2018

Mechanisms Of Phonological Change

Betsy Sneller
University of Pennsylvania, betsysneller@gmail.com

Mechanisms Of Phonological Change

Abstract
The traditional Philadelphia allophonic /æ/ system (henceforth: PHL shown in (1) below) is characterized by a set of complicated conditioning factors and a dramatic acoustic distinction between the two allophones. In recent years, some Philadelphians have begun to exhibit a new allophonic system (NAS, shown in (2) below). Like PHL, NAS is characterized by a dramatic acoustic distinction between tense and lax allophones. NAS is quickly overtaking PHL in the Philadelphia community, as demonstrated by Labov et al. (2016).

(1) PHL: æ→æh/ _ [+anterior] ∩([+nasal] ∪ [-voice + fricative)]σ

(2) NAS: æ→æh/ +nasal

This situation offers an exciting opportunity to observe phonological change in individual speakers. Most phonological changes involve the collapse or creation of a new phonological category; because of the large degree of acoustic overlap in these situations, it is difficult or impossible to identity individual tokens as having been produced by the old or the new phonology. In the current change in Philadelphia /æ/, however, both the old and the new system involve distinct acoustic targets, making it possible to identify which underlying system was used to produce a given word. It is therefore possible to test several distinct theories about phonological change: Whether change occurs through gradual phonetic incrementation (e.g. Ohala 1981), through individual speakers producing only the old or the new system (e.g., Janda and Joseph 2003), or whether change occurs via individual speakers probabilistically producing both the old and the new system in a process of individual grammar competition (e.g., Fruehwald et al. 2013).

In my dissertation, I examine natural speech production from 46 speakers who acquired language during the period of allophonic change, with a combination of topic-directed conversations and targeted natural language experiments. Using a glm classifier, I identify tokens of /æ/ as having been produced by either PHL or NAS. In concert with an analysis of speakers' social histories, I use these results to argue that the change from PHL to NAS in Philadelphia is driven by the mechanism of competing grammars, suggesting that both syntactic change and phonological change proceed in the same manner. My research provides one of the first pieces of direct empirical support for a unified theory of language change in which structural changes in syntax and phonology are implemented through the same mechanism of grammar competition (Kroch, 1989; Fruehwald et al., 2013).

In addition to the theoretical contribution to phonological change, my dissertation also traces the social patterns of the allophonic change, highlighting the effect of network structure and access to elite education on the adoption of the incoming allophonic system. I also employ experimental methods to demonstrate that the abstract allophonic rules of /æ/ are the target of social evaluation and contribute to social meaning. I find speakers producing surprisingly systematic evaluations of PHL and NAS, a result which only emerges when analyzing the evaluation of changing abstract parameters. Finally, to test whether the change from PHL to NAS was the inevitable result of phono- logical simplification, I developed a computational simulation built using a principle of language acquisition (Yang, 2016) to demonstrate that the allophonic restructuring in /æ/ was not the result of children simplifying their input data, but rather must have been the result of dialect contact with in-moving speakers of the new system.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Linguistics

First Advisor
William Labov

Keywords
Language Variation and Change, Philadelphia English, Phonological change, Phonology, Sociolinguistics

Subject Categories
Anthropological Linguistics and Sociolinguistics | Linguistics

MECHANISMS OF PHONOLOGICAL CHANGE

Betsy Sneller

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2018

William Labov, Professor of Linguistics, Supervisor of Dissertation

Eugene Buckley, Associate Professor of Linguistics, Graduate Group Chairperson

Dissertation Committee:
Meredith Tamminga, Assistant Professor of Linguistics
Josef Fruehwald, Lecturer in Sociolinguistics, University of Edinburgh

For Leah Sieplinga.

Acknowledgements

One of the most enjoyable parts of writing my dissertation has been reading the acknowledgement sections of the dissertations of people I admire. It's here in the outpouring of appreciation, love, and recognition of community that the truth of any Ph.D. journey comes out: it takes a village to raise a dissertation. This is certainly true of my own experience – this dissertation has come about through countless hours of mentorship, emotional and practical support[1], and well-timed distractions from an entire community of colleagues and friends.

I have been incredibly lucky to have Bill Labov as my advisor. He somehow manages to be kind, demanding, and brilliant all at once, and is a perfect model for what an advisor should be. His boundless energy for discussing ideas and his depth of insight on topics ranging from poetry to statistics is breathtaking. Bill's attitude toward research emphasizes the humanity of those around him, and encourages me to do the same. In an acknowledgement of his own advisor, Bill said: "Weinreich was the perfect academic: passionately interested in the ideas of others, brimming over with intellectual honesty, vigor and originality. He protected me from every academic evil. [...] I would like to think that my students are as lucky as I was, but I know better than that." In the spirit of academic honesty, I am compelled to disagree vehemently with this last claim.

A huge thanks must also go out to the rest of my dissertation committee: Meredith Tamminga, Joe Fruehwald, and Gene Buckley. Their feedback, intellectual curiosity and practical support have pushed this dissertation to the best version that it could be. It's a testament to the collaborative and generous nature of the Linguistics Department faculty that I also have Tony Kroch, Gillian Sankoff, Charles Yang, and Gareth Roberts to thank for their ideas, input and mentorship. Being at Penn Linguistics for me has meant being in a place where thoughts flow freely, new ideas grow from offhand comments, and students are treated like colleagues. It's been a vibrant and exciting place to be.

They say that getting a PhD is an isolating experience, and I feel very fortunate that my experience here at Penn has been anything but. Since before I actually started the program, the Linguistics Department has felt like home to me (thanks Aaron Ecay!). From abstract parties to clothing swap brunches, my fellow grad students have supported and inspired me. Thanks to Luke Adamson, Akiva Bacovcin, Ryan Budnik, Aaron Ecay, Haitao Cai, Spencer Caplan, Andrea Ceolin, Pleng Chanchaochai, Ava Creemers, Kajsa Djärv, Sabriya Fisher, Amy Goodwin Davies, Duna Gylfadottir, Anton Karl Ingason, Ava Irani, Helen Jeoung, Soohyun Kwon, Hilary Prichard, Ruaridh Purse, Milena Šerekaitė, Kobey Shwayder, Einar Freyr Sigurðsson, Lacey Arnold Wade, Yosiane White, Rob Wilder, David Wilson. A huge thanks goes out to the members of Phuntax, whose friendship and support have gotten me through more than just Syntax 1. To the extended

[1] The research for this dissertation was supported in part by National Science Foundation Grants #BCS-1628408 and #BCS-1251437.

Penn Linguistics family who have always made me feel like I am part of the community even though they graduated before my time at Penn, and whose ranks I look forward to joining: Joel Wallenberg, Caitlin Light, Suzanne Evans-Wagner, Laurel Mackenzie, Aaron Dinkin, Maciej Baranowski, to name just a few. My graduate career benefited enormously from the administrators of this department. To Amy Forsyth, whose foresight and strongly worded emails kept me on track to graduate. And to Sue Sheehan, who always made coming to work feel like coming home. Who somehow found time to connect with each us and make homemade treats for the whole lab, all while single handedly running Bill's lab. I look forward to many more lunches at Sitar to come.

I've been blessed to also build connections outside of Linguistics Department. To the wonderful women of 510, whose friendships nurtured my soul: Sarah Cate, Sarah Latimore, Hannah Weinstein, Marine Kambara, Madeline Smith-Gibbs. To Haley Pilgrim, Vinayak Mathur, Caitlin VanderWindt and Alex Miller, who have supported me through thick and thin. To my fellow and future IDEAL Council members – it's been an honor to stand in friendship and solidarity with you. May you always stand in victory.

I would not be where or who I am today without the love and support of my family. To my parents, who instilled in me just the right combination of relentless optimism and Protestant work ethic to make pursuing a PhD seem like a great idea. To my brother, who I've always looked up to. To my cousin Eli, who has been a great friend for as long as I can remember. And to Grandma Sip, who showed me how to walk in joy and in grace, and to whom this dissertation is dedicated.

Finally, to Matt. Your willingness to discuss any aspect of phonological change is unmatched even by most linguists, and your clarity of thought always makes my own work better. You've brought light and laughter to my life for the past six years. I could not be more grateful for your companionship. "Thank you" doesn't even come close.

ABSTRACT

MECHANISMS OF PHONOLOGICAL CHANGE

Betsy Sneller

William Labov

The traditional Philadelphia allophonic /æ/ system (henceforth: PHL shown in (1) below) is characterized by a set of complicated conditioning factors and a dramatic acoustic distinction between the two allophones. In recent years, some Philadelphians have begun to exhibit a new allophonic system (NAS, shown in (2) below). Like PHL, NAS is characterized by a dramatic acoustic distinction between tense and lax allophones. NAS is quickly overtaking PHL in the Philadelphia community, as demonstrated by Labov et al. (2016).

(1) **PHL:** $\text{æ} \rightarrow \text{æh} \; / \; \_[\ \text{+anterior}\ ] \cap ([\ \text{+nasal}\ ] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix})]\ \sigma$

(2) **NAS:** $\text{æ} \rightarrow \text{æh} \; / \; \_[\ \text{+nasal}\ ]$

This situation offers an exciting opportunity to observe phonological change in individual speakers. Most phonological changes involve the collapse or creation of a new phonological category; because of the large degree of acoustic overlap in these situations, it is difficult or impossible to identity individual tokens as having been produced by the old or the new phonology. In the current change in Philadelphia /æ/, however, both the old and the new system involve distinct acoustic targets, making it possible to identify which underlying system was used to produce a given word. It is therefore possible to test several distinct theories about phonological change: Whether change occurs through gradual phonetic incrementation (e.g. Ohala 1981), through individual speakers producing only the old or the new system (e.g., Janda and Joseph 2003), or whether change occurs via individual speakers probabilistically producing both the old and the new system in a process of individual grammar competition (e.g., Fruehwald et al. 2013).

In my dissertation, I examine natural speech production from 46 speakers who acquired language during the period of allophonic change, with a combination of topic-directed conversations and targeted natural language experiments. Using a glm classifier, I identify tokens of /æ/ as having been produced by either PHL or NAS. In concert with an analysis of speakers' social histories, I use these results to argue that the change from PHL to NAS in Philadelphia is driven by the mechanism of competing grammars, suggesting that both syntactic change and phonological change proceed in the same manner. My research provides one of the first pieces of direct empirical support for a unified theory of language change in which structural changes in syntax and phonology are implemented through the same mechanism of grammar competition (Kroch, 1989; Fruehwald et al., 2013).

In addition to the theoretical contribution to phonological change, my dissertation also traces the social patterns of the allophonic change, highlighting the effect of network structure and access to elite education on the adoption of the incoming allophonic system. I also employ experimental methods to demonstrate that the abstract allophonic rules of /æ/ are the target of social evaluation and contribute to social meaning. I find speakers producing surprisingly systematic evaluations of PHL and NAS, a result which only emerges when analyzing the evaluation of changing abstract

parameters. Finally, to test whether the change from PHL to NAS was the inevitable result of phonological simplification, I developed a computational simulation built using a principle of language acquisition (Yang, 2016) to demonstrate that the allophonic restructuring in /æ/ was not the result of children simplifying their input data, but rather must have been the result of dialect contact with in-moving speakers of the new system.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Since its inception, the field of language variation and change has made great progress, moving from the question "can sound change be observed?" (Labov et al., 1972, pg. 6) to the question of what observing sound change in progress can add to our theoretical understanding of language and how theoretical linguistics can add to our understanding of variation and change. In this dissertation, I analyze a phonological change in progress, with the goal of using this change to illuminate aspects of phonology that are most visible in an analysis of phonology in flux. In so doing, I also highlight the usefulness of a structural analysis of language variation and change.

