correctly predicts the population structure that facilitated the attested change.

1 Learning and Change as Selection

Language change, like biological change, cannot take place without variation. Language change can be observed when the distribution of linguistic forms changes over time. A novel syntactic pattern may gradually eliminate the old one, a new phonological system introduced by migrants can displace the indigenous system. Some linguistic changes take over a millennium to complete, while others develop rapidly in a span of several short years.

Suppose that language change involves two grammars A and a. The term "grammar" refers to any linguistic system and is not restricted to the conventional use of syntactic systems; they could mean two phonemic systems, two morphological rules for word formation, two opposite values of a linguistic parameter, etc. We also assume that the options of A and a are available to the learner, which may be through the innate linguistic system (Chomsky 1981) or inductively formed with (non-homogeneous) linguistic data.

Let the composition of the linguistic environment at some specific time be

$$\pi = p\mathsf{A} + q\mathsf{a} \text{ where } p + q = 1$$

That is, grammar A is used with probability p and a, q. The environment π can be approached through quantitative analysis of historical linguistic records and observations of language change in progress. The next generation of learner uses the algorithm of language acquisition **L** to learn from this mixture, and the resulting dynamical system is of the form (see Figure 1):

$$\pi' = \mathbf{L}(\pi)$$

And so on.

1.1 Learning

The framework of viewing change as learning over generations is quite general. Different learning models **L** can lead to different dynamics of change (e.g., Berwick & Niyogi 1995), though one should strive to find the more empirically appropriate models of learning when possible. There is now evidence that language acquisition by young children follows a *variational* process (Yang 2002; see also Labov 1994). In a direct homage to evolution (Mayr 1982, Lewontin 1983), the variational learning model views a major component of language learning as changes in the distributions of the learner's grammars in response to the statistical sample of input data in the environment. One instantiation of this approach is the Linear Reward Punishment model of Bush & Mosteller (1951), a simple probabilistic learning theory that has considerable support in many domains of learning. The learning algorithm is given below for A; the case of a is symmetrical.

Algorithm L: Given an input token (e.g., sentence) s in the environment π, the learner selects a grammar A with probability p:

$$\begin{aligned} &\text{if } \mathsf{A} \to s \text{ then } p' = p + \gamma q \\ &\text{if } \mathsf{A} \not\to s \text{ then } p' = (1-\gamma)p \end{aligned} \tag{1}$$

The learning rate γ controls the degree to which the learner adjusts the probabilities of the grammars. The notation "→" is used to describe an empirically motivated relation between grammar and data. For instance, $\mathsf{A} \to s$ could denote that A can successfully parse a sentence s or that A does not incur excessive cognitive load when processing s in real time, or that A causes no miscommunciation in the presence of s, etc. Its exact interpretation is to determined by the specific linguistic case.

Adapting Bush & Mosteller's formulation, the fitness of a grammar G can be defined as its *penalty probability*, the proportion of input tokens that G is incompatible with the linguistic environment π:

$$c = \Pr(G \not\to s \mid s \in \pi) \tag{2}$$

In other words, c is the probability of linguistic data in π which causes the probability of G to decrease during learning. Penalty probabilities of grammars, just like fitness of genotypes in evolution, are values that can be estimated by the scientist; they are not calculated by the language learner–or the organism under selection–explicitly or implicitly.

Let C_A and C_a be the penalty probabilities of A and a in π, it is possible to show (proof omitted) that if the learning rate γ is fairly small, the learner converges to the following distribution:

$$\lim_{t\to\infty} p_t = \frac{C_a}{C_A + C_a} \quad \lim_{t\to\infty} q_t = \frac{C_A}{C_A + C_a}. \tag{3}$$

It is clear that in a mono-grammar environment, the learner will converge to the target, which, by definition, has penalty probability of 0 while the competitor is penalized with a positive probability. This is the usual situation in first language acquisition; the evidence from child language can be found in Yang (2002, 2006) and numerous other sources. Significantly, the systematic errors children produce in deviation from the target grammar reflects the structural properties of the competitors–linguistically possible grammars though those to which children have no exposure. In this sense, language learning is, at least to a large extent, a process of unlearning or forgetting as has been suggested in the theoretical neuroscience literature (Changuex 1986, Edelman 1987). But in a monolingual environment without variation, there cannot be language change.

