A critical introduction to phonology:
Of sound, mind, and body

Daniel Silverman

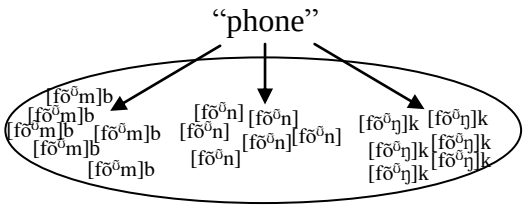
Part 2
I speak my mind

Chapter 5
Variation and probability

THE EXISTENCE OF VARIATION

When I first introduced set notation during our discussion of "phone books" and "foam books" in Chapter Three, I wrote, "Let's suppose for the moment that the word 'foam' has only one pronunciation, whereas 'phone' has…three…which are dependent on the context in which the word is found. While this is a simplification, for now, let's just suppose it's true". We're now ready to discuss the nature of this simplification. Although I have not called your attention to it, perhaps you have realized that my phonetic transcriptions, intended as they are to capture the physical properties of a single particular instance—or *token*—of a word spoken by a particular speaker at a particular time, are actually highly idealized in the sense that I have not linked these transcriptions to a specific utterance by a specific speaker. In fact, every speaker's pronunciation is slightly different from every other speaker's, and indeed, every *token* of every word is slightly different from every other token of that same word, even for a single speaker. In the history of the world, no two spoken utterances have ever been *exactly* alike. But the notation we have been employing up to now does not reflect this token-to-token variation at all. Instead, I have been using a single transcription that, at its very best, might represent either an average or an idealized pronunciation.

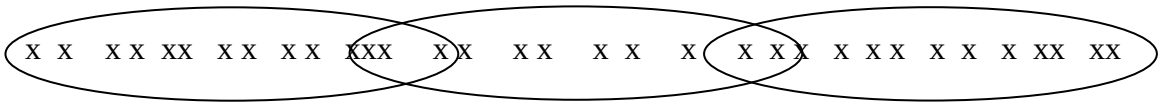
A far more representative display would involve a great number of transcriptions which all vary ever so slightly among one another in a manner that reflects the true variable nature of speech. Of course, even this sort of representation would not do genuine justice to reality, because we can never document the totality of realizations of any word. But we can, at least, employ a notational system that better approximates the true nature of speech. So look at the following revised display.



The *clouds* or *pools* of tokens in this figure do a modicum of justice to the genuine variability of speech production by suggesting that every token differs slightly from every other token. The idea is that each token falls in its own location in some (as yet undefined) multi-dimensional articulatory/acoustic/auditory space. For example, the formant values of the vowel differ slightly from token to token; the tongue position of the nasal's oral closure is slightly different from token to token as well. Now, you'll immediately notice that I've transcribed every token within each pool in an identical way.

So be it. This is just an impressionistic display, and is not intended to convey the actual properties of the variation. For now at any rate, this sort of display suffices.

So, in addition to the phonetic distinctions found among alternants such as [fõ$^{\breve{o}}$m]b, [fõ$^{\breve{o}}$n], [fõ$^{\breve{o}}$ŋ]k, this other sort of phonetic variation shows that categories which are clearly defined phonologically are not so clearly defined phonetically. Phonological categories are clear-cut and discrete in that meaning is not *gradiently* affected by sound substitutions: sound substitutions either change meaning, eliminate meaning distinctions, or maintain meaning distinctions, and that's it. By contrast, gradually lowering the tongue in going from "bid" to "bed" to "bad" does not produce corresponding intermediate meanings such as "biddish bed" or "beddish bad". That is, the phonetic variation that we may observe among speech tokens has no direct correlates in terms of the categories to which these tokens belong: the gradience is only phonetic, and never semantic. Indeed, as the linguist Andre Martinet remarked in 1975, "Linguistic identity does not imply physical sameness…Discreteness does not rule out infinite variety". So, while all tokens within a category are identical in linguistic terms, they are nonetheless phonetically diverse. So look at the following display.



Here, the "x"s represent tokens, the phonetic distinctions of which are suggested by their various locations along the one-dimensional scalar display. Moving from left to right, these tokens gradually change in terms of some phonetic property. Nonetheless, the categories which learners come to impose on this gradient distribution are completely clear-cut and discrete. Despite the phonetic variation, some tokens fall into one category, others fall into the other categories.

But what about the intersecting regions of the sets? Certain tokens find themselves in two categories, or at the boundary between one category and another. These tokens are ambiguous between one discrete category and another. (Later in this chapter, these tokens will be argued to play a very important role in the sound system.) But still, phonetic gradience has no correlate in terms of category gradience. It's certainly *not* the case that these tokens combine semantic elements of the two words, producing a *meaning* somewhere in-between one category and another. Again, discreteness does not rule out infinite variety, and infinite variety does not rule out discrete categoricity.

MODELS OF VARIATION

What might be the origin of the phonetic variation that is always present in speech? One possibility is that phonetic constraints on speech are not so strictly imposed, and so speakers engage in an approximation of sorts. That is, their speech more or less resembles the speech around them, with a little deviation here, a little deviation there, that is somehow "tolerated" at the cognitive level. We can refer to this approach as the *relaxed constraints model*.

Alternatively, there may indeed be strict cognitive constraints on speech. But if learners do have strict internalized constraints on their speech, then what accounts for the phonetic variation that is undeniably present? There are two common proposals for the cognitive organization of this variation, often referred to as the *prototype model*, and the *exemplar model*, respectively. Both of these models (and also the relaxed constraint model) allow for the possibility that linguistic sound categories emerge through

experience with individual examples or instances of perceptual events which come to self-organize themselves into distinct sets or categories. But these two approaches do have important differences. The prototype model of categorization proposes that speakers have very exact internalized phonetic "targets" for their speech (be these targets articulatory, acoustic, auditory, or some combination of these) but they don't hit these targets each and every time. This would be something like an expert dart player who inevitably misses the bull's eye at least once in a while: the darts are clustered around the target, but even an expert can't hope to get it just right every time. Some version of the prototype model has been assumed by many linguists at least as far back as the 19th century. For example, the linguist Hermann Paul wrote in 1880, "However much movement may be the result of training…it still remains left to chance whether the pronunciation be uttered with absolute exactness, or whether slight deviation from the correct path towards one side or the other manifests itself". While Paul does not specifically propose an abstract prototype, he nonetheless assumes that there is a single articulatory "target" that speakers aim for. In the prototype model then, there is an exact, abstract, category-defining value that emerges upon experience with individual tokens, and so any observed within-category variation is viewed as a deviation from this prototype.

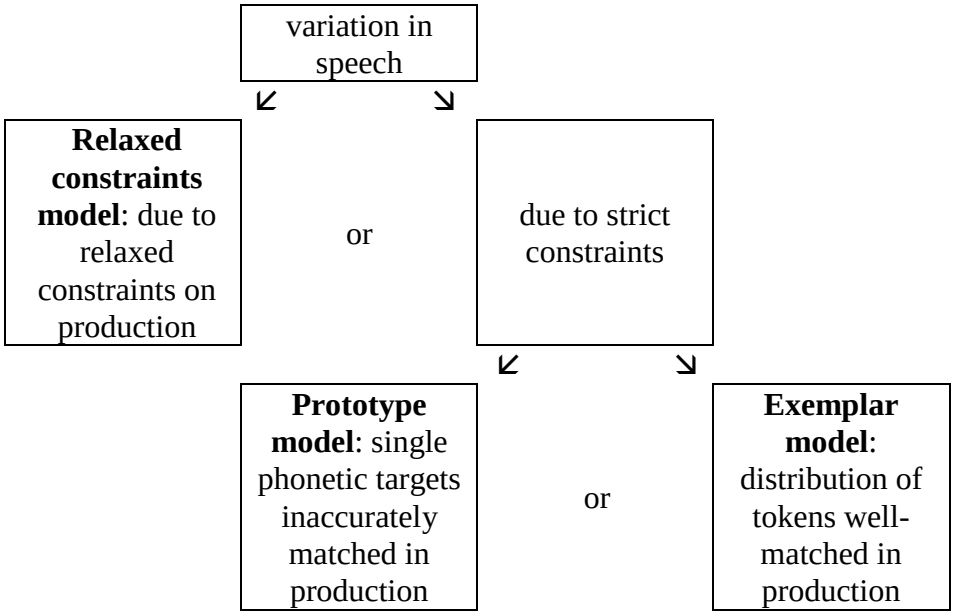
What distinguishes the exemplar model is that perceptual categories are defined as the set of all experienced instances of the category, such that variation among tokens actually contributes to the categorical properties themselves. That is, a given category is the culmination of the variable forms themselves, in that the distribution of tokens is not viewed as a deviation, but is instead viewed as a defining aspect of the category. So, within-category variation is thus part and parcel of the category itself. But what about the *origin* of variation under the exemplar approach? One idea is that speech patterns are copied again and again by generation after generation, but inevitably with very slight inexactitudes once in a while. According to the exemplar model, one generation's variation—inexactitudes included—serves as the next generation's template for copy. So variation may be viewed as the accumulation of very minor inexactitudes both within and between generations of speakers; the long-term product of excellent-though-imperfect copying of ambient speech patterns. The result is that tokens within a category cluster around each other, with each generation's distribution of tokens differing ever so slightly from the both the preceding and following generations' tokens. As we'll see momentarily, these slight differences may come to play a significant role in the way sounds change over time. If we further assume, as is reasonable, that more recently encountered tokens leave a stronger memory trace than do more remote tokens, then we can further account for the sound changes observable even across the lifetime of a single speaker.

This approach to linguistic categorization is hardly new, having been proposed in the 19th century by a number of scholars. Mikołaj Kruszewski, discussing a hypothetical case of a slightly fronted [k] (which he writes *k'*), with variants k'_1, k'_2, k'_3, etc., wrote in 1883: "Our characteristic, unconscious memory of the articulation of sound *k'* should be a complex recollection of all articulations of *k'* which we have performed. But not all of these articulations are arranged equally in the memory. For this reason, after performing the articulation of k'_3, the chances of performing k'_4 are much greater than they are for k'_1, etc".

