

Chapter 4

The Rate of Phonetic Change

When examining the effect that one speech segment has on an adjacent segment, there is a persistent problem involved in trying to determine whether that effect should be attributed to phonetic coarticulation, or to a phonological process, especially since the two can appear to be so similar (a so-called “duplication problem”, (Ohala, 1990; Cohn, 2007)). This problem is compounded when these effects are spread out across generational time. The difficulty in distinguishing between phonetic coarticulation and phonological processes with synchronic data, among other things, has led some to propose a much more phonetics-like model of phonology, where the phonology operates over much smaller granular primitives (e.g. Flemming, 2001), and where gradient phonetic realizations are subject to phonological considerations, like contrast.

Meanwhile, an increasing volume of research makes appeals to language change to explain phonological processes, and the apparent naturalness of phonology. Evolutionary Phonology, as proposed by Blevins (2004), is a good example. Blevins states the central premise of Evolutionary Phonology this way:

Principled diachronic explanations for sound patterns have priority over competing synchronic explanations unless independent evidence demonstrates, beyond reasonable doubt, that a synchronic account is warranted. (Blevins, 2004, p 23.)

A key problem for lines of research like this one is that few utilize evidence from language change *in progress* to support their arguments. Most of the argumentation in Blevins (2004), for

example, is based on comparative reconstructions, which provide us with proto forms A, and daughter forms B, C, and D, thus indicating 3 different sound changes: $A \rightarrow B$, $A \rightarrow D$, and $A \rightarrow C$. Blevins (2004) follows up the identification of sound changes like these with argumentation for the phonetic naturalness of each change, and how the change may have occurred given this phonetic naturalness. Blevins proposes some possible mechanisms for sound change (CHANGE, CHANCE, CHOICE), but these mechanisms are supported only by the conceptual plausibility that they may have generated changes $A \rightarrow B$, etc., not by direct evidence of these mechanisms at work in a sound change in progress.

Another good example of an appeal to sound change that lacks support from change in progress is Ohala (1990). In that work, Ohala examines the phenomenon of consonantal place assimilation. First, he identifies $C_1C_2 \rightarrow C_2C_2$ as a common sound change.

	Latin		Italian
(4.1)	<i>scriptu</i>	>	<i>scritto</i>
	<i>nocte</i>	>	<i>notte</i>

Then, he reports results of experiments where various manipulations to non-word sequences like [apta] can affect whether English listeners report hearing [apta], [atta], or [appa]. Unsurprisingly, subjects in the studies were much more likely to misperceive [apta] as [atta] (93%) than as [appa] (7%). The inference that Ohala (1990) makes is that these experimental subjects were, in some sense, recreating a sound change of the type in (4.1). However, even when taking together the experimental results with matching attested sound changes, the way in which the change took place remains underdetermined. The change from $C_1C_2 > C_2C_2$ could have been lexically gradual, slowly diffusing through the lexicon, or it could have been lexically abrupt. It could have started in one context (say kt > tt), then spread to other contexts, or it could have affected all contexts simultaneously. In the case of consonantal place of articulation, it's unlikely that this change would have been phonetically gradual, but in the case of post-coronal [u] fronting, another example from Ohala (1981), it's an open question whether it would progress in a phonetically gradual way, or abruptly. The fact that the way language changes like $C_1C_2 > C_2C_2$ are underdetermined by experimental work like Ohala (1990) is not just a descriptive gap, but an explanatory one. As

I argued in Chapter 2, the way in which language change progresses is determined by what part of speakers' linguistic competence is changing, meaning one's theory of linguistic competence defines possible paths of language changes, and vice versa. A solid result coming out of Ohala (1981, 1990) is that there appears to be a relationship between the kind of persistent errors listeners make and the outcomes of sound change, but I would argue that is a new fact to be explained, not an explanation itself.

Simulation of language change has also become an increasingly common tool for researchers interested in language change. Unfortunately, the success of these simulations is usually judged by comparing the initial and final states of the simulation to the initial and final states of attested sound changes, rather than by comparing the dynamics of change in the simulation to the dynamics of a known change in progress. For example, Boersma and Hamann (2008) try to model the fact that speech sounds tend to be maximally dispersed along acoustic dimensions by using agent based simulations of cross-generational language acquisition with bidirectional constraint grammars. Their results are interesting and compelling, but their conclusion that their model is a success is based on the fact that it produced maximally dispersed distributions, *not* that they produced realistic patterns of language change when compared to other language changes in progress.

I am proposing that theoretical models like the ones I've just mentioned must compare their predictions about the dynamics of language change to the dynamics of actual language changes in progress in order to claim definitive empirical support. Fortunately, there is a well established field of inquiry into the dynamics of language change in progress, Quantitative Sociolinguistics, with similarly well established methodologies for the study of language change in progress (e.g. Labov, 1994, ch. 3, 4). In order to compare the results of simulation and experiment to language changes in progress, it is crucial to hash out exactly what patterns of language change we ought to see based on different theories of phonology and phonetics, which is one of the partial goals of this dissertation.

The goal of this chapter is to both introduce a novel technique for distinguishing between phonetic and phonological influences on phonetic change, and to establish some basic facts about

the dynamics of sound changes which are subject to some kind of conditioning factors: comparing the rate of inter-generational sound change of vowels across different linguistic contexts. While these basic facts will be of considerable intrinsic interest to theories of language change, I believe I've made it clear that they will also be of considerable interest to phonological theory more broadly construed.

4.1 Phonetic Coarticulation vs Phonological Differentiation

Strycharczuk (2012, ch. 2) outlines a number of ways that researchers have attempted to distinguish between phonological processes and phonetic coarticulation.

- (4.2) Compare segments which are ambiguous between phonetic coarticulation and phonological assimilation to segments which are unambiguous. e.g. compare intervocalic /s/, which may undergo either categorical voicing assimilation or phonetic voicing coarticulation, to phonemic /z/.
- (4.3) Examine the coarticulatory effect over the duration of the segment. A phonetic cline, with its highest point adjacent to the coarticulatory source is indicative of phonetic coarticulation, while a phonetic plateau across the entire duration of the segment is indicative of a phonological process (Cohn, 1993).
- (4.4) Estimate the bimodality of the phonetic distribution of the ambiguous segments, with the hypothesis that strong bimodality is indicative of a phonological distinction.
- (4.5) Examine the coarticulatory effect's sensitivity to speech rate. The hypothesis is that phonetic coarticulation should be sensitive speech rate, but phonological assimilation should not be.

Both (4.2) and (4.3) appear to me to be reasonable approaches to the problem, but unfortunately not universally applicable. None of the cases studies I will be investigating involve neutralization, which is key for (4.2), comparing the phonetics of derived segments to underlying segments. For example, I will be looking at the effect of nasals on the /aw/ diphthong. The most

conservative realization for the nucleus of this diphthong is [æ], when followed by oral segments. However, even the most conservative realizations of the /aw/ nucleus are considerably fronter and higher when followed by nasal segments, [æ̣~ɛ̣]. I'm unable to utilize (4.2), because pre-nasal /aw/ isn't neutralized to a different segment which appears independently, so I have no unambiguously phonological form of [æ̣ʊ] to compare pre-nasal /aw/ to.

The next option of comparing phonetic clines to plateaus (4.3) is also difficult to bring to bear on the case studies at hand. To begin with, the Philadelphia Neighborhood Corpus, in the form I've had available for this dissertation, only contained point measurements for the nuclei of diphthongs. However, it is even difficult conceptually to determine what would constitute a cline, and what would constitute a plateau in the cases I will be looking at. Using the example of /aw/ again, its raising and fronting when adjacent to a nasal is undoubtedly related to nasality in some way. However, the dimension along which the effect of the following nasal plays out is in vowel height and frontness, which are only indirectly related to nasality. Moreover, /aw/ is an intrinsically dynamic speech segment with two targets. Determining whether the effect which fronts and raises the nucleus of /aw/ is somehow stronger in the glide, or whether it's a constant effect throughout the entire diphthong would be a complicated exercise indeed.

The remaining two options, examining bimodality of the distributions (4.4) and determining speech rate effects (4.5) could be feasibly applied to the cases I'm examining, but there are good reasons to call the diagnostic validity of these approaches into question. To begin with bimodality, it is trivial to come up with examples of bimodal distributions which clearly don't correspond to phonological differences. Figure 4.1 plots the distribution of mean F1 and F2 measurements for /ɪ/ for all speakers in the PNC. The distribution of /ɪ/ is strongly bimodal, but this bimodality is due to the sex of the speaker, since Figure 4.1 is displaying unnormalized data. There is no reason to believe that men and women have fundamentally different phonological representations or even different intended phonetic implementations for /ɪ/. Rather, men and women clearly have the same targets of phonetic implementation for the same phonological object, and those targets have then been filtered through phonetic contingencies (the sex linked differences in vocal tract length).

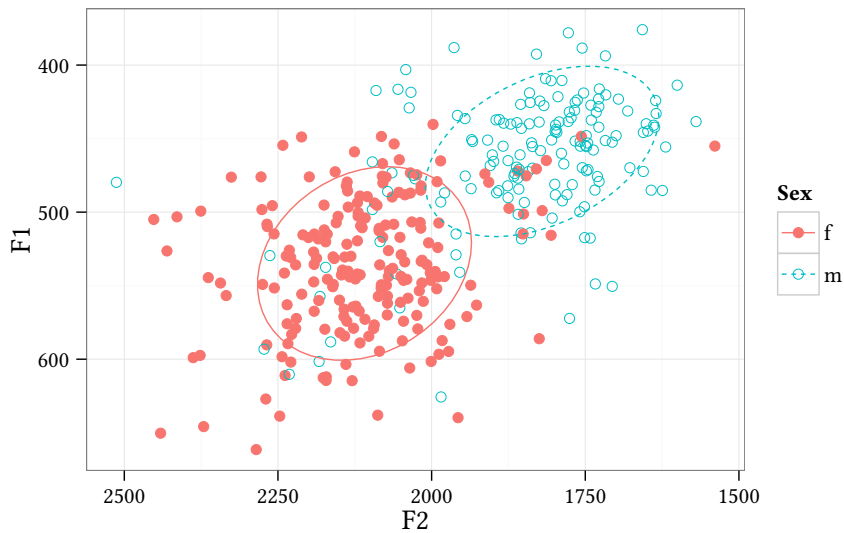


Figure 4.1: Sex differences in the acoustic realization of /ɪ/ in unnormalized F1 × F2 space.

The inter-speaker effect of vocal tract length on the realization of vowels is an extreme case of what I will henceforth be referring to as a “phonetic effect.” However, at the moment there is no theory of what the upper limit of intra-speaker phonetic effects due to coarticulation ought to be, especially if the degree of articulatory overlap is a language specific property, per the discussion in §2.1.5. Another case study in this chapter will be on the effect of a following /l/ on preceding /ow/ and /uw/. As /l/ in Philadelphia is frequently much more glide-like, especially in coda positions (Ash, 1982), with its primary place of articulation being dorsal, it is conceivable that it may have a considerable coarticulatory effect on /ow/ such that a bimodal distribution of [ow]~[owl] is the product. This is doubly so if the phonetic alignment constraints for the Philadelphia dialect allow for substantial gestural overlap of the /ow/ vowel and the dorsal /l/ gesture. Given these facts, it is not strictly necessary that strongly bimodal distributions are indicative of phonologically distinct targets.

Furthermore, there is also no theory for what the *lower* limit of phonetic difference is for two phonologically distinct targets. For example, Labov and Baranowski (2006) note that in the Inland North dialect region, the lowering and backing of /ɛ/ and the fronting of /ɑ/ has led to considerable

overlap between these vowels for many speakers, without resulting in merger. They argue that an average duration difference of 50ms is sufficient to maintain and signal the phonemic difference in this case. This is a relatively small difference. For comparison, I calculated the Median Absolute Deviation¹ for the duration of /ɑ/ for all speakers in the Philadelphia Neighborhood corpus. The median MAD across speakers is 45ms. While it is difficult to make direct comparisons between these two studies due to the drastic differences in the dialects, the fact remains that the size of between category differences in the Inland North is about the same size as the within category variation in Philadelphia.

So, it is both the case that strong phonetic bimodality is not necessarily an indicator of phonological differentiation, and the absence of strong phonetic bimodality is not necessarily an indicator of the absence of phonological differentiation. As such, I will not be utilizing bimodality as a diagnostic for distinguishing between phonetic and phonological effects.

The fourth option, determining whether the effect of one segment on another is sensitive to speech rate (4.5) would be possible to implement with the PNC data. However, the operating assumption behind this method that phonological processes should not be sensitive to speech rate does not stand up to the results of sociolinguistic research. The concept of a variable phonological rule was first introduced by Weinreich et al. (1968), and since then, variable linguistic processes of all sorts have been found to be sensitive to both grammatical and extra-grammatical variables, like speaking style.

Using the case of /ow/ followed by /l/ to make this argumentation concrete, we could imagine that there is a variable phonological process which spreads some additional dorsal feature from /l/ to /ow/, producing a phonetically fully back [oɹ]. This phonological process could be close to categorical at extremely fast speech rates, but as speech rate slows, its probability of application falls off. When /l/ doesn't spread its phonological features to /ow/, however, it might still be phonetically coarticulated with /ow/, an effect which itself might decrease as speech rate slows even further. The resulting data would appear to show a gradually decreasing effect of /l/ on /ow/ as speech rate decreases, and we would miss the generalization of a phonological process at work

¹The MAD is calculated by first calculating the distance of all data points from the sample median, then taking the median of their absolute values. i.e. $median(|x_i - median(x)|)$

if we were to interpret this to mean the effect of /l/ on /ow/ is purely coarticulatory.

4.1.1 Phonological vs Phonetic Processes in Sound Change

I would like to bring evidence from sound change to bear on the question of whether the influence of one segment on another is due to phonetic coarticulation or phonological differentiation. Let's assume that we are analyzing some hypothetical vowel, /V/, which appears in two different segmental contexts, /__x/ and /__y/. The distributions of [V_x] and [V_y] in F1×F2 space are given in Figure 4.2.

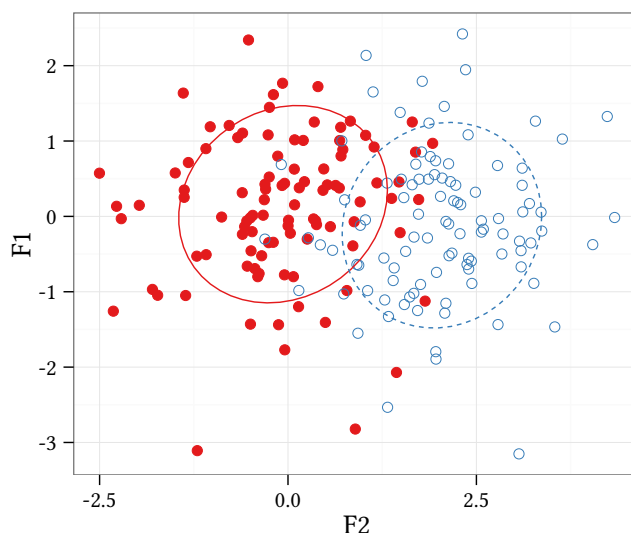


Figure 4.2: Distribution of contextual variants of a hypothetical vowel

There are two distinct ways in which the data in Figure 4.2 could have been generated. First, /y/ could have spread some feature *f* onto V, creating a featurally distinct, thus phonologically distinct allophone of /V/.

$$(4.6) \quad \textcolor{red}{V} \rightarrow \textcolor{blue}{V}_f / \text{ ___ } y$$

As phonologically distinct objects, [V] and [V_f] can have independent targets for phonetic implementation. The target of implementation for [V_f], in this case, happens to be further back along F2. In Figure 4.3, the two independent targets for [V] and [V_f] are represented as two larger points in the centers of their respective distributions.