When both A and a variants are in π, the learner will converge on to a statistical distribution of the two as specified in Eq (3). Presumably, language changes when a previously homogeneous environment (e.g., a) is punctured by the introduction of a variant (e.g, A). The present work is to establish the dynamics of change once A has been introduced. Once the basic model is in place, one can study the possible sources from which A arises, e.g., language contact, language acquisition errors, which have ready counterparts in biological processes (e.g., migration, mutation, probability of fixation, etc.).

1.2 Change

In many studies of syntactic and morphological change (Kroch 1989, Taylor 1994), language users during the course of change are in a sense bilinguals. This can be observed in historical records where a single writer alternates between two variant forms of grammatical use. And it is the distribution of these variants that change over time, which, in many cases, spans over hundreds of years or even over a millennium (Sun 1996). That is, the distribution of the grammars changes from one generation to the next, each transition described by the learning algorithm described above:

$$(p_0, q_0) \xrightarrow{\mathbf{L}} (p_1, q_1) \xrightarrow{\mathbf{L}} (p_2, q_2) \xrightarrow{\mathbf{L}} \ \ldots \ \xrightarrow{\mathbf{L}} (p_n, q_n) \tag{4}$$

We have studied such cases elsewhere. It turns out that the dynamics of change as measured by (p, q) is identical to that of Natural Selection in haploids. The fitness of grammars can be construed as probabilities with which A and a contradict each other. These values can be estimated from linguistic corpora and then used to predict the direction of change, which is gradual but deterministic regardless the initial environment (p_0, q_0), again similar to the basic models of selection. See Yang (2002) for details and an application to syntactic changes in the history of English and French.

However, other types of language change require modifications to the extended scenario of learning and change described in (4). Crucially, certain aspects of language, in particular the phonemic system, change at a far more brisk pace.[1] Herold's (1997) survey in the town of Tamaqua, PA reveals that all speakers 74 years or older (in 1988) maintained the distinction between the vowels in *cot* and *caught* while everyone under the age of 65 lost it, suggesting that the entire duration of change lasted no more than a decade. A recent study of these vowels at the Massachusetts–Rhode Island border (Johnson 2007) shows that the change from the distinct to the merged system in young children can be completed in only 2-3 years. Second, as also documented in Herold and Johnson's studies, the changes can be attributed to language contact, and can only be initiated – and completed rapidly – when there is a sufficient amount of migrant speakers with the new phonemic system. To accommodate these findings, we thus assume the learning in such cases follow a winner take all approach (Riesenhuber & Poggio 1999) that is sensitive to (p_0, q_0), the initial composition of the linguistic environment:

$$(p_0, q_0) \xrightarrow{\mathbf{L}} \begin{cases} (1,0) & \text{if } p_1 > q_1 \\ (0,1) & \text{if } q_1 > p_1 \end{cases} \tag{5}$$

We study the dynamics of phonemic learning and change described in Eq. (5), with specific attention to the requisite initial condition that facilitates, and indeed, necessitates, change.

---

[1] Ultimately, the pace of language change may reflect the fundamental differences across the components of the language faculty. Children routinely learn alternate forms of linguistic rules, the most obvious case being languages with flexible word orders, but an inconsistent or conflicting phonemic inventory is inconceivable for a functioning linguistic system. The specialization to the native phonemic system occurs very early in language acquisition, generally before the first birthday (Werker & Tees 1983; see Yang 2006 for reviews). Indeed, even bilingual acquisition from birth seems to show the dominance of a single phonological system (Cutler et al. 1989), where the quantity of experience plays a significant role (Bosch & Galles 2003)

2 Merger and Migration

2.1 Background

For many dialects of North American English, the low back vowels in words such as *cot* (/**o**/) and *caught* (/**oh**/) are indistinguishable. This phenomenon has been commonly referred to the *cot-caught* merger: "merger", as these vowels were once distinct (and still are, as in British English). Mergers have been extensively studied in the linguistics literature (e.g., Labov 1994, Labov, Ash & Boberg 2006). The very existence of mergers, which obliterate the differences between words, stands in contrast to the intuition long held by linguists and non-linguists alike, that the primary function of language (and language change) is to facilitate effective communication. Indeed, mergers are among the most widely documented linguistic changes. Moreover, once two vowels are merged, they are seldom reversed (Labov 1994).

