The regularity of regular sound change

William Labov

Language, Volume 96, Number 1, March 2020, pp. 42-59 (Article)

Published by Linguistic Society of America

THE REGULARITY OF REGULAR SOUND CHANGE

William Labov

University of Pennsylvania

The regularity of sound change as set out by the scholars of the late nineteenth century is a fundamental principle of historical linguistics. The principle as recognized by the Neogrammarian linguists states that once a sound change has begun, it affects every word in the vocabulary that contains the sound in question. The principle has been disputed by many linguists and especially dialectologists, who argue that 'every word has its own history'. This article demonstrates how the Neogrammarian principle operates in one prototypical change in progress, the raising of the mid front long vowel /ey/ before a consonant in Philadelphia English. Mixed-level regression analysis shows consistent phonetic constraints across the nineteenth century, with no effect of word frequency. The regularity of sound change is reflected in the common pattern of behavior of the most frequent words, those of moderate frequency, and words that occur only once in the corpus.*

Keywords: sound change, regularity, speech community

This is a study of the fundamental mechanism of sound change, which can be defined as the TRANSITION PROBLEM: how does a change proceed from one stage to another in the course of time (Weinreich, Labov, & Herzog 1968)? Or to put it more simply, what is it that changes? The history of linguistics has left us with two views of the matter: that it is the phonetic realization of a phoneme, or that it is the pronunciation of particular words. Though there are many sources of irregularity, this study argues that the fundamental process is regular. I develop this view by focusing on one change in progress from its earliest stages, in which it proceeded in the same direction for over a century with no trace of social awareness.

1. THE NEOGRAMMARIAN CONTROVERSY. All such discussions begin with the Neogrammarian position familiar from the 1878 statement of Osthoff and Brugmann: 'Every sound change, inasmuch as it occurs mechanically, takes place according to laws that admit no exception' (Osthoff & Brugmann 1878, translated in Lehmann 1967:204). This position was challenged early by dialectologists: 'The phonetic *law* does not affect all items at the same time: some are designed to develop quickly, others remain behind, some offer strong resistance and succeed in turning back any effort at transformation' (Gauchat, cited in Dauzat 1922), and more recently by Wang and his colleagues, who revived and reinvigorated the position that the fundamental mechanism of sound change is LEXICAL DIFFUSION: 'A phonological rule gradually extends its scope of operation to a larger and larger portion of the lexicon, until all relevant items have been transformed by the process' (Chen & Wang 1975:256).

This position does more than allow for lexical diffusion. It states that the Neogrammarians were fundamentally wrong, that the unit of change is never the phoneme but always the word: 'The lexically gradual view of sound change is incompatible, in principle, with the structuralist way of looking at sound change' (Chen & Wang 1975:257).

Some thirty years ago, I attempted a resolution of this controversy by arguing that both modes of change were to be found, and that the problem was to discover when lexical diffusion occurred and when regular sound change occurred (Labov 1981). It was

* This article has benefited from contributions and criticism from many colleagues, in versions dating from 2013, and in particular from Lyle Campbell, Joseph Fruehwald, Jennifer Hay, Donald Ringe, Gillian Sankoff, Betsy Sneller, Meredith Tamminga, and Charles Yang. Support from the National Science Foundation from 1974 to 2012 is gratefully acknowledged under contracts 75000245, 7680910, 781607, 8813415, 42061, and 921643.

proposed that regular sound change is the result of a gradual transformation of a single phonetic feature of a phoneme in a continuous phonetic space, and that lexical diffusion is the result of the abrupt substitution of one phoneme for another in words that contain that phoneme. Thus a characteristic case of lexical exceptions in the historical record is the shortening of the English long mid vowel /ɛː/ in *bread*, *head*, *instead*, *lead*, *read*, and so forth, leaving behind *bead*, *mead*, *plead*, and the present tense of *read* and *lead*. A characteristic regular sound change is the diphthongization of the high front long vowel /iː/ in *mine*, *wide*, and *try* (Labov et al. 2006:Ch. 12). No lexical exceptions have been recorded in the many British and American dialects where the sound change considered here has taken place.

This resolution fell short of success. In the years that followed, advocates of Chen and Wang's position continued to find cases of lexical diffusion, but I do not know of any reports of regular sound change that appeared in subsequent years. Following the collection of articles in Wang 1977, I find sixteen further papers reporting lexical diffusion (e.g. Phillips 1980, 1983, 1984, 1989, 2006, Wallace 1984, Lien 1987, Ogura 1987, Harris 1989, Shen 1990, Krishnamurti 1998, Bybee 2002). None of these authors include findings of regularity of the sound changes they studied.

Yet the practitioners of historical and comparative linguistics have not changed their view of the matter. In the face of so much evidence of lexical exceptions, it is common to reduce regularity to a working principle rather than a substantive statement. As judiciously stated in Campbell's assessment:

> The assumption of regularity is extremely valuable to the application of the comparative method. Knowing that a sound changes in a regular fashion gives us the confidence to reconstruct what the sound was like in the parent language from which it comes. If a sound could change in unconstrained, unpredictable ways, we would not be able to determine from a given sound in a daughter language what it may have been in the parent language, or, looking at a particular sound in the parent language, we could not determine what its reflexes in its daughter languages would be. (Campbell 2004:166)

This is the view of the historical linguists who contributed to the volume *The comparative method reviewed: Regularity and irregularity in sound change* (Durie & Ross 1996). In general, they recognize the existence of lexical irregularities but propose a separate accounting for them. Campbell lists a narrowly circumscribed set of conditions in which exceptions to the regularity of sound change are to be found, the most common being analogy and borrowing. Yet he concludes that 'the general assumption of regularity for sound change is necessary in order to recognize the potentially exceptional forms' (Campbell 1996:86). Blust (1996:153) argues that 'irregularity appears to be an integral part of the natural process of language change'. Recognizing the probabilistic character of lexically gradual sound changes, Durie (1996:131) concludes that some 'will not necessarily be exceptionless'.

Given the proliferation of reports of irregularity, we still have to determine what mode of change is more general in the historical record. Ringe and Eska (2013) address this issue vigorously by an examination of the first 200 words of the glossary in an Old English textbook, Moore & Knott 1955. They find that 88% of the contemporary reflexes can be derived by regular sound changes and known morphological changes. Considering the paths of all segments in these words, they report that no more than 3% per millennium are irregular.