Because of the difficult nature of observing phonological change in progress, most hypotheses regarding phonological change are drawn from a post-hoc analysis, with evidence of the language's phonology preceding the change and following the change but sparse or no data from speakers during the change. This set of facts results in necessary speculation about what individual speakers must have produced in order to cause change in a language. This speculation is by no fault of phonologists or sociolinguists: phonological change is difficult to observe in real time, since it occurs relatively infrequently in comparison to phonetic change, and because large scale corpora of speech are, relatively speaking, new sources of data. Add to this the fact that the study of sound change in progress itself is a young field, it is unsurprising that studies of real-time phonological change within individual speakers are rare. This logistical problem of capturing phonological change in real time is eloquently articulated by Hockett's discussion of the phonemic restructuring of /æ/ and /ɔ/ in early Middle English (Hockett, 1958, pg. 456–457, emphasis mine):

> Sound change itself is constant and slow. A phonemic restructuring, on the other hand, must in a sense be absolutely sudden. No matter how gradual was the approach of early ME [(Middle English)] /æ/ and /ɔ/ towards each other, we cannot imagine the actual coalescence of the two other than as a sudden event: on such-and-such a day, for such-and such- a speaker or tiny group of speakers, the two fell together as /a/ and the whole system of stressed nuclei, for the particular idiolect or idiolects, was restructured. *Yet there is no reason to believe that we would ever be able to detect this kind of sudden event by direct observation.*

Hockett points out that an abrupt change in phonological specification is an event so sudden and so difficult to observe that the chances of analyzing it are vanishingly small. In this dissertation, I attempt to do just that. Taking advantage of the large-scale and relatively new Philadelphia Neighborhood Corpus (PNC), I identify a phonological restructuring currently in progress in Philadelphia English /æ/. Using large-scale corpora as well as targeted interviews with the speakers most likely to be undergoing phonological change, combined with social evaluation experiments and a computational simulation of change, I attempt to provide a holistic sociophonological account of this allophonic restructuring.

I begin with a deceptively simple question: When phonological change occurs within a speech community, how do individual speakers contribute to that change? While different theories of phonology and phonological change make different specific predictions about the empirical outputs of individual speakers, it is only recently that our data sources have grown large enough to address this question for sound change; this dissertation represents one of the first large-scale investigations into phonological change in real time. The central drive of this project – determining how individual speakers drive community-wide phonological change – has in turn spawned its own related questions, which are the focus of Chapters 3, 5, and 6.

In §1.1, I outline the minimal theoretical assumptions necessary for my driving question. In §1.2 I describe the three primary theories of phonological change and the predicted outputs of these mechanisms of change for individual speakers. In §1.5, I provide an outline of the chapters in this dissertation.

1.1 Phonological Change

The broad purpose of this dissertation is to investigate how individual speakers' productions drive community-level phonological change. This granularity of investigation represents somewhat of a break from tradition in quantitative sociolinguistics. The empirical study of language change originated as a concept that exists on the level of the community rather than on the level of the individual, as articulated in Labov et al. (1972):

> The general position that we have taken is that no useful distinction can be made between a change and its propagation (Weinriech et al., 1968) as long as we continue to consider language an instrument of communication. The language does not change if one man invents an odd form or develops an idiosyncrasy, even if people understand and evaluate his behavior; it does change when others adopt his idiosyncrasy and use it as a new social convention for communicating their intent.

Historically, the program of analyzing language change has taken as its primary focus the pattern of the speech community as a whole, as it is at this level that the language can be most clearly said to exist and change. Nevertheless, when a language or a dialect undergoes a change, it is through the individual speakers who produce language. Herein lies an apparent contradiction: while the sometimes idiosyncratic and non-prototypical language produced by an individual is not the same as language change, any change in the community is itself made of individuals producing a difference in their own language from that of the previous generation. In the decades since Labov et al. (1972) asked whether sound change can be observed, sociolinguists have documented many sound changes occurring in different speech communities in different languages in real time. As a field, we know quite a bit about how language change works on the level of the community, but not as much about how individual speakers drive that change along. Given the decades of work on how language change operates on the level of the community, we can now turn to the question of how the production of individual speakers works in aggregate to produce the community-level change, which is the goal of the current dissertation.

1.1.1 Modular, Feed-forward Separation of Phonology and Phonetics

Throughout this dissertation, I assume a modular, feed-forward architecture of phonological grammar, following terminology in Pierrehumbert (2002) (see also Bermúdez-Otero, 2007). The basic modular architecture is as shown in Figure 1.1: lexical representation is stored with underlying categorical phonological representation. For example, *mitten* is stored as a lexical entry with the underlying categorical phonemic representation of /ˈmɪtɛn/. This underlying representation then undergoes abstract phonological rules, which are also categorical in nature. Our example *mitten* would undergo /t/ allophony, producing a surface phonological form /ˈmɪʔn̩/ for many American English speakers. At this point, the lexical entry has two categorical phonological aspects: (1) the underlying representation and (2) the abstract rules that result in the surface representation. From this surface representation, forms then undergo gradient phonetic processes to finally produce a phonetic output. The modular aspect of this model separates each process into a distinct level, while the feed-forward aspect means that each stage can only "see" what was given to it by the previous stage; a phonetic process can only make reference to the surface phonological representation it has been fed – it cannot make reference to any underlying representations.

A number of variations on this main architecture have been proposed (see, e.g., Keating, 1985, 1990; Cohn, 1993). Here, I adopt a stratal version of this architecture (Bermúdez-Otero, 2007), as shown in 1.2, which breaks phonology into a stem-level, word-level, and phrase-level module. The underlying phonological representation then may undergo phonological processes at each of these levels, resulting in a phrase-level surface phonological representation that gets fed into the phonetics modules. Under this variation, there are four targets for phonological change: (1) the underlying phonemic representation, (2) abstract phonological rules which produce a stem level representation, (3) abstract phonological rules which produce a word level representation, or (4) abstract phonological rules which produce a phrase level representation. Notably, sociolinguists have often found that the abstract rules applying to each of these phonological levels are the same.

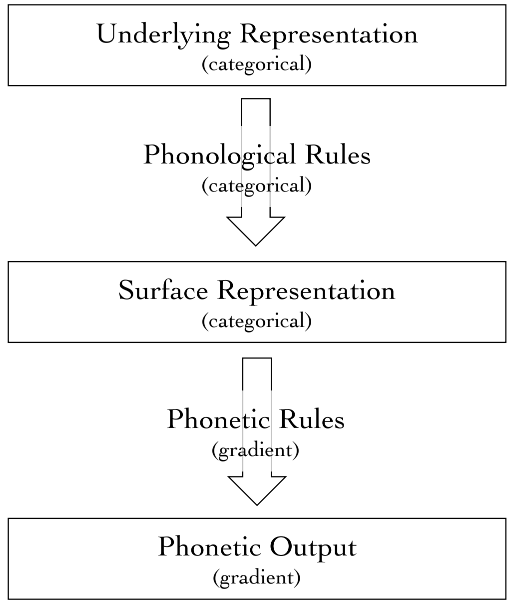


Figure 1.1: Modular, feed-forward phonology-phonetics interface.

In other words, a single rule may be repeated at each of these levels (see, e.g. Bailey, 2017, on /g/-retention in Mancunian English). However, because there are some processes which *only* apply at stem-level (e.g., the Scottish Vowel Length Rule Aitken, 1981) or at phrase-level (such as prosody), this must be representationally possible in the architecture.

The phonological modules in Figure 1.2 are boxed; any change occurring within one of the boxed modules constitutes phonological change. I note briefly that the phonetic components of the architecture are severely underdeveloped in the representation in Figure 1.2; this is fleshed out in several variations (Keating, 1990; Cohn, 1990), and often include distinct modules for language-specific phonetic processes and universal phonetic and articulatory processes. My exclusion of a more detailed phonetic framework in Figure 1.2 is not a theoretical stance, but rather intended to focus the dissertation on the levels of the architecture directly related to phonology. This architecture remains somewhat theory-neutral with regards to the formal specification of phonological processes. These specifications can be formalized under any theory of phonology compatible with

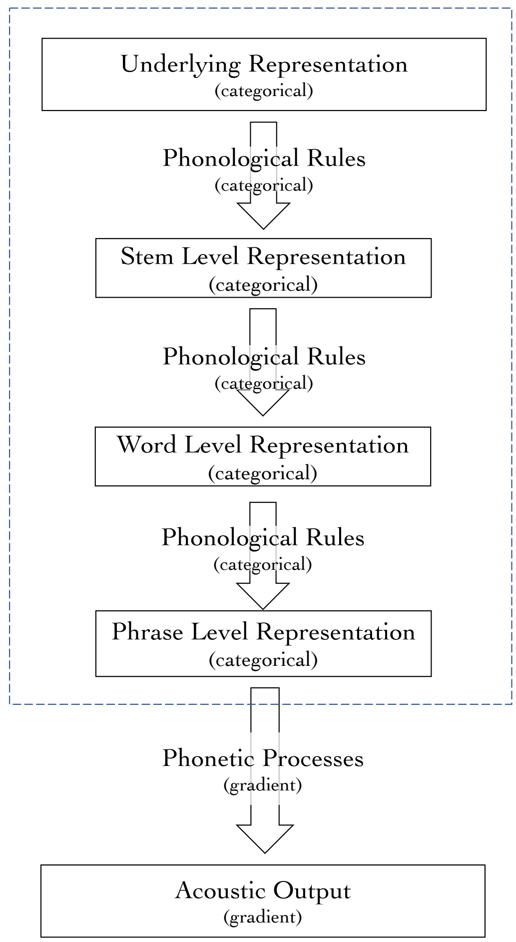


Figure 1.2: Modular, feed-forward phonology-phonetics interface with stratal phonology.

both underlying representations and categorical phonological processes.

Throughout the dissertation, I will refer to phonological processes as "rules," broadly adopting a broadly Generative Phonology framework. This terminology is not a theoretical stance; the phonological processes that take an underlying representation (like /t/) to the surface level phonological representation (like /ɾ/) can also be represented using most varieties of an Optimality Theoretic (Prince and Smolensky, 1993) as well as any version of an Exemplar Theory framework (Bybee, 2002; Pierrehumbert, 2001) that allows for categorical underlying specification which also undergo categorical processes (whether these features and processes are innate or emergent). Since I find rule-based notation easier to discuss, this is the terminology I adopt throughout the dissertation.

Defining Phonological Change

I consider phonological change to be any change to the phonological modules; this means either a change to (a) the underlying representation or a change to (b) any of the rules that produce a surface level representation.

1.2 Mechanism of Phonological Change

While the mechanism of phonological change is difficult to test in real time, there are three primary theories of how individual speakers contribute to community-level phonological change, which will be the focus of my dissertation. Here, I outline these three theories, the factors that govern them, and how they may be identified in the production of individual speakers.

1.2.1 Phonetic Incrementation

There is, to some degree, a level of conventional wisdom shared across a number of phonological frameworks which places the mechanism of phonological change on accruing errors in production or perception. This is the view espoused in Ohala (1981), which lays out a clear argument for the human body, rather than human cognition or abstract linguistic knowledge, as the locus of linguistic change. This is show in Figure 1.3, which provides a schematized illustration of a potential phonological change from /ut/ to /yt/.

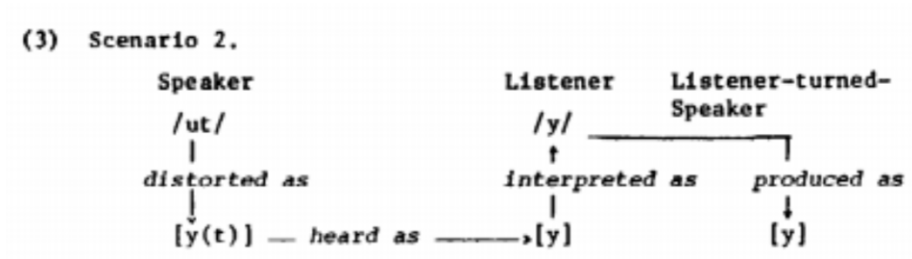


Figure 1.3: Accruing errors as the source of sound change. From Ohala (1981).