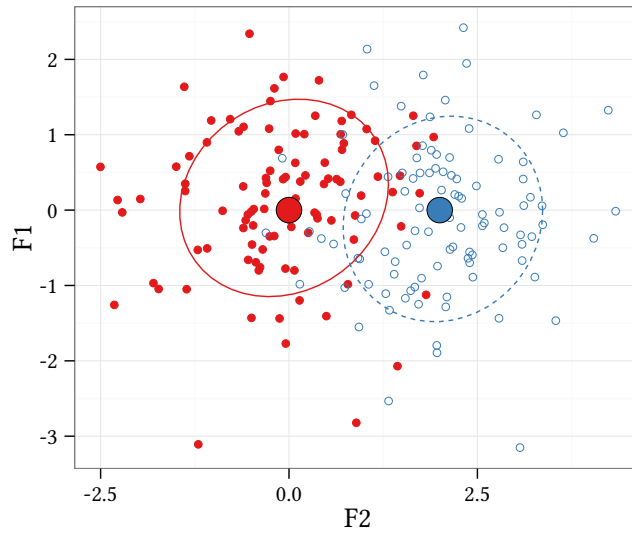


Figure 4.3: Independent targets of phonetic implementation produced by phonological differentiation.

Alternatively, there could be no phonological process involved here at all. Instead, the mapping from phonological representations to phonetic targets could produce only one target, that for [V_x]. However, segment [y] exerts a large coarticulatory pressure on [V], pulling the actual productions of [V_y] back from their intended target. This coarticulatory shift is represented by the arrow in Figure 4.4. The distribution for [V_y] does not have a larger point at the center of its distribution in Figure 4.4 in order to indicate that it does not have its own independent target for phonetic implementation.

As I have argued above, it is not possible to distinguish between these two scenarios given the most common methodologies, nor by just eyeballing the data. However, we should expect to see different patterns in *diachronic* change depending on which process is operating. The key difference is that in the case of phonological feature spreading, [V] and [V_f] have independent targets of phonetic implementation, while in the case of phonetic coarticulation, the realization of [V_y] is yoked to [V_x]. Thus, it should be possible for these contextual variants of /V/ to have separate diachronic trajectories *only* in the case of phonological feature spreading, while in the

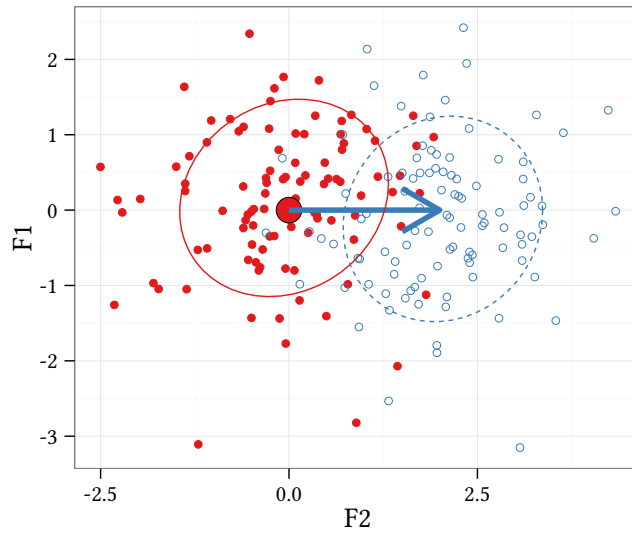


Figure 4.4: The effect of coarticulation on shifting productions from intended targets.

phonetic coarticulation case, the realization of one variant should be yoked to the diachronic trajectory of the other.

Figure 4.5 illustrates the interaction between phonological feature spreading and diachronic phonetic change. The data in this figure represents a shift in one generation from Figure 4.3 where the target of [V] has shifted frontwards along F2, but the target of [V_f] has remained stable. The target for [V] from the previous generation is represented as a large faint point. The important point is that [V] has shifted independently from [V_f], which contrasts sharply with Figure 4.6

Figure 4.6 represents the interaction of phonetic coarticulation and diachronic phonetic change. Again, the target for [V_x] has shifted frontwards along F2, but because the realization of [V_y] is the product of a coarticulatory shift, which has remained constant, [V_y] has also shifted frontwards along F2.

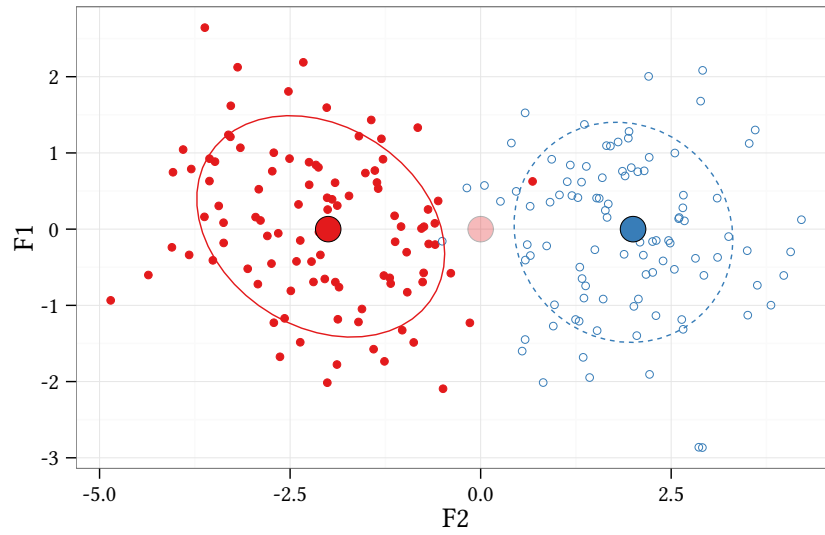


Figure 4.5: The interaction of phonological feature spreading and diachronic phonetic change.

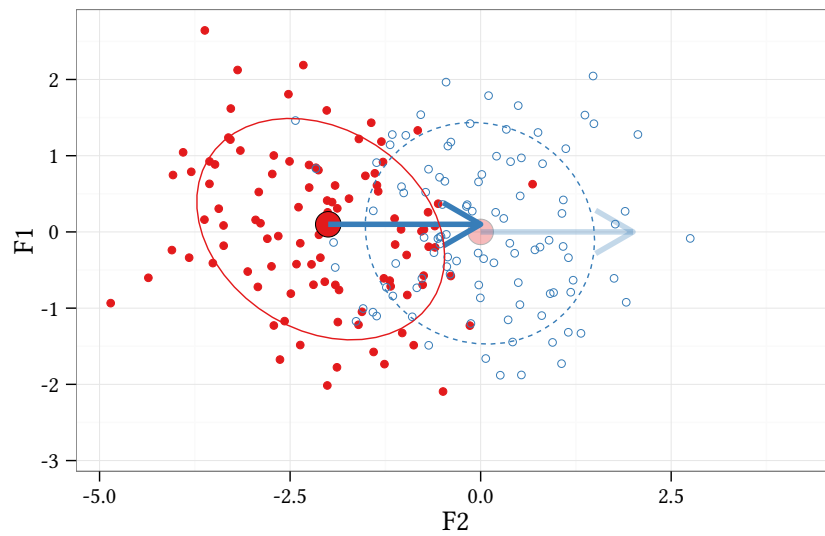


Figure 4.6: The interaction of phonetic coarticulation and diachronic phonetic change.

4.2 The Rate of Language Change

In this section, I'll be fleshing out more completely the way in which phonological feature spreading and phonetic coarticulation produce different predicted dynamics of sound change. Figure 4.5 illustrates the expected difference between two generations when phonetic change interacts with phonological feature spreading. [V] moves frontwards along F2, leaving [V_f] behind. Figure 4.7 presents a finer grained illustration of this effect over age cohorts. The top facet of Figure 4.7 illustrates the target for [V] moving along F2 from 0 to 2 along a classic S-shaped trajectory. The phonetic target for [V_f], on the other hand, remains constant at -2. The bottom facet of Figure 4.7 represents the year-to-year change in F2 for [V] and [V_f].

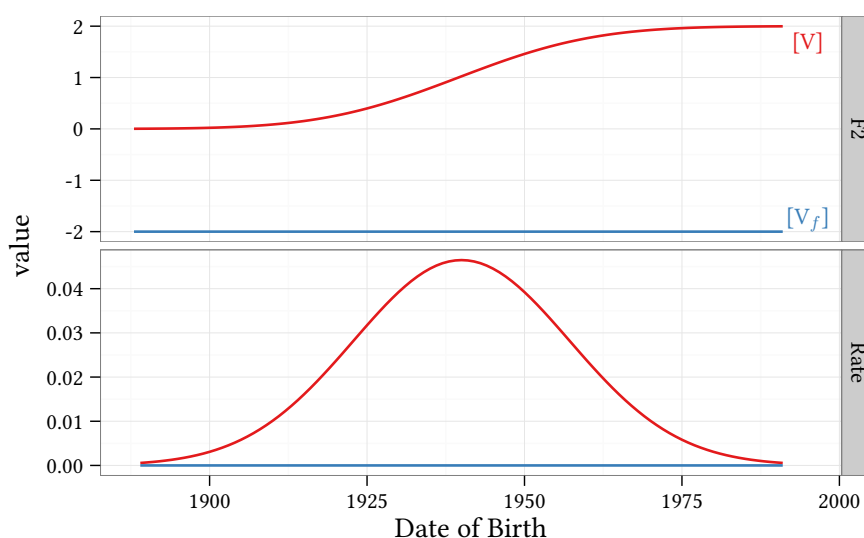


Figure 4.7: The rate of phonetic change in the context of phonological feature spreading.

By its very definition, the rate of change for [V] reaches its maximum in the bottom facet of Figure 4.7 at the midpoint of the S-shaped curve in the top facet, because it is at the midpoint of the S-shaped curve that the change is progressing at its fastest. The rate of change for [V_f] remains at 0 throughout, because it is not undergoing any phonetic change at all.

Another way to think about the relationship between the rate of change in the bottom facet of Figure 4.7 and the trajectory of change in the top facet is that the trajectory in the top facet represents the cumulative sum of values in the bottom facet. For example, the rate of change for

[V] in 1950 is approximately 0.04. This means that the predicted value of F2 in 1950 is equal to the value of F2 in 1949 plus 0.04. The value of F2 in 1949 was 1.42, so the predicted value of F2 in 1950 is $1.42 + 0.04 = 1.46$. To figure out how different the F2 of [V] is in 1950 from 1888 (the earliest point in time in these figures), we merely need to sum up all of the rates of change from 1888 to 1950, and add that to the value of F2 in 1888. In 1888, F2 was 0, and the sum of all by-year rates of change between 1888 and 1950 is 1.46, so the predicted value of F2 in 1950 is, again, $0 + 1.46 = 1.46$. Meanwhile, the rate of change for [V_f] in 1950 is 0, meaning the predicted value of F2 for [V_f] 1950 is the value of F2 in 1949 plus 0; $-2 + 0 = -2$. The sum of all by-year rates of change from 1888 to 1950 for [V_f] is also 0, meaning that [V_f] is expected to have the same F2 in 1950 as in 1888.

A more technically accurate description of the relationship between the rate of change and the trajectory of change is that the rate of change is the first derivative of the trajectory of change. I will continue to describe the rate of change in terms of year-to-year differences for the sake of interpretability. However, keeping in mind that I am really trying to model $f'(x)$, where $f(x)$ is the trajectory of change, could be useful for technical advancements of these methods in the future.

The key takeaway from Figure 4.7 is that [V] and [V_f] have different rates of change, and as I argued in §4.1.1, this is only possible because they are phonologically distinct objects, and thus have different targets of phonetic implementation.

Figure 4.8 illustrates the expected dynamics of phonetic change if the contextual variants of /V/ were due to phonetic coarticulation. The solid line represents the movement of the target for [V_x] along F2. The arrows indicate the coarticulatory effect, shifting the productions of [V_y] back along F2 from the target for [V_x]. This coarticulatory effect remains constant over time, producing a trajectory for [V_y] which is yoked to [V_x], thus parallel to it over time.

The rates of change of two parallel trajectories, even if these trajectories are displaced upwards or downwards, will always be the same. This is represented in the bottom facet of Figure 4.8. At all points in time, [V_x] and [V_y] have the same rate of change because they are moving in parallel, because [V_y] is yoked to [V_x] because they share a target for V.

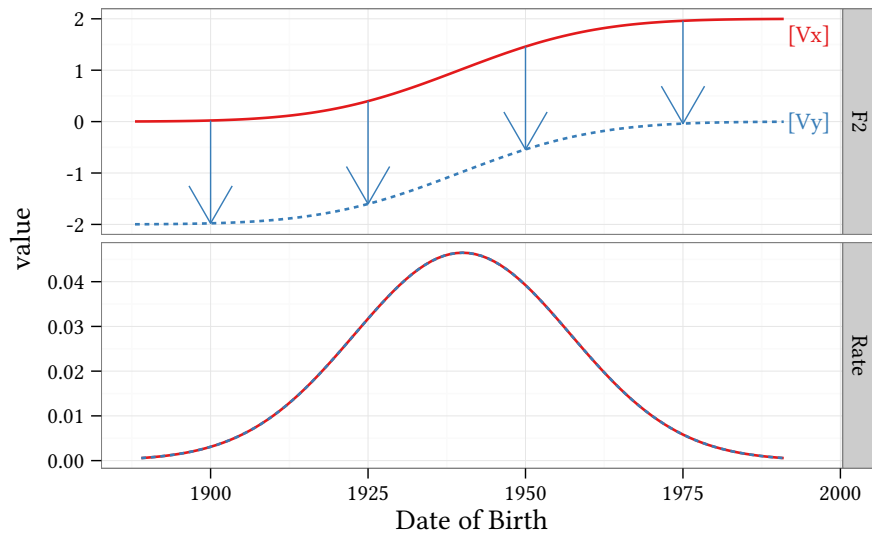


Figure 4.8: The rate of phonetic change in the context of phonetic coarticulation

The difference between phonological differentiation and phonetic coarticulation is large and qualitative. What I hope to have illustrated so far is that this qualitative difference can be connected to quantitative differences in the way the system changes over time. Specifically, for any given vowel which has two contextual variants, if we can estimate the rate of change of these two variants over time and determine whether they have a shared or different rate of change, then we can then use this information as an indicator of a qualitative difference.

Perhaps most importantly, we can utilize the comparison of rates of change to identify cases where phonetic coarticulation has been reanalyzed as phonological differentiation. That is, for some changes, the difference between [Vx] and [Vy] could have been originally due to phonetic coarticulation, but then speakers reanalyzed this difference as actually being due to a phonological process, with featurally distinct objects, [V] and [V_f], and targets. This process of reanalysis has been called “phonologization” (Hyman, 1976) or “stabilization” (Bermúdez-Otero, 2007), and is argued by some to be the primary source of naturalness in phonology (e.g. Cohn, 2006, 2007).

The effect this reanalysis would have on the dynamics of sound change is illustrated in Figure 4.9. At the beginning of the sound change, the difference in contextual variants of V is due to

phonetic coarticulation, and the trajectory of [Vy] is yoked to [Vx], causing them to have the same rate of change. The dark vertical line represents the time point when the coarticulatory effect is reanalyzed as being phonological. A process like (4.6) enters the phonological grammar, producing featurally distinct allophones, [V] and [V_f], which have independent targets of phonetic implementation. In this illustration, the trajectory of [V] continues along its previous path, but [V_f] ceases to undergo change.

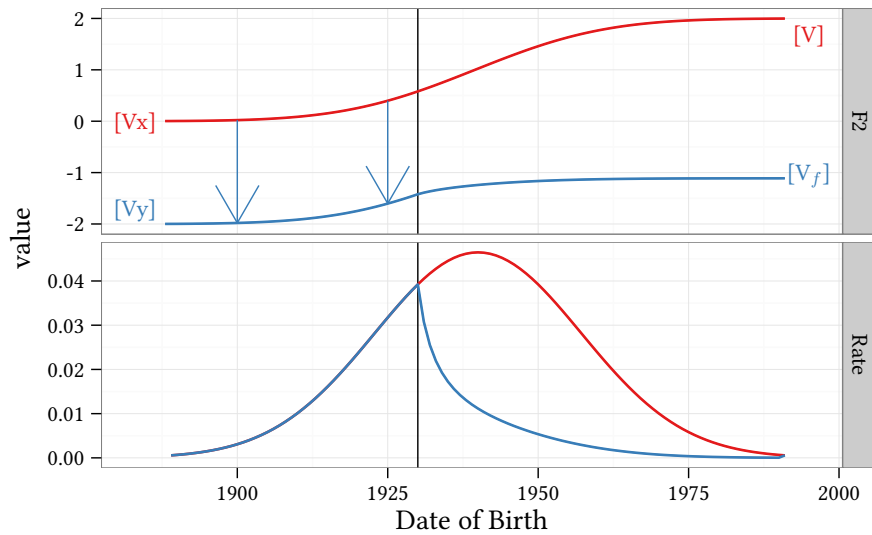


Figure 4.9: The reanalysis of phonetic coarticulation as phonological feature spreading, and its effect on the rate of phonetic change.