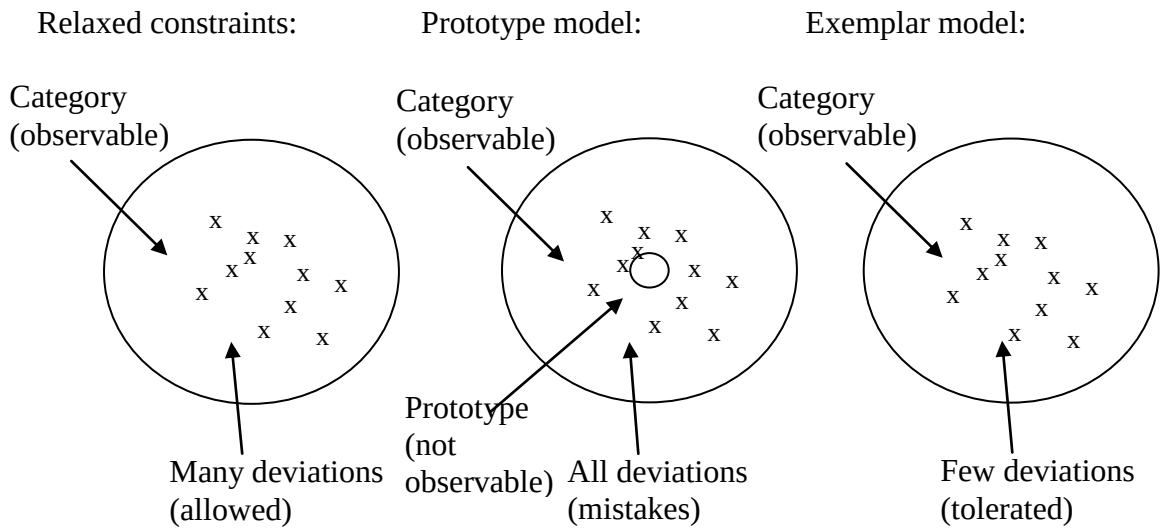
A few years later, in his book of 1890, Hermann Paul wrote,

"[V]ariability of production, which remains unnoticed because of the narrow limits in which it moves, gives the key to our comprehension of the otherwise incomprehensible fact that a change of usage in the sounds of a language sets in and comes to its fulfillment without the least suspicion on the part of those in whom this change is being carried out…If the motory sensation were to remain always unchanged as a memory-picture, the insignificant deviations would always centre round the same point with the same point with the same maximum of distance…[T]he later impressions always have a stronger always-influence than the earlier. It is thus impossible to co-ordinate the sensation with the average of all the impressions during the whole course of life; rather, the numerically-speaking inferior may, by the fact of their freshness, outbalance the weight of the more frequent…There thus gradually arises, by adding together all the displacements (which we can hardly imagine small enough) a notable difference…"

The three approaches to speech variation are presented in the following flowchart.



When we analyze the behavior of an individual, it would seem extremely difficult to figure out which approach—relaxed constraint, exemplar, or prototype—is best at characterizing the origin of variation, and the nature of sound category formation. The category itself is largely observable, since we can, at least in theory, investigate the phonetic properties of individual tokens, and also observe whether or not the tokens we are looking at correspond to a particular meaning. We can't directly observe the meaning that is associated with a given token of course, but we can probably determine if the speech signal was interpreted by a listener with the meaning intended by the speaker. In this way at least, meaning is observable. But since we are only dealing with the behavior of the individual speaker, there is really no reliable way to tease apart the different approaches to variation and categorization. In short, all three approaches make untestable predictions about the categories and variation of individual speakers. In the following displays, the distribution of elements is exactly the same within each set, and so there is no empirical evidence favoring any one approach over the others.



It now becomes apparent that, under the prototype model, *all* variation must be regarded as mistaken and unintended. And since virtually every token deviates from the abstract prototype in some way, this means that virtually *all* speech is to a certain extent *mistaken* speech. In this sense at least, both the relaxed constraint model and the exemplar model have a distinct philosophical advantage over the prototype model. However, in accounting for the variation itself, these latter two approaches are slightly different. The relaxed constraint model allows for some flexibility in the constraints on actual speech, and so all the variation is created anew by each speaker and each generation. The exemplar model assumes a particularly tight match-up between the cognitive constraints and actual speech, since most of the variation is copied intact.

Now let's talk about the individual as both a speaker and a listener. By doing so, we are incorporating the social context in which this categorization procedure takes place: listeners are listening to other speakers, and speakers are speaking to other listeners. Therefore, we can now *compare* the speech of various speakers in order to determine the similarities and differences of their within-category variation. This comparison could, in theory, open a window into how listeners' categories are similar to or different from the categories of those they are listening to. Moreover, the *extent* of their similarity or difference might help determine which approach—the relaxed constraint model, the prototype model, or the exemplar model—is best at characterizing the sound categories and their variation. If variation is extremely well-matched from speaker to speaker and from generation to generation, this would favor the exemplar model. This is because the exemplar model predicts very little difference in the nature and extent of variation from speaker to speaker and from generation to generation. By contrast, the other approaches allow for the possibility that the nature and extent of variation may be quite different from one speaker to another, and from one generation to the next, since under these theories variation is created anew by each speaker. We turn to this issue now.

PROBABILITY MATCHING

Both speaker-to-speaker and generation-to-generation comparisons are very important to our understanding of sound categorization. These are the areas of *variation* and *diachrony*, and it is here—as opposed to the investigation of individual speakers—where we might better understand the cognitive organization of the linguistic system. To

be sure, investigating variation and diachrony are indirect routes to understanding the nature of linguistic knowledge. But therein lay their greatest advantages. By analyzing variation and diachrony, we are not pretending to access the content of psychological states, which are inherently private and so unknowable to the outside world. Instead, we are comparing structural arrays of genuine physical objects—speech tokens—and so we harbor no illusions about the object of our inquiry. Although speaker-to-speaker variation is extremely important in this regard, our main focus here, and for the remainder of the book, is on generational comparisons; how sound categories remain stable, and how they change, as language passes from generation to generation. So the question we now turn to is this: what is the nature and extent of within-category *differences* in variation from generation to generation?

In order to answer this question, let's first talk a little bit about rats and ducks. It is well documented that animals appear to perform remarkably sophisticated statistical analyses as they navigate the world around them. For example, on the face of it, an animal foraging for food in the wild appears to be randomly searching high and low for a morsel here, a morsel there. However, it turns out that this behavior is remarkably well-matched in terms of actual payoff. What I mean is, the animal actually recapitulates the likelihood of payoff in terms of its foraging behavior, spending more time in a patch of ground that has a greater payoff, and less time in a patch of ground that has lesser payoff. So, if two-thirds of the available food is in one region, and one-third is in another region, the animal very quickly comes to spend two-thirds of its foraging time in the one area, and only one-third of its foraging time in the other area. This phenomenon is known as *probability matching*.

A number of ingeniously simple experiments have been performed which show that animals indeed engage in probability matching. In one of the simplest studies, a rat placed in a T-Maze is rewarded with food 75% of the time at one end, 25% of the time at the other. When provided with feedback, the rat's foraging behavior quickly comes to match the probability of reward—running to the one end 75% of the time, the other end 25% of the time. What's especially interesting is that the rat does not maximize its payoff. If it ran to the 75%-payoff end 100% of the time, it would be rewarded with food 75% of the time. But by distributing its foraging in a way that matches the probability of payoff, it actually reduces its food intake. So, 75% of the time it searches at the location where 75% of the food is found, and 25% of the time it searches at the location where 25% of the food is. This means that that it only receives 61.5% (.75 x .75 + .25 x .25) of the total available food, as opposed to the maximum of 75%.

As counter-intuitive as this result may seem, from a long-term, evolutionary perspective, the rat's behavior makes very good sense indeed. Remember that this experiment only involves a single rat. But rats in the wild, of course, live in packs. If all rats were to forage only in the location with the greatest payoff, then fierce competition would result in a rapid depletion of resources. After these supplies run out, these rats might very well move on to the location with less food, and again compete fiercely for the rapidly diminishing resources. But consider a rat which bucks this strategy, and instead quickly matches its foraging behavior to the probability of payoff. This rat would have less competition for resources at the location of lower payoff, guaranteeing itself a steadier intake of food. So those rats which engage in probability matching are in less competition for resources than those which forage exclusively in the locations of highest yield. Due to this reduced competition, these rats are more likely to survive, and so transmit their foraging proclivities to their offspring, who, in turn, are more likely to survive. So probability matching benefits the individual, and, as a by-product, enhances

the long-term stability and survival of the population as a whole. This behavior, then, is the long-term emergent result of variable feeding strategies across individual rats.

Experimental variations on the rat-in-a-T-maze theme have been employed, yielding similar results. For example, in a somewhat less controlled experimental setting, two experimenters, standing by a pond, set apart from each other some distance, throw food to ducks at two different rates. Very quickly, the ducks are able to calculate the distinct rates of feeding, and match their time near each experimenter accordingly, spending more time at the location of greater payoff, and switching to the location of lesser payoff for a percentage of time that matches the lower yield. I should point out that these ducks' is not merely a conditioned response to a reward schedule, since they do not necessarily receive any food before matching their behavior to the probability of payoff. Rather, they are able to predict the payoff before it is received! So it's clear that animals are sensitive to the probability of reward, and quickly match their behavior accordingly.

It turns out that similar statistical calculations underlie aspects of human linguistic behavior, in that the nature and extent of variation in speech is indeed largely matched as listeners become speakers. So let's consider an example of probability matching in phonology, in particular, how probabilities come to be matched during the course of language learning. Our focus is on the word-initial stops in English which we write "b, d, g". All along, I have been transcribing these as [b̥, d̥, g̥], the hollowed circles indicating that the stop closure is mostly voiceless, with voicing beginning just around the point when the closure is released into the next vowel. But now it won't surprise you to learn that these transcription conventions fail to capture the actual variation in these sounds' production. We find token-to-token variation, and also variation depending on the location of the stop closure itself. Typically, the farther forward in the mouth the stop closure is, the more often that tokens are genuinely voiced; the farther back in the mouth the closure, the less often that tokens are voiced. (We'll go into the phonetic motivation for this variation in Chapter Six). Research on young English-learning children shows that they initially produce all their word-initial stops—whether orthographic "p,t,k" (sometimes called the "fortis" category) or orthographic "b,d,g" (sometimes called the "lenis" category)—something like [p,t,k], with neither the aspiration ([p^h,t^h,k^h]) nor minimal voicing ([b̥,d̥,g̥]) that are characteristic of the adult fortis and lenis categories, respectively. Such young children may still lack the articulatory prowess to match the patterns they hear. Through three years of age, the two stop categories for English begin to take shape, in that some word-initial fortis stops are aspirated, but still, voicing during the stop closure is extremely infrequent in the lenis series, though less so for stops made at the lips. Even up to six years of age, children's lenis category involves fewer voiced tokens than adults'. Finally, only after six years of age do learners come to largely match the nuanced variability of their elders.

As with probability matching in lower animals, such behavior betrays an extremely sophisticated statistical analytic ability on the part of language learners. Moreover, children's eventual productions betray evidence that they are able to implement their calculated probabilities in their own speech with startling, though imperfect, accuracy. While we can never know for sure, a rather straightforward account of probability matching in speech production might consist of speakers randomly choosing one out of their pool of stored tokens each time they speak a word. So token variants which they hear often are more likely to be chosen, and token variants that they hear less often are less likely to be chosen. In this way, the overall distribution of tokens will be well matched from speaker to speaker and from generation to generation.