Ohala (1981) outlines a number of historical changes which can be accounted for by a perceptual bias of the surrounding phonetic environments that originally triggered such a change. This mechanism of sound change, however, still remains underspecified in terms of abstract linguistic properties. In Figure 1.3, the listener's failure to accurately account for the effects of coarticulation are shown in phonetic terms: the listener at this point has simply shifted their phonetic interpretation of the speaker's phonological content. At this point, phonological change as defined above cannot be said to have taken place. Furthermore, Ohala (1981) does not specify what the tipping point for phonetic incrementation turning into phonological change may be. Despite a lack of explicit specification of how or when this mechanism of sound change affects the abstract segments or rules, the mechanism of phonetic incrementation remains a possible driving force for phonological change; in the most general terms, this means that phonetic or perceptual processes drive sound change until it becomes phonologized either in the middle or at the end of the change (Kiparsky, 2015).

Phonological change via phonetic incrementation is also at the heart of many Exemplar Theoretic accounts of sound change (Bybee, 2002; Pierrehumbert, 2001; Hay et al., 2015). Here I set aside versions of Exemplar Theory that reject the notion of cohesive exemplar clouds altogether (e.g. Bybee and McClelland, 2005), and use the term Exemplar Theory to denote those frameworks that include some level of cohesive phonological identity, which in practice function as phonemes. Under this type of framework, the driving force of a sound change is also placed on listener misperception; here, the variation in the speech signal is caused both by physical reductive processes, such as increased coarticulation and decreased duration, as well as more abstract cognitive reductive processes such as decreased lexical retrieval found in more frequent words (Grainger, 1990; Goldinger, 1998).

While the specific motivation for misperception varies by framework, the crucial driving force for phonological change in both cases is that some level of phonetic misperception accrues, which at some point results in a phonological change.

Phonetic Mitigation

It is worth briefly drawing attention to the difference between phonetic incrementation and what I term *phonetic mitigation*. Phonetic mitigation here refers to a process by which speakers change their phonetic production in response to social stigmatization. Speakers are often found to produce unsystematic phonetic mitigation of stigmatized forms, particularly in settings that are more formal or induce higher attention paid to speech (Labov, 1989, 2001). The crucial distinction between phonetic incrementation and phonetic mitigation for the purposes of this dissertation is in the community-based outcome of the acoustic output: while the acoustic production of phonetic mitigation may look very much like the production of a speaker during phonetic incrementation, the main distinction between the two is in whether or not that output *drives* sound change in the community. While phonetic incrementation drives phonological change in the community, phonetic mitigation is a response to change or evaluation from the community.

To determine whether a speaker's production is phonetic mitigation or phonetic incrementation, a speaker's social environment and peer sociophonological production must also be taken into account. If we find phonetically mitigated output in a number of speakers in a subset of a speech community where the cohort of speakers older than them produce unambiguously non-mitigated tokens and the younger cohort of speakers produce a phonological change, we can conclude that sound change via phonetic incrementation has taken place. If, on the other hand, we find phonetically mitigated output in a speaker whose subset of the speech community already produces the new phonology, we can conclude that the phonetic mitigation of the outlier speaker is not *driving* sound change but rather is the socially motivated response to a change that has already happened.

1.2.2 Spontaneous Phonologization

The second theory of phonological change provides a dramatic foil to change via phonetic incrementation. As argued by Janda and Joseph (2003), this "Big Bang" mechanism of phonological change places the phonologization at a very early stage of the change, wherein speakers innovate phonological and sociolinguistic conditions on a pre-existing (but brief in timespan) phonetic condition. This is taken up more strongly in Fruehwald (2013), who argues that phonologization may occur even *before* any perceptible phonetic conditioning has occurred. This spontaneous phonologization, if independently innovated by enough speakers in a speech community, would then be able to acquire phonetic correlates of the already existing phonological innovation and become a sound change on the level of the community (Ringe and Eska, 2013).

In considering the mechanism of community-wide change, it is important to differentiate between spontaneous phonologization as the solution to the Actuation Problem (reproduced in (3)) and spontaneous phonologization as the solution to the Transition Problem (reproduced in (4), both from Weinriech et al. 1968)

(3) **Actuation Problem**: What factors can account for the actuation of changes? Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times?

(4) **Transition Problem**: [...] the intervening stage which defines the path by which Structure *A* evolved into Structure *B*

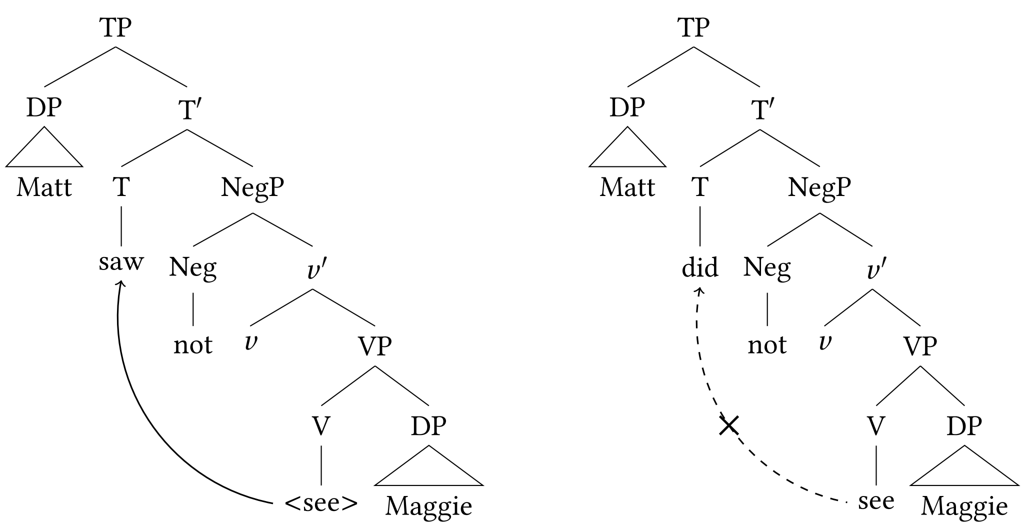
These problems can be thought of as the split between an individual change and a change on the level of the community. The *actuation* of a change asks what causes a change to be innovated by individual speakers. The *transition* of a change asks by what path does structural change then become propagated throughout the community. As a solution to the Actuation Problem, the mechanism of spontaneous phonologization defines how speakers may come to posit idiosyncratic structural changes, and it is largely in this vein that Janda and Joseph (2003) and Fruehwald (2013) discuss spontaneous phonologization. This does not prohibit speakers from also spontaneously positing multiple structural analyses for their input data, which may in fact be a critical aspect of the transition mechanism of competing grammars, which is discussed below. Here, I use the term *spontaneous phonologization* to describe the profile of Transition via spontaneous phonologization, and remain agnostic as to the Actuation of phonological change.

Under community level change via spontaneous phonologization, individual speakers posit a single phonological system and produce that system throughout their speech. In the beginning of the change, very few speakers in a given age cohort will have posited the change, but as time goes on, more speakers in each age cohort will produce the new system rather than the old system. As a mechanism of community-wide change, this predicts that what may look on a community scale to be intermediate productions between System A and System B is actually the result of some speakers producing A and some speakers producing B.

1.2.3 Competing Grammars

The third mechanism of phonological change is an adaptation of syntactic grammar competition to phonology. Grammar competition accounts for the optionality that arises when mutually exclusive parameter settings coexist within the grammar of a single speaker, as in Kroch (1989). While competing grammars grew out of analysis of syntactic change, here I apply this concept to phonological change as well. Under a competing grammars framework, the structured optionality found within each speaker results straightforwardly from variation in a single abstract parameter, providing empirical support for a theory of generative syntax with abstract functional heads. Kroch (1989) demonstrates abstract competition between two variants of a parameter for a number of changes crosslinguistically, including the replacement of *have* by *have got* in British English, the rise of the definite article in Portuguese possessive noun phrases, the loss of verb-second word order in French, and the rise of English periphrastic *do*.

The rise of periphrastic *do* in English provides strong support for a theory of syntactic change through competing grammars, partially due to the large amount of data and partially because analyzing this change as competition in an abstract syntactic parameter provides an explanatory account for a number of distinct surface phenomena which can be best be explained as underlying variation between an abstract syntactic parameter (Kroch, 1989; Pintzuk, 1996).



(a) Middle English V-to-T Raising. (b) Early Modern English V-to-T blocked.

Figure 1.4: V-to-T parameter resulting in DO-support.

The structural analysis of periphrastic *do* in English is analyzed as a consequence of the loss of V-to-T raising (see Figure 1.4b) in English. This abstract structural parameter can be most clearly seen in contexts with an intervening element, such as negation or subject-auxiliary inversion. The evidence for *do*-support arising from the loss of the abstract verb raising parameter in English is also supported by what appears on the surface to be unrelated changes. If verb raising is lost in English, this makes specific predictions about the placement of adverbial forms like *never.* In Modern English, *never* precedes finite verbs (as in *I never found that article*); a pattern that falls out straightforwardly from the loss of V-to-T raising. In a diachronic analysis, Kroch (1989) finds all contexts of V-to-T raising exhibiting the same rate of change (referred to as the Constant Rate Hypothesis), which stands in contrast to the previously received conventional wisdom that syntactic change proceeds context by context.

The concept of competing grammars has, to some extent, been present in the study of phonological change from the beginning of modern sociolinguistics. Empirical Foundations for a Theory of Language Change (Weinriech et al., 1968, pg. 184), describes the transition problem as occurring through speakers with *heterogenous systems*:

Context	Old	New
Intervening Negation	John saw not the cat	John didn't see the cat
Subject-Auxiliary Inversion	Went he to the store?	Did he go to the store?
Subject-Auxiliary Inversion	Where went Matt?	Where did Matt go?
Intervening Adverb	He eats always broccoli	He always eats broccoli

Table 1.1: Some contexts exhibiting differences between V-to-T raising and the loss of V-to-T in English.

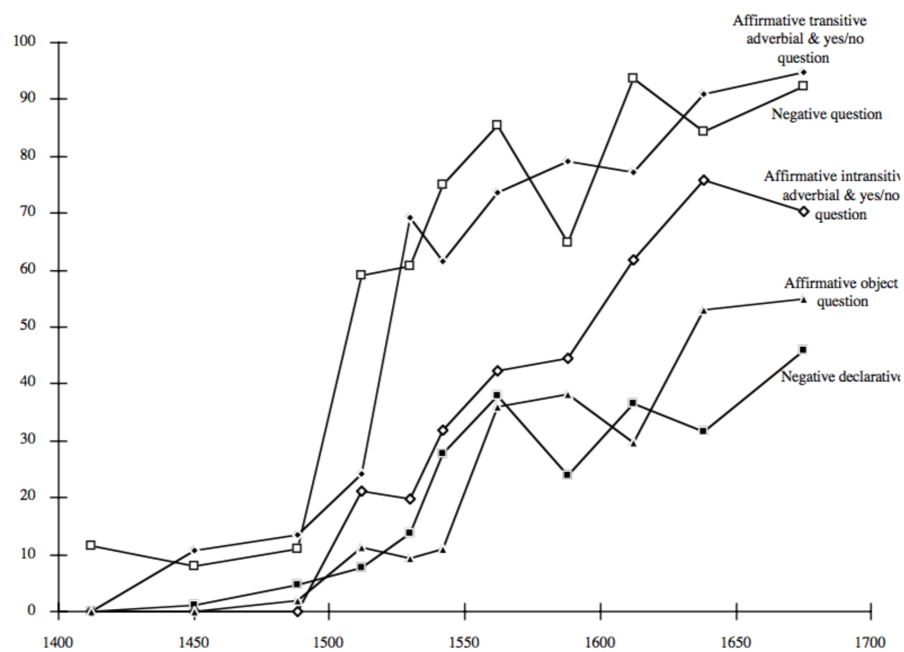


Figure 1.5: V-to-T loss increasing at the same rate across all syntactic contexts. From Kroch (1989).