Looking at the trajectories alone, it would be difficult to pinpoint with much accuracy when the reanalysis occurred if were not indicated on the graph. The rates of change, on the other hand, indicate rather unambiguously a sharp point at which [V_f] diverged from [V]. It is possible to model the trajectories directly using, for example, cubic regression splines, and comparing models where the trajectories are constrained to be the same to models where they are allowed to be different. This sort of modeling approach would tell us that in cases phonological feature spreading, like Figure 4.7, the trajectories differ significantly, while in the case of phonetic coarticulation, like Figure 4.8, they don't. However, this approach would also tell us that the trajectories differ significantly in cases where phonetic coarticulation has become reanalyzed as phonological feature spreading, like Figure 4.9. Given that we want to be able to disambiguate instances of all

three kinds of influences on sound change, and that in the case of reanalysis, we want to be able estimate a time point in the sound change when reanalysis occurred, a more complex approach is necessary, which involves directly modeling the rate of change.

I hope to have made clear, in this section, the possible diagnostic capacity of the rate of change. In principle, we should be able to not only identify qualitative differences through quantitative measure (i.e. the difference between phonetic coarticulation and phonological differentiation), but also identify the point in time where new qualitative options enter the grammar (i.e. the reanalysis of phonetic coarticulation as phonological differentiation).

4.3 The Model and the Data

This section will be devoted to the specifics of implementing a statistical model to estimate and compare the rates of change of different contextual variants, as well as the data behind the case studies I will be applying the model to.

4.3.1 The Model

As I stated above, we can conceptualize the rate of change as actually representing year-to-year differences along any particular phonetic dimension. Let's represent the rate of change for year l for a vowel in context k as δ_{lk} , which will be equal to the difference along the phonetic dimension between year $l - 1$ and l . This is the parameter of primary interest, specifically for particular years whether δ_{lk} is the same for different contexts. Contexts will be indexed by different values for k . The context $k = 1$ will always be some reference level context. For example, the first case study will focus on the effect of following nasals on /aw/. In this case, /aw/ followed by oral segments will be given index $k = 1$, and vowels followed by nasal segments will be given the index $k = 2$. Once we have estimated $\delta_{lk=1}$ and $\delta_{lk=2}$ for all l dates of birth, we will make the a quantitative comparison to see if $\delta_{lk=1} = \delta_{lk=2}$ or $\delta_{lk=1} \neq \delta_{lk=2}$. More precisely, we will be looking at the difference, $\delta_{lk=1} - \delta_{lk=2}$. There are three possible results for this comparison.

$$(4.7) \quad \delta_{lk=1} - \delta_{lk=2} > 0$$

This means that $\delta_{lk=1} > \delta_{lk=2}$ therefore $\delta_{lk=1} \neq \delta_{lk=2}$, therefore the vowel has different rates of change between contexts $k = 1$ and $k = 2$.

$$(4.8) \quad \delta_{lk=1} - \delta_{lk=2} < 0$$

This means that $\delta_{lk=1} < \delta_{lk=2}$ therefore $\delta_{lk=1} \neq \delta_{lk=2}$, therefore the vowel has different rates of change between contexts $k = 1$ and $k = 2$.

$$(4.9) \quad \delta_{lk=1} - \delta_{lk=2} = 0$$

This means that $\delta_{lk=1} = \delta_{lk=2}$, therefore the vowel has the same rate of change in contexts $k = 1$ and $k = 2$.

Now, δ_{lk} is not a directly observable variable in the data. Rather, it is a latent variable that we will be attempting to estimate from the data. For this reason, along with all of the usual constraints on statistical inference from a sample to a population, we will not be estimating precise values for $\delta_{lk=1} - \delta_{lk=2}$. Instead, we will be estimating *credible intervals* for the value $\delta_{lk=1} - \delta_{lk=2}$. If the credible interval excludes 0, then our inference will be that it is more likely than not² that $\delta_{lk=1} - \delta_{lk=2} \neq 0$. On the other hand, if the credible interval includes 0, our inference should be more cautious. It may actually be the case that $\delta_{lk=1} - \delta_{lk=2} \approx 0$, or it may be the case that the data is too sparse for either $k = 1$ or $k = 2$ to reliably determine otherwise.

As illustrated in Figures 4.7, 4.8 and 4.9, δ_{lk} should be modeled as a function of date of birth. However, I have no theoretically driven hypothesis about what the shape of that function ought to be. As such, I made the decision to model δ_{lk} using b-splines. I chose b-splines over other kinds of curve fitting because they are relatively easy to implement, conceptually easy to understand, and flexible in the kinds of curves they can approximate. Fitting a curve with b-splines begins by defining the “*basis*” of the curve. In the context of curve fitting, “basis” has a technical meaning of approximately a collection of curves which are then scaled and summed over to produce the final curve. Figure 4.10 displays the b-spline basis used in all of the models in this chapter, which was constructed with the `splines` package in R (R Core Team, 2012). This particular basis consists of three cubic polynomial curves which are evenly spaced along the time dimension, and one linear intercept term.

²In fact, for the credible intervals displayed in this work, it will be 95% more more likely than not.

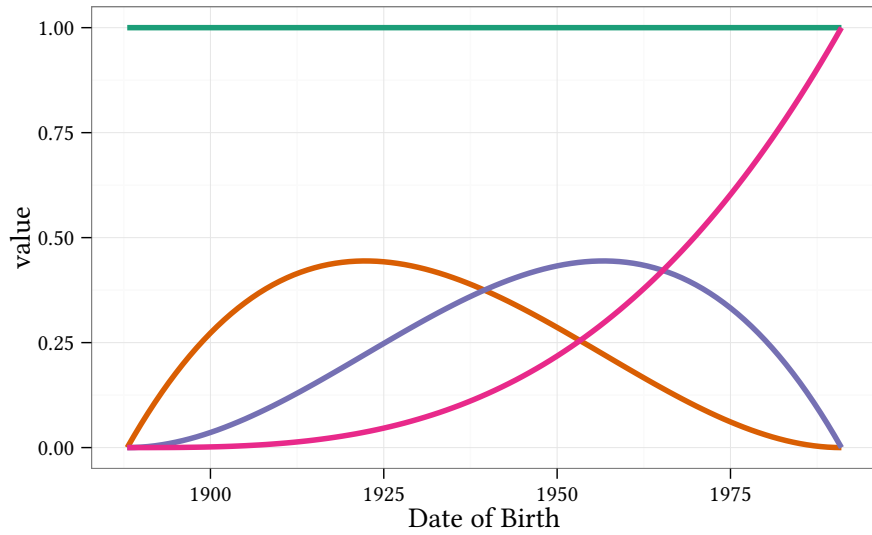


Figure 4.10: B-spline basis used in all following models.

After establishing the basis of the b-spline, you then estimate *weighting coefficients* for each curve in the basis. Usually, the weighting coefficients will be estimated from the data, but in this illustration, 4 coefficients were randomly chosen from a normal distribution. You then scale each polynomial by multiplying it by its corresponding weighting coefficient. The weighted form of the basis is represented in the top facet of Figure 4.11. In the final step, you sum across the polynomial along the x-axis, resulting in the final b-spline fit, which is represented in the bottom facet of Figure 4.11. Figure 4.12 displays five more b-spline fits based on more randomly generated weighting coefficients in order to provide a qualitative sense of how smooth b-spline fits with the basis in Figure 4.10 will be.

The degree of wiggleness of a b-spline fit is highly dependent on the size of the basis. For example, Figure 4.13 displays the kind of curve that a larger b-spline basis could fit. I will be restricting my modeling of δ_{lk} to the smaller basis displayed in Figure 4.10 for the following reasons.

- (4.10) As the size of the basis increases, the number of weighting coefficients increases, and the over all uncertainty about the final fit of the curve increases.

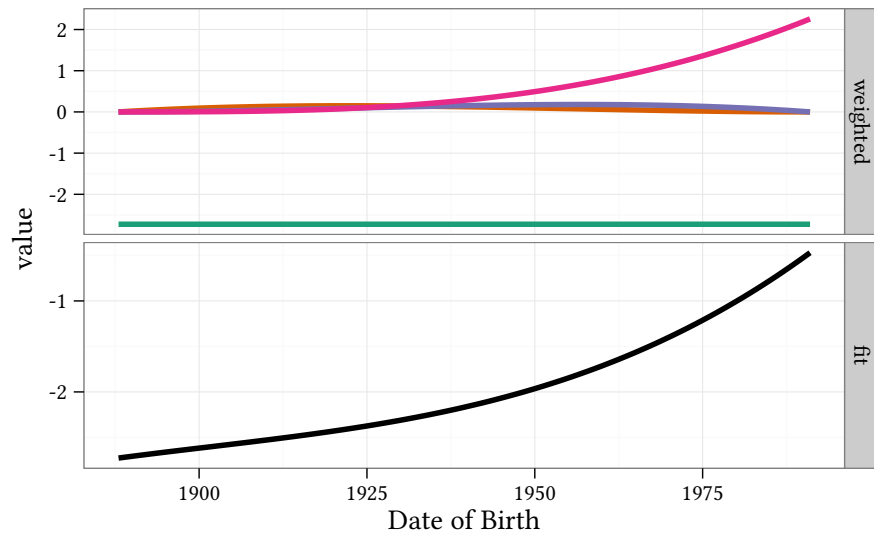


Figure 4.11: Weighted b-spline basis, and resulting spline fit.

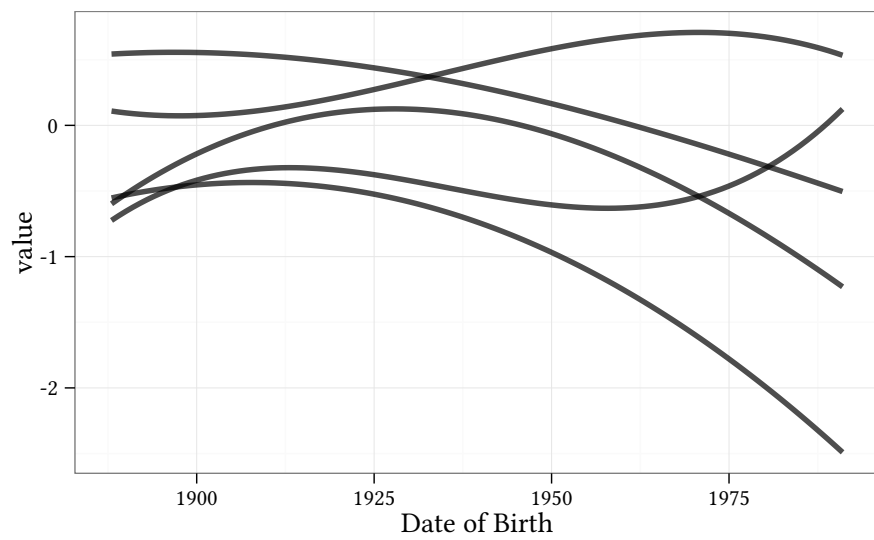


Figure 4.12: Five randomly generated b-spline curves

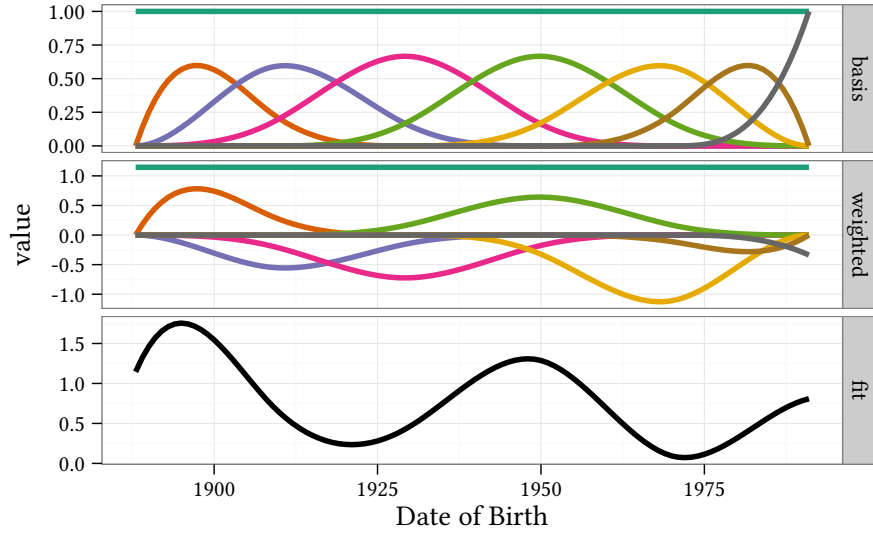


Figure 4.13: The effect of a larger basis on the fit's 'wiggleness'.

(4.11) Since δ_{lk} is a latent variable, there is already a higher degree of uncertainty built into its estimation.

(4.12) Additionally, since δ_{lk} represents the first derivative of the trajectory of change, it can afford to be relatively simpler than the actual trajectory, since $f'(x)$ is always one degree less than $f(x)$.

I will represent the fact that δ_{lk} is modeled by a b-spline smooth of date of birth, which is also the index for l , as follows.

$$\delta_{lk} = b.spline(l) \quad (4.13)$$

After estimating δ_{lk} for every date of birth, we then need to estimate the expected value along the phonetic dimension for that date of birth. That is, if the change we are modeling is /ow/ fronting along F2, δ_{lk} will represent how far /ow/ fronted along F2 between the years $l - 1$ to l , but we also need to estimate what the actual value of F2 is in year l . As was discussed in §4.2, this can be done by taking the cumulative sum of δ_{lk} from 1888 up to year l , then adding it to

the value of F2 in 1888. The cumulative sum will be represented by Δ_{lk} , the value in 1888 will be represented by β_k , and the expected value in year l will be represented by μ_{lk} .

$$\Delta_{lk} = \sum_{x=1888}^l \delta_{xk} \quad (4.14)$$

$$\mu_{lk} = \beta_k + \Delta_{lk} \quad (4.15)$$

At this point, μ_{lk} represents the expected phonetic target for a vowel in context k for a speaker born in year l . However, it would not be expected for *all* speakers born in year l to have the precise target of μ_{lk} . Obviously, inter-speaker variability exists for all manner of systematic reasons, some of which could be incorporated into the model, like socio-economic class, education, etc. Just as obviously, there are systematic causes of inter-speaker variation that we cannot include in the model because it didn't occur to us to document them, we have yet to operationalize measures for them, or they are in some sense immeasurable, related to the accidental personal history of every individual. Finally, even with a full accounting of all possible factors that predict inter-speaker variation, and well formulated operationalizations and measurements of those factors, there will always be some variation between individuals which is irreducible.

For these reasons, we will add an additional layer to the model, where we estimate phonetic targets for every individual speaker in the corpus, which will be represented as μ_{jk}^s , where j is an index for each speaker. These speaker-level parameters will be drawn from a normal distribution centered around μ_{lk} . The variance of the distribution will be another parameter in the model σ_k . The reason we want to include σ_k as a parameter in the model is that we want to allow speakers to be as similar to each other, or as different from each other as is warranted by the data. Notice that σ_k is also indexed by the context k . This means that inter-speaker variation can be greater or lesser for each context under question. In the following equations, DOB_j represents the date of birth for speaker j .

$$DOB_{1,2,\dots,n.speaker} \quad (4.16)$$

$$l = DOB_j \quad (4.17)$$

$$\mu_{jk}^s \sim \mathcal{N}(\mu_{lk}, \sigma_k) \quad (4.18)$$

Additionally, we should recognize that speakers will differ in the degree to which the individual tokens they produce are scattered around their target. Some speakers may have very small variance, with most of their productions being clustered tightly around their basic target, μ_{kj}^s , while other speakers may have much larger variance. As such, we will also be estimating speaker-level variances, which will be represented as σ_j^s .