Speaking, then, is not like playing darts at all. Even expert dart players can't hope to accurately match the variation of their opponents.

Of course, every person's linguistic experience is different from every other person's. This is even true among individuals with very similar linguistic experience such as siblings. Consequently, if variation is largely a consequence of experience, each individual's variation will be different—in some cases ever-so-slightly different—from every other person's. But within a *speech community*, such differences—by definition—are never sufficiently great to adversely affect communicative success.

"It is not a hypothesis that children do probability matching [during language learning]. It is simply a description of the observed facts". So writes William Labov in his 1994 book. The *fact* that children do probability matching during language fully supports the *hypothesis* that they also engage in exemplar modeling of variation and categorization, and casts strong doubt on both the relaxed constraint approach and the prototype approach. Neither of these approaches is properly equipped to handle the fact that variation is largely matched from generation to generation. Both of these approaches view variation as created anew by each speaker, unconstrained by the extent and nature of variation to which these speakers are exposed. They therefore predict that speaker-to-speaker and generation-to-generation variation will not be probability-matched. Finally, I should point out, again echoing Kruszewski and Paul, that the facts of exemplar-and-probability matching approach is consistent with the gradual nature of sound changes. Kruszewski's remarks, though not couched in the parlance of modern cognitive science, are perhaps all the more remarkable for exactly that reason, and warrant quoting at length:

> In the course of time, the sounds of a language undergo changes. The spontaneous changes of a sound depend on the gradual change of its articulation. We can pronounce a sound only when our memory retains an imprint of its articulation for us. If all our articulations of a given sound were reflected in this imprint in equal measure, and if the imprint represented an average of all these articulations, we, with this guidance, would always perform the articulation in question approximately the same way. But the most recent (in time) articulations, together with their fortuitous deviations, are retained by the memory far more forcefully that the earlier ones. Thus, negligible deviations acquire the capacity to grow progressively greater…

In other words, a prototype model does not readily allow for the possibility of gradual sound changes, since prototypes are presumably fixed. But allowing for both variation and probability matching, and differential sensitivity to recent versus remote tokens, the gradual nature of sound change may be accounted for quite straightforwardly.

But before wholeheartedly embracing the exemplar-and-probability-matching approach, I'd like to address a possible objection to its account of variation. Isn't it possible that the cross-generation stability of variation is not rooted in the nature of categorization, but is instead purely physiological in origin? That is, since we all have comparable speech apparatus, mightn't the similar distribution of variants across the generations simply follow as a natural physical consequence? Well, yes, this is certainly a possibility, but there are a few good reasons why we should be skeptical of this explanation.

First, if variation in speech were solely a consequence of physiological forces, then we might expect the nature and extent of variation to be nearly identical across languages with similar sound systems. For example, given two languages with similar vowel inventories, we might expect that the phonetic variation found in these languages vowels should be extremely similar. But in fact, this doesn't seem to be the case. Both the extent and the nature of variation is different from language to language, even among those whose sound inventories are otherwise quite comparable. A similar result emerges when we investigate nasalization on vowels when followed by a nasal consonant. Every language investigated has a certain amount of variable vocalic nasalization in this context, but different languages vary in different ways. The nasalization will always be there, but to different extents in different languages. So variation itself seems to be *conventionalized* on a language-specific basis. This sort of language-specific conventionalization is readily understandable under the exemplar-and-probability-matching approach, but is difficult to reconcile with a purely physiological account of speech variation.

Second, probability matching in language is found in domains that are surely not explicable in physiological terms. Some studies have shown that the optional use of certain morphemes—for example, agreement markers in certain grammatical constructions in Caribbean Spanish—is probability-matched across speakers: the rate of these morphemes' presence versus absence is conventionalized. For example, the Spanish plural marker is used on both nouns and adjectives. We may imagine the plural marker being used 95% of the time in the context where a plural meaning is intended, and so is not used 5% of the time. It turns out that this usage pattern won't significantly vary from speaker to speaker, but instead will be conventionalized throughout the speech community.

These sorts of results have also been reproduced in the speech laboratory. In one such study, subjects were taught a contrived mini-language in which nouns were optionally marked with a definite article (a morpheme meaning "the"). Subjects were divided into groups which differed in the extent to which the nouns they heard possessed this marker: one group was exposed to nouns, 75% of which had the marker, and another group was exposed to nouns, 25% of which had the marker. After sufficient exposure to the mini-language, subjects were asked to produce sentences in the taught language. Remarkably, subjects matched their usage to their exposure. That is, subjects in the 75% group produced about 75% of their nouns with the marker, and subjects in the 25% group produced about 25% of their nouns with the marker.

Since the exemplar-and-probability-matching approach offers a clear and satisfying account of conventionalized morphological variation—which cannot possibly be attributed to physiology—there would seem good motivation to propose a similar account of conventionalized phonetic variation as well. And after all, as William Labov writes in 1994, in a discussion of probability matching in language learning, "We should not be embarrassed if we find that systematic readjustments in…language are governed by the same cognitive faculty that governs the social behavior of mallard ducks…We are products of evolving history, not only our own but that of the animal kingdom as a whole, and our efforts to understand language will be informed by an understanding of this continuity with other populations of socially oriented animals."

PROBABILITY MATCHING PROMOTES CATEGORY SEPARATION AND PHONETIC STABILITY

Given the evidence for probability matching in language learning, it becomes quite understandable how phonological systems remain quite stable from generation to generation. Actually, I now want to argue that, in a seemingly paradoxical fashion, the excellent-though-imperfect matching of speech variation actually serves to *curtail* the very variation that is being matched! Let's consider how this can be so.

In Chapter Two I mentioned that front vowels are usually unrounded, like [i] and [e], and back vowels are usually rounded, like [u] and [o]. I wrote that this is probably because keeping the lips unrounded and keeping the tongue in a front position combine to create a very short oral cavity, while rounding the lips and backing the tongue combine to create a longer oral cavity. The difference in the lengths of the oral cavities corresponds to a difference in the acoustic qualities of front and back vowels. The second formant is significantly higher for front unrounded vowels, and is significantly lower for back rounded vowels. These differences, I suggested, are good from a functional standpoint, because they render the different vowel qualities less confusable with each other. Of course, there are languages that do have front rounded vowels, as our discussions of Finnish and Hungarian vowel harmony have shown us, for example. But the overwhelming tendency is that if a language has front rounded vowels, then the language has front unrounded vowels as well. So Hungarian has [y], but it also has [i]. [y]'s acoustic properties are somewhat intermediate between [i] and [u], since it involves lip-rounding (serving to lower F2) and tongue-fronting (serving to raise F2). The idea I'm getting at here is that the vowel qualities in any given language tend to be *dispersed* in terms of their acoustic qualities. The fewer the vowels, the more distinct from each other they tend to be. Consequently, as the vowel system gets more crowded, the acoustic distinctions among the vowel qualities necessarily decreases. Compare, for example, the Spanish vowel system—quite a common one in that it contains only five members—with that of many American English dialects.

Spanish:

i u
e o
ɑ

English:

i ɪ ʉ u
e ɛ ə ʌ ɔ o ʊ
æ ɑ

In a five vowel system like Spanish, the vowels are symmetrically dispersed quite widely in terms of their acoustic qualities, which for our purposes includes the first two formants (though there are many other phonetic differences as well). Since English has so many more vowel qualities than Spanish does, the vowel space is more tightly packed, but still, the vowels are symmetrically dispersed, and avail themselves of a comparable overall acoustic space. In fact, we can see that the Spanish system is merely a subset of the English system in that all the Spanish vowels have acoustically similar correlates in the English system: [i,e,ɑ,o,u] are present in both languages.

The particular subset relation isn't the only conceivable one, however. We could imagine that Spanish or another language might have one of the following subsets of the English system:

i ɪ
e ɛ
æ

or

ɑ ɔ o ʊ u

or

ɛ ə
æ ɑ ɔ

In fact, no language has anything remotely approximating these lopsided distributions. In all likelihood, it is neither by design, by intention, nor by chance that vowels systems take the dispersed forms that they do. Rather, it is most likely due to a form of *evolution*, specifically, of *natural selection*. Vowels systems take the forms they do exactly because there are *selectional pressures* to keep vowel categories dispersed, so that words are rendered distinct from one another.

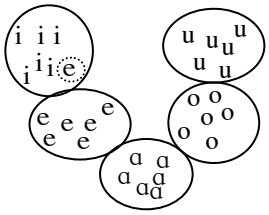
Before going even one step further, there are important aspects of the arguments I will be developing that require clarification. Specifically, when I say "natural selection", what exactly do I mean, or, more to the point, what exactly do I *not* mean? First, I expressly do not mean that vowel systems, or any other structural properties of language, are genetically transmitted from generation to generation, and as such, are subject to the genuinely evolutionary pressures which genetic mutations allow. I don't mean this *at all*. Second, I am *not* proposing that the dispersion we observe in vowel systems today historically derives from vowel systems that did *not* have this quality of dispersion. There is no reason to believe that the general characteristics of vowel systems have ever been significantly different from what they are today. Third, I am not proposing a theory of evolution that allows for goal-directed behavior on the part of the individual speaker. Indeed, the theory of natural selection does not admit this possibility. Probability matching suggests that speakers are primarily engaged with copying the speech patterns that they hear around them, and not with actively modifying their speech patterns so that one sound is rendered more distinct from another.

So what *do* I mean then? If vowel systems are not genetically endowed, and if they are not a consequence of design, intention, or chance, then what is the nature of this "natural selection" that I propose is influencing their symmetrical shape? Well, we now know that there is inherent variation in speech, such that no token is ever identical to any other token. Inevitably, tokens deviate from each other in terms of their articulatory, acoustic, and auditory properties. Nonetheless, tokens of particular vowels do cluster together—and *away* from other vowels—so that the speech signal is transmitted quite effectively to listeners, and so variation in production is well-matched from generation to generation. I said "well-matched", but not "perfectly matched"; *any* system of reproduction—genetic or otherwise—is subject to imperfect copy.

Where is the locus of this imperfection? In fact, both language perception and language production are demonstrably imperfect. Most tokens, of course, are perceived accurately, in the sense that the meaning intended by speakers is recovered by listeners, because the tokens sound remarkably similar to previous tokens of the same word. These correctly perceived tokens are also usually produced as accurate copies. However, once in a while, the production of one vowel might stray a little too close to the phonetic quality of some other vowel. For example, every once in a while, a Spanish word which usually has [e] might be made with a somewhat higher tongue position, and end up sounding like [i]. Such stray tokens are inevitable; systems of reproduction are *never* perfect.