Perhaps the most detailed study of mergers in progress is Johnson's University of Pennsylvania dissertation on several speech communities at the Massachusetts and Rhode Island border (2007). In the Eastern Massachusetts area, including Boston, the *cot-caught* merger is firmly in place: the vowels in minimal pair words such as *cot-caught*, *Don-Dawn* etc. sound the same. We introduce the notation <MEANING, vowel> to describe the intended meaning and the articulatory realization of words. For instance, <CAUGHT, **oh**> stands for the word with the meaning CAUGHT (uppercase)and the vowel is /**oh**/. In the Boston system, denoted as M_+, /**o**/ has merged with /**oh**/, we thus have words <DON, **oh**>, <DAWN, **oh**>, <COT, **oh**>, <CAUGHT, **oh**> etc.[2]

The neighboring state of Rhode Island, however, traditionally has the distinct system (M_-). Johnson's field work focuses especially on the Southwestern Massachusetts towns of South Attleboro and Seekonk at the dialectal boundary. In these communities, adults are aligned with Rhode Island and maintain the distinct vowel system. However, Johnson discovers that in the past two decades, young children have acquired the merger as evidenced by both production and perception studies. In some families surveyed, the younger children have M_+ but their older siblings have M_-, suggesting that the merger has taken place in a space of few short years. Johnson traces the spread of the merger to the recent migration of speakers from the M_+ region of Massachusetts. The sociolinguistic situation is as follows. Young children first acquire their parents' M_- system at home. They then encounter, and acquire, the M_+ system from their peers who have the M_+ system. Johnson conjectures that the merger is acquired only when there is a critical mass of M_+ peers, and hence differences between older and younger siblings in the family survey.

2.2 Processing mergers

The basic model of variational learning can be extended to provide a mechanistic account of the *cot-caught* merger. We will derive the numerical condition – the critical mass in Johnson's conjecture – that would provide sufficient conditions for the merger to spread. But several linguistic

---

[2] As it will become clear, our formulation requires no modification, but only different empirical parameters values, to account regions (e.g., Canada) where /**oh**/ has merged into /**o**/.

matters need to be addressed before the formal details of the model are laid out.

First, the competition is between the merged system M_+ and the distinct system M_-, which must be accessible for young learners and subject to linguistic evaluation. The effect of the critical period for language learning is well known; it is thus assumed that children initially exposed to M_- can acquire M_+ if exposed to the latter at a sufficiently young age, even replacing the former under certain conditions (see Chambers 1992; see also Sankoff & Blondeau 2008).

Second, while many aspects of the phonological system have significant social effects—Labov's NYC department store study being the classic example–there is no reliable evidence that mergers are subject to social evaluation (Labov 2001). Furthermore, children appear to acquire the linguistic system in the environment first, while attaching social values to it at a later age (Romaine 1984). We thus focus only on the linguistic consequences of competitions between M_+ and M_-. Of course, if future research were to identify precise effects of social conditioning on phonemic acquisition in child language, one could include a quantitative measure of external fitness in conjunction with the internal linguistic factors.

Finally, and most crucially, we need to develop an empirically motivated interpretation of fitness for the grammars under competition. One option is to relate a grammar's fitness to its tendency of miscommunication Labov (2009) provides a catalog of naturally occurring cases; a typical example goes as follows:

Speaker A (M_+) : It would be even better if **Don** could take her to the airport.

Speaker B (M_-) : (Wondered for some time about how **Dawn**, who is visually impaired, could take her.)

An interesting asymmetry is revealed by Labov's examples, that the distinct system M_- causes more instances of miscommunication than M_+; we return to this matter momentarily. Here we pursue an alternative and potentially more general approach than miscommunication *per se*; both direct and indirect evidence from the study of the lexicon can shed light on the processing of phonologically similar words, which may provide the ultimate cause for miscommunication in language use.