An additional point of complexity to the data is that not only is it generated by many different speakers, but also represent many different lexical items. Whether or not lexical items play an important role in sound change over and above environmental conditioning is a long, and ongoing debate (Labov, 1981, 1994, 2010; Pierrehumbert, 2002; Bybee, 2002 *inter alia*). Regardless of whether or not lexical items can have individualized phonetic targets, I will be including by-word random effects in this model for much the same reason as why by-speaker random effects were included. It is certainly the case that there are systematic properties of lexical items which affect their phonetic realizations which we have not accounted for, and are therefore missing from the model. Therefore, we will be estimating by-word random effects drawn from a normal distribution centered around 0, with a variance parameter which will be estimated on the basis of the data. The random effect for each word will be indexed by m , and will be represented as μ_m^w . As can be seen in the equation below, μ_m^w is not sensitive to any properties of the speaker, including date of birth, making it time insensitive. It would be ideal to model the effect of a word as being variable over time, to see if it changes or remains stable, but the model as I've laid it out up to this point is already very complex, and making the by-word effects time sensitive would *minimally* involve adding two parameters to the model for every lexical item: slope and intercept. Therefore, I will be backing off from an ideal model of lexical effects to a merely sufficient one.

$$\mu_m^w \sim \mathcal{N}(0, \sigma^w) \quad (4.19)$$

Finally, we come to the raw data layer of the model. The raw acoustic data will be represented by y_i , where i is an index for every observation. The total number of observations is represented

as n , J is a vector of speaker indices, K is a vector of context indices, and W is a vector of word indices. We will be adding speakers' expected phonetic target for a vowel in context k , represented by μ_{jk}^s , to the word level effect, μ_m^w , to arrive at the expected target for observation y_i . Of course, any particular observation from a particular speaker of a particular word will not precisely be equal to $\mu_{jk}^s + \mu_m^w$ for all of the reasons which have already been stated, so we will actually be presuming that y_i is drawn from a normal distribution centered around $\mu_{jk}^s + \mu_m^w$ with a speaker specific variance, σ_j^s , which was mentioned above.

$$y_{1,2,\dots,n} \tag{4.20}$$

$$J_{1,2,\dots,n} \tag{4.21}$$

$$K_{1,2,\dots,n} \tag{4.22}$$

$$W_{1,2,\dots,n} \tag{4.23}$$

$$j = J_i \tag{4.24}$$

$$k = K_i \tag{4.25}$$

$$m = W_i \tag{4.26}$$

$$y_i \sim \mathcal{N}(\mu_{jk}^s + \mu_m^w, \sigma_j^s) \tag{4.27}$$

Human Readable Form

This model of the rate of change has three levels. At the highest level, the year-over-year differences are estimated using non-linear curve fitting. I didn't assume that the rate of change was constant across the lifespan of the phonetic change because, in fact, all three of the changes I look at in this chapter move in one direction, stop, then reverse, and also, the relative timing of when contextual variants diverge in their rate of change is of key interest. At the next level, the estimated phonetic targets of each speaker are estimated. The expected phonetic target for a speaker born in a particular year is estimated by summing up the year-over-year differences from the first layer of the model. The phonetic targets of the actual speakers in the model are assumed to be

normally distributed around the expected target for their date of birth. By-word random errors are also assumed to be normally distributed around 0. The third, and lowest level, treats each individual measurement as being drawn from a normal distribution centered around the specific speaker’s phonetic target plus the particular word’s random error.

Some readers may be more familiar with the syntax of mixed effects linear models as implemented in the `lme4` R library. Faux-`lme4` syntax for this rate of change model is provided in (4.28) and (4.29). It includes random intercepts for speaker and word.

(4.28) `rate_of_change ~ b_spline(DOB)`

(4.29) `F2 ~ sum(rate_of_change) + (1|Speaker) + (1|Word)`

4.3.2 Implementing the model

The model I have just described does not easily submit to a reformulation as a linear regression, or even in terms of modeling techniques like generalized additive models. As such, I have implemented it in Stan (Stan Development Team, 2012). Stan is a package designed to implement graphical Bayesian models with Hamiltonian Monte Carlo (Hoffman and Gelman, 2011). Providing a precise description of HMC is well beyond the scope of this dissertation. Generally speaking, HMC is closely related to Markov chain Monte Carlo methods of model estimation, for which Kruschke (2011) is an excellent introduction. In an iterative process, the system samples possible values for the parameters it’s trying to estimate from a probability distribution which is in part determined by its prior probability, the probability of the other parameter values estimated so far, and the observed data. After a sufficient number of iterations, the samples produced by the system will approximate the posterior probability distribution of the parameters, which is what we will use for our inferences. As it is an iterative process, we want to be sure that it is not sensitive to its initial values, so the model will be fit multiple times with different random initializations, and the results compared across the fits (or chains) to verify that they have converged on the same values. Figure 4.14 illustrates the convergence of three chains to stable distribution. The parameter being estimated in this case is δ_{lk} (the rate of change in year l for context k) for women in 1940 for /aw/ in pre-oral contexts. The top facet represents the full trace of the three chains. As can be seen,

in the first few iterations, the values being estimated for δ_{lk} vary broadly, but rapidly converge to a narrower range. Not all parameters converge this quickly, so as a general practice, the first half of the samples are discarded as a “burn-in.” The second half of the samples are taken to be representative of the posterior distribution, which is represented in the bottom facet of Figure 4.14.

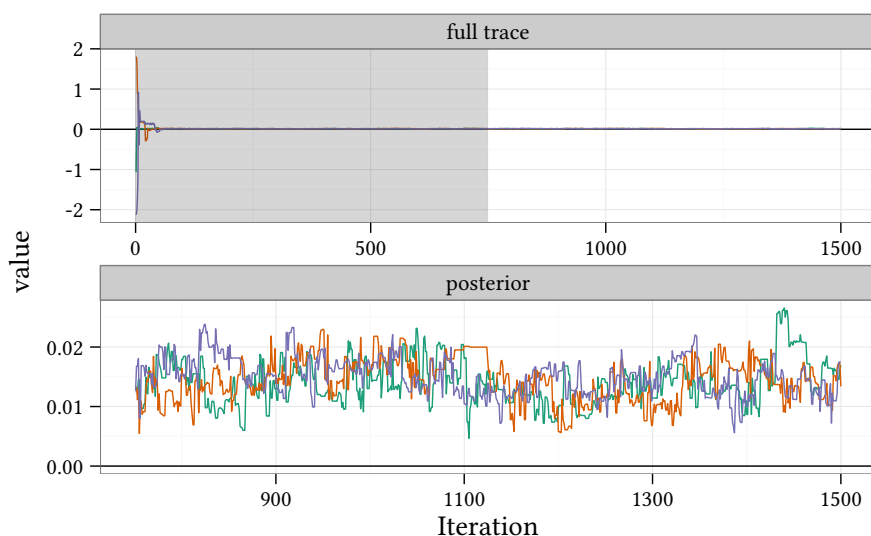


Figure 4.14: The full trace for three chains estimating δ_{lk} for $l = 1940$, and the sample approximating the posterior.

There are a number of different diagnostics for determining how well converged a model is. Note, these are not diagnostics of how well the model estimates match reality, which is unknown, but rather, how *consistent* the model’s estimates are. I will be employing the Gelman and Rubin Potential Scale Reduction Factor, represented by \hat{R} , which compares the between-chain variance to the within-chain variance. Values of \hat{R} close to 1 indicate good convergence, and the example in Figure 4.14 has $\hat{R} = 1.06$.

As a Bayesian model, it’s necessary to define prior probability distributions over the parameters it is going to estimate. I’ve already specified some of the model priors above. For example, I specified that the speaker-level target estimates, μ_{jk}^s should be normally distributed around the community-level estimate for that speaker’s date of birth, μ_{lk} . However, there are many parameters for which I have not mentioned what their prior probability distribution should be, like

the variance parameters, σ , σ_j^s , σ^w , or the b-spline weighting coefficients. These parameters, and any others not explicitly mentioned in the description above, were given non-informative, or weakly informative priors. Specifically, scale and variance parameters were given a uniform prior between 0 and 100, $\sim U(0, 100)$, and all other parameters were given a normal prior with mean 0 and standard deviation 1,000, $\sim \mathcal{N}(0, 1000)$. Given the scale of the data, which is z-score normalized Hz measurements, these constitute, at most, weakly informative priors.

4.3.3 I have just described a generative model

The model I have described is called a generative model in statistical terminology, because it describes a model of how the observed data was generated. That is, it models observations as being drawn from speakers speaking specific words, and speakers as being drawn from a larger and dynamically changing population. However, I believe there is also a felicitous convergence of terminology here with “generative” as it is used in Linguistics. To begin with, the specification of any statistical model is theory laden, and the reason I have specified the model above is driven primarily by the linguistic theory I want to evaluate, which is based on generative phonology and phonetics. Moreover, some of the parameters in the model correspond nicely to theoretical concepts in linguistics. Specifically, μ_{lk} , which represents the expected phonetic target of a speaker born in year l , could be understood as representing the “community grammar,” in the sense of Weinreich et al. (1968). Alternatively, it could just as easily be conceived of as representing the knowledge of

... an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky, 1965)

In the model, speaker-level factors such as memory limitations, distractions etc. are factored out by σ_{jk}^s to arrive at the idealized knowledge of each speaker, μ_{jk}^s . Community level factors, such as unaccounted for heterogeneity, is factored out by σ_k , to arrive at the idealized knowledge of an idealized speaker, μ_{lk} . The goal of this model is to determine what factors can account for

the idealized knowledge of an idealized speaker, which is also a goal of generative linguistics as I understand it.

4.4 Case Studies

All of the the cases studies presented here are based on data drawn from the Philadelphia Neighborhood Corpus (Labov and Rosenfelder, 2011). The measurements used are those produced by the FAVE suite (Rosenfelder et al., 2011), but additional contextual information has been collected from the PNC raw data.

Figure 4.15 is presented as background, and represents the trajectories of sound change in the 1970s as determined by the LCV study in Philadelphia. I will be examining the conditioning factors on /aw/, /ow/ and /uw/ here. Table 4.1 provides a broad IPA transcription for these vowel classes, their corresponding Wells Lexical Set labels, and an approximate transcriptions defining the range of phonetic variation.

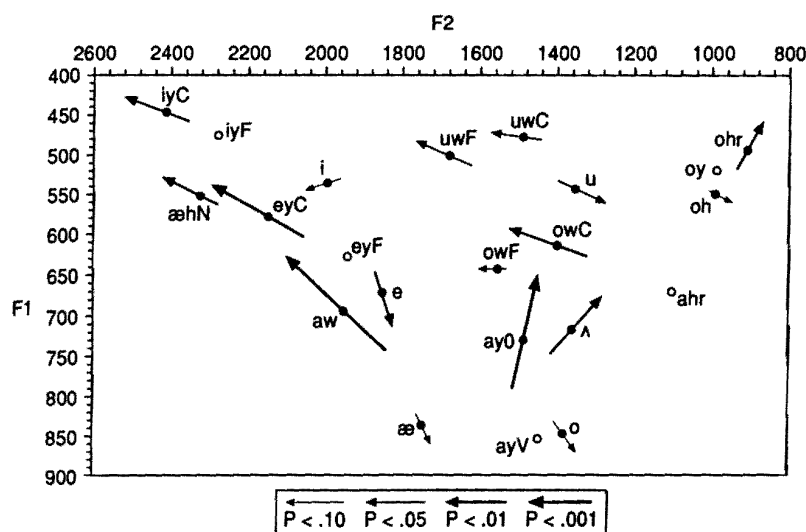


Figure 4.15: The Philadelphia Vowel System in the 1970s. From Labov (2001).

Before moving forward, I should note that for all of the following analyses, vowel tokens that were either word initial, or co-extensive with the word were excluded. I did this, in part, to exclude as many cases which could be attributable to errors in forced alignment, but also to reduce

Label	Broad IPA	Wells Lexical Set	Range of Variation
/aw/	[æw]	MOUTH	[eɔ]~[æʊ]
/ow/	[ow]	GOAT	[əʊ]~[o:]
/uw/	[uw]	GOOSE	[iʊ]~[u:]

Table 4.1: The case studies in this chapter, a broad IPA transcription, Wells Lexical Set labels, and IPA transcription of the range of variation.

the number of cases being examined. As we will see, the vowel in each case study is already subdivided in many ways, and including parameters for word initial and co-extensive tokens would have expanded the size of the statistical models further, without any clear advantages in return.

4.4.1 /aw/

In the 1970s, the fronting and raising of /aw/ along the front diagonal of the vowel space was identified as a vigorous change in progress (Labov, 2001). In the PNC, /aw/ raising has been found to be reversing, starting with speakers born around the 1960s. It also exhibits strong sociolinguistic conditioning, with a large difference between women, who are more advanced, and men (Labov et al., 2013). Figure 4.16 plots the basic, smoothed trajectory for /aw/ in $F1 \times F2$ space, overlaid on the full vowel triangle for context.

The PNC group has found that the best way to capture movements along the front diagonal of the vowel space is with a diagonal measure given as $(F2 - \beta F1)$, where β depends on the transformed scale of the data. In the z-score normalized space, which I am presenting here, the optimal value for β is 1, so Diag, in all future figures and statistics, is simply $(F2 - F1)$. Figure 4.17 displays the basic trajectory of /aw/ over time. Each point represents the mean value of Diag for one speaker.

The largest conditioning factor on the raising and fronting of /aw/ is whether it is followed by a nasal or oral segment (Labov et al., 1986). Figure 4.18 displays the mean values for speakers for /aw/ in pre-oral and pre-nasal contexts. Throughout the entire change, [awN] is consider-

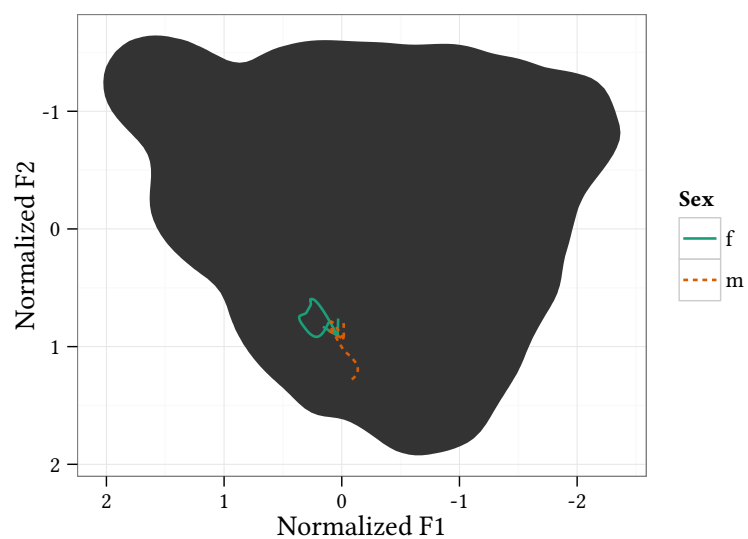


Figure 4.16: /aw/ Trajectory in $F1 \times F2$ Space

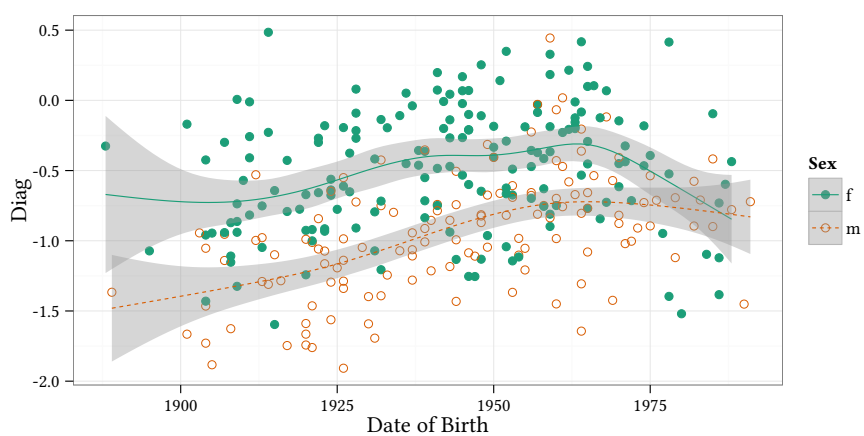


Figure 4.17: /aw/ change trajectory

ably more advanced along the Diag scale for both men and women. At this point, it appears impressionistically that the effect of following nasals remains consistent across the 20th century.