With this in mind, let's reconsider the vowel inventory of Spanish, this time employing the "cloud of tokens" notation I introduced at the beginning of this chapter, but allowing for the presence of these stray tokens.

Vowel production:



Tokens situated well within a given cloud keep a safe distance from all the other vowel qualities, and so are sufficiently distinct from vowels in the other clouds so that misinterpretation is not a problem. In all likelihood, these will be unambiguously communicated to listeners, and quite accurately reproduced. The pooling together of these variable tokens into a single category is indicated, as always, by the circles. (Of course, learners do not come pre-equipped with these categories, these circles. Rather, they emerge from experience with pairing sound and meaning.) However, once in a while there will be tokens that should be grouped with one vowel quality, but stray into the region of another vowel quality. Look at the stray token of [e] in the dashed circle. Although the word with which this token is associated almost always has a mid front vowel, this particular token was made with a slightly higher tongue position, so that it is largely indistinguishable from words that usually have [i]. As we know, listeners are usually able to overcome any ambiguities in the speech signal, because the context—real-world or grammatical—will serve to clarify meaning. So, if listeners encounter such a stray token, chances are fairly good that they nonetheless supply the word with the meaning intended by the speaker. (More on these correctly interpreted strays in the next section.) But learners, who are still getting the hang of pairing sound with meaning, are still developing their knowledge of the real world and their knowledge of grammar. Consequently, they are less able to recover the intended meaning of these stray tokens. It's been shown that adults are also found to misinterpret these stray tokens, more often than you might imagine, in fact. By my reckoning, there are at least three different ways that learners might misinterpret this confusing token: (1) if the stray vowel quality results in another word of the language (for example, *mella* [mejɑ] "notch" is produced as *milla* [mijɑ] "mile"), they could conceivably pair the token with the wrong meaning (2) they might assign the token to more than one meaning (3) if the stray token results in a meaningless word, the token might remain uninterpreted. Each of these sorts of misinterpretation has the potential to induce confusion on the part of the listener, since the meaning intended by the speaker is not recovered by the listener. So, almost all tokens will be unambiguous, but *some* tokens will be confusing to listeners, and will remain uninterpreted or assigned to the wrong meaning.

The great Hermann Paul, by the way, would having nothing of this argument, though he places the locus of confusion on the *hearing* mechanism itself, rather than solely on the recovery of word *meaning*:

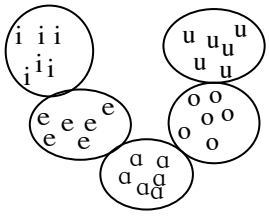
> The attempts which have been made to explain sound-change as dependent on individual caprice or on an inaccurate ear are hardly worth mentioning. A single inaccuracy cannot possibly have any lasting results for the history of language. If I do not accurately catch a word from any one who speaks the same dialect as myself, or another with which I am well acquainted, but I guess his meaning from the context of his discourse, then I supply the word in question according to the memory-picture which I have in mind. If the connexion is not sufficient to explain clearly the meaning, it may be that I shall supply a wrong meaning, or I may supply

> nothing at all, and satisfy myself with understanding nothing, or I may ask again. But how should I come to think that I have heard a word and still to set this word in the place of the one I understand, is to me incomprehensible."

Paul goes on to suspect that, perhaps, young children, or more likely, second-language learners may be susceptible to such misapprehensions, but certainly not full-fledged adults speakers of their native language. Paul's reservations aside though, let's continue with our own proposals. How do stray tokens affect the probabilities which learners come to match in their own speech? Consider the pool over which learners determine the phonetic distribution of tokens. Within-pool variants are clustered together, but stray tokens—those that fall within the phonetic space of some other value, and also, ambiguous tokens that are at the outer reaches of the cluster—might be ignored, since they may not have been categorized properly. Consequently, these confusing tokens will not be pooled with the vowel quality which is normally employed for that particular word. Since these will be not be pooled with other tokens of these vowels, this results in categories consisting of distinct pools of tokens with fairly sizeable phonetic buffer regions separating them. And since listeners can only match probabilities to their *perceptions* of speakers' productions, and not to speakers' productions directly, they might conclude that the variation in the speech signal is *not as extensive* as it actually is. That is, they *overestimate* the percentage of speakers' non-stray tokens, and match this estimate in their own speech.

So now let's consider how a learner might perceive the array of tokens that were produced by our Spanish speaker.

Vowel perception:

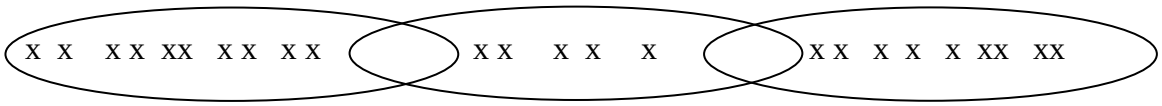


The token of the [e] word which had strayed into [i]'s territory has not been perceived as such by the listener. So, as these listeners become speakers, their productions—which largely match the distribution of variation that they perceive—also consist of pools of tokens with fairly sizeable buffer regions separating them (with, of course, new strays now and again). So, the uninterpretability of stray tokens actually serves to reinforce the distinctions among the categories themselves, driving one category farther away from others, and thus rendering the linguistic system more effective in fulfilling its communicative function. These uninterpreted tokens also serve to promote the long-term stability of the phonetic qualities of the vowel system, since usually, only tokens that are similar to the norm will be perceived, and, in turn, produced. So, under the exemplar-and-probability-matching account, production errors *create* variation in speech, but consequent perception errors *curtail* variation in speech.

Now, before going on, it's important to keep something in mind. The confusions induced by stray tokens are *not* comparable to the ambiguities of standard neutralizations. Neutralized forms are part of a regular linguistic pattern, and so listeners encounter them all the time, and so rather easily come to master their distribution. By contrast, stray

tokens are not regular or patterned in their occurrence; they are genuinely aberrant, and so listeners are not equipped to deal with them in a comparable way.

Let's revisit the figure I provided at the beginning of this chapter. We can see now that the array of tokens in that display was not a realistic one, in that the tokens were dispersed very uniformly across the acoustic space. But we now know that, due to the uninterpretability of strays, tokens in the border regions may very well be eliminated, and probability matching will maintain the separation of categories. The following revised display reflects this more realistic distribution of tokens.



Of course, we will inevitably find a few strays located in the border regions, but still, the distribution of tokens across the acoustic space is probably far less regular than our first figure indicated, with most tokens falling into well-separated pools.

To summarize, in general, speakers do a remarkable job of matching the variability that is present in the speech signal, and listeners do a remarkable job of perceiving this variability. However, the system isn't perfect: there are both stray tokens and consequent perception errors that influence the categorization procedure. The passive filtering out of these strays enhances the phonetic distinction among tokens belonging to different sound categories. The result is that vowel systems tend to avail themselves quite well of the phonetic space, dispersing their members into well-defined, well-separated regions. So the dispersion of vowel qualities in the phonetic space, and the buffer regions between them, may be seen as the natural, passive consequence of the miscommunication of stray tokens. The idea then, is that our excellent-though-imperfect ability to engage in probability matching both *causes* and *inhibits* variation in speech. And the phonetic separation and stabilization of categories is best viewed as a *consequence*—as much as it is a *cause*—of effective communication.

We can now see how imperfect copy may lead to the symmetrical distributions that we observe time and time again in vowel systems. Contrary to the assertion of some linguists (including, it must be said, the great Andre Martinet), I don't think this derived symmetry should be viewed as some sort of cognitive pressure in the minds of individual language users that favors the symmetrical distribution of elements. I don't think the symmetry of the system is relevant at any psychological level by language users; it's only appreciated by linguists. Indeed, these processes are extremely slow-acting and so cannot be attributed to individual speakers—speakers are excellent in mimicking what they hear, and so changes are very gradual. Rather, the symmetry evolves passively, as a function of language use within a community of speakers.

This sort of system may be viewed as both self-organizing, and self-sustaining. It is self-organizing because its structural properties are a consequence of its use, requiring no outside monitor, guide, or force, to affect its organization. It is self-sustaining because, by its very use, it repairs and maintains itself. So once again, language form is inseparably intertwined with language use and language function.

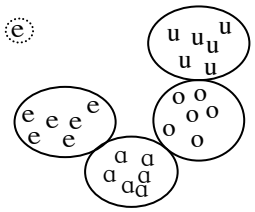
PROBABILITY MATCHING PROMOTES CATEGORY SEPARATION AND PHONETIC *CHANGE*

We've just seen how imperfect copy might contribute to the phonetic separation and phonetic stability of sound categories. However, sounds *do* change, and these

changes are embodied in the slightly different distribution of tokens which are observable as the generations proceed. We'll now consider a rather different effect: imperfect copy might lead not to stabilization, but to an *increased* separation of sound categories. This mechanism, in fact, is already built into the system as we have characterized it. Since tokens of one category which are more distinct from tokens of other categories are more likely to be perceived correctly, then sound categories may drift farther apart over the generations, but only provided that this drift does not come to encroach on the phonetic character of yet *another* category.

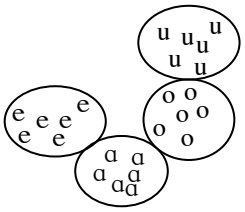
Given the enormously complex interaction of forces that come into play in phonological systems, asymmetric subsystems are bound to develop, at least temporarily. So let's imagine a situation in which contrastive categories, for one reason or another, are not fully dispersed in the perceptual space. Under these circumstances, one category may increase its phonetic distance from another, and no third category is present to provide a limiting counterforce. For example, we might imagine a hypothetical language like Spanish, except that it lacks a high front vowel (an admittedly unlikely system).

A wildly stray [i]-like token of a word that usually possesses [e] may well induce confusion on the part of listeners, since it is so different from the vowel qualities that they are used to.



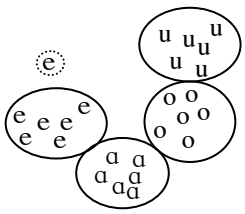
Vowel production

Such tokens will probably be thrown away—filtered out—regarded as mere speech errors. If noticed, they might be laughed at by both speaker and listener.



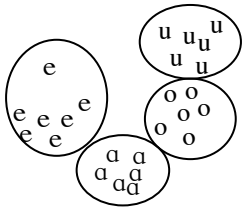
Vowel perception

However, a token of an [e]-word that is only marginally [i]-like might not induce confusion at all, but on the contrary, might be better at communicating the intended message to listeners, since this token is actually further dispersed from the other vowels of the system ([ɑ,o,u]), though not outlandishly distinct from other [e]s.



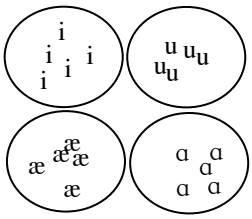
Vowel production

In this language, since such tokens marginally drift farther and farther away from other categories are *not* encroaching on a third category, then it's these tokens that are most effective in conveying the meaning intended by speakers.