Despite their support for regular sound change, this finding of Ringe and Eska is nevertheless consistent with a view that the basic mechanism of change is lexical diffusion across the vocabulary but that the sweep through the lexicon eventually leads to a regular outcome when the choice of lexical items is exhausted. It is argued that when some factor interrupts the sound change in progress, the end result is not regular but irregular, since items not yet selected are stranded in an earlier form (Wang 1969). On this view, the outcome of this word-by-word process may be regular, but the process is not. A review of the regularity of completed changes cannot illuminate the process itself. Here I follow the lead of Chen and Wang in their original statement:

> Oddly enough, one of the most neglected aspects of historical linguistics, which professes to be a study of language evolving across time, is the time element itself. Of course much discussion has been devoted to the relative chronology of phonological processes, but this concerns the EXTERNAL relation between rules in terms of time sequence; the INTERNAL time dimension has not received equal attention until fairly recent times. By internal dimension we mean the CHRONOLOGICAL PROFILE … of the gradual evolution, and expansion or regression, of a single phonological process. (Chen & Wang 1975:256, emphasis in the original)

2. THE ROLE OF FREQUENCY IN LEXICAL DIFFUSION. The issue addressed here goes beyond the mechanism of sound change and involves the question of how language is produced and perceived through the implementation of rules operating on abstract categories. The Neogrammarians were aware that the exceptionless character of sound change demanded an explanation, which is to be found in the rule-governed nature of language as a whole. Thus Paul: 'A particular sensation does not form itself specially for every word, but in every case where the same elements recur in language their production is guided by the same sensation' (1891:59).

According to recent, exemplar-based theories of language production and perception (Kruschke 1992, Pierrehumbert 2001, 2002, Bybee 2002, Todd et al. 2019), lexical effects on sound change involve not the discrete SELECTION of a word to participate in the change but a more continuous correlation of change with the frequency of the word involved. In the words of an anonymous referee of the current article, 'exemplars of words are added to the lexicon. Since frequent words are heard more often, there will be more exemplars of frequent words. Since production samples from the same space as perception, then an exemplar is more likely to be chosen that contains the characteristics of frequent words'. This was indeed the position of the authors of the study of change in the short front vowels of New Zealand English (Hay et al. 2015), a study comparable in size and scope to the Philadelphia project to be discussed here. This view of change projects a continuum across the entire word class in which all words are affected to a degree depending on their frequency, as opposed to the Wang and Chen view of abrupt irregularities in the development of word classes.

Accounts of sound changes have documented many cases of irregular outcomes. But it remains to be shown whether any one of the changes that exhibit regularity in the outcome was regular in process—that it did not in fact proceed word by word to achieve that result.

3. SEARCHING FOR REGULAR SOUND CHANGES. From the outset the Neogrammarians hoped to find support for their position in local dialects. Thus Osthoff and Brugmann (1878):

> only that comparative linguist who for once emerges from the hypotheses-beclouded atmosphere of the workshop in which the original Indo-European forms are forced, and steps into the clear air of tangible reality and of the present … only he can arrive at a correct idea of the way in which linguistic forms live and change … (translated in Lehmann 1967:202)

But in fact dialect geographers as a whole moved in the opposite direction, arguing that each word has its own history. This follows naturally from the question-and-answer format of most dialect questionnaires, which focuses on conscious choices at a high level of awareness. With the development of computational methods of processing acoustic analysis, it is possible to capture hours of speech from hundreds of speakers in a form not very far from the vernacular of everyday life.

How then would such an empirical study establish that any one sound change was progressing in a manner that affected each and every member of the word class according to its phonetic environment? First of all, we want to study a change in progress, rather than a stable distribution that we infer must be the result of change some time in the past (as in Bybee 2002). It should be a unidirectional change, in which the whole community participates. And we want to focus on a change below the level of social awareness, to avoid the confusion of sound change with the subsequent development of lexical markers as social stereotypes (Labov 1975).[1]

In the approach to be followed here, the defining phonetic features of the sound change are examined to see if they prevail across the data set. To begin with, mixed-level multiple regression is used to define these parameters and detect lexical diffusion on the basis of word frequency. We then plunge into the more difficult task of searching for the absence of exceptions among the least common as well as the most common words.

3.1. SOUND CHANGE IN PHILADELPHIA. We pursue this search for regularity in the city of Philadelphia, where sound change in the local community has been studied for over a century, ranging from the phonetic transcriptions of dialect geographers (de Camp 1939, Tucker 1944, Kurath & McDavid 1961, Ferguson 1975) to large-scale quantitative studies of change in progress (Labov 1976, 2001, Conn 2005, Fruehwald 2013, Labov et al. 2013).[2]

From the outset, the changes appeared to be likely candidates for irregularity rather than regularity. In Philadelphia:

- The tensing of short-*a* before /d/ is found only in *mad*, *bad*, and *glad*, laxing before front nasals is found only in *ran*, *swam*, and *began*—and a host of other details that show that lexical identity and frequency have been at work.
- Philadelphia shows the same phonetic condition for the raising of /ay/ as in 'Canadian raising'—before voiceless obstruents. But when the voicing condition became obscured by flap formation, as in *writer* and *rider*, the Philadelphia rule became extended irregularly to a number of frequent words, such as *spider*, *tiger*, *Snyder* (Fruehwald 2013).
- The mid back ingliding vowel /oh/ in *lost*, *off*, *all*, and so forth is raised to upper mid or lower high position in Philadelphia and the Middle Atlantic States generally. But Philadelphians have become acutely aware of this vowel and label it with the lexical stereotype *wooter* (for *water*), which is pronounced with a vowel considerably higher than that used in everyday speech. Advance or retreat of this variable is high on the scale of social awareness.

[1] See, for example, *Hoi Toider* in the Outer Banks of North Carolina. The New York City dialect is often labeled by the stereotype *Toity Toid* Street, with the diphthong [ɔɪ] in place of the common [əɪ].

[2] The notation used for the vowel system of Philadelphia English discussed here is that of the *Atlas of North American English* (*ANAE*; Labov et al. 2006), as follows.

(i) *ANAE*	IPA	J. C. Wells	/b _ t/
iy	[i]	FLEECE	beet
ey	[eɪ]	FACE	bait
ay	[aɪ]	PRICE	bite
aw	[aʊ]	MOUTH	bout

- Among the prominent features of the Philadelphia dialect are the fronting of the back upgliding diphthongs /uw, ow, aw/, as in *root*, *coat*, *out*. Although no lexical variation has been reported for these variables, they are not likely candidates for the study of regularity, since they all show a reversal of direction in time for those born around 1960 (Labov et al. 2013). Evidence of regularity will depend upon the abruptness of the reversal.

Given these complexities, it might seem likely that the Philadelphia studies of change in progress would once again leave the Neogrammarian principle as a working hypothesis, without the clear demonstration hoped for. However, a promising site for the investigation of regularity appears in a change in progress that had not been recognized prior to the quantitative studies of the 1970s: the raising of /eyC/ before obstruents and nasals in *came*, *gave*, *bake*, *wait*, *ache*, and so forth. A significant correlation of /eyC/ with age was presented as a new discovery in the 1977 report to the National Science Foundation on project #7500245, ‘The quantitative study of linguistic change and variation’ (LCV).[3]

In the decades that followed, a steady flow of data on this variable was obtained from the *Atlas of North American English* (Labov et al. 2006), Conn’s dissertation on the role of gender in Philadelphia sound changes (2005), and by students in the University of Pennsylvania course L560, ‘The Study of the Speech Community’. The course was taught yearly from 1972 to 1994, and every two years from 1994 to 2012. In each class, groups of three to six students selected a neighborhood, observed the use of language in the local setting, made initial contacts, and carried out recorded interviews of speakers in local social networks. Recordings from LCV and L560 were combined to form the Philadelphia Neighborhood Corpus (PNC), with 379 speakers, transcribed with ELAN (2014) and analyzed acoustically by FAVE (Rosenfelder et al. 2014).[4] The distribution of speakers by age and year of interview is shown in Figure 1.