> This transition or transfer of features from one speaker to another appears to take place through the medium of bidialectal speakers, or more generally, speakers with heterogenous systems characterized by orderly differentiation. Change takes place (1) as a speaker learns an alternate form, (2) during the time that the two forms exist in contact within his competence, and (3) when one of the forms becomes obsolete.

Using the framework of competing grammars more specifically, it becomes possible to make additional predictions about the time when two forms exist within a single speaker's competence. This has been done explicitly by Fruehwald et al. (2013), in an investigation of stop fortition in Middle High German. Using two corpora of written Early New High German, Fruehwald et al. (2013) find evidence for intraspeaker variation between a stop-fortition grammar and a non-stop-fortition grammar, which exhibits a Constant Rate Effect across all potential application contexts. In general terms, applying competing grammars to phonology as a mechanism of phonological change hypothesizes that variation on the level of the community may be the result of individual speakers exhibiting optionality between two options of a single abstract parameter.

Competing Grammars as a Single Parameter

In both syntactic change as well as phonological change, we conceive of the locus of variation being a single abstract parameter that governs surface-level output. Here, an example will be useful. Take, for example, the merger of the vowels in LOT and THOUGHT to LOT which is spreading geographically across the U.S. (Labov et al., 2006) as an example of phonological change to the underlying phonemic representation. A competing grammars mechanism of this change would consider there to be an abstract parameter governing the selection of LOT and THOUGHT for canonical THOUGHT words; within an individual speaker, this parameter would probabilistically select the LOT (merged) phoneme or the THOUGHT (unmerged) phoneme each time the speaker goes to produce a word. Different phonological contexts, such as following or preceding segment, are encapsulated under this single parameter. While the rate of usage across these contexts may differ, a competing grammars analysis requires that these contexts still exhibit the same rate of change, following the Constant Rate Hypothesis (see, e.g. Fruehwald, 2013, for an account of phonological change analyzed using this Constant Rate Hypothesis).

That many contexts are classified under a single parameter which is realized in two competing ways is particularly important when the object of consideration encompasses multiple discrete contexts, as in the case of the phonological change I focus on in this dissertation. The use of *grammar* here in place of *parameter* has occasionally been the source of confusion for readers who are not operating under a Chomskyan theory of syntax, as it may be read to imply that the object under competition is a speaker's entire linguistic competency rather than a single parameter. I highlight here that the term *grammar* in the context of Kroch (1989) is drawn from a Principles and Parameters or Minimalist framework (Chomsky and Lasnik, 2008; Chomsky, 1995), in which syntactic items – both lexical and functional head – are selected by a merge function. In the case of syntactic change, *merge* has the option of selecting between two functional heads. Under this framework, the term *grammar* refers to the objects that are selectable by *merge* and not to a complete description of linguistic competency. In this dissertation, I use the terms *grammar* and *parameter* interchangeably.

Similarly, I refer at times to the allophony of /æ/ under investigation here as an *allophonic system* as well as an *allophonic rule*. As I argue in Chapter 3, any allophonic rule also includes any lexical exceptions to that rule, meaning that *system* and *rule* are synonymous, both referring to one of the two parameters in competition.

1.3 Transition Cohort Speakers

Finally, here I briefly define the target research population of this dissertation, which is the *Transition Cohort Speakers*. I've defined phonological change as a difference in phonology between older speakers and younger speakers within a given speech community. In the time period before any change, every speaker in the community produces the old phonology; after the change is completed, every speaker produces the new phonology. It is the speakers acquiring language in between these two time periods who are of the most interest to the mechanism of phonological change. The phonetic outputs of these transitional cohort speakers are what, in the aggregate, produce the overall community shift. The primary question in this dissertation is whether the transition cohort speakers produce a community-level phonological change via phonetic incrementation, spontaneous phonologization, or competing grammars.

1.4 Disambiguating Evidence

While the three proposed mechanisms of phonological change result in clearly distinct trajectories of a change, it is not necessarily straightforward to disambiguate between the three mechanisms by the production of a single speaker. In this section, I discuss some of the evidence that must be drawn on in order to disambiguate potentially ambiguous data.

1.4.1 Phonetic Evidence for Competing Phonological Parameters

It is occasionally assumed that phonological competing parameters will manifest in a phonetically obvious manner (see, e.g. Dinkin and Dodsworth, 2017). Unfortunately, this is not the case. Phonological change occurring via a mechanism of competing grammars refers *only* to variation in the abstract linguistic parameters. Assuming a modular and feed-forward model of phonology, as I do here (and in fact, as do Dinkin and Dodsworth 2017) means that the phonetic manifestation of the phonological input is not within the domain of phonology. It is possible, in other words, for competing parameters to be active in a speaker's cognitive representation of the language without that competition resulting in an easily measurable output. It could even be active without any difference at all in output, in a situation where the phonetics interpret two distinct surface representation as having the same phonetic output. Setting aside this case, which results in a theoretical distinction without an empirical difference and is therefore a moot point, it remains that phonological competing grammars may not be easily discrete. This is particularly true for phonological mergers and splits, which although produce structurally radical differences, may not be easily identifiable in the phonetic implementation of those abstract differences.

The main point here is that a theory of competing grammars makes no assumptions about the phonetic output of those competing grammars. Of course, a grammar competition that is completely imperceptible to other speakers will not last beyond the speaker(s) who innovated that change. Phonetically distinct but similar outputs, on the other hand, may require an extremely large data set to analyze the underlying mechanism of change. One potential method of distinguishing between competing grammars and phonetic incrementation in a case where the phonetic targets of the two parameters are similar is in the expected standard deviations for conditioning factors under each theory. In general, we would expect change via competing grammars to exhibit higher standard deviations for each conditioning factor (because speakers are actually producing two targets) than we would expect for change via phonetic incrementation (where speakers produce only one target per conditioning factor). Unfortunately, the amount of data required to make a strong distributional case for competing grammars is out of reach for most phonological variables in current sociolinguistic corpora. While advances in recording technology are making it easier to obtain relatively large-scale data sets from speakers, the sheer volume of data needed to distinguish the significance of standard deviations of phonetically similar outputs is, at this point, prohibitive.

Fortunately, phonological mergers or splits are not the only type of phonological change that can be investigated. In this dissertation, I analyze the mechanism of phonological change for an allophonic restructuring in Philadelphia English. The nature of this restructuring means that both the old system (which I call PHL) and the new system (which I call NAS) produce outputs that are phonetically distinct. This means that the amount of data required to identify a competing grammars speaker is relatively small, compared to a merger or a split, making it opportune for investigating the mechanism of phonological change.

1.4.2 Social Evidence for Spontaneous Phonologization

In first-wave sociolinguistics, a speech community is generally thought of as a relatively monolithic entity exhibiting an "enigma of uniformity" (Labov, 2009). And in fact, generally speaking, the level of uniformity in both production of and evaluation of features found across millions of speakers in a single speech community is difficult to explain given speakers' lack of contact with the entirety of their speech community. Layered above this backbone of general uniformity, however, smaller communities of practice (Wenger, 1998; Eckert and McConnell-Ginet, 1992) introduce local socially defined loci of linguistic variation that is often itself socially meaningful in nature.

With change rather than variation in mind, these local socially meaningful community of practice divisions in a broader speech community raise the possibility that phonological change may be introduced or innovated differently across communities of practice within the broader community. This is particularly of importance in trying to determine whether phonological change has occurred via spontaneous phonologization or not. In spontaneous phonologization, speakers produce either the old phonology or the new phonology. However, this output is also consistent with the beginning and end stages of phonetic incrementation and competing grammars. If change has occurred via phonetic incrementation or competing grammars but is affecting different subsets – or communities of practice – within the larger speech community at different times, the output of these speakers as a whole will show some speakers with the old system and some speakers with the new system. Taking only speakers' phonetic outputs into account will not allow us to disambiguate between different mechanisms of phonological change. Instead, the social divisions within a larger speech community must also be taken into account; if all speakers within a subset of the community produce only one system, this suggests that community of practice is not in flux and has either not undergone the change or has already completed the change. If, on the other hand, some speakers within a single community of practice produce the old system and some produce the new system, this suggests change via spontaneous phonologization.

It will therefore be necessary to obtain information on the relevant social divisions within a broader speech community in order to disambiguate whether phonological change has occurred via spontaneous phonologization or another mechanism of change.

1.5 Roadmap

In this chapter, my goal has been to outline the motivating theoretical question of this dissertation and the minimal theoretical assumptions I make. As highlighted in §1.4 above, a full investigation of the mechanism of phonological change must bring social, phonological, and phonetic evidence to bear, which is what I aim to do in this dissertation. The dissertation is organized as follows.

In Chapter 2, I describe the phonological change that serves as the case study in this dissertation, which is the allophonic restructuring of /æ/ in Philadelphia English. I outline the community-level pattern of this change, specifically highlighting the meaningful social divisions produced by the educational system in Philadelphia. I argue that the social divisions produced by the educational system results in communities of practice that either promote the change (in the case of Special Admissions non-Catholic schools) or inhibit the change (in the case of Open Admissions Catholic schools). A bipartite network diagram visualizes the distinct fragmentation in Philadelphia's school system and the subsequent linguistic consequences. Chapter 2 also presents an analysis of the intergenerational pattern of change, finding that the allophonic restructuring of /æ/ occurs in three stages.

Because the allophonic status of /æ/ in traditional Philadelphia English has often been the topic of phonological debate (e.g., Ferguson, 1972; Labov, 1989; Kiparsky, 1995; Dinkin, 2013; Labov et al., 2016), I devote some considerable space in Chapter 3 to a theoretical account of traditional Philadelphia /æ/ as a productive allophonic rule with limited lexical specificity. I propose more generally in Chapter 3 that productive phonological rules, much like productive morphological rules, can tolerate a limited number of lexical exceptions. I specifically appeal to the Tolerance Principle formula from Yang (2016) as a way to define the upper limit of lexical exceptions that a productive process may tolerate. This solution provides a resolution for a number of phonological relationships that have been set aside as troubling or puzzled over as intermediate between phonemic and allophonic under the classic definitions of contrastiveness, without needing to add any additional categories such as *quasi-phonemes* or *fuzzy contrasts* to the phonological architecture.

In Chapter 4, which provides the main evidence for the mechanism of phonological change for the allophonic restructuring of /æ/ in Philadelphia, I take a close look at the speech of transitional cohort speakers to determine which mechanisms of change are at play. I find evidence that the change in /æ/ is occurring via competing grammars in Philadelphia, suggesting that phonological change and syntactic change proceed in the same manner. In this chapter, I also present evidence that the lexical exceptions discussed in Chapter 3 participate in the intraspeaker variation, supporting the claim in Chapter 3 that lexical exceptions are in fact stored as part of the productive phonological rule.

The findings in Chapter 4 suggest the existence of a single parameter governing the choice of allophonic system: in Chapter 5, I investigate whether this abstract parameter may be the target of social evaluation. Using a Matched Guise task, I find Philadelphian participants rating a speaker with the old /æ/ system as more *accented* than a speaker with the new /æ/ system. I follow this with a modified Magnitude Estimation task, which finds Philadelphians evaluating the pronunciation of /æ/ under different conditioning factors in a surprisingly systematic (rather than phonetic) way. My results suggest that not only are speakers able to socially evaluate phonological structure, but that an investigation of evaluation during a period of phonological change may reveal an underlying abstract reason for apparent surface-level results.

Finally, in Chapter 6, I tackle the question of the inevitability of this change, asking whether the complex traditional /æ/ system was destined to be replaced by the simpler, surface-true nasal /æ/ system. Using a computational simulation of acquisition given mixed input, I find that Philadelphian children could not plausibly produce this change through a reanalysis of their input and that instead it is most likely through dialect contact with outside speakers that the nasal /æ/ system entered the Philadelphian speech community.

In Chapter 7, I provide some concluding remarks and directions for future research.