Vowel perception

Over time, the whole pool of tokens may gradually drift farther and farther away from the other categories, and a more symmetrical four-vowel system might emerge; other values may now spread out to exploit the entirety of the available acoustic space. The result is that the system will evolve towards a symmetrical dispersion of its categories.



Newly evolved system

So let's move away from the hypothetical realm, and consider a real-world example of this sort of dispersion effect.

TRIQUE

Trique (pronounced [ˈtɻikeˌɪ], though the Spanish spelling Triqui is often seen, pronounced [ˈtɾiki]) is a language spoken in the southern regions of Mexico. It is a member of the Otomanguean language family. In Trique, whenever a round vowel precedes a tongue-body consonant ([k,g]), it is immediately followed by [w]. So look at the following examples; I've underlined the relevant sounds ([ʒ] as in azure).

[nukwɑh]	strong	[dukwɑ]	possessed house
[dugwɑh]	to twist	[zugwi]	(name)
[ʒugwɑ]	to be twisted	[dugwe]	to weep
[dugwɑne]	to bathe (someone)	[rugwi]	peach
[rugwɑh]	hearth stones	[dugwi]	together with, companion

You'll notice that when we have [u] followed by either [k] or [g], there is [w] immediately following: the language never has sequences such as [ukɑ] or [ugɑ]. You can readily see how the [w] is merely a continuation of the [u], such that the lip-rounding gesture is realized both immediately before and immediately after the consonant. The pattern, then, can be conceived of as a minor form of vowel harmony.

It wasn't always this way, however. Both the comparative and the internal reconstruction methods converge on the same conclusion about the history of Trique. At an earlier stage of the language, the round vowel was *not* realized on both sides of the tongue-body consonants. Instead, there was *ukɑ and *ugɑ, the patterns that are completely absent today. (Reconstructed forms are traditionally indicated with an asterisk preceding them.) But at some point in the language's history, round vowels began to "unhinge" from their position, and continue across [k] and [g]—the tongue body consonants—eventually turning these into [kw] and [gw], respectively. Employing the comparative method, when we investigate other languages that are closely related to Trique we do *not* observe the "unhinging" of round vowels. Related languages usually have [ukɑ] and [ugɑ] in words for which Trique has [ukwɑ] and [ugwɑ]. And since less common patterns among closely related languages often reflect more recent changes, Trique was probably innovative in this sense. Evidence from the internal reconstruction method comes from the simple fact that these [w]s are completely predictable in their presence such that we can "undo" their present distribution and recover an earlier stage of the sound pattern.

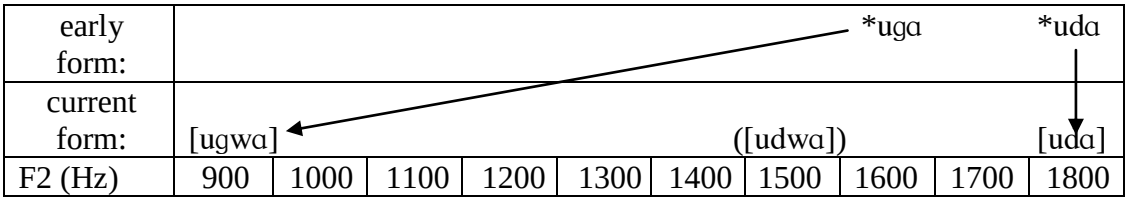
But now look at the next set of words. Here—when the consonant which follows the round vowel is made with the tongue tip—the [w] is not present. In fact, it's *never* present here. So, we never find [utwɑ] or [udwɑ], for example.

[rune]	large black beans	[utɑh]	to anoint
[utʃe]	to get wet	[utʃi]	to nurse
[utɑ]	to gather	[dunɑ]	to leave something
[gunɑh]	to run	[rudɑʔɑ]	stone rolling pin
[ʒutʃe]	hens, domestic fowl	[gunɨ]	to hear

The question that a phonologist must now ask is, why did the Trique pattern arise? Why did [u] harmonize across [k,g], but not across [t,d]? The answer I'd like to pursue is that this minor form of vowel harmony *enhances* the acoustic distinction between the tongue-body and tongue-tip consonants. Recall that tongue-tip consonants like [d] bring F2 toward about 1800Hz as the closure is being released. By contrast, when releasing a tongue-body consonant like [g] into a vowel, F2 begins at about 1600Hz. (We'll just be discussing [d] and [g] from here on out, but all arguments apply to [t] and [k] as well.) This means that the difference in F2 between, say, [dɑ] and [gɑ] is about 200Hz at consonantal release. (There are several other acoustic differences between these two sounds, and so they are not terribly likely to be confused with one another.) Now, if the tongue-body consonant is altered such that a [w] is superimposed onto its release, the oral cavity becomes longer, and so F2 lowers. In fact, F2 lowers rather significantly, to about 900Hz. This means that the difference in F2 between [dɑ] and [gwɑ] is about 900Hz (1800Hz minus 900Hz). Clearly, the superimposition of the [w] increases the acoustic distance between the release quality of tongue-tip and tongue-body consonants. Importantly, since [gw] and [kw] sequences were elsewhere absent in the earlier stage of the language, harmonizing lip-rounding across [g] increased the acoustic distinction between these consonants and the tongue-tip consonants, without encroaching on another sound category. So [ugɑ] could become [ugwɑ], and there were no other words in the language like [ugwɑ], and so there was no functional counter-pressure acting to inhibit the sound change.

Harmonizing across the tongue-tip consonants, by contrast, would serve to diminish the tongue body – tongue tip acoustic distinction. Why is this so?

Superimposing a [w] onto the release of a tongue-tip consonant would change the F2 onset from about 1800Hz to about 1500Hz, decreasing the difference in F2s to a mere 100Hz (1600Hz minus 1500Hz). So, an accompanying change from [udɑ]-to-[udwɑ] would have undone the functional benefits of the [ugɑ]-to-[ugwɑ] change.



In Trique, the diachronic harmonizing of lip-rounding onto the release of the tongue body consonants ([ugwɑ]) increased their F2 distinctions with the tongue tip consonants ([udɑ]), and didn't encroach on the perceptual space of another category. Harmonizing lip-rounding onto the release of the tongue tip consonants ([udwɑ]) would have had counter-functional consequences.

Of course, this [w] didn't just pop out of the ether in order to help increase the acoustic distinction between [udɑ] and [ugɑ]. So, for example, we wouldn't expect an [s] or an [m] to arise in order to enhance the perceptual distinctness in the [ugɑ] context. Instead, these sorts of changes exploit the sounds that are already loitering in the neighborhood, so to speak, their properties harnessed, co-opted, or, in the parlance of Stephen Jay Gould and modern evolutionary biologists, *exapted* to fulfill new functional roles: [gɑ] and [dɑ] are not especially confusable with each other, but since [u] was right next door, and since its harmonizing across [g] served only to increase the acoustic separation of the elements without jeopardizing another contrast, there was nothing to inhibit the beneficial change. So [u] served a contrastive role on its own, and was passively recruited to assist in distinguishing another, neighboring sound. Over time, the number of [ugwɑ] variants was likely to increase, since these forms increased the acoustic distance from [udɑ], and so were more likely communicated correctly to listeners. Meanwhile, [udɑ] remained largely stable over time: [udwɑ]-like variants were confusable with both [uga] and the increasing number of [ugwɑ] tokens, and so were not likely to take hold. The proposal then, is that due to the acoustic and consequent functional advantages of harmonizing lip-rounding across the tongue-body consonants, and the disadvantages of harmonizing across tongue-tip consonants, the Trique system evolved towards its present state.

Consider how the exemplar-and-probability-matching approach may account for sound changes like this. There is inherent gradience and variation in speech production, and so [ugɑ...ug̭ɑ...ugwɑ], and [udɑ...uḓɑ...udwɑ] are among the possible variants that any speaker might produce. (The subscripted hook indicates partial rounding). In the earlier stages of the language, productions and subsequent probabilities leaned heavily toward [ugɑ] and [udɑ], just as they still do in the languages related to Trique. However, stray [ugwɑ]-like variants rendered these words more distinct from their [udɑ] counterparts. This is especially true since words with [ugwɑ] were *not* previously present in the language. Consequently, there was no counterforce inhibiting a change toward [ugwɑ]. Therefore, those variants with [w] were more likely communicated unambiguously to listeners. Ambiguous tokens were sometimes confusing to listeners. Specifically, [udwɑ]-like variants of words that usually had [udɑ] may be confused with [ugwɑ], and so weren't added to the pool of tokens over which probabilities were calculated. They were

"repelled" due to the presence of [ugwɑ] forms. Consequently, as the generations proceeded, listeners were more likely to perceive [ugwɑ] and [udɑ] as unambiguously belonging to different categories, and so they were more likely to produce [ugwɑ] and [udɑ] in their own speech, as a consequence of probability matching.

So, the variation engaged in by elders was largely matched by learners, but nonetheless, due to the greater likelihood of unambiguous perception of certain variants over others—[ugwɑ] over [ugɑ]; [udɑ] over [udwɑ])—learners' calculated probabilities may have differed slightly from their elders', in that the variants which were more dispersed from the opposing value were more often perceived correctly, and so, in turn, more often produced. In essence, the presence of ambiguous tokens may result in listeners *overestimating* the prevalence of more distinct tokens. This overestimation, in turn, may result in more distinct tokens being produced, and, eventually, the better separation of phonological categories.