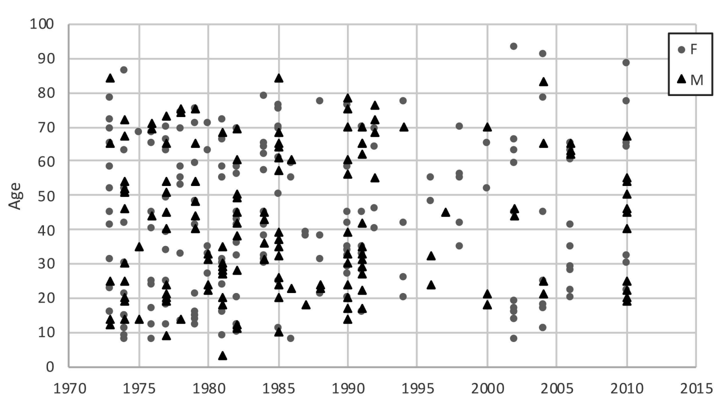


FIGURE 1. The distribution of the 379 speakers in the Philadelphia Neighborhood Corpus by age and year of interview.

[3] This finding was published as figure 4.8 in Labov 2001:143.

[4] Fifty of the PNC interviews in the early 1970s are drawn from the study of Philadelphia carried out by the LCV project (Labov 2001), supported by NSF grants 7500245 and 7680910, and thirteen interviews with African American speakers are drawn from the project on the Influence of Urban Minorities on Linguistic Change, supported by NSF grant 8023306.

The data that emerged indicated that the sound change given in 1 is most likely to provide the setting in which the basic process of regular change would emerge, free of the processes of analogy, borrowing, reversal, and social selection, which can obscure that regularity.

(1) /eyC/ → <peripheral> / __ /obs, nas/

This sound change is marked by four features:

- PHONETIC CONDITIONING: The raising process selects only words with consonants following stressed /ey/ (*made*, *eight*, *main*, *days*) as opposed to /ey/ in final position, before hiatus, and before liquids (*day*, *mayor*, *hale*).[5] The allophone so defined is designated here as /eyC/, and all others of the /ey/ class as /eyF/. /eyC/ includes vowels in open syllables (*baby*, *neighborhood*) as well as closed syllables (*babe*, *raid*) (Fruehwald 2013).[6]
- INSENSITIVITY TO MORPHEME BOUNDARIES: The allophonic definition applies across morpheme boundaries, so that *days* and *rays* are not differentiated from *daze* and *raise*. The raising effect is also found, to a lesser degree, across word boundaries (Fruehwald 2013:137–40).
- ABSENCE OF SOCIAL AWARENESS: Though the Philadelphia speech community shows considerable sensitivity to many features of the metropolitan dialect, the raising of /eyC/ was never reported prior to the publication of the acoustic measurements of 1976. In 2012, the IHELP project at the University of Pennsylvania conducted interviews with 108 friends and families of Philadelphia college students that explored social attitudes toward features of the Philadelphia dialect: no mention was made of the raising of /eyC/.[7] YouTube currently has several dozen videos focused on 'Philly tawk', transcribing *can* as 'cean' and *right* as 'roit', but the raised *came* is never represented as 'ceem'.
- MONOTONIC INCREMENTATION: While other Philadelphia sound changes have developed nonlinear patterns across time, the raising of /eyC/ displays a markedly monotonic, almost linear relation to time, which may be displayed in several ways. Figure 2 is the locally weighted regression (LOESS) analysis that first appeared in Labov, Rosenfelder, & Fruehwald 2013. It shows the raising of the mean position of /eyC/ by date of birth for 264 white PNC subjects. The vertical axis is distance along the front diagonal, calculated as $F2 - 2 * F1$. Each symbol represents the means for one speaker. Darker gray areas indicate the limits of the 95% confidence interval, so that any light space separating lines is a significant difference. The upper dashed line for /eyC/ shows a monotonic incrementation from date of birth 1888 to 1991. The lower line shows a very different pattern—almost level—for the free allophone /eyF/. There is a small but significant raising of /eyF/ with a date of birth coefficient of 11.5 per decade as against 33.9 per decade for /eyC/; the difference between the two allophones increases steadily over the century.

The temporal development of the raising of (eyC) can be displayed in several ways: by the year of the interview, the age of the speaker, or the date of birth of the speaker.

[5] The mid front vowel before the other liquid /r/ is strongly centralized in Philadelphia, so that *ferry* and *furry* are engaged in a near merger. In the *ANAE* notation it is represented by /ehr/.

[6] A parallel differentiation is made with /iyC/ vs. /iyF/. However, the low member of the Vy subset shows a different partition, distinguishing nuclei before voiceless obstruents from all others, with the notation /ay0/ vs. /ayV/.

[7] In contrast to what is seen in the exaggerated raising of *water* to [wʊɾɚ] as a widespread stereotype of upper mid /oh/ in Philadelphia.

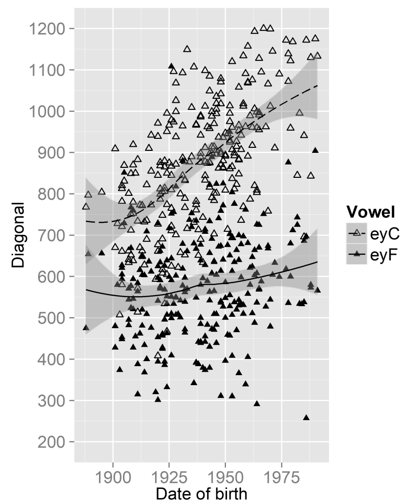
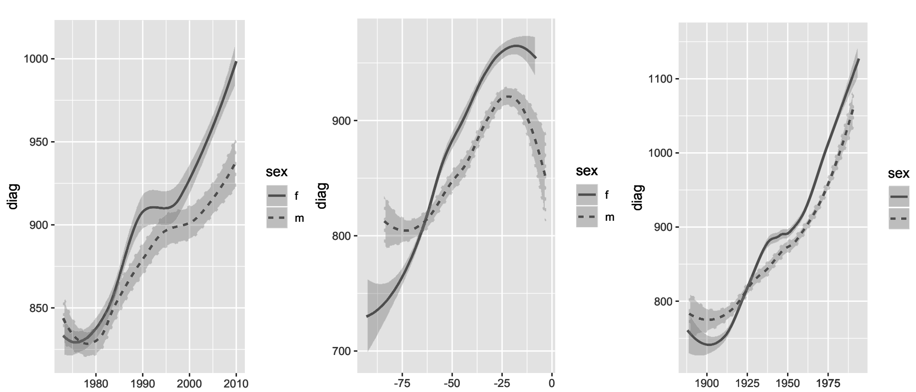


FIGURE 2. Locally weighted (LOESS) regression analysis of /eyC/ and /eyF/ along the front diagonal (F2 – 2 * F1) by date of birth for white adults in the PNC. Each symbol represents the mean value for one speaker.