Chapter 2

Phonological Change in Philadelphia

/æ/

While the scale of sociolinguistic data has increased dramatically, given technological advances in recording, transcription, and measurement, most corpora still fall short of the necessary data to analyze the mechanism of phonological change. The problem of capturing the elusive timing of a change, as outlined in Hockett (1958) “on such-and-such a day, for such-and such- a speaker or tiny group of speakers, the two fell together [. . .] and the whole system [. . .] was restructured” requires any empirical investigation into phonological change to contain data from speakers before this sudden restructuring as well as data from speakers after this sudden restructuring. Because phonological restructuring does not occur as frequently as phonetic change, any corpus that encompasses the entire change – before and after – must either be specifically targeted towards a potential change (as in the case of Johnson’s 2010 investigation of the spread of the low-back merger in Massachusetts) or contain enough longitudinal data to capture a change. As sociolinguistic corpora continue to be built up (e.g., Buckeye Corpus, Origins of New Zealand English Corpus, Philadelphia Neighborhood Corpus, Voices of California Project, *inter alia*) this longitudinal data will become more and more possible. In addition to requiring a large longitudinal corpus to capture a change, any phonological change resulting in a merger or a split also will require a massive amount of per-speaker data in order to disambiguate between the three mechanisms of phonological change: since phonetic similarity renders individual tokens difficult to classify phonologically, an analysis of the mechanism of change for a merger or a split will rely primarily on the distribution of the data.

In this dissertation, I focus on a phonological change currently under way in Philadelphia English. This change has two important benefits for investigating the mechanism of phonological change. First, because it is occurring in Philadelphia English, we have a wealth of apparent-time data from before and during this change from the Philadelphia Neighborhood Corpus (PNC). Second, this change is an allophonic restructuring between two /æ/ systems, where both the old system and the new system have two allophonic targets. This means that (1) we have the relevant apparent-time data on the community level to identify the sociolinguistic nature of this change, and (2) it will require less data per speaker to identify the mechanism of change.

Here, I provide an analysis of the community-level pattern and social divisions within this larger community that have an effect on the spread of this allophonic restructuring. I end with an analysis of the intergenerational pattern of this change, analyzing the production of two different families that represent different stages in the allophonic change. Versions of my work presented in this chapter have appeared in Labov et al. (2016) and Fisher et al. (2015). With the goal of limiting a reiteration of previously published work, here I focus on and expand the analysis of those aspects of Labov et al. (2016) and Fisher et al. (2015) that are the most relevant to the question of the mechanism of phonological change.

2.1 Philadelphia /æ/

Philadelphia English, like a number of dialects along the Mid Atlantic region of the United States, traditionally contains a split in the /æ/ phoneme into a lax form and a tense form. Lax forms are produced as a short low front nonperipheral [æ], while tense forms are raised and typically inglided, resulting in one of the following productions: [ɛː^{ə},eː^{ə}, iː^{ə}]. The tense forms, but not the lax, have been found via matched guise test and self-reports to be socially salient and stigmatized (Labov, 2001). The distribution of tense and lax forms can be largely described by a single productive allophonic rule, shown in 5. I will henceforth refer to this traditional /æ/ split as PHL. The phonological nature of PHL – as an allophonic split or a phonemic one – has been the topic of some debate in the literature (see, e.g. Ferguson, 1972; Payne, 1980; Labov, 1989; Kiparsky, 1995; Dinkin, 2013). Here and in recent work (Labov et al., 2016; Sneller, 2018), we have taken the position that PHL is an allophonic split with some lexical specificity. In Labov et al. (2016), this position is based on the empirical pattern of community level variation (see §2.3); in Chapter 3, I expand on this by providing a more detailed theoretical account of PHL as a productive allophonic rule. Here, I represent the tense allophone of /æ/ as æh, following the conventions of Labov (1989).

(5) **PHL:** $\text{æ} \rightarrow \text{æh}\left[\ \text{+anterior}\ \right] \cap (\left[\ \text{+nasal}\ \right] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix})]\ \sigma$

Encroaching on the centuries-long stability of PHL in Philadelphia, there has also been emerging evidence of a new allophonic split governing /æ/ documented in the geographic area surrounding Philadelphia (Ash, 2002) and in more recent years in younger Philadelphian speakers as well (Labov et al., 2013; Prichard and Tamminga, 2012; Labov et al., 2016). This incoming allophonic system, which I refer to as NAS, is shown in 6 below, in which /æ/ is tensed preceding any nasal token. NAS can be found in speech communities across America, including New Haven (Johnson, 1998), the Midland region (Boberg and Strassel, 2000), Ohio (Durian, 2012), Indiana (Fogle, 2008), the St. Louis Corridor (Friedman, 2014), New York City (Becker and Wong, 2009), the West Coast (Hall-Lew et al., 2010), and Michigan (Wagner et al., 2015). Socially, NAS holds the position of being a supraregional standard, which is exemplified by its use in national media outlets such as NPR.

(6) **NAS:** $\text{æ} \rightarrow \text{æh} \ / \ \_\left[\ \text{+nasal}\ \right]$

Here, I've used featural representations to describe both PHL and NAS; this is partially to highlight the fact that NAS can be seen as a featural subset of PHL, and partially because our investigation into the inevitability of NAS replacing PHL in Chapter 6 relies on a featural analysis. For PHL, this rule is represented as a tensing process triggered by a disjoint set of phonological conditions: nasals or voiceless fricatives which are also anterior and syllable final. This produces tense *hand*, where /æ/ is followed by a syllable final anterior nasal /n/, but lax *manner*, where the following /n/ is syllabified as the onset of the following syllable. For clarity of exposition, both PHL and NAS may also be represented by simply listing the set of segmental triggers, as in (7) and (8). As discussed in Chapter 1, I adopt a stratal theory of phonology, enabling us to stipulate that PHL is a phonological rule that applies only at stem-level but not also at word- or phrase-level. This accurately captures the fact that an /æ/ followed by an open syllable in the stem (e.g., *manage*) is produced as lax but that any open syllable created by an inflectional morpheme (e.g., *man+ning the ship*) is invisible to the PHL, resulting tense *manning the ship*. See Chapter 3 for a more detailed account of the phonology of PHL.

(7) **PHL:** / _ æ → æh / _{m, n, f, θ, s}] σ

(8) **NAS:** æ → æh / _{m, n, ŋ}

While NAS appears to be on the rise in dialects across the country, it is worth noting that the phonological effects of NAS as an incoming allophonic system will differ by the regional dialect it is usurping. In many dialects, NAS replaces a phonologically simple system, as in the raised single-target Northern Cities Shift system or the continuous /æ/ system of Eastern New England (Labov et al., 2006). For the White speakers in Philadelphia whose speech is the focus of this dissertation, the incoming NAS system is in community-level competition with one of the most complex allophonic /æ/ systems in English dialects. This provides a particularly interesting case study for the question of the mechanism of phonological change: a changing complex system will enable us to see more clearly whether change does in fact affect all aspects of a complex system simultaneously, as we would expect to find in cases of phonological change via intraspeaker grammar competition or spontaneous phonologization but not necessarily for phonetic incrementation.

There are several additional points to make here about the differences between PHL and NAS, which I will return to throughout the dissertation. First, unlike PHL, NAS is typically a surface-true rule that does not have any lexical specificity (though anecdotal evidence has found some NAS speakers with lexical specificity, particularly in highly frequent words such as the speaker's name adhering to PHL rather than NAS). This makes NAS a phonologically simpler rule, which is often thought to be an inevitable direction for sound change to occur. Not only is NAS a surface-true rule and therefore presumably easier for a language learner to acquire, NAS is also a featural subset of PHL; if we removed three conditions from PHL ([+anterior], [+voiceless fricative], [σ]) this would result in NAS, suggesting a potential route by which PHL could be restructured into NAS. This set of facts raises an important question: whether the allophonic change from PHL to NAS was an inevitable simplification; I return to this question in Chapter 6.

Secondly, because NAS is a featural subset of PHL, there are some tokens that would be produced the same under both PHL and NAS: tense /æ/ in words like *hand*, in which /æ/ precedes a tautosyllabic anterior nasal, and lax in words like *cat*, which fall into the elsewhere condition for both allophonic systems. Tokens belonging to either of these shared conditioning factors will be referred to as *shared* or *training* tokens throughout the dissertation, while tokens belonging to any of the four primary distinguishing factors will be referred to as *test* tokens. Table 2.1 displays the six primary conditioning factors for PHL and NAS, along with their expected realization under each system, their type frequency and their token frequency (see Chapter 3 for a full run down of all conditioning factors and lexical exceptions). For expositional ease, I will refer to each conditioning factor as a lexical set or class of words, following the example in Table 2.1. For instance, a token of the word *path* is considered to be a LAUGH class word, since it has a tautosyllabic anterior voiceless fricative. I refer to the four conditions that differentiate between PHL and NAS (LAUGH, MAD, MANAGE, HANG) as *test conditions* and the tokens that fall under these conditions as *test tokens*. I briefly note that the MAD class in Table 2.1 represents a somewhat strange "conditioning factor," as it is a class of three lexical exceptions produced as tense (*mad, bad, glad*). This list of exceptions remains useful as a condition for PHL, because of its stability across speakers. In contrast, the lexical exceptions produced as lax vary somewhat from speaker to speaker; for this reason, I use the MAD class as a reliable test condition but do not rely on the more unreliable lax exceptions as a test condition. As shown in Table 2.1, the vast majority of /æ/ words, as measured either by token frequency or type frequency, fall under the HAND class or the CAT class, which are the two classes of words that are produced the same under PHL and NAS.

Most critically for a dissertation investigating the mechanism of phonological change, this shift from PHL to NAS in the Philadelphia speech community is a change in the abstract phonological rules governing /æ/ allophony. This allophonic restructuring is a phonological change.

Conditioning Factor	Class	PHL	NAS	Token frequency	Type frequency
Tautosyllabic anterior nasal	HAND	tense	tense	.20	.19
Tautosyllabic anterior voiceless fricative	LAUGH	tense	lax	.16	.07
Lexical exceptions as tense	MAD	tense	lax	.05	.001
Intervocalic nasal	MANAGE	lax	tense	.06	.10
Velar nasal	HANG	lax	tense	.03	.04
Elsewhere	CAT	lax	lax	.5	.6

Table 2.1: The six primary phonological conditioning factors between PHL and NAS. Token and Type frequency obtained from the IHELP corpus.

2.2 Why this change is particularly useful

The allophonic change from PHL to NAS provides a uniquely convenient testing ground for investigating the mechanism of phonological change, for several reasons. First and perhaps most importantly, we have an unprecedented scale of data from speakers born before, during, and even from some speakers after the change. This means that we have unprecedented access to data from transitional cohort speakers, which will allow us to test the mechanism of phonological change using data from speakers during the actual change, providing insight that a post-hoc analysis cannot give us.

Secondly, the structure of PHL and NAS results in both shared and test tokens, enabling us to more easily identify whether any particular token is consistent with either PHL or NAS: both systems have two distinct targets, and the differences in conditioning environments governing which tokens belong in which target between PHL and NAS enables us to identify the underlying system for a given test token of /æ/. For example, a token of *manage* produced in the acoustic space of a speaker's lax allophone is consistent with PHL conditioning but not NAS conditioning, allowing us to identify that specific token as adhering to PHL.