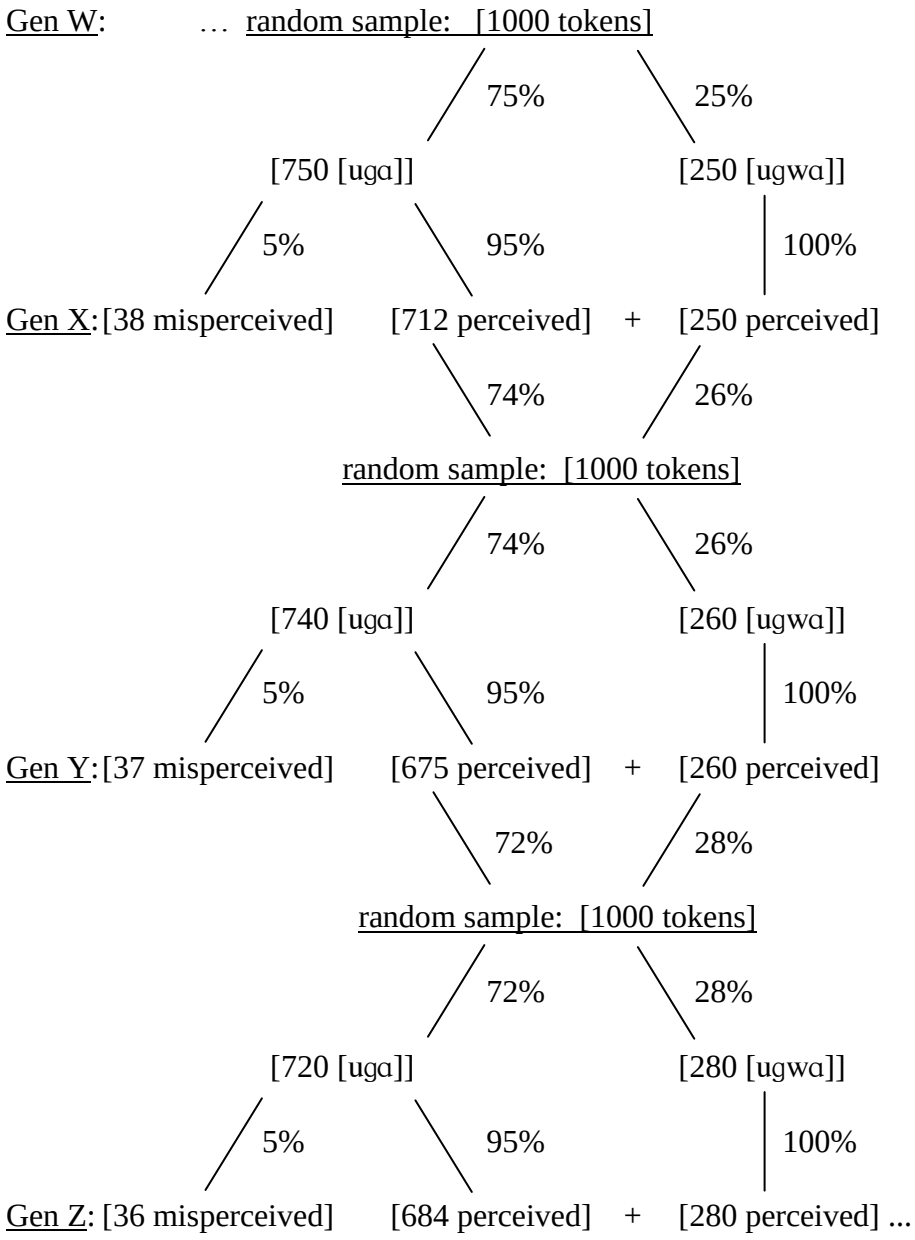
These proposals are summarized in the following table, which demonstrates how very minor phonetic tendencies, coupled with the confusion they might induce or eschew, may eventually have far-reaching consequences for the sound system.

* ugɑ...ugɑ...ugwɑ		* udɑ...uḍɑ...udwɑ	
↙	↘	↙	↘
less distinct from [udɑ]	more distinct from [udɑ]	more distinct from [ugɑ]	less distinct from [ugɑ]
↓	↓	↓	↓
less likely perceived unambiguously	more likely perceived unambiguously	more likely perceived unambiguously	less likely perceived unambiguously
↓	↓	↓	↓
less likely produced	more likely produced	more likely produced	less likely produced
∴ gradual move towards [ugwɑ]		∴ stability of [udɑ]	

*The fates of *uga and *uda*

Let's consider this in a bit more detail. Entering the sound change midstream, we might take a 1000 token sample from one generation of speakers. Let's call them Generation W. Of these tokens, 750 are [ugɑ], while 250 are [ugwɑ]. Most of these tokens are produced as a consequence of learners' matching their probability of occurrence to the productions of Generation V. In turn, Generation X perceives *all* [ugwɑ] tokens unambiguously. Among [ugɑ] tokens however, let's suppose that a full 5% of these 750 tokens (38 in all) are confusing to listeners, since their acoustic separation from [uda] is not as sharp. These 38 misperceived tokens will not be pooled with those over which Gen X-ers calculate their probabilities. Now we iterate the process: if we take a random sample of 1000 of Generation X's *productions*, we should observe that they largely match the probabilities that they *perceive* their elders to have produced. Generation X perceived 712 out of 962 tokens as [ugɑ] (38 tokens were misperceived); this constitutes a rate of 74%. So, out of 1000 tokens produced by Generation X, 740 will be [ugɑ], and 260 will be [ugwɑ]. And again assuming that 5% of the [ugɑ] tokens will be misperceived by Generation *Y*, *these* children will perceive only 72% tokens as [ugɑ] (675 of 935 tokens), and so on down the generations. We may now see, given the small tendency to

better perceive [ugwɑ] tokens, how, over the course of time, the conventions of the language may change.



Schematic diachrony of [ugɑ]-to-[ugwɑ]

A model like this does not perfectly or exhaustively predict specific language patterns. As already noted, we can no better predict the future direction of a sound than we can the future direction of a species. Indeed, one of the best advantages of this account is that it effectively captures the *probabilistic* nature of sound change. Trique's relations did not undergo the sound change that Trique did. There simply exists a probability that any given sound change will take hold in any given language.

These sorts of proposals for the origin and development of sound changes may actually be reproduced in a laboratory setting. A laboratory condition may serve to recapitulate elements of the hypothesized historical scenario in "sped-up" form if we find a way of inducing a high rate of perception errors on the part of listeners. How might we

do this? I had subjects listen to [udɑ], [udwɑ], [ugɑ], [ugwɑ] in various levels of "white noise" (computer-generated noise across a broad frequency range which decreases the signal-to-noise ratio, making the signal harder to decipher). Noise introduced into the speech signal might induce a "sped-up" rate of misperception in certain contexts, and thus reflect one origin of real-world sound change. I found, indeed, that listeners were far more likely to hear [udɑ] as [ugɑ] than they were [udɑ] as [ugwɑ]. Among the four forms presented, these latter two forms ([udɑ] and [ugwɑ]) were the least often confused with each other.

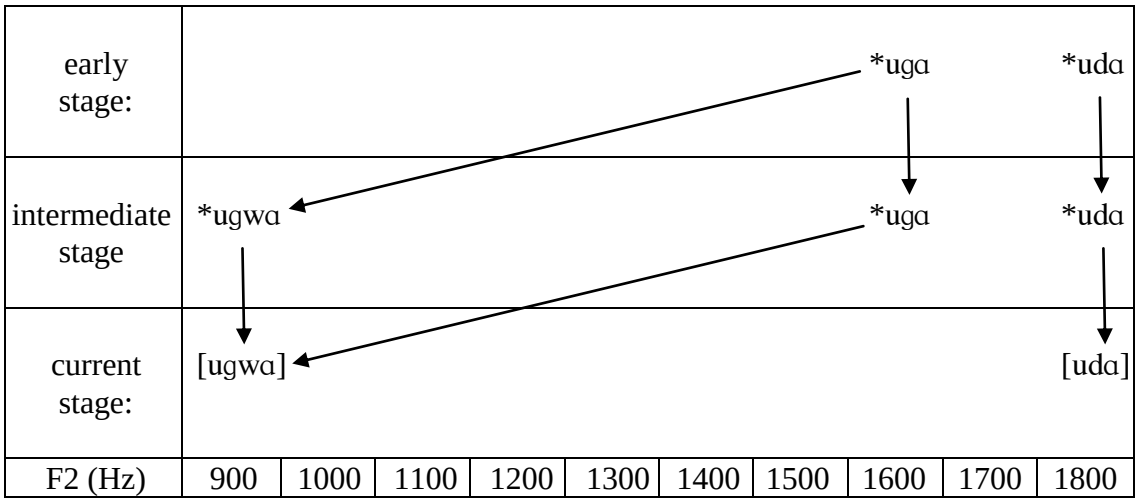
Now, this sort of result doesn't immediately translate into a real-world context that unfolds over generations of speakers, but nonetheless, it is probably more than mere coincidence that in my experiment, the least confusable forms ([udɑ] and [ugwɑ]) are exactly those which actually seem to have evolved in Trique from more confusable forms ([udɑ] and [ugɑ]). So, given that language learners largely (though imperfectly) match the variation they perceive, the sorts of perceptual errors induced in my experiment might only reflect the culmination of a slow, generation-to-generation accretion of such errors, rather than offering any major insights into the online processing of natural speech. Nonetheless, the results are consistent with the hypothesis that the gradience and variation inherent in speech production may be the fodder for these sorts of sounds changes: the more distinct the variant from an acoustically similar word, the more likely that it will be interpreted correctly, and so the more likely the system will wend towards this value.

The past is not only reflected in extant phonological alternations; the same sorts of forces that gave rise to the present state of the system may, at least in theory, be brought to the fore under the proper laboratory conditions, not necessarily by modifying the natural speech *signal*, but instead modifying the *noise* that accompanies this signal. As remarked by Baudouin de Courtenay in 1910 (p.267), "I must emphasize the importance of errors in hearing (*lapsis auris*) when one word is mistaken for another, as a factor of change at any given moment of linguistic intercourse and in the history of language as a social phenomenon. Experimental methods can help to define the types and directions of these errors…".

Before concluding our discussion of Trique, there is an important point to consider. In both my discussion of the actual Trique system, and in the experiment I performed, I have been operating under the assumption that [ugɑ] and [udɑ] constituted the critical distinction between the words that drove the sound change. However, it's simply not the case that a huge inventory of Trique words were originally differentiated *solely* in terms of whether they had [udɑ] or [ugɑ]. Usually, words with these sequences had additional elements that rendered them distinct, such as the presence of word-initial consonants (for example [utɑh] "to gather" versus [nukwɑh] "strong", which have the voiceless stop counterparts), and/or different tones (I haven't been indicating tones, but Trique is a tonal language). Moreover, if some words were indeed solely differentiated by [udɑ] versus [ugɑ], couldn't Trique have evolved the [ugwɑ] pattern only in those specific cases in which homophony might otherwise be the result?

Indeed, it may be that [ugwɑ] in Trique first arose in those very [ugɑ] words that were minimally distinct from [udɑ] words, that is, in those words that were identical except for their [g] versus [d]. But these few pioneering [ugɑ] words that evolved toward [ugwɑ] may have opened the floodgates of change: as *some* words were now implemented with [ugwɑ], more and more words may have quickly fallen in line with the emerging pattern. Why might this have happened? Due to the pioneering [ugwɑ] words,

the language now possessed three relevant patterns: [udɑ], [ugɑ], and [ugwɑ]. Of these three patterns, [udɑ] and [ugɑ] are phonetically much more similar to each other than either is to [ugwɑ]. At this point, when an [ugɑ] word was now heard, it was more likely to be confused with [udɑ] than with the newly-developed [ugwɑ] words; since [ugwa] has now entered the language, new stray [ugwa] variants were more likely to be recognized, and so were communicated more effectively to listeners. We might even say that the new presence of new [ugwɑ] words *attracted* [uga] words toward them. In sum, functional pressures may have induced the change to [ugwa] in some words. But now that [ugwɑ] was present in the system, it was far more likely that additional [uga] words would fall in line with the new pattern, since [ugɑ] is far more confusable with [udɑ] than it is with [ugwɑ].