Consider first the case of the listener using M_+: since there is only one vowel **oh**, the learner will encounter words such as <DON, **oh**> and <DAWN, **oh**>, i.e., words that are intended to mean DON and DAWN but with the same vowel /**oh**/. One might initially suppose that, despite the phonological confound, the linguistic contexts will nevertheless provide sufficient information such that no ambiguity *Don* and *Dawn* will arise. It is thus perhaps surprising to discover that the initial activation of word meanings is based largely on the phonological form of words alone. In a classic study (Swinney 1979 and much subsequent work), subjects hear utterances such as:

> Rumor has it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other bugs$_\Delta$ in the corner of his room.

A lexical decision task is carried out at the position of Δ, whereby the subject is instructed to press a button as soon as he/she recognizes the word on the screen. One group of subjects is visually presented with the word *ant*, a contextually relevant meaning, while the other group is presented with the word *spy*, a contextually inappropriate meaning. In both cases, however, strong facilitation effects (speedup) are observed, as compared to an unrelated control word (e.g., *sew*). But most importantly, there is no significant facilitation difference between *ant* and *spy*, where the context preceding Δ unambiguously favors the former reading. Thus, the initial stage of lexical access is driven by the phonological form, however briefly, and the integration of contextual cues come in at a later stage.

How, then, are homophones such as <DON, **oh**> and <DAWN, **oh**> processed in real time? Camarazza et al. (2001) find that homophone processing reflects the frequency of individual words, in line with the findings of lexical processing in general. For instance, the word *nun* has roughly the same frequency of *owl*, and these words have similar reaction time in picture-naming tasks. A more direct test comes from a study in French (Bonin & Fayol 2002). In a production latency task, speakers process the higher frequency member of homophones (*verre*, glass) consistently faster than the lower frequency member (*ver*, worm). In other words, homophonous words that are less frequent face competition, and thus delay of access, from their more frequent counterparts.

The lack of immediacy in contextual disambiguation, and the frequency effects in homophone processing, together prove hazardous for the M_+ grammars. When probabilistically selected during the learning process, the M_+ grammar will always lead to listener to perceive the vowels in both *Don* and *Dawn*, whether they are produced by M_+ or M_- speakers. Assume, without loss of generality, that the use of the meaning DON is more frequent than that of DAWN. Thus, when either *Don* or *Dawn* is uttered, the M_+ listener will initially access DON due to its high frequency. Now if DON is indeed intended, the listener will have correctly interpreted the utterance. However, if DAWN is intended, the listener must carry out *reanalysis*,[3] overriding DON to retrieve DAWN instead. We take the expected probability of reanalysis–i.e., the probability of encountering a lower frequency item in a minimal pair–to be the fitness of the grammar M_+.

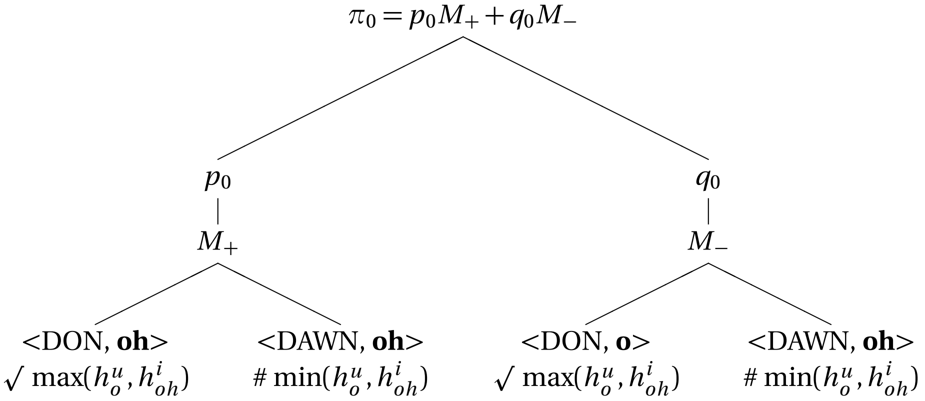
A detailed calculation can be carried out as follows. We obtain a list of 11 minimal pairs from Labov (2009) and then estimate their frequencies from a large English word corpus.[4]. We use h^i_o and h^i_{oh} to denote the frequencies of the **o** and **oh** variant of the ith minimal pair, e.g., the frequency of *Don* (meaning DON) and the frequency of *Dawn* (meaning DAWN). Let $H = \sum_i (h^i_o + h^i_{oh})$, the total frequency of homophone pairs.