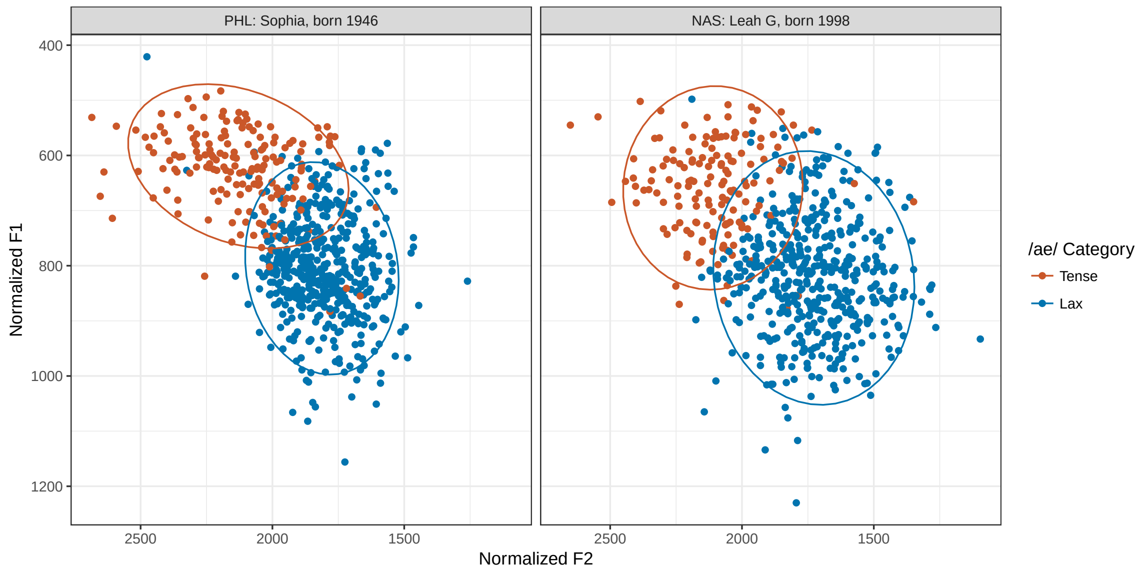


Figure 2.1: PHL (left) and NAS (right) have similar phonetic targets for tense and lax

Thirdly, the acoustic targets of tense /æ/ and lax /æ/ are very similar for the PHL speakers and NAS speakers. Figure 2.1 shows the acoustic output of a PHL system speaker (left) and a NAS system speaker (right) with normalized values of F1 along the y-axis and normalized values of F2 along the x-axis. That the phonetic realizations of the tense and lax allophones of both systems are similar means that the community-level acoustic variation presented here in Chapter 2 is most attributable to phonological change rather than idiosyncratic phonetic implementation of each rule.

2.3 Community Level Pattern

2.3.1 Corpora

The data in this chapter come from two main data sources. The first is the Philadelphia Neighborhood Corpus (henceforth: PNC), which has been thoroughly described in previous literature (Labov et al., 2013; Fruehwald, 2013). The second is the Influence of Higher Education on Local Phonology corpus (henceforth: IHELP), which was previously described in Labov et al. (2016).

The IHELP corpus was designed specifically to obtain data on the reorganization of /æ/ by the population of college students most affected. In contrast to the PNC, which was developed over a period of forty years and was designed to obtain a representational sample of Philadelphia speech, the IHELP corpus was developed between September of 2013 and September of 2016 and was designed to target speakers who acquired language during the period of allophonic restructuring. For the IHELP corpus, twelve undergraduates were recruited from different colleges in Philadelphia, and were trained to conduct sociolinguistic interviews following the classic protocol outlined in Labov (1984). Interviewers primarily targeted their high school and college friends, but also obtained some data from family members. The resulting corpus comprised 170 speakers ranging in date of birth from 1922 to 2006, with the majority of speakers born after 1983. To date, 106 speakers have been transcribed and analyzed using the Forced Alignment and Vowel Extraction (FAVE) program.

2.3.2 Diachronic Acoustic Pattern

Diachronically, PHL has been stable in Philadelphia for over a hundred years; it is only within the last few decades that echoes of a NAS invasion come into play. On the community level, Labov et al. (2016) demonstrate an abrupt shift towards NAS, where all phonological contexts affected begin to shift simultaneously rather than one phonological context at a time. Figure 2.2, adapted from (Labov et al., 2016) depicts this synchronization for the six primary conditioning factors of PHL and NAS, for all White speakers from the PNC and IHELP corpora who produce more than ten tokens of /æ/ in each conditioning environment. To mitigate the possible effect of a talkative speaker skewing the results, each point on the plot represents a single speaker's mean phonetic production of one of the six conditioning factors. F1 and F2 measurements were z-scored by participant, and y-axis represents the measure of the front diagonal (F2-2*F1), with a higher value representing a tenser token. Date of birth is displayed along the x-axis.

The diachronic stability of PHL in Philadelphia is immediately clear: the three traditionally tense main conditioning factors (HAND, LAUGH, MAD) remain tense for much of the recorded data while the three traditionally lax main conditioning factors (MANAGE, HANG, CAT) remain lax. We see the four test conditions exhibit a sudden reanalysis beginning with speakers born

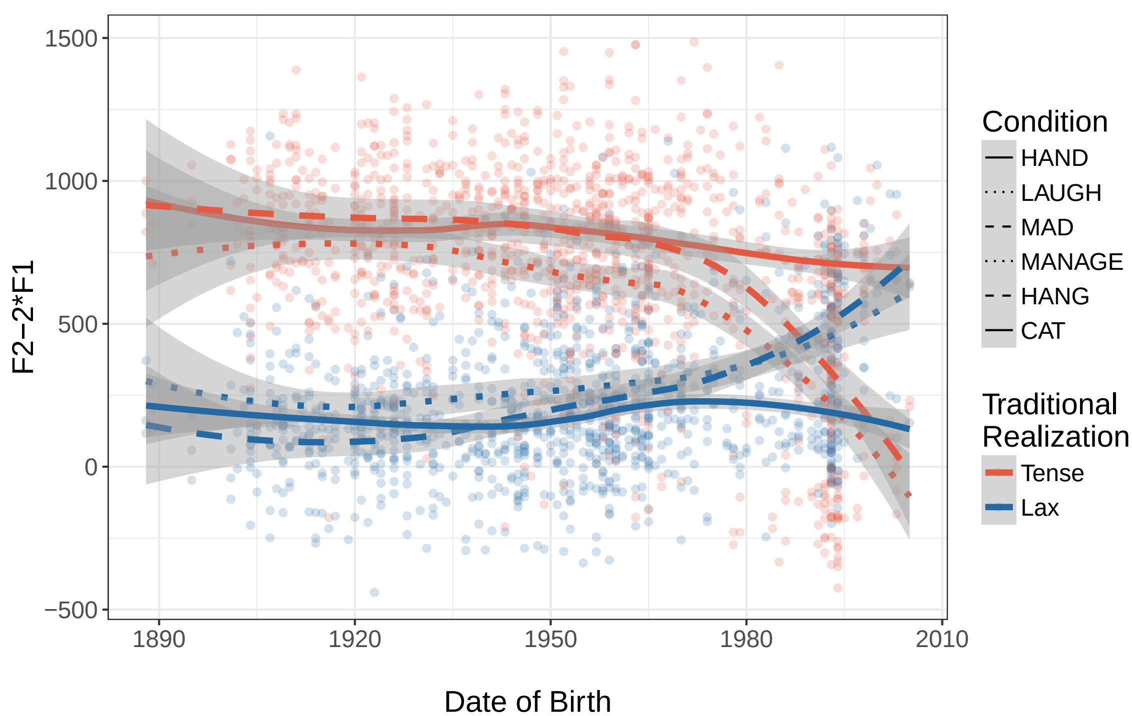


Figure 2.2: Transition of traditional PHL to NAS: LOESS diagram of height along the front diagonal (F2-2*F1) by date of birth. Allophonic restructuring begins around 1983.

around 1983, when MANAGE and HANG begin to rise in average tenseness for the community while MAD and LAUGH begin to plummet in average tenseness. A change point analysis was run separately for each of these four test conditions using the changepoint package in R. This analysis selected 1985 as the change point date for the MAD class, 1983 for LAUGH, 1981 for MANAGE, and 1983 for HANG. These dates are remarkably close, and suggest a wholesale change between two systems on the community level rather than a piecemeal change affecting one conditioning factor at a time.

2.3.3 Measuring Conformity to PHL and NAS by Pillai scores

In this chapter, I analyze the degree of conformity to PHL and NAS for each speaker using the Pillai-Bartlett statistic, following the analysis done in Labov et al. (2016). Each /æ/ system defines a cluster of tense and a cluster of lax vowels, resulting in a bimodal distribution of nearly separate clusters for those speakers who exhibit maximum conformity to either system. In this chapter, I report individual speakers' overall conformity to PHL or NAS using the Pillai-Bartlett statistic (Hay et al., 2006; Hall-Lew, 2010); in Chapter 4 I will take a closer look at each speaker's production of individual tokens. The Pillai-Bartlett statistic uses MANOVA to measure separation, evaluating both the distance between two distributions as well as their variances.

The output is mathematically bounded by 0 (no difference in either mean or variance between the two distributions) and 1 (maximum separation). Used as a measure of acoustic separation for vowels, the maximum separation score lies around .8. To provide a frame of reference, I've included normalized F1-F2 vowel plots of two phonemic distinctions along with their corresponding Pillai scores in Figure 2.3, which displays the separation scores for Leah Green's phonemic distinction between two front phonemes /ɪ/ and /ɛ/ (left) as well as the separation score for her two most distinct vowels /i/ and /a/. As shown in Figure 2.3, a robust phonemic distinction produced in the front vowel space reaches a Pillai separation score of 0.5, while the most acoustically distinct vowel separation in Leah's inventory achieves a 0.8 score.

In comparison then, we see in Figure 2.4 that the acoustic distinction between the tense and lax allophones of /æ/ is relatively robust for both the PHL speaker (left) and the NAS speaker (right).

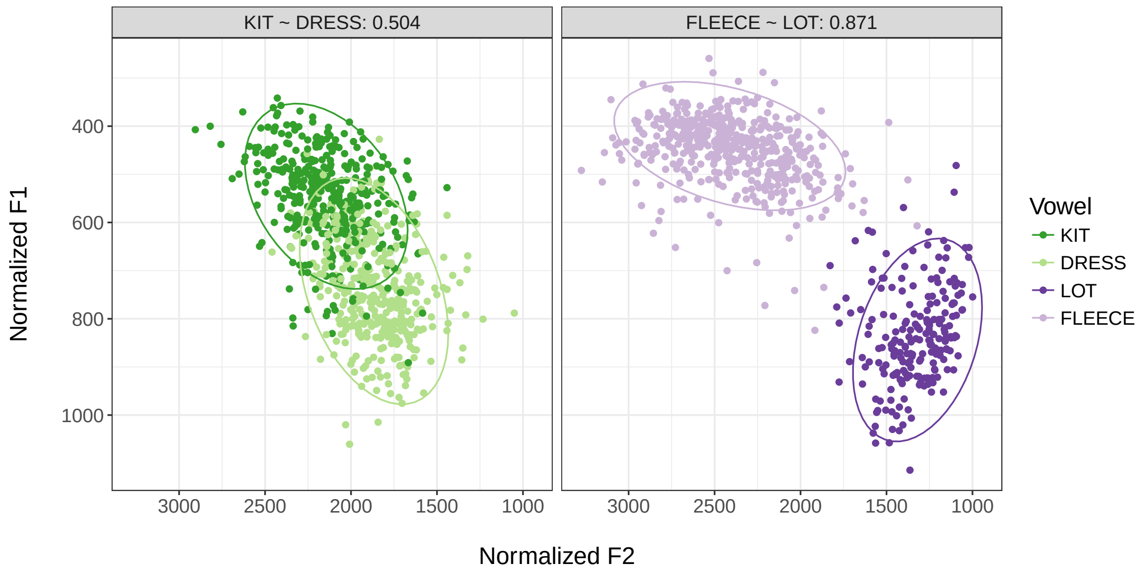


Figure 2.3: Pillai scores for Leah Green's phonemic distinction between KIT and DRESS (left) and FLEECE and LOT (right).

The left panel of Figure 2.4 displays the normalized F1-F2 distribution of /æ/ vowels for the IHELP subject with the highest Pillai score for PHL, 55-year-old Antonio Lyons who has a PHL Pillai score of .769, and the right panel displays the distribution of /æ/ vowels for 16-year-old Leah Green, the IHELP subject with the highest Pillai score for NAS (.727).