*The proposed Trique change: the first words to become *ugwa may have been *uga words that minimally contrasted with *uda words. With these *ugwa forms now in place, the pattern was more likely to generalize, changing* all **uga words to [ugwa]. Meanwhile, *uda remained stable.*

We actually observe this sort of scenario time and time again in phonology. The linguist Joan Bybee has demonstrated that sound changes often begin in a word here, a word there, but eventually come to permeate the language. Inspired by the proposals of nineteenth century scholar Hugo Schuchardt, Bybee observes that it is the most frequently used words that might change first (more on this in Chapter Seven), but what I'm suggesting here is that passive pressures toward homophone avoidance may also trigger individual words to undergo pioneering changes.

COMALTEPEC CHINANTEC

Like a classical Darwinian approach to evolution, I've just suggested that the origin of lip-rounding harmony in Trique is rooted in two related phenomena. First, random, minor inexactitudes of speech production slowly amass over generations of speakers, such that one generation's inexactitudes serve as the next generation's template for copy. The result is that variation is largely—though imperfectly—matched over generations of speakers. Second, beneficial variants are more likely to be perceived correctly by listeners, and so it's these variants which are more likely to survive and propagate as listeners become speakers. These beneficial phonetic variants may come to be generalized throughout the language.

In Trique, change was initiated by purely random, directionless, *isotropic* chance, sine variation potentially proceeds in a radially symmetrical fashion. Variation may have proceeded in any direction, but s*ome* tokens just happen to have better functional success over others, and so the sound change moved in that direction. Maintaining rounded lips through a tongue body consonant is no more phonetically natural than *not* maintaining this lip rounding. Rather, it is due to the functional advantages of lip-rounding harmony that the Trique sound system began its new trajectory.

In this section, I'd like to consider a slight variation on the Trique theme. Some sound changes, although also subject to the sorts of functional pressures discussed for the Trique pattern, are actually "helped along" by certain natural phonetic tendencies. What I mean is, certain variants may be more likely than others due to purely phonetic pressures. And if these variants are *functionally* beneficial as well, then a sound change is more likely to be channeled in that direction. The variation which leads to sound change in this scenario is not *isotropic*, but is instead *anisotropic.*

Chinantec, like Trique, is a member of the Otomanguean language group. The dialect we are interested in is spoken in the beautiful mountainside village of Santiago Comaltepec ([koˌmɑlteˈpɛk], a four hour bus ride north of the city of Oaxaca, Mexico. The Comaltepec dialect of Chinantec, like all Otomanguean languages, is tonal. Comaltepec Chinantec words may have a low tone (L), mid tone (M), high tone (H), low-to-mid tone (LM), or low-to-high tone (LH), along with a few allophonic alternants which we'll discuss momentarily. I'll now try once again to get you to love—and not fear—tones. It might help to use the music scale as a guide. Let's translate these five tones into a do-re-mi notation: L=do, M=re, H=mi, LM=do-re, and LH=do-mi. For the contour tones, don't just sing one note followed by the next. Instead, glide your pitch from the first note to the second. It might help if we use more iconic symbols to represent the tone values. Simply hum along to the non-vertical line of the following symbols: ⌋=L, ┤=M, ␦=H, ⟋=LM, ⟋=LH. These represent relative pitch values, whereas the vertical line itself represents the entirety of the pitch range, from L at the bottom, to H at the top. Now, to complete the picture, let's attach these five tonal melodies to some consonants and vowels, say, "la": la⌋, la┤, la␦, la⟋, la⟋.

In Comaltepec Chinantec, there is a rather complicated tone substitution pattern, aspects of which we'll be considering now. First, when a LH word precedes a word that otherwise has a L-tone, then HL (⟍) is found instead of L.

toː⌋	banana	kwɑ⟋ toː⟍	give a banana
ŋɨh⌋	chayote	kwɑ⟋ ŋɨh⟍	give a chayote

Second, when a LH word precedes a word that otherwise has M, then HM (□) is found instead of M.

kuː┤	money	kwɑ⟋ kuː┐	give money
dʒuː┤	jug	kwɑ⟋ dʒuː┐	give a jug

Finally, when a LH word follows another LH word, it changes to MH (⟌).

ʔŋɑ⟋	forest	heːh⟋ ʔŋɑ⟌	in the forest
bʌʔ⟋	ball	kuɑ⟋ bʌʔ⟌	give the ball!

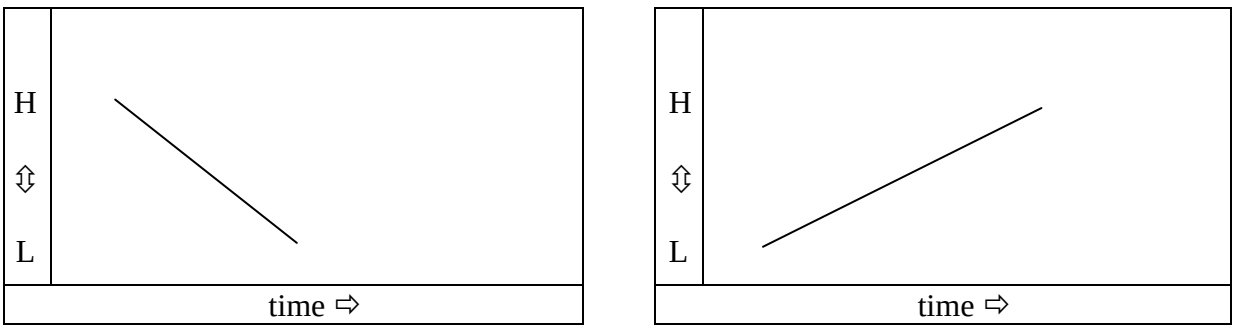
Now, there are two interesting generalizations we can make about these sound substitutions, one generalization about their phonetic character, and one about their functional character. Phonetically, we can characterize the sound substitution in much the same way we did the lip-rounding harmony process of Trique. Specifically, the H component of the HL and HM contour tones may be viewed as being a mere extension of the preceding H tone, moving across the intervening consonant, and continuing into the first part of the next vowel. So the substitution of HL for L, and HM for M, is a consequence of the preceding H tone being implemented both before and after the intervening consonant. The MH tone may be viewed in similar terms, the preceding H tone serving to at least partially raise the first portion of the following LH-tone.

The second interesting generalization is a functional one. Recall that I've listed five tone values for Comaltepec Chinantec—L, M, H, LM, LH. these five tones may occur on words that do *not* follow words with LH tones. However, we've just discussed three more tones that may *only* occur on words that follow LH tones: HL, HM, and MH (H and LM may occur here as well, but L, M, and LH do not). In other words, L allophonically alternates with HL, M allophonically alternates with HM, and LH allophonically alternates with MH; this is a non-neutralizing sound substitution.

L	M	H	LM	LH	HL	HM	MH	
[toː˩]					[toːˇ˩]			banana
	[kuː˧]					[kuːˇ˧]		money
		[li˥]						flower
			[kiˊ]					garbage
				[bʌʔˊ]			[bʌʔˇ]	ball

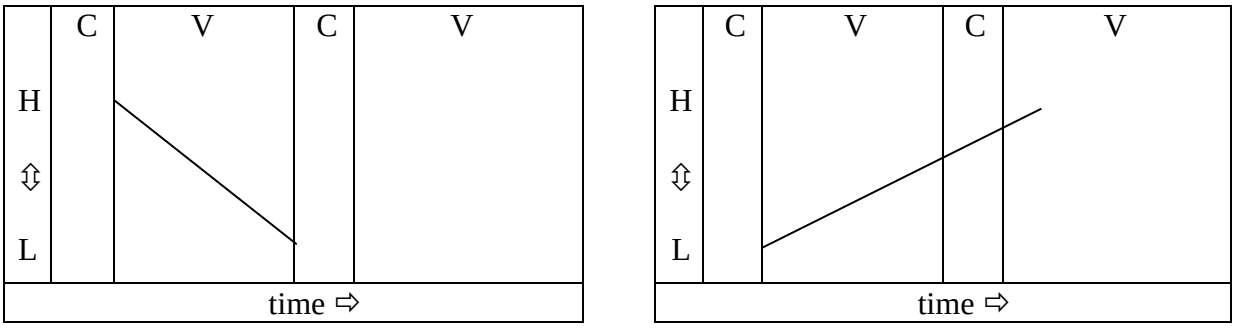
The allophonic nature of Comaltepec Chinantec tone substitution

Given these two generalizations, we're now in a position to understand the origins—the *explanation*—for this aspect of the Comaltepec Chinantec sound system. The first point to consider is a phonetic one. It's been shown experimentally that pitch rises take longer to implement than do pitch falls.



Pitch rises take longer to implement than do pitch falls

Given the sluggishness of pitch rises in comparison to pitch falls, a consonant may already be made *before* a pitch rise is fully achieved: upon the release of this subsequent consonant, finally, maximum pitch height is achieved on the next vowel. The idea then, is that rising tones are more likely than falling tones to spill their high component on to a following vowel. Since falling tones can be produced faster than rising tones, they might be less likely to spill over onto the next vowel. In the following figure I've superimposed consonants ("C") and vowels ("V") on the pitch patterns. The potential for H "spill-over" should be clear.

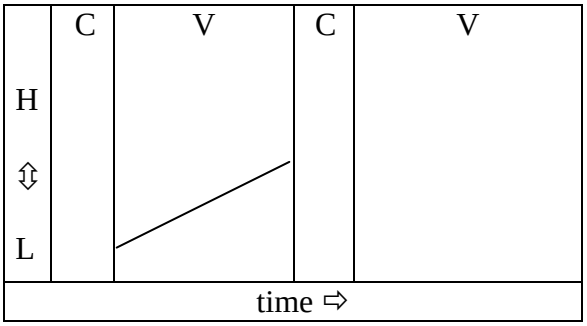


High tone "spill-over" from rising tones

Comaltepec Chinantec has conventionalized this phonetic tendency is a way that fairly hugs the physical limitations of the speech apparatus. The H component of LH contour tones is implemented both at the end of the first vowel, and into the beginning of the second vowel.

But just because speakers' physiological limits might be encountered in an experimental context doesn't mean that these limits will play a role in natural linguistic contexts. Indeed, only if it can be shown that speech patterns *exactly match* experimentally-determined physical limitations can we establish a direct link between phonetic limitations and phonological patterning. In fact, as far as I know, an *exact* match between physiological constraints and linguistic conventions has *never* been established in linguistic research. For example, it's been found that women can increase their rate of pitch rise more quickly than can men, but no language is sensitive to such sex-based differences. Nonetheless, physiological constraints might come to constrain phonological patterning at a historic distance. That is, the conventions of sound systems might not push the absolute limits of physiology, but might nonetheless come to be paleophonetically shaped by them.