We draw the outcome tree for a listener using M_+. We use the pair *Don* and *Dawn* to illustrate the calculation, again assuming that the former is more frequent than the latter. We use # to denote the event of reanalysis due to phonological ambiguity, and $\surd$ otherwise (i.e., the initial processing requires no modification); the probabilities of these events are also given.

---

[3] A type of process similar to recovery from other linguistic ambiguities such as garden path sentences "The horse raced past the barn fell".

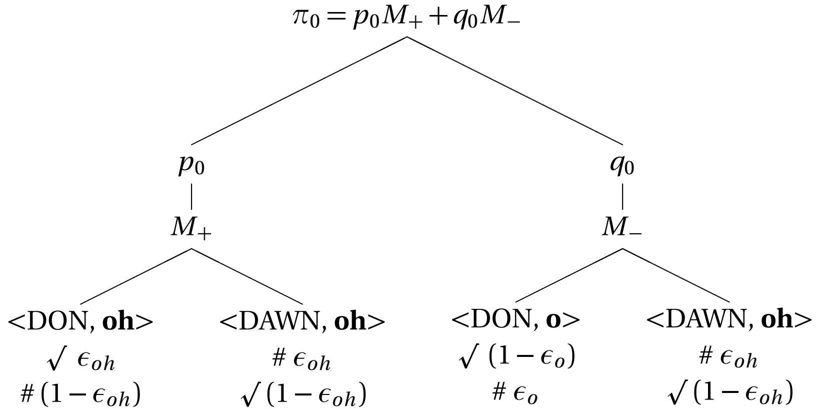
[4] These words and their frequencies are: don (1052), dawn (736), collar (403), caller (23), knotty (25), naughty (195), odd (830), awed (80), Otto (67), auto (260), tot (9), taught (1327), cot (39), caught (2444), pond (258), pawned (31), hock (25), hawk (127), nod (180), gnawed (53), sod (30), sawed (37). The corpus is from the Wortschatz project `http://corpora.informatik.uni-leipzig.de/`, which consists of 24 millions words.



Note the left branch under M_-. The word is <DON, **o**>; a listener using M_+ will hear the vowel as **oh**, the only vowel available, but since DON is the more frequent homophone, no reanalysis is necessary despite the incorrect perception of the vowel. It is easy to see that M_+ triggers reanalysis only when the low frequency member of minimal pairs is encountered. Its penality probability, i.e., the probability with which it is punished in the linguistic environment π_0, is thus

$$\begin{aligned} c_+ &= \frac{1}{H}\sum_i \left[(p_0 + q_0)\min(h^i_o, h^i_{oh})\right], \text{ where } H = \sum_i (h^i_o + h^i_{oh}) \\ &= 0.15 \end{aligned} \tag{6}$$

The case for M_- is slightly more complicated, due to the inherent confusability between /**o**/ and /**oh**/ on acoustic and perceptual grounds. It has been well known that the phonetic similarity between these vowels may cause miscomprehension. The classic study of Peterson & Barney (1958; see also Labov 2009 for refinements) provides a confusion matrix of the American English vowels through a listening test. The /**o**-**oh**/ pair has the highest probability of confusion of all vowel pairs in American English. Specifically, the probability of /**o**/ perceived as /**oh**/ is $\epsilon_o = 0.06$, and the probability of /**oh**/ perceived as /**o**/ is $\epsilon_{oh} = 0.10$. The late onset of contextual factors in online processing entails that when using M_-, misperception cannot be preempted by contextual cues and thus may lead to reanalysis and disambiguation. The probability tree for listening with the M_- system is given below:



Of interest here is the leftmost branch, where a listener using M_+ hears <DON, **oh**>. If it commits a perception error in hearing /**oh**/ as /**o**/, he nevertheless will have correctly interpreted the utterance with no reanalysis necessary! Collecting the terms we obtain the fitness of M_-:

$$\begin{aligned} c_- &= \frac{1}{H}\sum_i \left[p_0 [(1-\epsilon_{oh})h^i_o + \epsilon_{oh} h^i_{oh}] \right] + \left[q_0(\epsilon_o h^i_o + \epsilon_{oh} h^i_{oh}) \right] \text{ where } H = \sum_i (h^i_o + h^i_{oh}) \\ &= (0.294 p_0 + 0.086) \end{aligned} \tag{7}$$

From the convergence properties of the learning model Eq. (3), we obtain the value of p_1 as a function of p_0, by using the fitness values of M_+ and M_- from Eq. (6) and (7) along with minimal pair word frequencies and the probabilities of vowel misperception:

$$p_1 = \frac{0.294 p_0 + 0.086}{(0.294 p_0 + 0.086) + 0.15} \tag{8}$$

Figure 2 plots the value of p_1 as a function of p_0. Given the decision process in phonemic learning described in Eq. (5), we obtain the critical value of p_0 by setting p_1 above to 1/2, which is 0.217. That is, if 21.7% of the peer population for young children are speakers with the merger system, then the merger system characteristic of Eastern Massachusetts will necessarily replace the distinct system characteristic of Rhode Island.

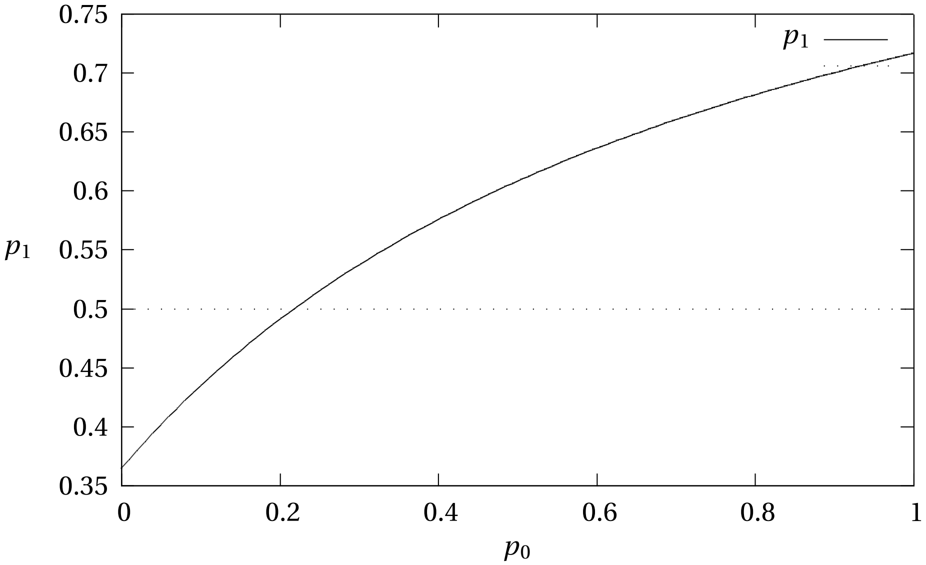


Figure 2. The dynamics of learning from p_0 to p_1 for the *cot-caught* merger. The critical threshold for p_0, which is 0.217, is obtained when $p_1 \geq 1/2$.

2.3 Tipping point

Confirmations of our numerical analysis can be found in Johnson's fieldwork in South Attelboro and Seekonk. Additional evidence comes from broader considerations.

The first major observation of Johnson (2007) is that children born to one M_+ parent generally acquire the M_+ system at home, without needing to be in contact with merged peers. This seems extremely likely provided that the M_+ parent contributes at least 21.7% of the input to the child learner in the home environment, where the child primarily acquires parental speech patterns.