Figure 3 shows these patterns for male and female speakers in the PNC. The linear pattern is most striking for date of birth (Fig. 3c), and only slightly less regular for year of interview (Fig. 3a). But age shows a notably curvilinear pattern (Fig. 3b), with a peak in the youngest age range and a maximum advantage for females at seventeen to eighteen years old. This is the pattern of adolescent incrementation first identified in Labov 2001:Ch. 14 and Denis et al. 2019. The peak of a change in progress is found among adolescent speakers, who advance the change one step beyond their older peers, while preadolescents are still under the influence of their more conservative parents.



a. Year of interview. b. Age of speaker. c. Date of birth.

FIGURE 3. Temporal development of (eyC) by LOESS correlation with (a) year of interview, (b) age of speaker, and (c) date of birth.

Figure 4 shows this pattern of incrementation by separating the data into five age groups. Across the century, adolescents thirteen to nineteen years old are in the lead. Following them are young adults twenty to thirty-nine years old, then mature adults forty to fifty-nine years old, and at the lowest level are speakers sixty years and older. The only group that does not fit into this pattern of steady incrementation is the youngest, eight to twelve years old, reflecting the fact that they are still under the influence of their relatively conservative parents. The thirteen-to-nineteen group plainly exhibits the adolescent peak.

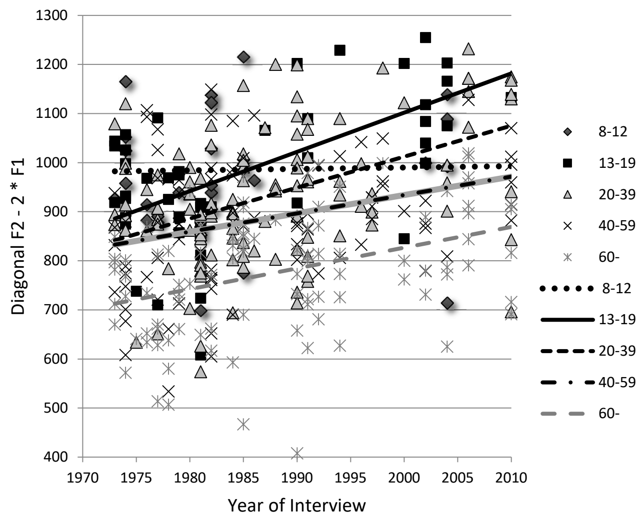


FIGURE 4. Linear regression of /eyC/ along the front diagonal for five age groups in the Philadelphia Neighborhood Corpus. *N* = 392.

The raising of /eyC/ is therefore a promising site to test the Neogrammarian position that sound changes select all lexical items in a phonetically defined class, corresponding to the concept of regular change as INNOVATION in Janda and Joseph's model of regular sound change (2003). Granted, the deck is stacked in favor of regularity: I have chosen a site in which the factors that most favor lexical irregularity are absent and argue that the end result should be lexical regularity.

3.2. MIXED-LEVEL REGRESSION ANALYSIS OF /eyC /. Table 1 includes the data from all 320 white speakers in the PNC, a total of 25,458 observations. The chief indicator of change in progress, date of birth, is the predominant fixed effect. It amounts to a mean increment of 3.44 units per year along the front diagonal (F2 − 2 * F1) over the century. The second fixed effect in Table 1 tests for the influence of word frequency in the PNC.[8] It is not significant even at the 0.05 leve1. The next line examines the interaction of date of birth with word frequency: here too we find no significant effect.

SOCIAL FACTORS. Several social factors that are found to be prominent in more developed stages of sound change prove to be minimal in Table 1. The most common social factor in linguistic change—female sex—favors the change at the 0.05 level. Education (not listed) was not significant.

PHONETIC FACTORS. Table 1 shows that syllable onsets have stronger and more significant effects on the raising of /eyC/ than syllable codas. The strongest single effect is the positive coefficient of 227 for velar onsets. This is similar to what is seen in Fig. 5

[8] The advantage of using frequency in the PNC itself is that the interviews come closer to the topics and style of everyday life than other corpora, but with some skewing toward those topics that are frequently introduced by the interviewer (as shown by a relatively high frequency of *neighborhood*).

RANDOM EFFECTS

GROUPS	NAME	*SD* Corr
Token	(intercept)	305.6133
DOB		0.2025−1.00
Name	(intercept)	47.1721
Neighborhood	(intercept)	275.8690
Residual		250.2897

Number of observations: 25,458

Groups: word: 1,704 speakers: 320 neighborhood: 52

FIXED EFFECTS	EST	*SE*	*t*-VALUE	Pr(>|*z*|)	
(intercept)	−5773.26	338.67	−17.05	0.00E+00	
Date of birth	3.44	0.17	19.93	0.00E+00	*****
Frequency	−0.44	0.23	−1.92	5.50E−02	
DOB : frequency	0.00	0.00	1.77	7.71E−02	
Sex female	17.57	6.89	−2.55	1.08E−02	*
Coda nasal	−81.49	28.04	−2.91	3.66E−03	**
Coda labial	94.99	43.69	2.17	2.97E−02	*
Coda velar	−113.67	14.89	−7.63	2.31E−14	***
Onset labial	−43.98	11.98	−3.67	2.40E−04	**
Onset palatal	112.56	21.20	5.31	1.10E−07	****
Onset obs-liquid	−172.81	13.25	−13.04	0.00E+00	*****
Onset apical	−171.83	13.81	−12.44	0.00E+00	*****
Onset velar	227.06	14.71	15.44	0.00E+00	*****
Onset zero	105.17	15.84	6.64	3.12E−11	*****

TABLE 1. Mixed-level regression analysis of the raising of /eyC/ along the front diagonal (F2 − 2 * F1) for the Philadelphia Neighborhood Corpus. (* < 0.05, ** < 0.001, *** < 0.0000, ***** < 0.00001.)[9]

below, where three of the ten highest words had velar onsets. (Obstruent-liquid onsets as in *break* and *great*, however, strongly inhibit the sound change, showing a negative effect almost identical to that of onset apicals.) Velar, palatal, and zero onsets strongly favor the raising of /eyC/, while coda velars disfavor it.