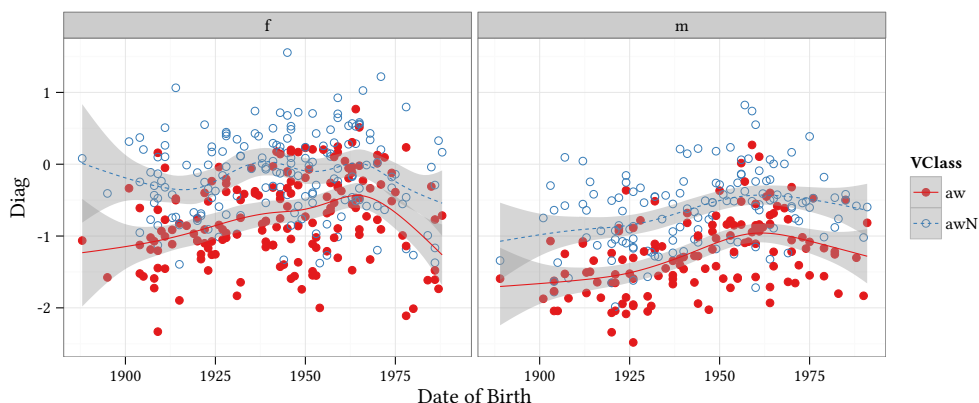


Figure 4.18: The effect of following nasals on /aw/.

Given that following nasals have such a strong effect on /aw/, I also coded /aw/ according to whether it was preceded by a nasal (to be represented by [Naw]), and whether it was sandwiched by nasals (to be represented by [NawN]). Also, for maximal parallelism between the rest of the analyses in this section, I also coded /aw/ for whether or not it was word-final, or followed by an /l/. Table 4.2 describes the coding criteria.

Variant	Criteria
awN	Is followed by /m, n, ŋ/ within the same word
Naw	Is preceded by /m, n, ŋ/, and is not word final
NawN	Is preceded and followed by /m, n, ŋ/ within the same word
awF	Is word final, and not preceded by /m, n, ŋ/
NawF	Is word final, and preceded by /m, n, ŋ/
awL	Is followed by /l/ within the same word
aw	Remaining cases

Table 4.2: Coding criteria for /aw/

Table 4.3 displays the token counts of each variant in the corpus. Given that [awL], [Naw] and [NawN] are relatively rare, and that the parameter δ_{lk} is already fairly abstracted away from the data, I will be excluding these contexts from further analysis in this section. That leaves [aw], [awN], [NawF] and [awF]. The results for [NawF] should be taken with some caution, however,

Variant	N	Word Types
aw	6382	150
awN	5494	147
NawF	2504	1
awF	1377	11
NawN	183	21
Naw	92	12
awL	14	5
total	16046	343

Table 4.3: Token counts of each variant

because even though it is relatively frequent, it consists entirely of tokens of the word *now*. As I mentioned in §4.3, a reference level must be chosen to compare the other variants to. In this case, that will be /aw/. As a first pass visualization of the data, Figure 4.19 plots cubic regression splines over speaker means for each of these /aw/ variants. For the most part, all variants seem to follow the same trajectory transposed up and down with the possible exception of [awF] for women. The extreme wobbliness of [awF] for women is almost certainly not “real”, but is rather a common result of sparse data for these curve fitting methods.

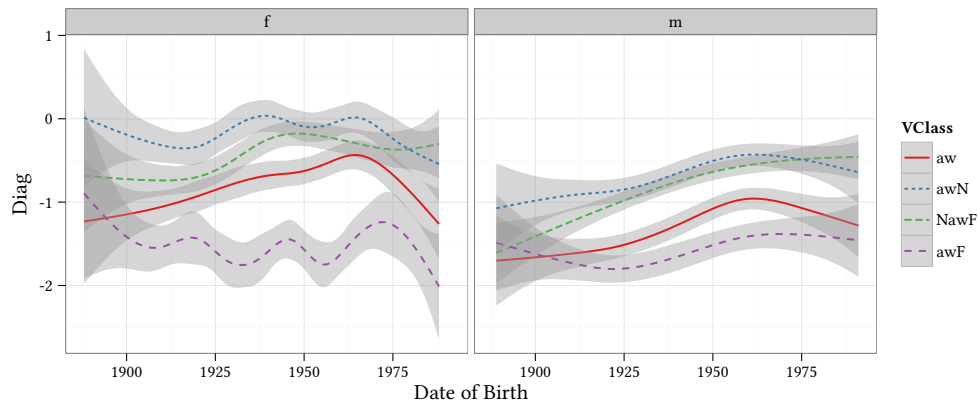


Figure 4.19: The /aw/ variants to be modeled.

Model Fit

As a first step to evaluating the quality of the model fit, we'll first examine the trajectory of change it predicted. Figure 4.20 displays the distribution of \hat{R} values for the estimated trajectories of [aw], [awN], [awF] and [NawF]. Values of \hat{R} indicate good convergence of the model. Unfortunately, the model appears not to have converged well at all for [NawF], with most of its \hat{R} values being greater than 2. As mentioned before, this is probably due to the fact that [NawF] is represented by just 1 lexical item in the corpus: *now*. For this reason, the data is actually sparser for [NawF] than the raw data might suggest, and the model estimation would face considerable ambiguity between attributing the target of [NawF] to its specific μ_{lk} value, or to its word-level effect, μ_m^w . Moving forward, [NawF] will be excluded from the analysis.

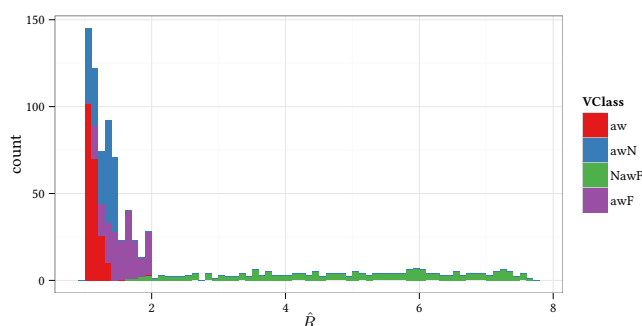


Figure 4.20: \hat{R} for the predicted trajectories of /aw/ variants.

Figure 4.21 displays the 95% Highest Posterior Density intervals for the predicted trajectories of change for [aw], [awN] and [awF]. There is a slightly larger probability range for the trajectory of [awF], but over all, these trajectories seem to fairly well fit, approximating closely the trajectories in Figure 4.19, suggesting that the model has not “blown up.”

Figure 4.22 plots the estimates of the parameter of central interest, δ_{lk} , representing the year-to-year differences. All three variants appear to share approximately the same rates of change, but [aw] appears to be slightly more exaggerated than the other two. In this figure, and the ones that follow, the color of the lines along the edges of the 95% HPD indicate whether or not 0 is excluded. As we can see in Figure 4.22, all three variants of /aw/ have significantly positive rates of change starting somewhere around the mid 1910's and continuing into the 1950's. For [awF]

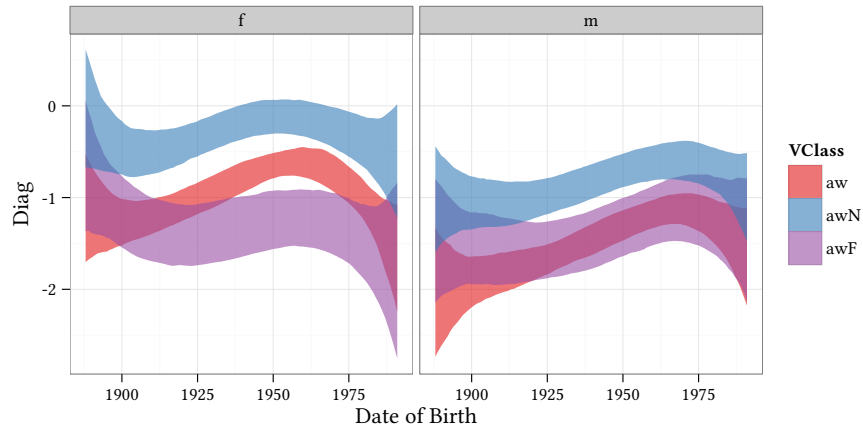


Figure 4.21: Predicted trajectories of change for /aw/ variants.

for women, the 95% HPD never actually excludes 0 in this time period, but its over-all trend is the same. For women, the rate of change for [aw] and [awN] turns negative as the change begins to reverse in the early 1960's.

However, the key comparison to make is whether [awF] and [awN] have reliably different rates of change from [aw]. This comparison is made in Figure 4.23, and as can be seen, neither [awF] nor [awN] exhibit considerable differences from [aw]. There is a brief period of about 5 years where [awF] seems to be changing more slowly than [aw], and in fact, looking at Figure 4.21, this is because [awF] appears to be moving downwards. Given the fact that this trend is so brief (less than 10 years), and that it is located so early in the sample, where there are fewer speakers, I'll attribute this blip to the the idiosyncrasies of a few speakers' data, rather than to a real trend.

Discussion

In the cases where there was enough data to make the comparison, it appears as if the different contextual variants of /aw/ share the same rate of change. Over the course of 100 years, [aw], [awN] and [awF] follow parallel trajectories, a remarkable fact in and of itself. My conclusion for /aw/ is that its most notable conditioning factor, the presence of a following nasal, is due to

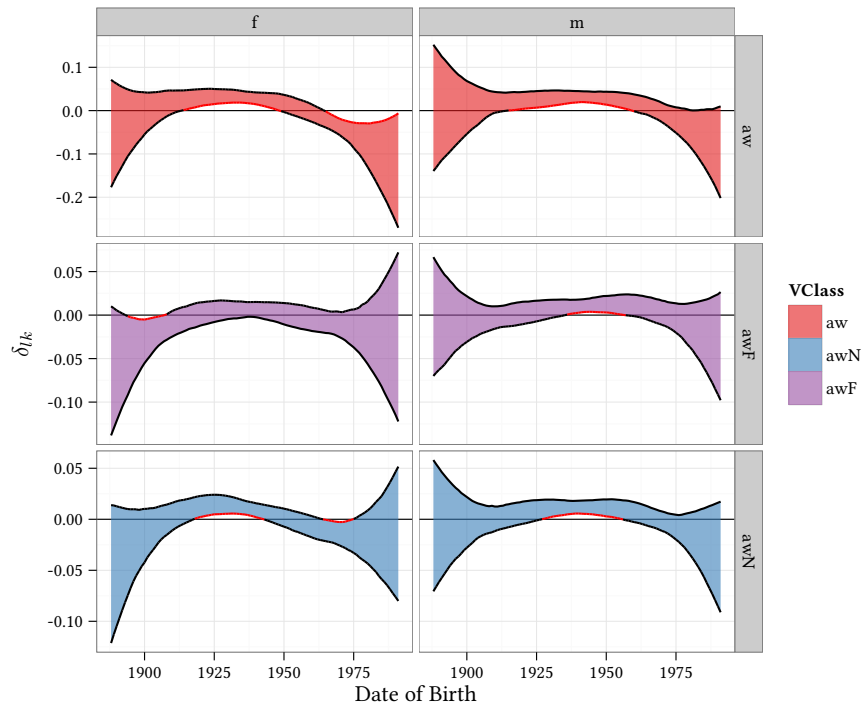


Figure 4.22: Year-to-year differences for variants of /aw/. Note: y-axis ranges differ across each horizontal set of facets.

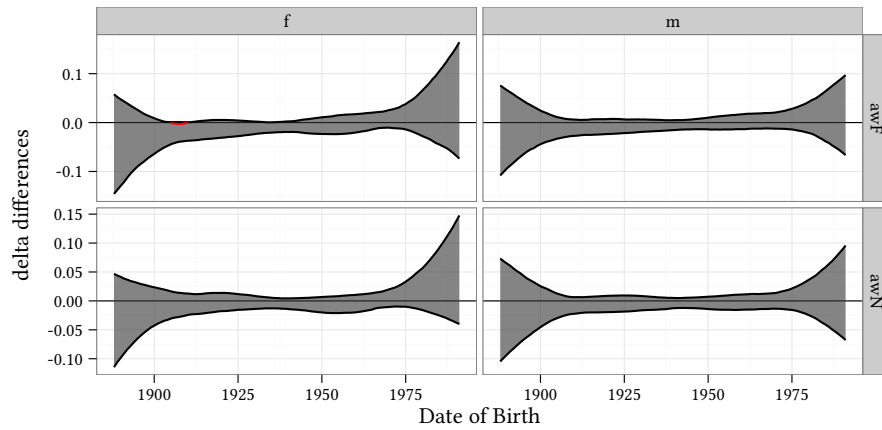


Figure 4.23: Rate of change differences from /aw/. Note: y-axis ranges differ across each horizontal set of facets.

phonetic coarticulation, not due to any categorical phonological process. That is, there is just one target for /aw/ that changes over time, and that target is merely shifted upwards in production by the presence of a following nasal.

4.4.2 /ow/

The next case study in this chapter is /ow/ fronting. Again, /ow/ fronting was found to be a change in progress in Philadelphia in the 1970's that has since began reversing (Labov, 2001; Labov et al., 2013). Figure 4.24 plots the trajectory of /ow/ in the F1×F2 space. Women underwent a fronting change which has reversed, but the pattern for men is a bit more ambiguous.

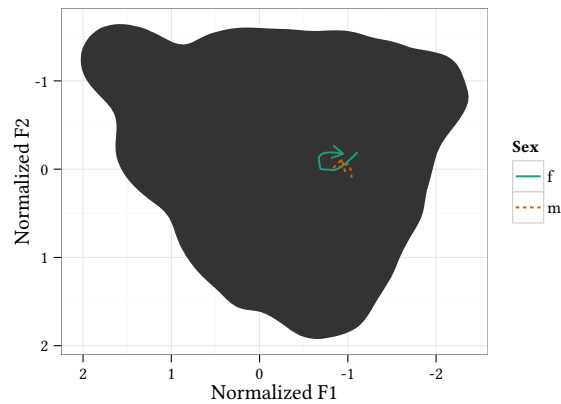


Figure 4.24: /ow/ Trajectory in F1×F2 Space

As a fronting change, the acoustic measure being used in this section is simply normalized F2. There are two major conditioning factors on /ow/ of note. First is whether or not /ow/ is absolute word final, which favors a more fronted form of /ow/, and whether or not it is followed by /l/, which favors a backer form of /ow/.

Figure 4.25 plots speaker means for these /ow/ variants. The first thing to note is that there is much stronger sociolinguistic differentiation between men and women for /ow/ than there was for /aw/. For /aw/, men lagged behind women, but were still participating in the change. For /ow/, it does not appear as if men undergo any change at all. The directions of the contextual effects are the same between men and women, but there is not much diachronic pattern to speak of for men. The second thing to note is that for women, the differentiation of [ow] and [owL] looks

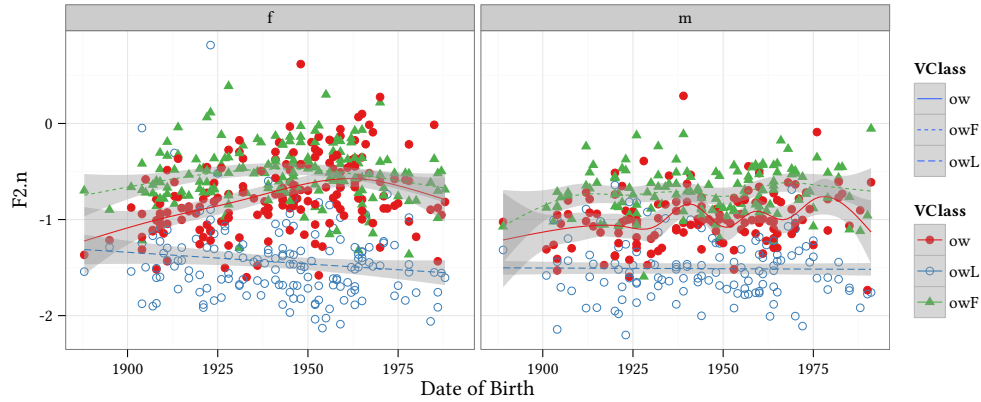


Figure 4.25: The effect of following /l/ and a word boundary on /ow/.

strikingly similar to the hypothetical patterns of phonological processes interacting with sound change from §4.2.