We apply the Pillai-Bartlett statistic to the /æ/ distributions of each of the 106 IHELP speakers that have been transcribed and analyzed in FAVE individually, assigning each speaker two Pillai scores: one to measure their conformity to PHL and one to measure their conformity to NAS. These overall conformity results are shown in Figure 2.5, which shows the PHL Pillai score along the x-axis and the NAS Pillai score along the y-axis for each speaker. The higher each score, the better a participant's data conforms to either PHL (along the x-axis) or NAS (along the y-axis). Participants are broken into White speakers (left panel) and Black speakers (right panel).

Each speaker in the IHELP corpus is represented by a single point on the plot. As we will see in §2.3.4, high school education plays a major role in the likelihood that a speaker will conform to PHL or NAS; the two primary educational factors are represented here by color and shape. Catholic

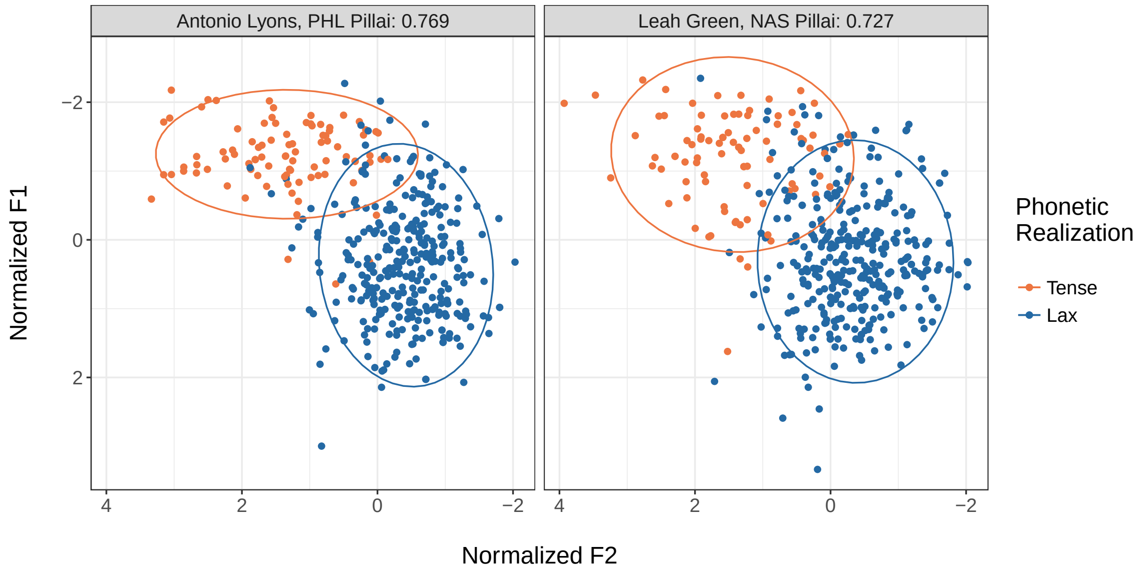


Figure 2.4: Antonio Lyon’s 0.77 Pillai production of PHL (left); Leah Green’s 0.73 Pillai production of NAS (right).

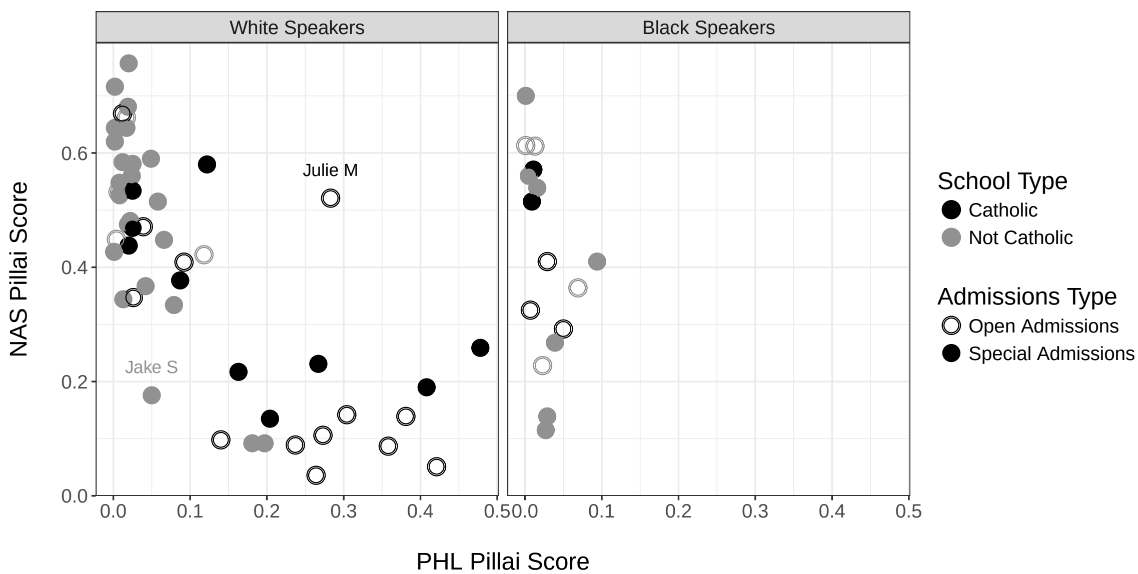


Figure 2.5: PHL and NAS Pillai scores for White speakers (left) and Black speakers (right)

high schools are represented in black while non-Catholic high schools are represented in gray; additionally, Special Admissions schools are represented with solid points while Open Admission schools are represented with open points. The role of Open Admissions Catholic high schools (black open points) in maintaining PHL for White speakers can clearly be seen in the congregation of these speakers along the x-axis.

I've highlighted two White speakers in Figure 2.5, whose Pillai scores stand out as exceptional: Julie M., who exhibits high conformity to both NAS and PHL, and Jake S., who exhibits low conformity to both NAS and PHL. I examine these speakers in some detail in §2.3.4 below.

2.3.4 Social factors conditioning the use of /æ/ systems

In this section, I provide some discussion on the major social factors conditioning conformity to PHL and NAS.

Ethnicity

The separation of White speakers and Black speakers in Figure 2.5 has a theoretical underpinning. The traditional African American Philadelphia /æ/ system is not a split system like PHL or NAS, but rather an /æ/ system with a single phonetic target typically realized acoustically as a long /ɛː/. We see clearly in Figure 2.5 that Black speakers in Philadelphia are also participating in the shift to NAS, employing this change in the service of social mobility alongside White Philadelphians (Labov et al., 2016). However, because the traditional African American /æ/ pattern is a single target, those speakers with a more traditional African American /æ/ show up in the lower left corner with a low separation score for both PHL and NAS.

Contrast this to the White speakers, who for the most part show a PHL-conforming cluster along the x-axis and a NAS-conforming cluster along the y-axis, with almost no speakers in the lower left hand space. Aside from Jake and Julie, the White speakers fall into two clear groups: predominately PHL, with PHL Pillai scores above .15 and NAS Pillai scores lower than .3, and predominately NAS, with NAS Pillai scores above .3 and PHL Pillai scores lower than .15. In Chapter 4 I take a closer look at the production of each speaker; here, I will take an overarching view and bin the White speakers according to these two groups.

The participation of Black Philadelphians in the supraregional change to NAS is an an important example of cohesion across historically distinct dialect groups, and is explored in further detail in (Labov et al., 2016). However, because the phonological change at play in the speech of Black Philadelphians is between the traditional African American Philadelphia English /æ/ system with a single phonetic target and the incoming NAS system with two targets, this change is not useful for analyzing the mechanism of phonological change within individual speakers: analyzing any individual token as conforming to the old neutralized system or the new NAS system will require far more data than we have access to. A change involving two phonetic targets in both the old and the new systems, such as the change from PHL to NAS, enables the classification of each token as conforming to the old system or the new system, making it easier to determine which mechanism of phonological change is at play. For this reason, I focus on the White speakers throughout the rest of the dissertation, whose allophonic change is between two two-target systems and whose output is most likely to bear on the mechanism of phonological change.

Education

For the White speakers, there is ample evidence that we have encountered a systematic "change from above" (Labov, 2001) in which education plays a major role, and here we examine in some detail how the structure of educational institutions in Philadelphia also structures linguistic change, by simultaneously maintaining and exaggerating social class differentials. There is already evidence that speakers with higher education produce less phonetically extreme forms of the salient aspects of the Philadelphia dialect, and in particular less phonetically extreme forms of the tense traditional PHL system (Labov, 2001; Labov et al., 2013). Prichard and Tamminga (2012) and Prichard (2016) demonstrated an effect of a hierarchy of national, regional, and local institutions of higher education (colleges and universities). While these studies suggest a strong effect of the *type* of higher education on the production of local phonology, the data from our IHELP subjects suggest an earlier social impetus for linguistic change. We see, for instance, that even the youngest subjects of the IHELP corpus already display differentiation by high school even though they have not yet enrolled in college (see, e.g., our prototypical NAS speaker, 16-year-old Central High School student Leah Green).

A closer look at the structure of high schools in Philadelphia reveal two main dimensions along which high schools contribute to social stratification. The first of these dimensions is a Catholic vs. non-Catholic distinction. Catholic schools in Philadelphia, particularly in the inner city, historically served the working and middle classes and are seen by many residents as an alternative to neighborhood public schools. While Catholic schools in Philadelphia are open to students from any cultural background, many diocesan schools waive the tuition fee for students whose parents are a member of the local Catholic diocese; this results in a social pattern where Catholic schools, practically speaking, have traditionally served as the White alternative to the predominately Black-serving public schools. This reality can be seen in the relative proportions of White and Black students between Catholic and neighborhood public schools: in Philadelphia Catholic schools today, roughly a third of Catholic high schools are predominately (> 70%) White and one third are overwhelmingly (< 10%) non-White. In comparison, only 1% of the district public schools are predominately white, while two-thirds of local public schools are overwhelmingly non-White.

Among the White speakers in our corpus, those who attend non-Catholic schools typically attend either a Quaker school or an elite public school. Admittance into a Quaker school relies on expensive tuition or on academic scholarships for students whose family can not afford the tuition fee. Admittance into elite public schools is similarly difficult, as I outline below.

Differentiation by Special Admission

In addition to an effect of Catholic vs. non-Catholic school, there is a second educational trait that we find associated with the preference for NAS. A pilot study of high school students in J. R. Masterman High School found all students, regardless of ethnicity, adopting NAS.[2] Masterman holds the position of being an elite Philadelphia high school: it has the highest SAT scores in the state of Pennsylvania, with highly competitive admission procedures and a high rate of success in

[2] The data for this pilot study are not available to be reported here in detail, as the study was conducted by a high school student at Masterman and is not IRB approved for detailed dissemination.

sending graduates to nationally oriented and Ivy League universities like the University of Pennsylvania. Eighteen of the 106 IHELP subjects are graduates of Masterman High School, with 16 of these speakers exhibiting clear NAS productions, one exhibiting potential variation between NAS and PHL (Jerry P.), and one whose data is discussed below as an outlier (Jake S.). A second elite public high school, Central, also shows consistent NAS systems for the seven Central students in the IHELP sample. The three Quaker private schools represented in the IHELP sample similarly show high academic achievement levels overall, along with 4/4 White graduates of thse schools demonstrating high conformity to NAS in our sample.

The academic success that we are associating with the label "elite" here can also be found in several of the Catholic schools. Two schools found in our sample – Nazareth Academy and Roman Catholic – rival the elite public schools in terms of college admissions; these two schools also show a preference for NAS. Much of our background understanding of schools' academic achievement was drawn from the greatphillyschools.org website, which displays high academic ratings for many of the high schools in our sample that have high levels of NAS speakers. However, this site is not useful as a way to operationalize school eliteness, because many of the elite high schools in our sample are not rated on the site. We turn instead to the concept of "special admissions" as a way to distinguish "elite" from "non-elite" schools.