Second, Johnson observes a major transitional point in the population structure. Prior to this point, children from M_- families do not acquire the merger despite interacting with some M_+ children; this corresponds to the older siblings in the families studies. This is followed by a stage where "the proportion of natively merged exceed Y [CY: a threshold]. While distinct children may not be in the minority, they have enough contact with merged peers that they lose their inherited distinction" (Johnson 2007, p248). This stage was reached in South Attleboro around 1990 and about a decade later in Seekonk. School survey data establishes the demographic composition of the student body at the time, and is most directly relevant to our work. Conducted in 2005, Johnson's shows that in Seekonk, 18% of 12th graders come from families with at least one merged parents–and are thus merged themselves–and the percentages are 20% for 8th graders and 23% for 4/5th graders. Notice that these percentages of merged peers are extremely close to the theoretical expectation of 21.7%, which is derived *entirely* and *solely* from word frequencies and vowel mutual confusability measures, which we assume to be constant.[5] In this sense,

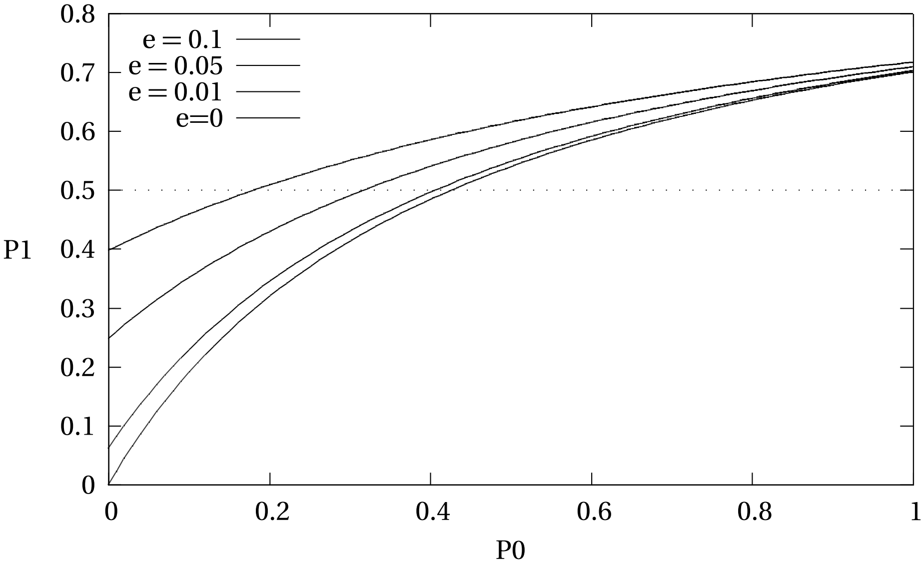
---

[5] The ratios for South Attleboro, which the merger took hold a decade earlier, are 48% (12th grade), 38% (8th grade)

the critical threshold under which the M_+ merger spreads into and replaces the M_- system is predicted.

Third, and beyond the field study of Johnson, we have an account for the asymmetry observed in Labov's naturally occurring examples of miscommunication, that the M_- listener commits more errors than the M_+ listener. Let us assuming that in normal discourse, M_+ and M_- speakers produce speech with comparable frequencies (e.g, $p, q \approx 1/2$). The expressions in (6) and (7) show the probability of reanalysis for 0.15 and 0.233 for the M_+ and M_- system respectively. Presumably, most instances of reanalysis is successful and do not will not cause miscommunication, as the contextual cues generally provide sufficient disambiguation–and not everyone talks about minimal pair individuals named "Don" as well as "Dawn". Nevertheless, the asymmetry in miscomprehension follows from the theoretical calculations.

Finally, and very tentatively, the foregoing discussion may provide an explanation for the observation that mergers spread at the expense of phonemic distinctions and are generally irreversible. The spread of the *cot-caught* merger requires only about 20% of speakers migrating into a distinct community; for the trend to reverse requires four times of M_- speakers migrating into a M_+ community, which seems a priori less likely to occur; see Labov (2009) for additional discussion. The reanalysis probability for M_+, as can be seen in Eq. (6), is correlated with the sum of the lesser frequencies from homophone pairs. By contrast, the reanalysis probability for M_-, from Eq. (6), is correlated with the sum of the frequencies of minimal pairs that contain one of two phonemes. The first quantity is of necessity smaller than the second, and the asymmetry is thus inherent. Even when the two phonemes are not mutually confusable at all, a minority of M_+ speakers is still sufficient to eliminate the M_+ system. Figure 3 plots the cases for the mutual confusability (both directions) from a range of values, while keeping the word frequencies as before:



---

and 40% (4th grade).