To keep the results for /ow/ as comparable to the results for /aw/ as possible, I coded /ow/ in the same way. Table 4.4 displays the coding scheme for /ow/, and Table 4.5 displays the total number of tokens for each variant.

Variant	Criteria
owN	Is followed by /m, n, ŋ/ within the same word
Now	Is preceded by /m, n, ŋ/, and is not word final
NowN	Is preceded and followed by /m, n, ŋ/ within the same word
owF	Is word final, and not preceded by /m, n, ŋ/
NowF	Is word final, and preceded by /m, n, ŋ/
owL	Is followed by /l/ within the same word
ow	Remaining cases

Table 4.4: Coding criteria for /ow/

As with /aw/, the /ow/ variant sandwiched by nasals is too low frequency to include in the model, and will be excluded from here on out. All of the other variants are relatively high frequency. Unsurprisingly, the highest frequency variant, [NowF], is dominated by the lexical items *no* and *know*, but is not exclusively constituted of them, so it should not exhibit the same ill fitting that [NawF] did. Figure 4.26 displays cubic regression splines over speaker means for the /ow/

Variant	N	Word Types
NowF	12956	33
owF	11092	200
owN	6091	164
ow	5629	592
owL	1946	153
Now	1836	123
NowN	95	11
total	39645	1253

Table 4.5: Token counts of each variant

variants to be included in the model. In the model, as with /aw/, the reference variant will be [ow], which corresponds to non-word final /ow/ that is not followed by /l/, and is neither preceded nor followed by a nasal consonant.

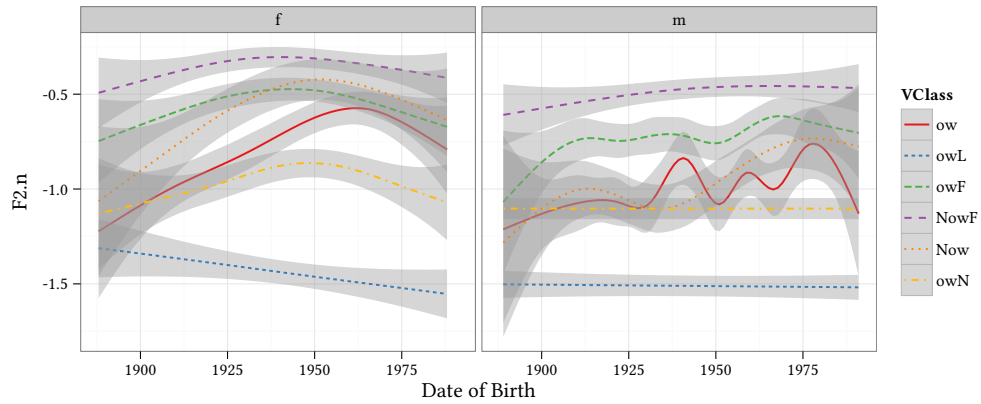


Figure 4.26: The /ow/ variants to be modeled.

Model Fit

Again, we'll examine how well the model estimated the basic trajectories of /ow/. Figure 4.27 displays the \hat{R} convergence estimates for the trajectories, broken down by variant. As with /NawF/, /NowF/ is the least well converged variant, but in this case, its \hat{R} values are not dire, so we will keep /NowF/ in for the rest of the analysis.

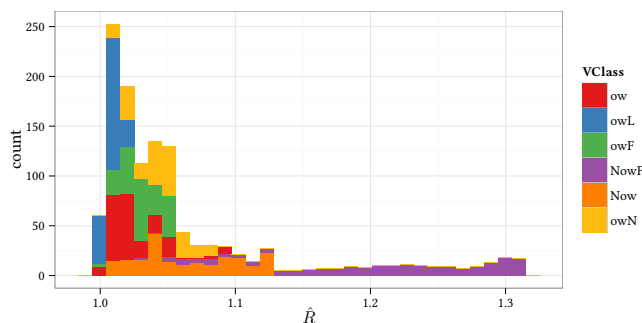


Figure 4.27: \hat{R} for the predicted trajectories of /ow/ variants.

Since there are six different variants of /ow/, I will not be plotting them over each other. Instead, Figure 4.28 plots a row of facets for each /ow/ variant and superimposes the predicted trajectory for [ow] on each one. All of the trajectories appear to be fit to comparable degrees of certainty, and look very similar to the trajectories in Figure 4.26. Again, there seems to be a striking difference between [owL] and all other variants. Every other variant of /ow/ (for women at least) appears to undergo some sort of fronting and subsequent backing, and [owL] is completely divergent from this pattern. The divergence of [owL] appears equally strongly when looking at the rates of change in Figure 4.29.

For women, [ow], [owF], [NowF], [Now] and [owN] all exhibit a very clear pattern of a positive rate of change in fronting beginning somewhere near the turn of the century, followed by a reversal starting around the 1960's. In contrast, the rate of change for [owL] is plausibly 0 throughout the entire century. The pattern for men is much more ambiguous. There is some hint of fronting for some variants centered approximately around the 1950's, but it is very subtle.

Finally, we come to the differences in the rates of change between [ow] and the other variants in Figure 4.30. Men have virtually no difference between [ow] and the other variants, so the rest of this discussion will focus exclusively on women.

As expected, [owL] has a reliably different rate of change from [ow] almost from the very beginning of /ow/ fronting. The earliest date of birth where [ow] has a rate of change reliably greater than 0 (as depicted in the top row of facets in Figure 4.29) is 1906, and the earliest [owL]

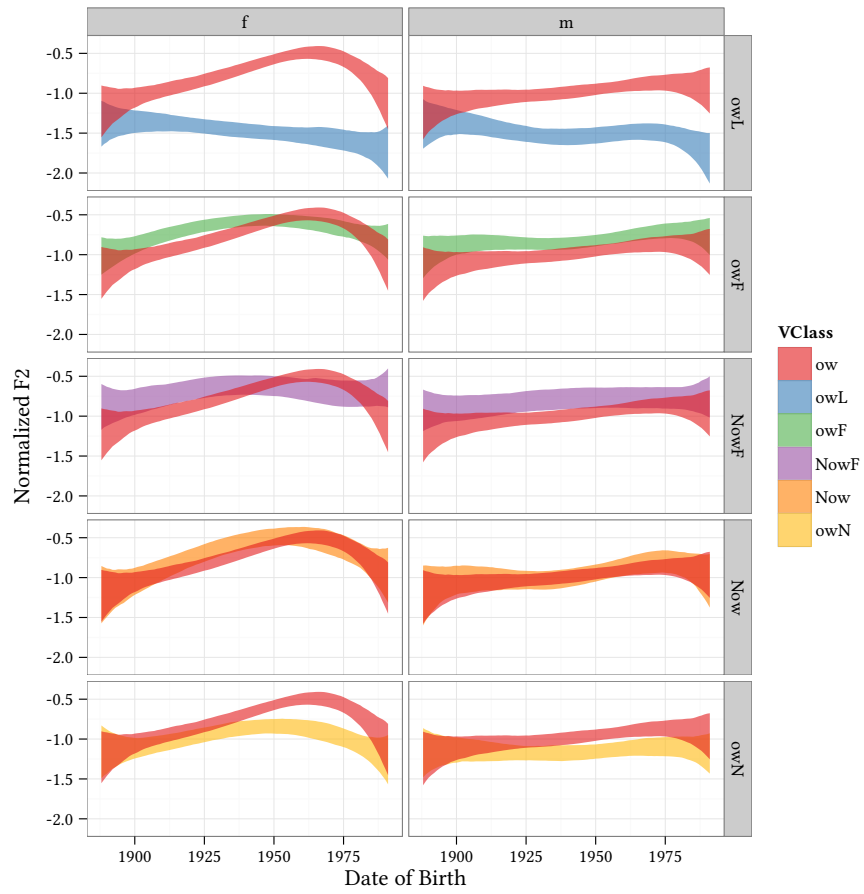


Figure 4.28: Predicted trajectories of change for /ow/ variants.

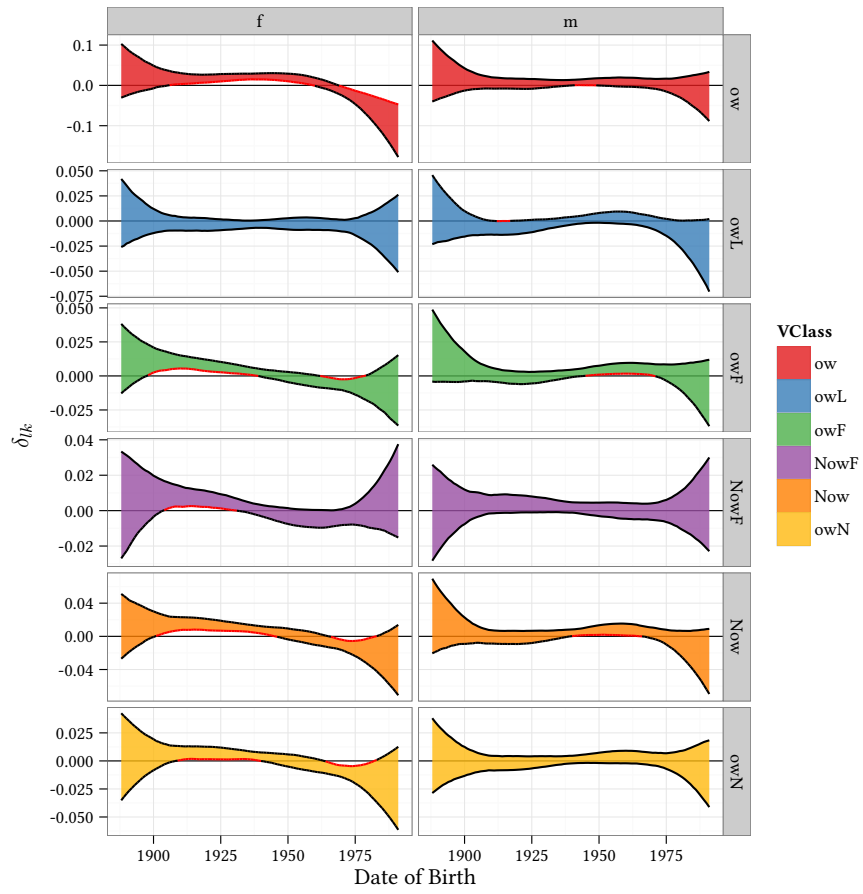


Figure 4.29: Year-to-year differences for variants of /ow/. Note: y-axis ranges differ across each horizontal set of facets.

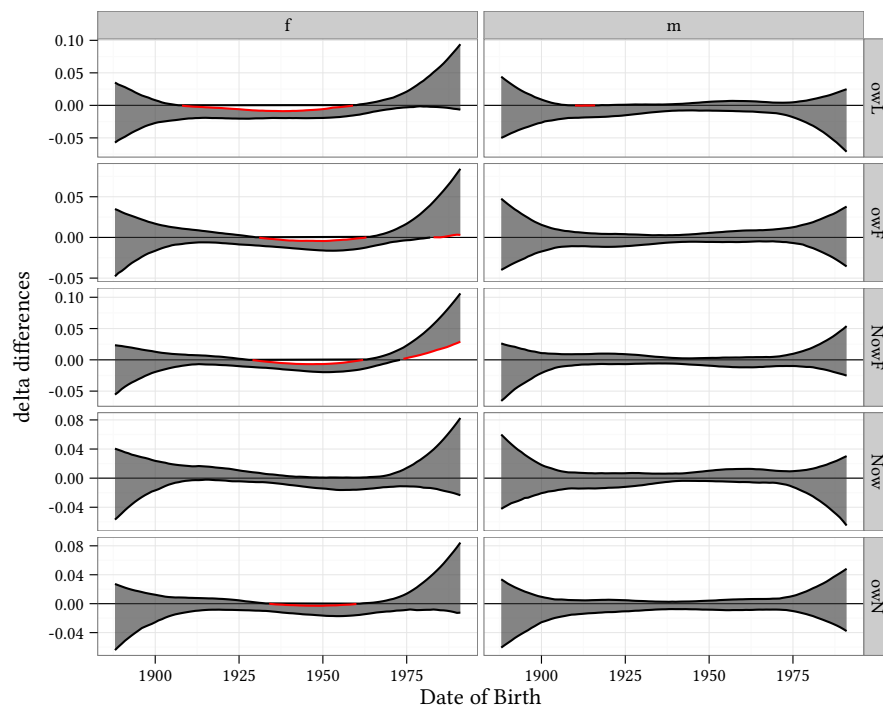


Figure 4.30: Rate of change differences from [ow]. Note: y-axis ranges differ across each horizontal set of facets.

exhibits a reliably different rate of change from [ow] is 1908. These dates are obviously overly precise, but support the interpretation where [ow] and [owL] have been categorically differentiated from the very beginning of /ow/ fronting.

Importantly, it is highly unlikely that the differentiation of [ow] and [owL] could be due to changing degrees of coarticulation between /ow/ and /l/. In this change, [ow] is undergoing a fronting change, and [owL] is being left behind. The estimated rate of change for [owL] (Figure 4.29) contains 0 across the entire century, meaning that if the difference between [ow] and [owL] were due to phonetic coarticulation, the strength of the coarticulation effect would have to be increasing exactly in proportion to the degree of frontness of [ow]. This is highly unlikely, and positing a categorical differentiation between [ow] and [owL] is the simpler explanation.

Surprisingly, even though [owF], [NowF] and [owN] share the same profiles as [ow] in both their over all trajectories and in their rates of change, there are some reliable differences between their rates of change. All four of these variants began fronting at approximately the same time around the turn of the century, but [ow] continued fronting until about 1960, while [owF], [NowF] and [owN] stopped fronting in the 1930's. Table 4.6 contains the model estimates for the dates when fronting began and ended, based on when the lower bound of the HPD for δ_{lk} excluded 0. [owF] and [NowF] also appear to be sluggish in participating in /ow/ retraction which began in the mid-1960's for [ow].

Variant	Began Fronting	Stopped Fronting
ow	1906	1959
owF	1898	1938
NowF	1904	1930
owN	1909	1939

Table 4.6: Dates that /ow/ variants began and stopped fronting, based on δ_{lk}

I think a reasonable analysis for [owF] and [NowF] is that we are observing a ceiling effect. Looking at the trajectories of change, [owF] and [NowF] in Figure 4.28 are the most fronted /ow/ variants, an effect that can be probably be simplified to simply being word final. If, for reasons which are unclear, there were a maximal degree to which /ow/ could be fronted phonetically, then it would make sense that [owF] and [NowF] would hit this limit first, and bottom out. In fact,

if we look at the trajectories of change for [ow], [owF] and [NowF], and examine the estimates for the maximum values of F2 these variants reached, we see that all three variants reached very similar peaks, even though the times at which they reached these peaks are spread out over 25 years. This suggests that there is, in fact, some ceiling value around $F2 \approx -0.5$ which [owF] and [NowF] hit first, slowing down their rates of change.

Variant	Maximum F2	Date of maximum
ow	-0.49	1964
owF	-0.56	1948
NowF	-0.61	1940

Table 4.7: Maximum /ow/ F2 values, and dates they were reached.

This ceiling analysis does not extend to [owN], however, which also stops fronting much earlier than [ow], but reaches a much lower peak value. The fact that [owN] could not possibly be slowing down due to a ceiling effect, and that its rate of change is reliably slower than [ow] starting in the mid 1940's, means that this may be a candidate example for the reanalysis of a phonetic effect as a phonological process.

Discussion

This analysis of /ow/ illustrates some interesting limitations of the rate of change diagnostic. First, in order to be able to diagnose anything at all, there must be a change occurring. In this case, the non-participation of men in /ow/ fronting meant that nothing can be said with much certainty about the phonetic and phonological status of contextual variants of /ow/ in their speech. Second, the analysis will be sensitive to ceiling and floor effects. Both [owF] and [NowF] phonetically favored fronter /ow/, and therefore reached the ceiling in the phonetic space (at approximately -0.5) first, which flattened out their trajectory of change, reducing their rate of change. I believe saying that the reduction in the rate of change of [owF] and [NowF] relative to [ow] is due to a ceiling effect is well founded, because [owF], [NowF] and [ow] all reached the same peak F2 value, but at different points in time. The fact that the rate of change for /owN/ slowed at a point which could not be considered a ceiling means that it should be held out as a potential case of a phonetic

bias becoming reanalyzed as a phonological process.