On the whole, these figures display the fundamental regularity of Neogrammarian sound change. Table 1 presents a picture of a phonetically governed change at an early stage of social awareness, without any evidence of an effect of lexical frequency. The upward movement of the tongue body in the raising of the vowel is plainly governed by the articulatory postures of the neighboring consonants. If the tongue body is raised for these consonants, the vowel is relatively higher. If the tongue margins or the tongue center is lowered for the initial consonant, the raising of the tongue is strongly inhibited. The continuous movement of the tongue body is governed by the interaction of these categorical settings of phonological parameters. By contrast, no effects associated with word membership have appeared upon the scene.[10]

4. THE SEARCH FOR LEXICAL EXCEPTIONS. Table 1 addressed the question of whether the raising of /eyC/ in Philadelphia is governed by the frequency of lexical forms. The

[9] Linear mixed model fit by REML ['lmerMod']. Formula: diag ~ DOB + frequency + DOB:frequency + sex + education + ethnicity + Manner + Place + Onset + (1 | name) + (DOB | wordLa) + (1 | neighborhood). REML criterion at convergence: 354865.9.

[10] The New Zealand study used measures of frequency (F1 and F2 for /i/, /e/, and /æ/) and the intersection of each of these with date of birth (Hay et al. 2015, table A1). An effect of frequency at the 0.01 level of significance was found in a number of variables but in the opposite direction from that predicted. Hay et al. (2015) suggest that this reverse effect is a property of a push-chain situation in New Zealand that relates the three vowels /i, e, æ/. Since the raising of /eyC/ in Philadelphia is not part of a chain shift, a negative effect of word frequency would not be predicted here based on the New Zealand findings.

result was negative, supporting the Neogrammarian position on this issue. We are now ready to attack the more difficult question: are there among the 1,704 different words of Table 1 any exceptions to the application of the raising rule in 1? Is the raising of /eyC/ regular in that every word in which the /eyC/ allophone occurs is produced under the phonetic constraints of Table 1?

A search for regular behavior is a search for the absence of irregularities—that is, a search for nothing in particular. Yet the history of languages has had no difficulty in finding these particulars. Exceptions to the evolution of English /ɛː/ troubled the historians of Early Middle English. The two words *great* and *break* relinquished any link to the ongoing raising of the lower mid long vowel—spelled *ea* as in *eat* and *meat*—and were reassigned to the class of long low vowels spelled with *a*, as in *grate* and *brake*. Less common words of similar structure like *creak*, *freak*, *bleak*, *streak* followed the general pattern.

No such clear exceptions have been found for the raising of /eyC/. How might we go about finding them? It would hardly be practical to address each of the 1,704 words in the data set of Table 1 and ask if it was consistently different from Table 1's predictions over the course of the 25,458 observations. Our strategy is instead to first examine a small number of very large numbers—very common words—and see if their behavior is fully accounted for by the phonetic constraints of Table 1. We then consider a large number of a very small number—words that occur only once in our corpus—which will be a key to the inquiry as it touches on the problem of learnability. To complete the survey, we examine an array of typical words of intermediate frequency.

4.1. SLOPES OF THE MOST COMMON WORDS. Let us first extract from the data set of 25,458 /eyC/ words the trends by date of birth for the ten most common words with CVC structure, shown here with their frequencies.[11]

(2) Ten most common /eyC/ words with CVC structure in the PNC

came	1,893
take	1,746
make	1,424
made	980
name	936
days	708
eight	521
gave	467
game	348
age	346

Figure 5 plots the linear trend lines for the most common /eyC/ words across the 110-year time span. We see that they all move together, with no sign of random selection. Instead, they follow the logic of their phonetic composition. The estimates of phonetic factors in Table 1 showed that the onset was more influential than the coda, with zero (105), palatal (113), and velar (227) most heavily weighted. Of the most common words, three in particular with velar or vocalic onsets confirm the ordering of the strongest coefficients of Table 1: *gave*, *came*, and *eight* lead in the raising process.[12] Conversely, *take* and *make* are much lower, reflecting the negative weighting of velar

[11] Very common polysyllabic words like *always* (1,187) and *maybe* (1,286) were not included in order to avoid questions of colloquial reduction in less stressed forms.

[12] Lemmas with initial /h/ as in *hate* were classified with those with no consonantal onset, as in *eight*.

coda (−114). One word, *game*, shows a slope markedly steeper than the others, moving from the lowest to the highest relative position over this time span.

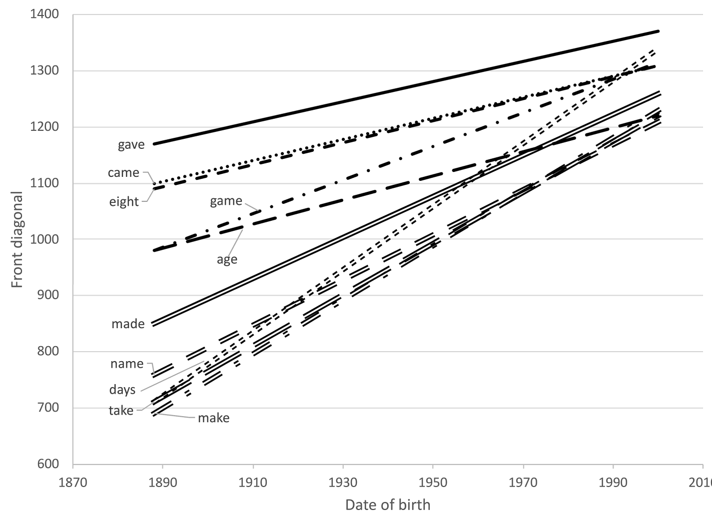


FIGURE 5. Linear regression trend lines of the ten most common monosyllabic /eyC/ words by date of birth.

4.2. THE LEAST COMMON WORDS. A referee of an earlier draft of this article raised the objection that the demonstration of regularity was clearest for the most common words, and that to prove the point, one would have to show that regularity also applied to words that occurred only once. The 1,704 different words in the /eyC/ vocabulary follow Zipf's law (1949): the frequency of a word is inversely proportional to its rank. Whereas fifty-six words occur more than 100 times in the data set, a large number—662—occur only once. I thought that these 662 words did raise a heavy objection and had no immediate response. How could one show the direction of change for a word that occurs only once? At this point Charles Yang called to my attention the fact that a similar situation prevails throughout the study of the acquisition of variable elements of language. His argument (Yang 2013, 2016) is that:

- more frequent items are more resistant to sampling effects and are thus more likely to reflect the full range of variation;
- the statistical evidence for a rule is stronger when evaluated over a smaller number of types;
- once a categorical rule is acquired, the learner can generalize it to any lexical item.

The same logic applies to continuous distributions like /eyC/ just as it does to the determiner-noun combinations discussed by Yang. Once a child has learned that the membership of the /eyC/ class includes *shameful* as well as the earlier-acquired *came*, he or she is ready to assume that the productive parameters that govern this less common word are the same as those that are used so much more frequently in *came*. Thus the observation of the regularity of the most frequent elements of /eyC/ supports the argument in favor of the regularity of the /eyC/ lexicon as a whole.