Figure 3. The effect of phoneme mutual confusability on the spread of the merger. Note that even when perception error rate is zero, there still needn't be a majority of M_+ speakers for the merger to spread; the critical value here is 0.43.

The considerations of lexical processing make it clear that the obliteration of phonemic contrast cannot go on forever. Mergers create homophones, and as the number of homophones increases, the probability of reanalysis–the consequence of frequency effects in lexical processing–will increase as well, resulting in the higher reanalysis probability for the merged system, which at some point will exceed that of the distinct system. We are currently exploring the limit of homophony through stability analysis of changes predicated on the present model.

3 Conclusion

It must be pointed that the models developed here as well as in previous work represent significant simplifications of reality. Such inherent limitations are at the same time inherent virtues as well. We have taken findings, some direct while others direct, from language learning and language processing, abstracted them into simple mathematical models and extrapolated their properties into a dimension of change. To the extent such efforts advance our understanding of historical linguistics, it is also hoped that they will sharpen the methods and provide guidance for the study of language in a synchronic setting.

Reference

Berwick, R. & Niyogi, P. (1995). The logical problem of language change. MIT AI Lab Memo.

Bonin, P. & Fayol, M. (2002). Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology*, 14, 289-313.

Bosch, L. & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46, 217-243.

Bush, R. & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, 68, 313-323.

Caramazza, A., Costa, A., Miozzo, M. & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 27, 1430–1450.

Cavalli-Sforza, L. L. (1997). Genes, peoples and languages. *PNAS*, 94, 7719-7724.

Cavalli-Sforza, L. L. & Feldman, M. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton, NJ: Princeton University Press.

Chambers, J. K. (1992). Dialect acquisition. *Language*, 68, 673-705.

Changeux, J.-P. (1986). *The neuronal man*. Oxford: Oxford University Press.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Cutler, A., Mehler, J., Norris, D. & Segui, J. (1989). Limits on bilingualism. *Nature*, 20, 229-230.

Edelman, G. (1987). *Neural Darwinism: The theory Of neuronal group selection.* New York, NY: Basic Books.

Gray, R. & Atkinson, Q. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426, 435-9.

Herold, R. (1997). Solving the actuation problem: Merger and immigration in eastern Pennsylvania. *Language Variation and Change*, 9, 165-189.

Johnson, D. E. (2007). Stability and change along a dialect boundary: The low vowels of Southeastern New England. Ph.D. dissertation, Department of Linguistics. University of Pennsylvania. Forthcoming as an American Dialect Society monograph.

Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1, 199-244.

Labov, W. (1994). *Principles of language change: Internal factors.* Basil: Blackwell.

Labov, W. (2001). *Principles of language change: Social factors.* Basil: Blackwell.

Labov, W. (2009). *Principles of language change: Cognitive factors.* Basil: Blackwell. (To appear).

Labov, W., Ash, S. & Boberg, C. (2006). *The Atlas of North American English.* Mouton: de Greuyter.

Lewontin, R. (1983). The organism as the subject and object of evolution. *Scientia*, 118, 65-82.

Mayr, E. (1982). *The growth of biological thought.* Cambridge, MA: Harvard University Press.

Peterson, G. & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America.* 24, 175-184.

Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.

Ringe, D., Warnow, T. & Taylor, A. (2002). Indo-European and computational cladistics. *Transactions of the Philological Society*, 100, 59-129.

Romaine, S. (1984). *The language of children and adolescents: The acquisition of communicative competence.* Basil: Blackwell.

Sankoff, G. & Blondeau, T. (2008). Language change across the lifespan: /r/ in Montreal French. *Language*, 83, 560-588.

Sun, C. F. (1996). *Word-order change and grammaticalization in the history of Chinese.* Stanford: Stanford University Press.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659.

Taylor, A. (1994). Variation in past tense formation in the history of English. Penn Working Papers in Linguistics, 1, 143-158.

Werker, J. & Tees, R. (1983). Developmental change across childhood in the Perception of Non-native Speech Sounds. *Canadian Journal of Psychology*, 37 (2), 278-286.

Yang, C. (2002). *Knowledge and learning in natural language.* New York: Oxford University Press.

Yang, C. (2006). *The infinite gift.* New York: Scribner's.