Finally, I will be analyzing the categorical exemption of /owL/ from fronting as a phonological distinction. For now, I will propose the process in (4.30) which I will further support in §4.5.

(4.30) $ow \rightarrow o:/__l$

Under this analysis, only /ow/ which has a phonological glide target is affected by /ow/ fronting, while the long monophthong remains fully back.

4.4.3 /uw/

The third and final back-upgliding diphthong I'll be analyzing is /uw/. /uw/ has also undergone a fronting change which has also been reversed. As with /ow/, men have had very limited involvement in /uw/ fronting. Figure 4.31 displays the basic trajectories of /uw/ for men and women in $F1 \times F2$ space. As with /uw/, the primary acoustic dimension describing the change is F2, so the following models will focus on normalized F2.

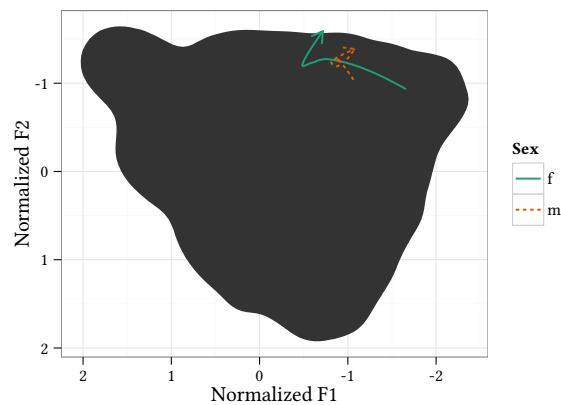


Figure 4.31: /uw/ Trajectory in $F1 \times F2$ Space

The two biggest conditioning factors on /uw/ fronting are following /l/, which favors backer /uw/, and preceding coronals, which favor fronter /uw/. The possible reasons for /l/ favoring backer /uw/ carry over from the discussion of /ow/. The effect of preceding coronals, however, is new. It is, in fact, a more general property of North American English that /uw/ tends to be fronter when preceded by a coronal, as discussed in the Atlas of North American English (Labov

et al., 2006, ch 12). The ANAE proposes that this phenomenon is related to the merger of /juw/ and /uw/ post-coronally in North America. For example, there are the differences between RP and “Standard American” presented in Table 4.8, with very broad transcriptions. The ANAE argument is that before /j/ was lost in this context in North America, it had the effect of fronting the nucleus of /uw/, which has persisted. Furthermore, /tu:/ sequences which never had a /j/ (e.g. *do* vs. *dew*) have merged to the fronted version. This is a classic phonologization argument, in the sense if Hyman (1976), and is briefly sketched out in Table 4.9.³

	RP	North American
<i>tube</i>	/tju:b/	/tu:b/
<i>tune</i>	/tju:n/	/tu:n/

Table 4.8: Examples of post coronal /j/ loss in North America

tju:b	>	tjiub	>	tiub
Initial state: little coarticulation		Coarticulation of /j/ and /uw/		Phonologization

Table 4.9: The phonologization of post-coronal /uw/ fronting

Another possible account for the coronal effect on /uw/ fronting is that it is simply a case of coarticulation. Figure 4.32 is an illustration of the coarticulatory effect of coronals on back vowels from Ohala (1981). This illustration is of anticipatory coarticulation, where the coronal follows the vowel, but it could extend in principle to the case here where the coronal precedes the vowel.

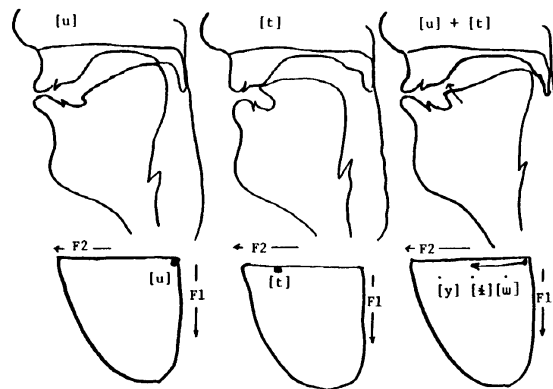


Figure 4.32: Illustration of the effect of coarticulation on /uw/ from Ohala (1981)

³A very similar account could be given the development of [tjub]>[tʃub] for many British speakers.

The difference between the ANAE account of the coronal effect and Ohala's is, in fact, precisely the difference between phonological and phonetic conditioning that I would like to use the rate of change analysis to resolve.

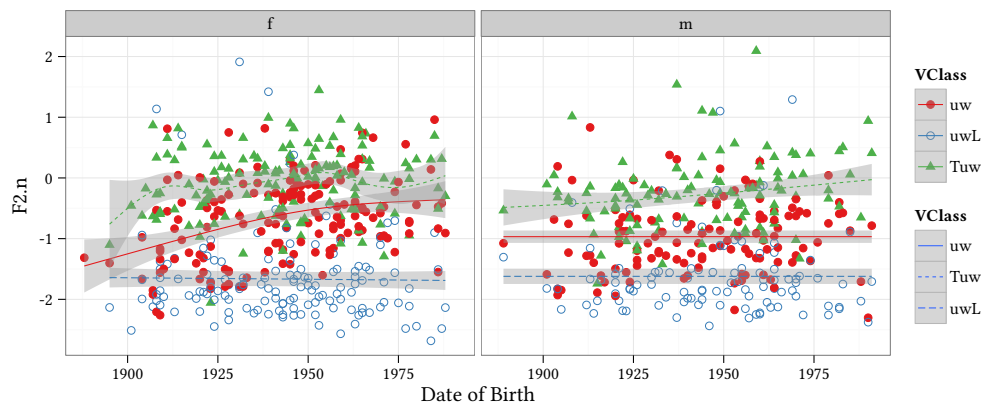


Figure 4.33: The effect of following /l/ and preceding coronal on /uw/.

Figure 4.33 displays the trajectories of the three basic variants to be investigated in this section. Again, it looks as if the [uwL] variant is categorically exempted from the change, but the pattern is more ambiguous for [Tuw]. Table 4.10 lists the coding criteria for /uw/ variants, and Table 4.11 displays the counts for each variant. I coded for whether /uw/ was followed by a nasal, /l/, or was word final, and for whether or not it was preceded by a coronal.

Variant	Criteria
TuwN	Is preceded by a coronal and followed by a nasal within the same word
TuwF	Is preceded by a coronal and is word final
TuwL	Is preceded by a coronal and followed by /l/ within the same word
Tuw	Remaining cases which are preceded by a coronal
uwN	Is followed by a nasal
uwF	Is word final
uwL	Is followed by /l/
uw	Remaining cases

Table 4.10: Coding criteria for /uw/

The most frequent variant, [TuwF], is composed mostly of the lexical items *do*, *to*, *two*, *too*, but also consists of a number of other lexical items, so it should be well modeled. The variants [uwN] and [TuwL] are too infrequent to model in this way, which is especially unfortunate for [TuwL],

Variant	N	Word Types
TuwF	9564	27
uw	2035	149
uwL	1474	31
Tuw	1461	180
uwF	883	24
TuwN	589	56
uwN	148	27
TuwL	78	16
total	16232	509

Table 4.11: Token counts of each /uw/ variant

since it could be crucial to see how a favoring and disfavoring context interact. Figure 4.34 plots the trajectories of the remaining variants which were fitted by the model. The reference variant in this model will be [uw], the variant which is not post-coronal, pre-nasal, pre-/l/, nor word final.

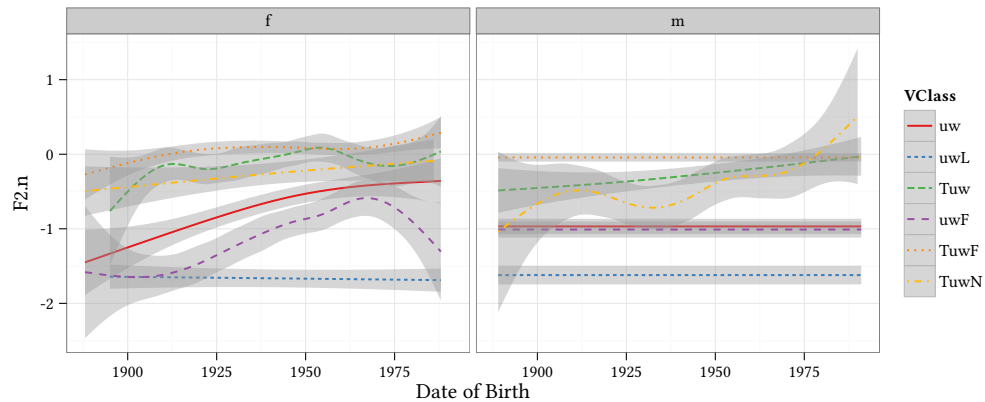


Figure 4.34: The /uw/ variants to be modeled.

Model Fit

Figure 4.35 displays the distribution of the \hat{R} convergence diagnostic for the estimated trajectory of /uw/ variants. [TuwF] and [uwF] have the largest \hat{R} values, but they are still acceptably close to 1 to include them in the analysis.

Figure 4.36 plots the estimated trajectories of /uw/ variants, following the same convention

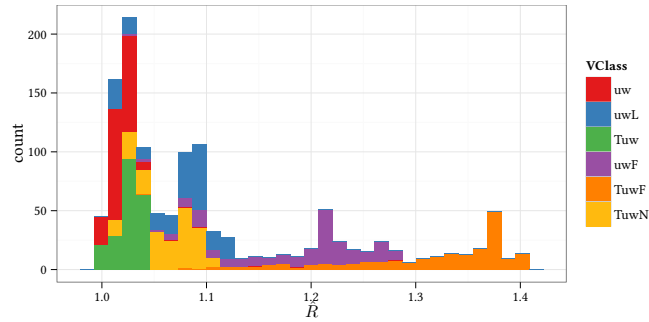


Figure 4.35: \hat{R} for the predicted trajectories of /uw/ variants.

used for /ow/, where the trajectory of the reference level, [uw], is superimposed over the trajectory of each other variant. The trajectories fitted by the model replicate the disfavoring and diverging effect of [uwL], as well as the favoring effect of a preceding coronal, which is stably in place regardless of the following segment, since [Tuw], [TuwF] and [TuwN] are all displaced upwards along F2. The sociolinguistic difference between men and women is also on display in Figure 4.36, where men appear to be only minimally participating in the change. Based on the results for /ow/, this will mean that the rate of change diagnostics for men will be only minimally informative.

Figure 4.37 plots the estimated rates of change for these /uw/ variants. [uw] has a positive rate of change starting at the turn of the century, as does [uwF] and most post-coronal variants. Before /l/, it looks like [uwL] has been completely flat over the course of the century, just like /ow/, and may have even undergone some retraction in the 1970s. A notable pattern for many of the post-coronal variants is a double dip, where they start out with a positive rate of change, level out, and then begin fronting again. There is a positive rate of change around the mid 1960s and 1970s for [TuwF] for women, and [Tuw] and [TuwN] for men, which is not present in either [uw] or [uwF].

Turning now to the crucial comparison of the rate of change of [uw] to the other variants, we see that [uwL] has had a reliably slower rate of change than [uw], unsurprisingly. A number of post-coronal variants also appear to have a period of time where their rate of change is slower

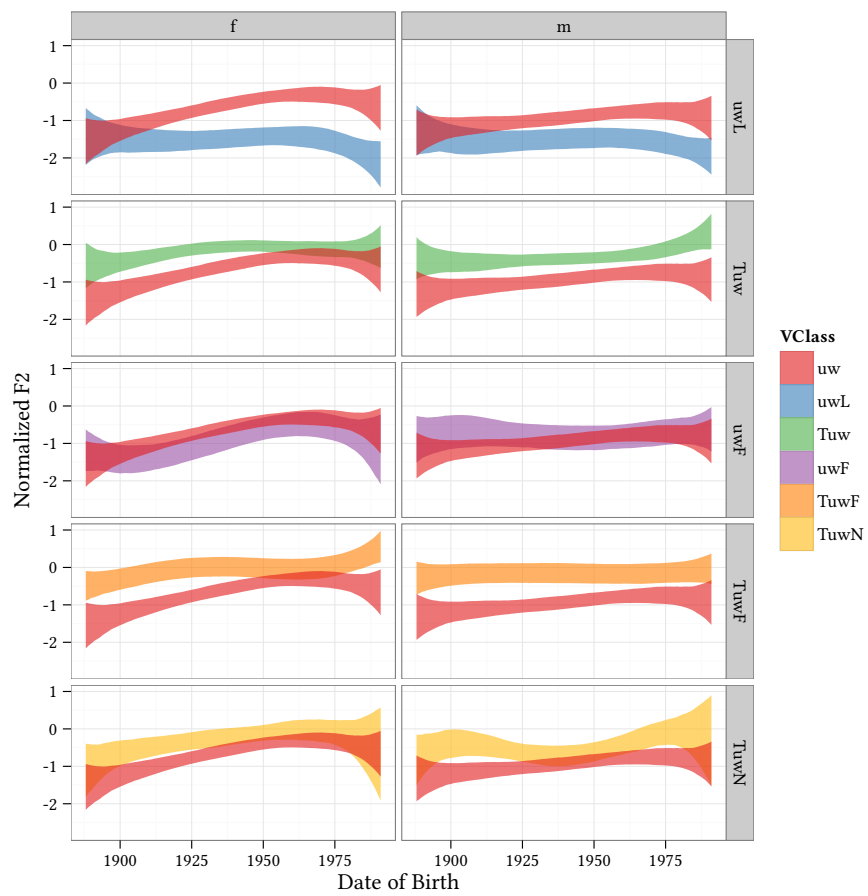


Figure 4.36: Predicted trajectories of change for /uw/ variants.

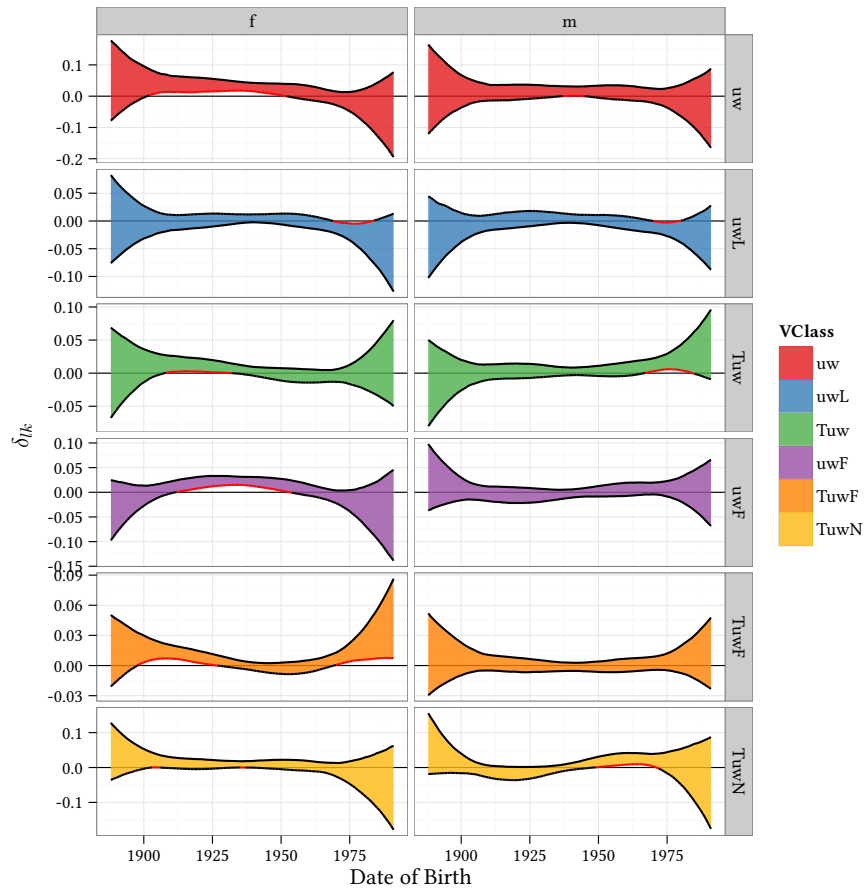


Figure 4.37: Year-to-year differences for variants of /uw/. Note: y-axis ranges differ across each horizontal set of facets.

than [uw]: [Tuw] and [TuwF] for women, and [TuwN] for men.