Though more frequent words support the regularity of change, there always remains the possibility that less frequent words are subject to lexical diffusion, granted the fact that frequency was not significant in the regression analysis of the whole body of /eyC/ words (Table 1). For any one of the 662 words that occurred only once, it is hardly possible to measure the rate of raising along the front diagonal, any more than a child might learn from it to what degree change is in progress. We may, however, assess the participation of these 662 words in the raising of (eyC) by determining the trend of the body of singletons as a whole. Figure 6 shows the distribution of these singleton words by date of birth and position on the front diagonal. The trend line shows a slope of 4.19, not very different from the 3.42 figure for the data set as a whole.

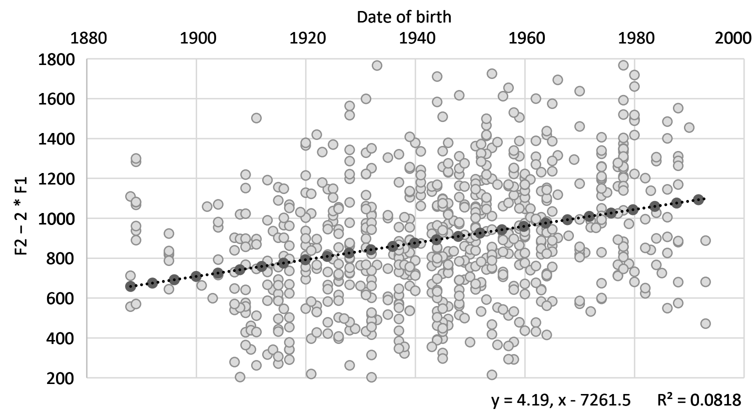


Figure 6. Distribution along the front diagonal of 662 /eyC/ words that occur only once in the PNC data.

4.3. Words of moderate frequency. It is of course always possible that there is a curvilinear distribution of raising: that words of moderate frequencies are different from those with low or high frequencies, just as it is sometimes found that word frequency has a positive effect on change, and sometimes a negative. Figure 7 considers the possibility of such a curvilinear range by charting the same distribution for all words with frequencies between 200 and 400. The trend line is close to the two others we have seen, with a y-value of 3.35. All three trend lines begin at about 650 on the front diagonal and end at a little more than 1,040.

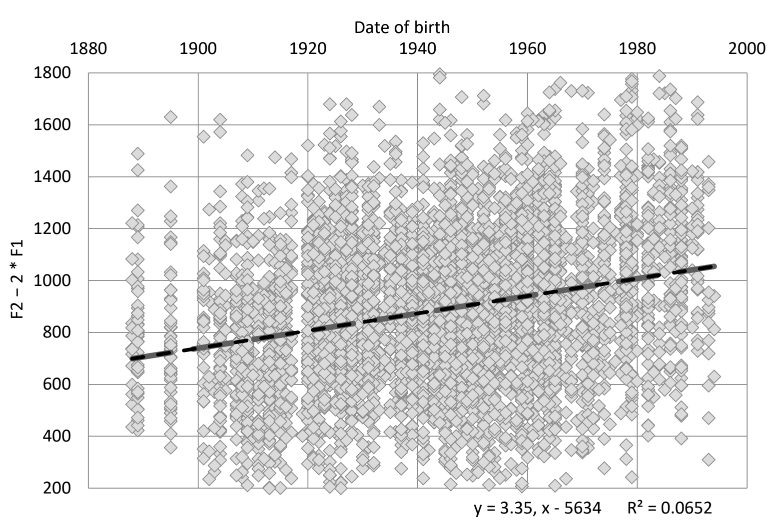


Figure 7. Distribution along the front diagonal of all /eyC/ words occurring with frequencies of 200–400.

5. Reexamining the history of obstruent-liquid onsets. If we have not succeeded in finding exceptions through this overview of the PNC, we might benefit from hints from the history of English in §4 combined with the phonetic analysis of §3.4. Figure 8 adds to the display of Fig. 5 the following words with obstruent-liquid onsets.

(3) /eyC/ words with obstruent-liquid onsets in the PNC

place	784
grade	525
great	521
break	236

These four words form a cluster concentrated around 900 on the front diagonal in the most recent period. They may indeed represent the exceptions we have been looking for—words that do not follow the rule for raising /eyC/—or they may represent a modification of that rule to exclude obstruent-liquid onsets.

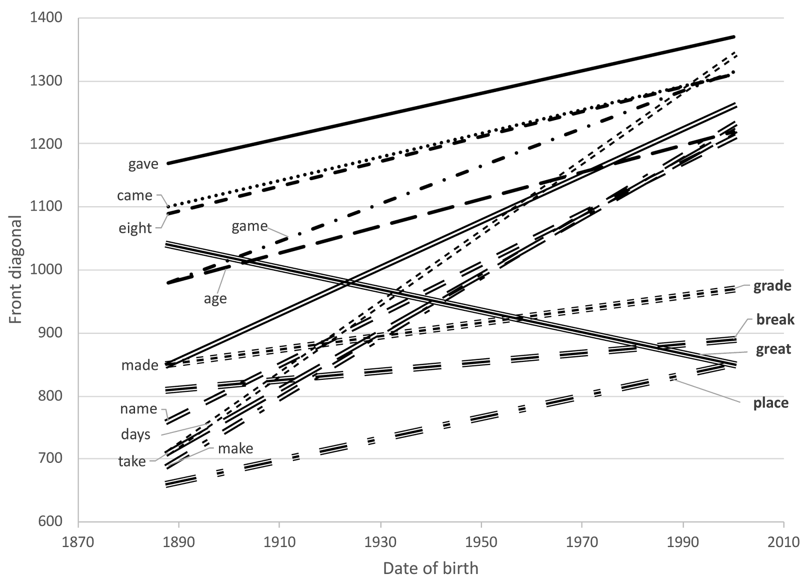


FIGURE 8. Phonetic deviation of words with obstruent-liquid onsets from the main group of /eyC/ words.

Figure 9 maps the development of liquid onsets by a LOESS tracing across time. It shows that the four deviant words of Fig. 8 are not exceptional, but rather exemplify a general effect of liquids /r/ and /l/ on the raising of /eyC/. The uppermost line on the right (YY) represents the general raising without liquids. The line just below shows the trend for following laterals (XL), which were not differentiated in the earliest period but are left behind by the raising process (/ey/ with a following /r/ is a completely different allophone, as noted above). The lower two lines represent the position on the front diagonal of onset laterals (LX), as in *late*, and obstruent-liquid clusters, as in *great* and *place* (XX). They both follow a slightly upward path from about 700 to almost 850 on the front diagonal. The figure as a whole may be said to represent the phonological creation of a raised allophone of /eyC/ with a reorganization of the liquids within that allophone. There is no doubt that the liquids are a conditioned subset of /eyC/, rather than /eyF/, which is firmly located at 600 on the front diagonal (Fig. 2).