This is where functional forces on the system become relevant, which may, over generations of speakers, crucially interact with phonetic pressures. If H tones did not spill over in Comaltepec Chinantec, then they might be misperceived by the listener as belonging to the LM tone category, due to the only limited temporal domain in which the pitch rise is implemented. The pitch rise may be cut off as the second consonant is beginning, and so does not achieve nearly as high a value. With its lower ending point, the tone might be confusable with LM-tones.



Early cut-off of a low-high tone may be confusable with a low-mid tone

As greater pitch increases can be effected *after* the following consonant, and since there is a natural tendency for pitch rises to "spill over" anyway, the LH-tone may be better cued when it spills over. Also, since HL, HM, and MH are elsewhere absent in the language, high-tone spill-over better conveys the high tone value without the possibility of neutralization. So, variant forms in which the H component of LH forms spills over on to a following vowel is functionally advantageous in two ways: (1) The LH-tone is now less likely to be confusable with LM-tones (and maybe L-tones too), and (2) the spilled-over component can never neutralize distinctions, since HL, HM, and MH are purely allophonic sound substitutions.

So, I am suggesting that H-tone spill-over has its origins in anisotropic variation: there is an intrinsic phonetic influence on high tone spill-over in Comaltepec Chinantec. And exactly because the high tone spill-over has functional value—meaning that distinct words are more readily conveyed to listeners—this tone value has been conventionalized in its present form.

There are complications, however. M tones on syllables which lack post-vocalic [h] or [ʔ] induce the same tone substitutions on following syllables as LH tones do.

hi˩	book	miː˧ hi˥˩	I ask for a book
mohʔ˩	squash	miː˧ mohʔ˥˩	I ask for squash
kuː˧	money	miː˧ kuː˦	I ask for money
ʔoː˧	papaya	miː˧ ʔoː˦	I ask for papaya
ŋi˩˥	salt	miː˧ ŋi˧˥	I ask for salt
loh˩˥	cactus	miː˧ loh˧˥	I ask for a cactus

Also, these allophonic MH and HM tones may themselves induce tone changes on following syllables! Since they induce the tone change when they are LH and M, they also do so even when they have been substituted by MH and HM, respectively. That is, the tone substitution pattern may iterate itself across a string of words. In theory then, a string of words may possess nice smooth M-tones in most contexts, but suddenly encounter a very bumpy road if they end up next to each other: [˧], [˧], [˧], [˧], but [˧ ˦ ˦ ˦]!

This is really unusual, to say the least! Why should a H-tone suddenly pop up after an M-tone? Clearly, the proposed explanation cannot offer an immediate account of this pattern. Indeed, I'm really at a loss to offer any sort of proximate, phonetically-rooted account here. But of course, proximate phonetic forces are only of limited help

when trying to fully understand synchronic phonological patterns. Instead the answer may be recoverable from an investigation of language history, where phonetic factors interact with cognitive and functional factors.

As in Trique, both the comparative and internal reconstruction methods converge to offer a compelling account of the present-day pattern. Forgoing the details, the linguist Calvin Rensch has reconstructed an earlier stage of Chinantec, and suggests that the M-tones which induce a H-tone on following syllables are historically derived from H-tones.

Present-day Comaltepec Chinantec:	Historic Chinantec:	
kuːꜜ	*kuːꜛ	money
ⁿdʒœːꜜ	*dʒuːꜛ	earthen jar/jug
ʔwɨːŋꜜ	*ʔwɨːꜛ	Ojitlán (a large Chinantec village)

Historic H-tones which were immediately followed by [h] or [ʔ] have remained H, according to Rensch.

Present-day Comaltepec Chinantec:	Historic Chinantec:	
lihꜛ	*lihꜛ	flower
huːhꜛ	*huːhꜛ	word
huːhʔꜛ	*huːʔꜛ	pineapple

With this in mind, it's not too hard to reconstruct the origins of this superficially strange pattern. We have established that the high component of LH-tones naturally spills onto a following vowel. Over time, this pattern may have *generalized* to include other tones which ended on a high pitch. In particular, H-tones on vowels that were not immediately followed by a laryngeal sound ([h] or [ʔ]) were the most likely tones to be recruited into the pattern.

Why should this be so? Glottal opening ([h]) and glottal closing ([ʔ]) typically make demands on the vocal folds that are in conflict with tone production. If a laryngeal sound immediately follows the H-tone, the vocal folds are not likely to maintain the posture necessary for the production of this tone. *Without* a following laryngeal sound however, a pitch may be prolonged into the following vowel without interference. So, *level* H tones also came to be associated with the appearance of a H-tone on the first portion of a following vowel, provided no [h] or [ʔ] immediately followed. But then, these H-tones lowered to M, and so all these historic H-tone forms disappeared from the language. There are reasonable aerodynamic reasons for this sort of pitch differential, since both shutting down and opening up the glottis may both—for rather different reasons—be accompanied by pitch raising: [h] involves increased airflow, which might raise pitch, and [ʔ] involves tensed vocal folds, which also might raise pitch. (Both these patterns are found in other languages, by the way.) In Comaltepec Chinantec, these historic H-tones are lower today, but the following H tone remains, as a relic, or vestige, of the past pattern. So whenever one of these former H-tones (which had now become M) came up against a following L, M, or LH tone, these following tones were still substituted with HL, HM, and MH, respectively: the preceding H-tone has lowered to M, but the substitution pattern on the following vowel remains solidly in place.

As a result of all this, former H-tone words which *lacked* a post-vocalic laryngeal are now M-tone words, and they induce the presence of an H tone on a following vowel. But he H-tones which *possessed* a post-vocalic laryngeal remained H, yet these do *not* spill their H on to the following vowel! The following timeline summarizes the proposed sequence of changes.

What: H spreads rightward from LH syllables.	What: H spreads rightward from H-final vowels which lack postvocalic [h] or [ʔ].	What: Level H *without* postvocalic [h] or [ʔ] lowers to M; the H tone on the following vowel remains.	What: Level H *with* postvocalic [h] or [ʔ] remain H; there is no H tone spread
Why: Functionally beneficial anisotropic variation leads to the conventionalization of H-tone spill-over in this context.	Why: The pattern is generalized to include those H-final vowels most susceptible to spill-over: those lacking [h] or [ʔ].	Why: Lack of post-vocalic laryngeals lead to a phonetically natural pitch split, while the allophonic substitution remains unchanged.	Why: Presence of post-vocalic laryngeals ([h] and [ʔ])make demands on the vocal folds that may be in conflict with tone production.
Example: toː˩ banana kwɑ˧˥ toː˥˩ give a banana	Example: hi˩ book miːˀ˥ hi˥˩ I ask for a book	Example: kuːˀ˦ money miːˀ˦ kuːˀ˥ I ask for money	Example: ʔnehˀ˥ need ʔnehˀ˥ ni˩ kihˀ˦ We need to pay
Time→			

Proposed timeline of Comaltepec Chinantec tonal allophony

To summarize, the present-day tone patterning Comaltepec Chinantec, as superficially strange as it is, can be understood as the culmination of a series of small, local, and emphatically *natural* incremental changes. First, anisotropic variation of LH tones may have become conventionalized, since those LH tones which, quite naturally, spilled over onto a following vowel were better at keeping words distinct from each other that differed in meaning. These variants were naturally selected, and the spill-over came to be conventionalized.

The pattern seems to have generalized to include level H-tones as well, but only those that lacked post-vocalic laryngeals. Without conflicting demands placed on the vocal folds (due to the absence of a following [h] or [ʔ]), it was these tones that were most naturally incorporated into the pattern. And although these tones subsequently lowered to M for rather well-understood aerodynamic reasons, the tone-change process had been fully conventionalized by this time, and so we still observe—up to today—a H-tone on the first portion of the following vowel. The result is that this pattern has neither functional nor phonetic motivation, at least for present-day speakers of Comaltepec Chinantec.

The appearance of H tones following M tones in Comaltepec Chinantec exemplifies something quite remarkable about the nature of sound substitution. Even when a phonological pattern seems to be downright bizarre, lacking any reasonable phonetic or functional motivation at all, there are *always* perfectly natural, incremental processes that have unfolded over time that may account for the pattern. A series of small, local, interactions of phonetic, functional, and cognitive pressures may, over generations of speakers, render alternants quite distinct from each other: *alternations in the present—even when phonetically unnatural and superficially counter-functional—are the long-term product of small, local, and perfectly natural processes that play themselves out over generations of speakers.*

Still, phonological systems tend to remain remarkably natural and phonetically plausible, even though the ravages of time logically allow for bizarre patterns to slowly emerge. This disparity between reality and logical possibility can be reconciled quite intuitively, however: with every utterance by every speaker in every language, phonetic pressures exert force on the system. As unnatural patterns slowly emerge (always, of course, as a consequence of slow, natural, and local steps), phonetic pressures will always be exerting themselves, due to the simple fact that each speech utterance is an actual physical event which unfolds in real time, and so is subject to genuine physical forces. In time, irregular, unusual patterns may once again be slowly shaped by raw physiology. Consequently, phonetically implausible patterns are constantly under pressure to fall back in line, and so those that do survive are not only phonetic oddities, but are statistical oddities as well.

In the case of Comaltepec Chinantec at least, we may have successfully uncovered some of the major pressures on the system that have led to its present state. But even in those cases when the present state of the system is hopelessly obscured by long-forgotten, undocumented historical changes, we should not just throw up our hands and give up the notion that *all* patterns are—at least in theory—explainable by real-world forces that are known to act in any number of natural circumstances. The optimistic nature of scientific pursuit demands us to operate under the assumption that broadly applicable, locally active principles go far in explaining the complex world around us, in phonology, and elsewhere.

SUMMARY

In this chapter we've explored in some detail how variation in speech can sometimes lead to confusion for listeners, and how this confusion may lead to the better separation of phonological categories. We've seen how, under some circumstances, variation may induce the phonetic stability of categories, but under other circumstances, variation may induce sound change. Under both sets of circumstances however, I attributed the variation inherent in speech production to the accumulation of minor, chance errors over generations of speakers. Sounds in alternation in the present, which undergo quantum leaps of change in phonetic quality as they shift from context to context, have evolved in the absence of the user who newly comes to possess them. Allophonic alternants can now be viewed as the culmination of a series of small, natural changes to the system that take place over generations speakers. Even when a pattern does not lend itself to a compelling explanation in the present, we should not abandon the idea that phonetic, functional, and cognitive pressures are ultimately responsible for its linguistic comportment.

Carla L. Hudson and Elissa Newport. “Creolization: could adults really have done it all?