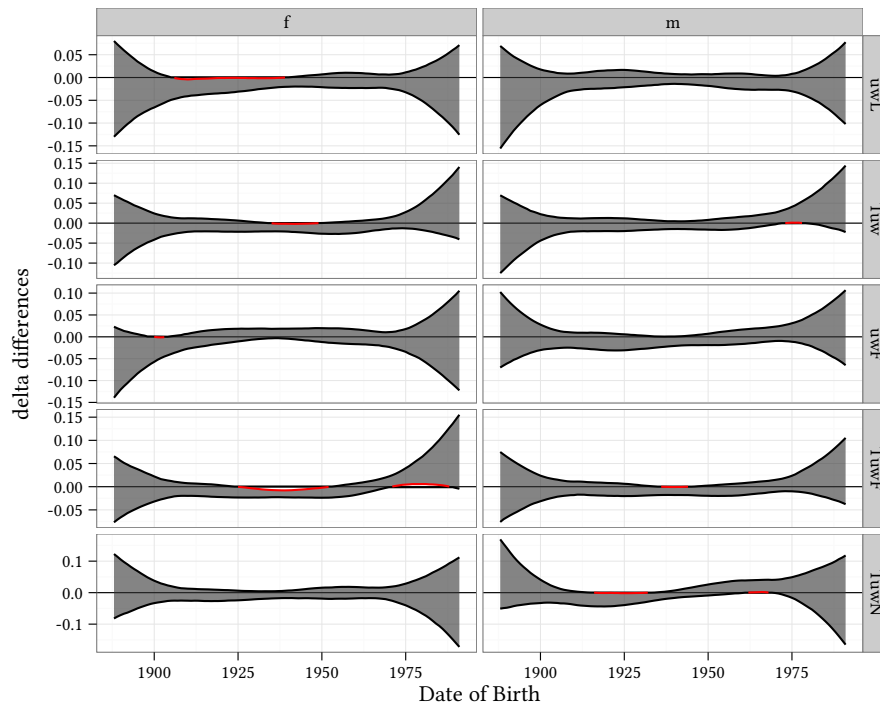


Figure 4.38: Rate of change differences from [uw]. Note: y-axis ranges differ across each horizontal set of facets.

Given that a preceding coronal favors /uw/ fronting, can we attribute the reliably slower rates of change for these post-coronal variants to a ceiling effect, like I proposed for word final /ow/? It is a possibility, but it seems less likely in this case. My argument for a ceiling effect on /ow/ rested on the fact that all of the /ow/ variants reached very similar maxima, but at different times. In the case of post-coronal /uw/, these variants had clearly not reached their maxima in the 1930s and 1940s, because in the 1960s and 1970s they began to front some more, as was seen in Figure 4.37. Perhaps this second phase of fronting for post-coronal /uw/ could be interpreted as being a phonological reanalysis of the coronal coarticulatory effect. If so, however, it occurred at the strangest time: when [uw] and [Tuw] were actually minimally different, as can be seen in Figure 4.36. It seems clear from the rate of change diagnostic that post-coronal /uw/ is phonologically distinguished from /uw/ in other contexts, but unfortunately, the timing of when this phonological differentiation entered the grammar relative to the phonetic fronting of /uw/ is ambiguous.

The effect of following /l/, on the other hand, is unambiguous. [uwL] has a reliably slower rate of change from [uw] almost as soon as [uw] begins fronting. Just like it was for /ow/, the pre-/l/ variant of /uw/ is being categorically exempted from ever undergoing the change.

Event	Year
[uw] is reliably fronting	1902
[uwL] is reliably slower than [uw]	1906

Table 4.12: Comparing the timing of [uw] fronting and the differentiation of [uwL]

4.5 Summary of /Vw/ results

Table 4.13 summarizes the results of the case studies just presented. The conditioning factors which were labeled in previous literature as having substantial effects are bolded. There is only one case which fits the profile of potential phonological reanalysis of a phonetic effect, [owN]. The rest of the unambiguous cases either exhibit parallelism throughout the change (with the exception of [owF], which is explicable by a ceiling effect), or were divergent from the very start of the change.

	/aw/	/ow/	/uw/
VwN	Always Parallel	Potential Reanalysis	–
NVw	–	Always Parallel	–
VwL	–	Divergent from start	Divergent from start
VwF	Always Parallel	Ceiling Effect	Always Parallel
TVw	–	–	Ambiguous

Table 4.13: Summary of /Vw/ Results

The possible source of the /l/ effect

The stand out effect here has been following /l/. For both /uw/ and /ow/, a following /l/ has categorically blocked fronting since the very beginning of the change. It is true that we would expect a following /l/ to disfavor the fronting of back vowels phonetically, especially in Philadelphia

where /l/ undergoes darkening and vocalization at a higher rate and in more environments than other North American dialects (Ash, 1982). But based on the argumentation in §4.1.1 and §4.2, the *categorical* nature of /l/ blocking means that /l/ must have had a categorical phonological effect on these vowels that was present in the grammar either before, or concurrent with the phonetic fronting of the other /ow/ allophones. In fact, I believe it is plausible to argue that this effect is due to glide simplification.

A rather salient feature of the Philadelphia dialect is that the glide of /aw/ is deleted when followed by an /l/. This is a rather old feature of the dialect, as it was mentioned by Tucker (1944), who said:

When *ou*, pronounced [æʊ], loses its second element, the result simply ‘flat a’: *hour* [æ:r], *owl* [æ:l], *Powell* [pæ:l], the latter two hardly to be distinguished from *Al* and *pal*.

In this same description of the Philadelphia dialect, Tucker also explicitly notes that Philadelphians make no distinction in the vowel quality of /ay/ before voiceless segments. The PNC data shows pre-voiceless /ay/ raising entering into the Philadelphia dialect with speakers born in the 1920s, meaning that we can reasonably place this process of glide deletion as being present in the dialect well before that. We could formulate the process of /aw/ glide deletion as follows.

(4.31) æw → æ: / ___l

This effectively captures the argument of Dinkin (2011) that in Philadelphia /æ/ has merged with /awl/, leading to the raising and tensing of /æ/ before /l/. Dinkin (2011) argues that rather than this being an extension of the Philadelphia split short-a pattern, this tensing and raising of /æ/ before /l/ is actually because its phonetic target is that of the nucleus of /aw/.

It would be reasonable to extend this process to cover both /ow/ and /uw/. To begin with, the phonetic realizations of the fully back /ow/ and /uw/ before /l/ is as long monophthongs: *bowl* [bo:ʷ],⁴ *school* [sku:ʷ]. Secondly, it is reasonable to categorize /aw/, /ow/ and /uw/ as belonging to a phonological natural class. The common patterning of these vowels cross-dialectally was the

⁴I’m using [ʷ] to represent the extreme reduction of /l/ to a velar approximant. /l/ is not always vocalized in these contexts, and the coda can be occasionally accompanied by light velar frication.

motivating force behind the Labovian transcription conventions for these vowels, which I have largely adopted, and in Philadelphia, we observe all three undergoing a simultaneous fronting and reversal in parallel, a phenomenon which will be covered in depth in later chapters. The extension of glide deletion to /ow/ and /uw/ could be formulated as follows.

$$(4.32) \text{ Vw} \rightarrow \text{V:} __\text{ l}$$

This could be reformulated in moraic terms. We start out with an underlying form like (4.33). The /Vw/ glide is then delinked, triggered by the following /l/ (4.34). This may be an OCP effect of some sort, especially if /l/ is really a velar approximant in this position. Crucially, the mora originally associated with the glide becomes associated with the nucleus, creating a long monophthong (4.35). This would have the effect of exempting /ow/ and /uw/ from fronting in this context, because fronting only affects the nucleus of these vowels, i.e. the first mora.

$$(4.33) \begin{array}{ccc} \mu & \mu & \mu \\ | & | & | \\ \text{V} & \text{w} & \text{l} \end{array}$$

$$(4.34) \begin{array}{ccc} \mu & \mu & \mu \\ | & / & | \\ \text{V} & \text{w} & \text{l} \end{array}$$

$$(4.35) \begin{array}{ccc} \mu & \mu & \mu \\ | & / & | \\ \text{V} & & \text{l} \end{array}$$

4.5.1 Connection to Broader Theory

I am proposing that the glide deletion process discussed above was categorical, and must have been present in speakers grammars at the very start of the phonetic change that fronted /ow/ and /uw/ in all other contexts in order to categorically block fronting. Since neither [uwL] nor [owL] ever underwent fronting, and since the phonetic difference between [uw]~[uwL] and [ow]~[owL] was still very small at the time that categorical blocking was in place, this phonological process was not the reanalysis of phonetic coarticulation. This supports my general argument that phonetic change operates over the representations produced by a distinct phonology, and

that theories of sound change based solely in phonetics are insufficient to capture the facts of all, or most sound change.

4.6 Conclusion

In this chapter, I have laid out my definition of a phonetic effect, or phonetic coarticulation, in contrast to phonological differentiation, and examined how these different phenomena ought to interact with sound change. Importantly, my definition of an effect being phonological or phonetic is based on which domain of the sound system the effect originates, *not* on its size. It is possible under the model of the phonology-phonetics interface I have adopted for effects originating in the phonetics to be large, and produce discrete non-overlapping distributions, and for effects originating in the phonology to be small, and produce partially overlapping distributions. Before even examining empirical case studies, the phonology-phonetics interface model first described in Chapter 2, and fleshed out in more detail in this chapter, when combined with diachronic change, produces what I'll call the "Unity Principle," and is very similar to what Kroch (1989) called the "Constant Rate Effect."

(4.36) THE UNITY PRINCIPLE

If two contextual variants have the same surface phonological representation, then they must shift in parallel diachronically. Contrapositively, if two contextual variants have divergent diachronic trajectories, they must not have the same surface phonological representation.

The Unity Principle is not itself a falsifiable hypothesis, but rather a logical consequence of the phonology-phonetics model I have adopted. It serves as a tool for investigating the interaction of phonology and phonetics over the course of phonetic change.

Applying the Unity Principle to the fronting of /aw/, /ow/ and /uw/ was successful in terms of discovering new details about these particular changes and the broader generalizations they imply, as well as a demonstrating the utility of the Unity Principle for phonological investigation. Some conditioning effects, like that of following nasals on /aw/, appear to be strictly phonetic,

despite for their large effect size, because they move in parallel over the entire course of the change. Many apparent exceptions to parallelism are reasonably understood in terms of ceiling effects. The biggest exception to parallelism was for /ow/ and /uw/ when followed by /l/. These variants appeared to be categorically exempt from fronting.

Faced with the divergent trajectories of [ow ~ owL] and [uw ~ uwL], it follows from the unity principle that these variants must have different surface phonological representations, meaning we must posit either a phonological process to differentiate them, or a different underlying form. The fact that it is *necessary* to posit a phonological analysis in the face of the diachronic data combined with the Unity Principle speaks to both its utility, and the importance of diachronic data for phonological investigation. Any given snapshot of [ow ~ owL] [uw ~ uwL] based on a demographically narrow set of speakers would be ambiguous, and open for reasoned argument for either a purely phonetic or purely phonological explanation. It is the diachronic dimension in combination with a moderately articulated model of the phonology-phonetics interface which disambiguates the two sources of explanation, and opens the door for more detailed inquiry.

The more surprising result with broader implications for language change in general is the relative timing of the phonetic change which began fronting [ow] and [uw], and when the phonological process differentiating [owL] and [uwL] must have been in the phonological grammar. Rather than the slow and gradual reanalysis of coarticulated [owL] and [uwL] as being their own phonological allophones over the course of the change, they behaved as categorically distinct allophones from the very beginning of the change. This effect of early phonological differentiation in phonological change will reappear in Chapter 5, and is unexpected under most accounts of conditioned sound change where the introduction of a phonological process follows a period of accumulated phonetic errors (Ohala, 1981; Blevins, 2004). Even Janda and Joseph (2003) who propose a “Big Bang” model of phonologization of sound change include a “brief” period of purely phonetic conditioning. The results presented in this chapter and at the beginning of Chapter 5 suggest that if there is a brief period of pure phonetic conditioning, it is too brief to be detectable by statistical methods. In fact, the available data is equally consistent with phonologization occurring *simultaneously* with the onset of the phonetic change.

Bibliography

- Ash, Sherry. 1982. The Vocalization of /l/ in Philadelphia. Ph.d., University of Pennsylvania.
- Bermúdez-Otero, R. 2007. Diachronic Phonology. In *The Cambridge Handbook of Phonology*, ed. Paul de Lacy, chapter 21, 497–517. Cambridge: Cambridge University Press.
- Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. New York: Cambridge University Press.
- Boersma, Paul, and Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25:217–270.
- Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14:261–290.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Cohn, A. 2006. Is there Gradient Phonology? In *Gradience in grammar: Generative perspectives.*, ed. Gisbert Fanselow, Caroline Féry, Ralf Vogel, and Matthias Schlesewsky, 25–44. Oxford: Oxford University Press.
- Cohn, Abigail. 1993. Nasalisation in English: Phonology or Phonetics. *Phonology* 10:43–81.
- Cohn, Abigail. 2007. Phonetics in Phonology and Phonology in Phonetics. *Working Papers of the Cornell Phonetics Laboratory* 16:1–31.
- Dinkin, Aaron. 2011. What’s Really Happening to Short A Before L in Philadelphia?
- Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:7–44.
- Hoffman, Matthew D, and Andrew Gelman. 2011. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv.org* URL <http://arxiv.org/abs/1111.4246>.
- Hyman, Larry M. 1976. Phonologization. In *Linguistic studies offered to Joseph Greenberg on the Occasion of His Sixtieth Birthday*, ed. Alphonse Juilland, A.M. Devine, and Laurence D. Stephens, volume 4, 407–418. Anima Libri.
- Janda, Richard D, and Brian D Joseph. 2003. Reconsidering the Canons of Sound-Change: Towards a “Big Bang” Theory. In *Historical Linguistics 2001. Selected Papers from the 15th International*

- Conference on Historical Linguistics, Melbourne*, ed. Barry Blake and Kate Burridge, 205–219. John Benjamins.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.
- Kruschke, John K. 2011. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Oxford: Elsevier.
- Labov, William. 1981. Resolving the Neogrammarian Controversy. *Language* 57:267–308.
- Labov, William. 1994. *Principles of linguistic change. Volume 1: Internal Factors*. Oxford: Blackwell.
- Labov, William. 2001. *Principles of linguistic change. Volume 2: Social Factors. Language in Society*. Oxford: Blackwell.
- Labov, William. 2010. *Principles of Linguistic Change. Volume 3: Cognitive and Cultural Factors*. Oxford: Blackwell.
- Labov, William, Sherry Ash, and Charles Boberg. 2006. *The Atlas of North American English*. New York: Mouton de Gruyter.
- Labov, William, and Maciej Baranowski. 2006. 50 Msec. *Language Variation and Change* 18:223–240.
- Labov, William, David Graff, and Wendell A. Harris. 1986. Testing listeners' reactions to phonological markers. In *Diversity and Diachrony*, ed. David Sankoff, 45–58. Philadelphia: John Benjamins.
- Labov, William, and Ingrid Rosenfelder. 2011. The Philadelphia Neighborhood Corpus.
- Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One hundred years of sound change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language* 89:30–65.
- Ohala, John. 1981. The listener as a source of language change. In *Papers from the parasession on language and behavior*, ed. C S Masek, R A Hendrick, and M F Miller. Chicago Linguistics Society.
- Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In *Papers in laboratory phonology 1: Between the grammar and physics of speech*, ed. John Kingston and Mary E Beckman, 258–282. Cambridge: Cambridge University Press.
- Pierrehumbert, Janet B. 2002. Word-specific phonetics. In *Laboratory Phonology VII*, 101–139. Mouton de Gruyter.
- R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL <http://www.r-project.org/>.
- Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite. URL <http://fave.ling.upenn.edu/>.

- Stan Development Team. 2012. Stan: A C++ Library for Probability and Sampling. URL <http://mc-stan.org/>.
- Strycharczuk, Patrycja. 2012. Phonetics-phonology interactions in pre-sonorant voicing. Ph.d., University of Manchester.
- Tucker, R Whitney. 1944. Notes on the Philadelphia dialect. *American Speech* 19:37–42.
- Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics*, ed. W Lehmann and Y Malkiel. U. of Texas Press.