The apparent exceptions *grade*/*break*/*great*/*place* are phonetically motivated by obstruent-liquid onsets, an inhibiting influence on the raising of vowels as defined in Table 1 and Fig. 9. A similar phonological displacement appears in a well-known chapter in the history of English involving the general sound law that merged /eː/ and /ɛː/ in the six-

FIGURE 9. LOESS plot of /eyC/ categories involving liquids.

teenth to seventeenth centuries: *meat* merging with *meet* (Labov 1975). This was embedded in the progress of the Great Vowel Shift as it raised /eː/ to /iː/ and /æː/ to /eː/. Two words with Middle English /ɛː/, *great* and *break*, have long been considered exceptions to regular sound laws by historians of English and a challenge to Neogrammarian doctrine. Jespersen rejected the notion that the preceding /r/ accounts for the lowering, given *treason*, *breach*, *preach*, and so forth. He notes that John Walker argues for sound symbolism 'as deeper and more expressive of the epithet *great*', while others looked to the preservative influence of the short /e/ in *breakfast* (Jespersen 1954:338–39).[13]

In our search for lexical exceptions, we encountered exceptional behavior from *great* and *break*, words with obstruent-liquid onsets and voiceless obstruent codas. We note that these two words are members of the same phonetically defined class that appeared to be breaking away from the /eyC/ class in Fig. 9. A dictionary search shows nine other words that fit this phonological description, listed in 4.

(4) treat, creak, streak, freak, bleach, preach, pleat, bleat, bleak

Not one of these nine appears in the PNC. But *break* and *great* constitute 24% of the total number of words with obstruent-liquid onsets in the PNC, as detailed in 5.

(5) Words with obstruent-liquid onsets in the PNC

a.	GREAT		BREAK		
	great	521	break	236	
	greater	13	breaks	35	
	greatest	35	breaking	38	
	greatly	13	breaker	35	
			breakdown	35	
			breakdowns	1	
	TOTALS	582		380	962
b.	other obstruent-liquid onsets			3,080	

It seems that in the seventeenth century, high frequency and strong coarticulation combined to disrupt the regularity of change. The allophone most strongly differentiated by

[13] A second such reassignment can be observed in the parallel chain shift in the back vowels, where long open /ɔː/, spelled *oa* as in *boat*, *moat*, *goat*, and so forth, rose in the back section of the vowel shift to /oː/. One word remained behind, *broad*, now the only English word spelled with *oa* and pronounced as long open /ɔː/. It is, of course, no accident that it too has an obstruent/liquid onset.

coarticulatory effects may be reassigned to a neighboring category. In the seventeenth-century case this involved phonemic change, but in the twentieth century we have no more than a refinement of allophones.

In twenty-first century Philadelphia we may find the *break*/*great* class joining with the residual /eyF/ allophone, which has resisted the raising process for over a century. Figure 10 merges the phonetic features of this lexical class with the rest of the words with obstruent-liquid onsets. This class with obstruent-liquid onsets is more distinct from the other /eyC/ classes, though it clearly follows the upward trend of Figs. 3, 4, and 5, reaching 1100 on the front diagonal. The *break*/*great* lexical class does not appear to participate in this upward movement, hovering at 800, with a slight upward turn in the most recent years. It shows no tendency to join the /eyF/ allophone, which remains below 600 on the front diagonal. It appears that the long-term application of coarticulatory factors has preserved the regularity of change in progress, although the features that once defined lexical subclasses are still available to speakers of the language.

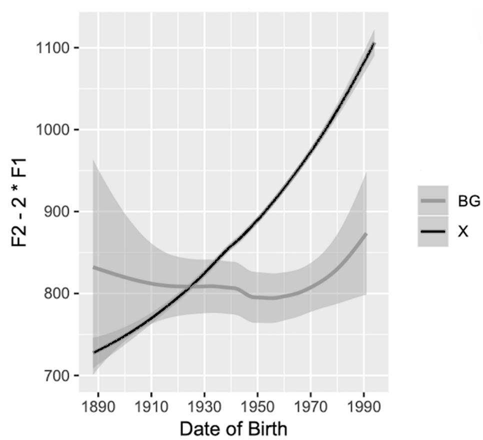


FIGURE 10. Raising on the front diagonal by date of birth of words with obstruent-liquid onsets. BG = *break*, *great*; X = other words.

6. WHAT HAS BEEN LEARNED. We have examined the raising of /eyC/ as a prototypical sound change in a continuous phonetic space, below the level of conscious awareness, which has continued in the same direction for over a century. It was selected in order to test the Neogrammarian principle that change affects all words in which the target phoneme is found. The extended time period of 110 years, the volume of data, and the monotonic incrementation of /eyC/ have facilitated the comparison of the rates of change of individual words at different stages. A few candidates for lexical effects were detected in the course of this study, but by the end none were found to have escaped the tyranny of phonetic constraints. Like a child learning a language we impose this pattern on the vocabulary, as it imposes regularity on us.

What then has been learned from the 25,458 measurements of the /eyC/ nucleus that we could not have learned from a more limited data set? Or to put it another way, what are the implications for linguistic theory of the FAVE suite for automatic vowel analysis? So far we have learned that this sound change has continued in the same direction for a hundred years, that it is found in all neighborhoods of the city and at all educational levels, and that it is defined phonologically from the outset.

Is the raising of Philadelphia /eyC/ a regular sound change? The answer provided here is ‘yes’. This says to the historical linguist that you can apply your working principle with good confidence that you are in touch with linguistic reality. The regularity of sound change is the link between the student of the speech community and the student of language history. We have a common ground in our understanding of what it means to know a language. It involves knowing a vast number of particular things. But at bottom it is reaching down to something very deep, very abstract, and very satisfying.

REFERENCES

Blust, Robert. 1996. The Neogrammarian hypothesis and pandemic longevity. In Durie & Ross, 135–56.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14.261–90. DOI: 10.1017/S0954394502143018.

Campbell, Lyle. 1996. On sound change and challenges to regularity. In Durie & Ross, 72–89.

Campbell, Lyle. 2004. *Historical linguistics: An introduction*. 2nd edn. Cambridge, MA: MIT Press.

Chen, Matthew Y., and William S.-Y. Wang. 1975. Sound change: Actuation and implementation. *Language* 51.255–81. DOI: 10.2307/412854.

Conn, Jeffrey C. 2005. *Of 'moice' and men: The evolution of male-led sound change*. Philadelphia: University of Pennsylvania dissertation. Online: https://repository.upenn.edu/ircs_reports/5/.

Dauzat, Albert. 1922. *La géographie linguistique*. Paris: Flammarion.

de Camp, L. Sprague. 1939. fıləde̞lfıə əmɛrəkən ɪŋglɪʃ [Transcription of 'The north wind' as spoken by a Philadelphian]. *Le Maître Phonétique* 17(54).50–51. Online: https://www.jstor.org/stable/44704742.

Denis, Derek; Matt Hunt Gardner; Marisa Brook; and Sali A. Tagliamonte. 2019. Peaks and arrowheads of vernacular organization. *Language Variation and Change* 31.43–67. DOI: 10.1017/S095439451900005X.

Durie, Mark. 1996. Early Germanic umlaut and variable rules. In Durie & Ross, 112–34.

Durie, Mark, and Malcolm Ross (eds.) 1996. *The comparative method reviewed: Regularity and irregularity in sound change*. New York: Oxford University Press.

ELAN. 2014. Nijmegen: Max Planck Institute for Psycholinguistics. Online: https://tla.mpi.nl/tools/tla-tools/elan/.

Ferguson, Charles A. 1975. 'Short a' in Philadelphia English. *Studies in linguistics in honor of George L. Trager*, ed. by M. Estellie Smith, 259–74. The Hague: Mouton.

Fruehwald, Josef. 2013. *The phonological influence on phonetic change*. Philadelphia: University of Pennsylvania dissertation. Online: https://repository.upenn.edu/edissertations/862/.

Harris, John. 1989. Towards a lexical analysis of sound change in progress. *Journal of Linguistics* 25.35–56. DOI: 10.1017/S0022226700012093.

Hay, Jennifer B.; Janet B. Pierrehumbert; Abby J. Walker; and Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139.83–91. DOI: 10.1016/j.cognition.2015.02.012.

Krishnamurti, Bhadriraju. 1998. Regularity of sound change through lexical diffusion: A study of *s* > *h* > Ø in Gondi dialects. *Language Variation and Change* 10.193–220. DOI: 10.1017/S0954394500001289.

Kruschke, John K. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99.22–44. DOI: 10.1037/0033-295X.99.1.22.

Kurath, Hans, and Raven I. McDavid, Jr. 1961. *The pronunciation of English in the Atlantic states*. Ann Arbor: University of Michigan Press.

Janda, Richard D., and Brian D. Joseph. 2003. Reconsidering the canons of sound-change: Towards a 'Big Bang' theory. *Historical linguistics 2001: Selected papers from the 15th International Conference on Historical Linguistics, Melbourne, 13–17 August 2001*, ed. by Barry J. Blake and Kate Burridge, 205–19. Amsterdam: John Benjamins.

Jespersen, Otto. 1954. *A modern English grammar on historical principles, part 1: Sounds and spelling*. Copenhagen: E. Munksgaard.

Labov, William. 1975. On the use of the present to explain the past. *Proceedings of the 11th International Congress of Linguists*, ed. by L. Heilmann, 825–51. Bologna: Il Mulino.

Labov, William. 1976. The relative influence of family and peers on the learning of language. *Aspetti sociolinguistici dell' Italia contemporanea*, ed. by Raffaele Simone and Giulianella Ruggiero, 11–56. Rome: Bulzoni.

Labov, William. 1981. Resolving the Neogrammarian controversy. *Language* 57.267–308. DOI: 10.2307/413692.

Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Oxford: Wiley-Blackwell.

Labov, William; Sharon Ash; and Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology, and sound change*. Berlin: Mouton de Gruyter.

Labov, William; Ingrid Rosenfelder; and Josef Fruehwald. 2013. One hundred years of sound change: Linear incrementation, reversal, and reanalysis. *Language* 89.30–65. DOI: 10.1353/lan.2013.0015.

Lehmann, Winfred P. (ed.) 1967. *A reader in nineteenth-century historical Indo-European linguistics*. Bloomington: Indiana University Press. Online: https://liberalarts.utexas.edu/lrc/resources/books/reader/index.php.

Lien, Chinfa. 1987. *Coexistent tone systems in Chinese dialects*. Berkeley: University of California, Berkeley dissertation. Online: https://escholarship.org/uc/item/8q33g8tm.

Moore, Samuel, and Thomas A. Knott. 1955. *The elements of Old English*. Ann Arbor, MI: Wahr.

Ogura, Mieko. 1987. *Historical English phonology: A lexical perspective*. Tokyo: Kenkyusha.

Osthoff, Hermann, and Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, vol. 1. Leipzig: Hirzel.

Paul, Hermann. 1891. *Principles of the history of language*. London: Longmans. [First published as *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer, 1886.]

Phillips, Betty S. 1980. Lexical diffusion and Southern *tune*, *duke*, *news*. *American Speech* 56.72–78. DOI: 10.2307/454480.

Phillips, Betty S. 1983. Middle English diphthongization, phonetic analogy, and lexical diffusion. *Word* 34.11–23. DOI: 10.1080/00437956.1983.11435734.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. *Language* 60.320–42. DOI: 10.2307/413643.

Phillips, Betty S. 1989. The diffusion of a borrowed sound change. *Journal of English Linguistics* 22.197–204. DOI: 10.1177/007542428902200203.

Phillips, Betty S. 2006. *Word frequency and lexical diffusion*. New York: Palgrave Macmillan.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan L. Bybee and Paul J. Hopper, 137–57. Amsterdam: John Benjamins.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. *Laboratory phonology 7*, ed. by Carlos Gussenhoven and Natasha Warner, 101–39. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110197105.1.101.

Ringe, Donald, and Joseph F. Eska. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge: Cambridge University Press.

Rosenfelder, Ingrid; Josef Fruehwald; Keelan Evanini; Scott Seyfarth; Kyle Gorman; Hilary Prichard; and Jiahong Yuan. 2014. FAVE (Forced Alignment and Vowel Extraction) program suite. Version 1.2.2. DOI: 10.5281/zenodo.22281.

Shen, Zhongwei. 1990. Lexical diffusion: A population perspective and a numerical model. *Journal of Chinese Linguistics* 18.159–201. Online: https://www.jstor.org/stable/23767133.

Todd, Simon; Janet B. Pierrehumbert; and Jennifer Hay. 2019. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition* 185.1–20. DOI: 10.1016/j.cognition.2019.01.004.

Tucker, R. Whitney. 1944. Notes on the Philadelphia dialect. *American Speech* 19.37–42. DOI: 10.2307/486530.

Wallace, Rex E. 1984. Variable deletion of *-s* in Latin: Its consequences for Romance. *Papers from the XIIth Linguistic Symposium on Romance Languages*, ed. by Phillip Baldi, 565–77. Philadelphia: John Benjamins.

Wang, William S.-Y. 1969. Competing changes as a cause of residue. *Language* 45.9–25. DOI: 10.2307/411748.

WANG, WILLIAM S.-Y. (ed.) 1977. *The lexicon in phonological change*. The Hague: Mouton.
WEINREICH, URIEL; WILLIAM LABOV; and MARVIN HERZOG. 1968. Empirical foundations for a theory of language change. *Directions for historical linguistics*, ed. by Winfred P. Lehmann and Yakov Malkiel, 97–195. Austin: University of Texas Press.
YANG, CHARLES. 2013. Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences* 110.6324–27. DOI: 10.1073/pnas.1216803110.
YANG, CHARLES. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.
ZIPF, GEORGE K. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley.

2048 Rittenhouse Sq.
Philadelphia, PA 19103
[labov@comcast.net]

[Received 19 October 2017;
revision invited 24 March 2018;
revision received 19 February 2019;
revision invited 23 May 2019;
revision received 20 July 2019;
accepted 18 August 2019]