This rating system relies on the social stratification inherent in the structure of the Philadelphia public school system which distinguishes between three types of schools: "Neighborhood", "City Wide", and "Special Admissions." Neighborhood schools have an attendance boundary that gives admission priority to students living within that boundary. Students living outside of the neighborhood boundary are able to submit an application for acceptance consisting of a request to join, and final acceptance is selected by lottery. For these Neighborhood schools, academic performance does not factor into admissions. Both City Wide and Special Admissions schools require a more extensive application to attend, and admission is based upon entrance requirements that include both behavioral and academic performance. Although City Wide schools–which include technical and vocational curricula–have an element of competitive entrance requirements, the final selection for admission is made via computerized lottery. Special Admissions schools, on the other hand, select successful candidates based upon a rigorous set of requirements that include behavior records, test scores, and in-person interviews. For a child applying to a Special Admissions elementary school, the vetting process may include several trial "play dates" with the child and a parent in attendance, as an assessment tool. Higher level schools, both middle and high school, often require a formal interview.

The two elite Catholic high schools in our sample are distinguished by the same criterion. The Nazareth Academy admissions page features a 7th grade Practice Test as well as an 8th grade entrance examination. The Roman Catholic admissions process advertises a 1st, 2nd, and 3rd Entrance Test and warns that "any student wishing to attend Roman Catholic is required to take the High School Placement Test." In this process, academic admissions tests are kept separate from any scholarship examinations which determine how much financial aid will be offered to accepted students. Contrast this to Father Judge, a non-elite Catholic school, which begins the admissions page by stating that "all 8th grade students who would like to compete for an academic scholarship must take the Scholarship/Placement test." For Father Judge, this test is not required for admission but rather only taken in the event that a student wishes to apply for financial aid.

Regression Analysis of Social Factors

Table 2.2 shows the results of two separate linear regression models for the 71 IHELP subjects who were enrolled as undergraduates during the period of data collection, predicting Pillai score for each of the two systems. Although the effect of college type on retreat from local dialect features is a significant indicator for the speakers and features analyzed in Prichard (2016), including students' choice of college (whether Locally, Regionally, or Nationally-oriented) did not significantly improve either model fit here, and therefore was taken out of the model.

We find in Table 2.2 overall confirmation of the patterns described above. The Catholic status of a speaker's high school is the strongest predictor of their overall conformity to /æ/ system, with Catholic schools favoring PHL and Non-Catholic schools favoring NAS. We see also an effect of Special Admissions for both Catholic and Non-Catholic schools, with the elite Special Admissions schools favoring NAS. There is an effect of Ethnicity on conformity to PHL; this is unsurprising,

	PHL Pillai		NAS Pillai	
	Estimate	p-value	Estimate	p-value
Non-Catholic	-0.18	$p < 0.001***$	0.19	$p = 0.04*$
Special Admissions	-0.04	$p = 0.03*$	0.07	$p = 0.03*$
Ethnicity (Black)	-0.09	$p = 0.005**$	-0.00	$p = 0.99$
Gender (M)	0.037	$p = 0.17$	-0.08	$p = 0.11$
Non-Catholic:Special Admissions	0.04	$p = 0.45$	-0.07	$p = 0.5$

Table 2.2: Social factors conditioning PHL and NAS Pillai scores among college students in the IHELP data set.

since we find traditional White Philadelphia English producing PHL but traditional Black Philadelphia English producing a neutral /æ/ system. In contrast, we find no effect of Ethnicity on conformity to NAS; this is unsurprising, as we have seen in Figure 2.5 that Black speakers participate in this change to NAS. It is worth briefly pointing out that we do not find any effect of Gender on conformity to PHL or NAS; this is somewhat surprising given that changes from above typically exhibit an effect of gender, with females leading in the use of the incoming standard (Labov, 2001).

Outliers

Here I return to the two speakers whose Pillai scores for NAS and PHL make them outliers amongst the White speakers. The first outlier is Julie M, whose short interview yielded a total of 171 /æ/ tokens (avg. 324 per speaker in the IHELP data set) which consisted of a higher than average proportion of training tokens (72% in Julie's interview, avg. 49.4% in the IHELP data set). Because Julie's Pillai scores were based on tokens that primarily conformed to both systems (being predominately training tokens), her PHL Pillai score and NAS Pillai score are both high. Julie's output is displayed in Figure 2.6, where her HAND class tokens, represented in red, display her tense target and her CAT class tokens, represented in blue, display her lax target. Julie's test tokens are plotted in black lettering above her plot.

In terms of Julie's Pillai scores, her high conformity to both PHL and NAS is driven by the proportion of training tokens to test tokens (153 training: 18 test). The fact that Julie produces some of

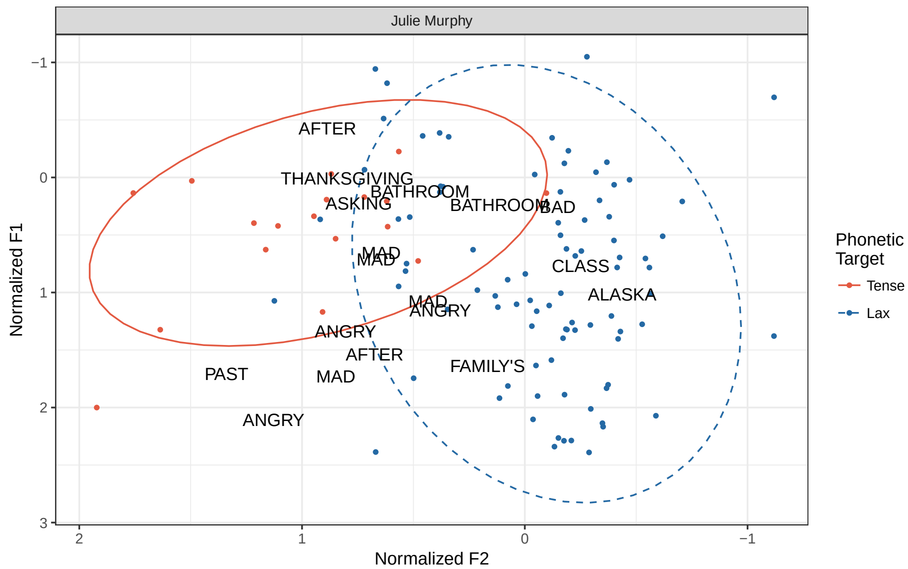


Figure 2.6: Julie Murphy's production of /æ/.

her LAUGH class tokens as tense (e.g. *after, asking, bathroom*) and some as lax (e.g., *class, Alaska*) suggests the operation both PHL and NAS as competing grammars in Julie's production, since the tense tokens conform to PHL while the lax tokens conform to NAS. However, we note that even in Labov (1989)'s analysis of traditional PHL speakers, conducted before the incursion of NAS into Philadelphia, participants produced up to 15% of their tokens as incongruent with their dominant traditional PHL conditioning. In other words, Labov (1989) found participants laxing traditionally tense words up to 15% of the time. Of Julie's 18 test tokens, this proportion would predict roughly 2.7 incongruous tokens, of which only *class* and *Alaska* are unambiguous examples. In other words, Julie simply does not produce enough data for us to analyze any particular mechanism of phonological change. In fact, as we will see in Chapter 4, a paucity of test tokens per speaker in the IHELP data is a common problem for our program of determining the mechanism of phonological change. As it stands, we must simply set aside Julie's data as too sparing to be useful.

The second outlier in Figure 2.4 is Jake S, whose production is displayed in Figure 2.7. Unlike

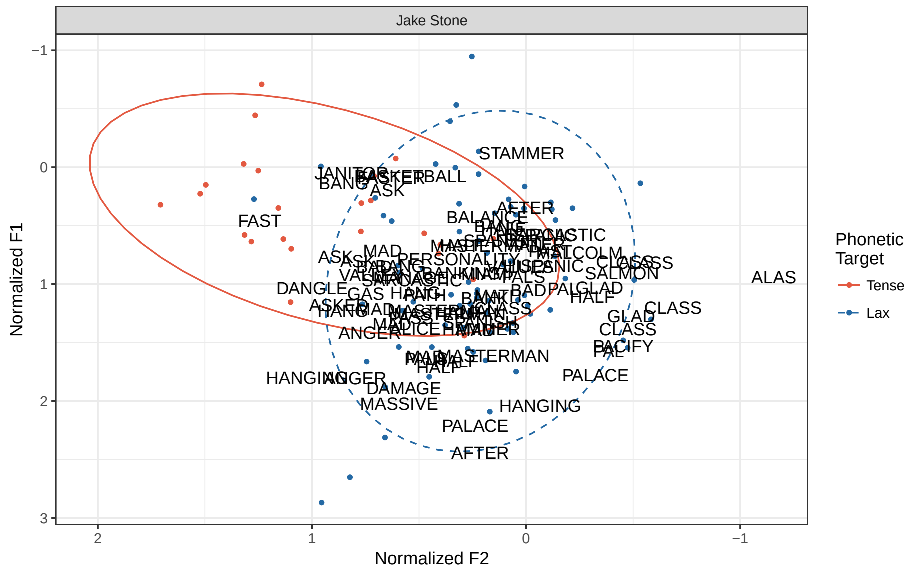


Figure 2.7: Jake Stone's production of /æ/

Julie, and in fact unlike the rest of the IHELP speakers, Jake produces a clear phonetic lowering of his traditionally tense PHL test tokens (LAUGH and MAD classes). This production is predicted by a phonetic incrementation mechanism of phonological change; if speakers habitually laxed their stigmatized tense productions of LAUGH and MAD, this would result in transition cohort speakers producing outputs similar to Jake's. This cohort would then be followed by a cohort of speakers that reorganizes the apparent merger, and begin to tense NAS test tokens (MANAGE and HANG classes).

In phonological change via phonetic incrementation, productions like Jake's would drive sound change and result in incremental steps toward NAS. However, as we have seen in Figure 2.5 and in the results of the regression analysis presented in Table 2.2, Jake's age and social cohort predicts that he would produce NAS. As we have seen, nearly all of his classmates at Masterman produce a NAS system, and Jake emerges as an outlier given his education. This social situation suggests that Jake's production is *phonetic mitigation* rather than phonetic incrementation. In other words, Jake's production is more likely the result of phonetically laxing his underlyingly tense PHL test tokens in *response* to his NAS-speaking environment than it is the driving factor in his peers adopting NAS. As Jake's data is an outlier due to phonetic mitigation rather than phonetic incrementation, I set his data aside.

2.3.5 Network Analysis

As we have seen in the outputs from the regression analyses of Pillai scores, a Philadelphian's educational history has a clear impact on their adoption of local or supraregional /æ/. In Figure 2.8, a bipartite social network (Dodsworth, 2014) provides a visual representation of the impact of school networks. Unlike typical social network diagrams, which place individuals as nodes on the graph and link these nodes together with edges to represent personal connections or interactions between two individuals, bipartite social networks have two types of nodes. One type of node is the individual. These individuals are linked to the second type of node, which here is the educational institutions they attended. This method has been used successfully by Dodsworth (2014) to demonstrate the importance of school affiliation and centrality in the retreat from the Southern Vowel Shift in Raleigh. One of the benefits of a bipartite social network diagram such as this is that it can capture the socialization effects that an institution typically has on individual speakers; while two speakers in our sample who graduated from the same school may not be connected personally, these two speakers will have both been strongly influenced by the norms of that institution.

Because I find school type (Catholic vs. not Catholic) and admissions type (Special Admissions vs. Open Admissions) to be the strongest effect on which /æ/ system the White IHELP speakers conform most to, in Figure 2.8 I bin our school nodes along these two dimensions. Each point in the network diagram represents a single speaker, and the edges in the diagram connect each speaker to the type of middle school they attended as well as the type of high school they attended. This plot only traces the White speakers, which is the community that varies between PHL and NAS. Speakers have been binned according to their location on the PHL-NAS Pillai score plot (Figure 2.5): all speakers with a PHL score above 0.17 and a NAS score below 0.27 have been classified as PHL-dominant speakers and are represented in orange while speakers with a PHL score below 0.17

and a NAS score above 0.27 have been classified as NAS-dominant speakers, represented in green. Note that Julie M. and Jake S., the exceptions from Figure 2.5, are excluded from Figure 2.8.

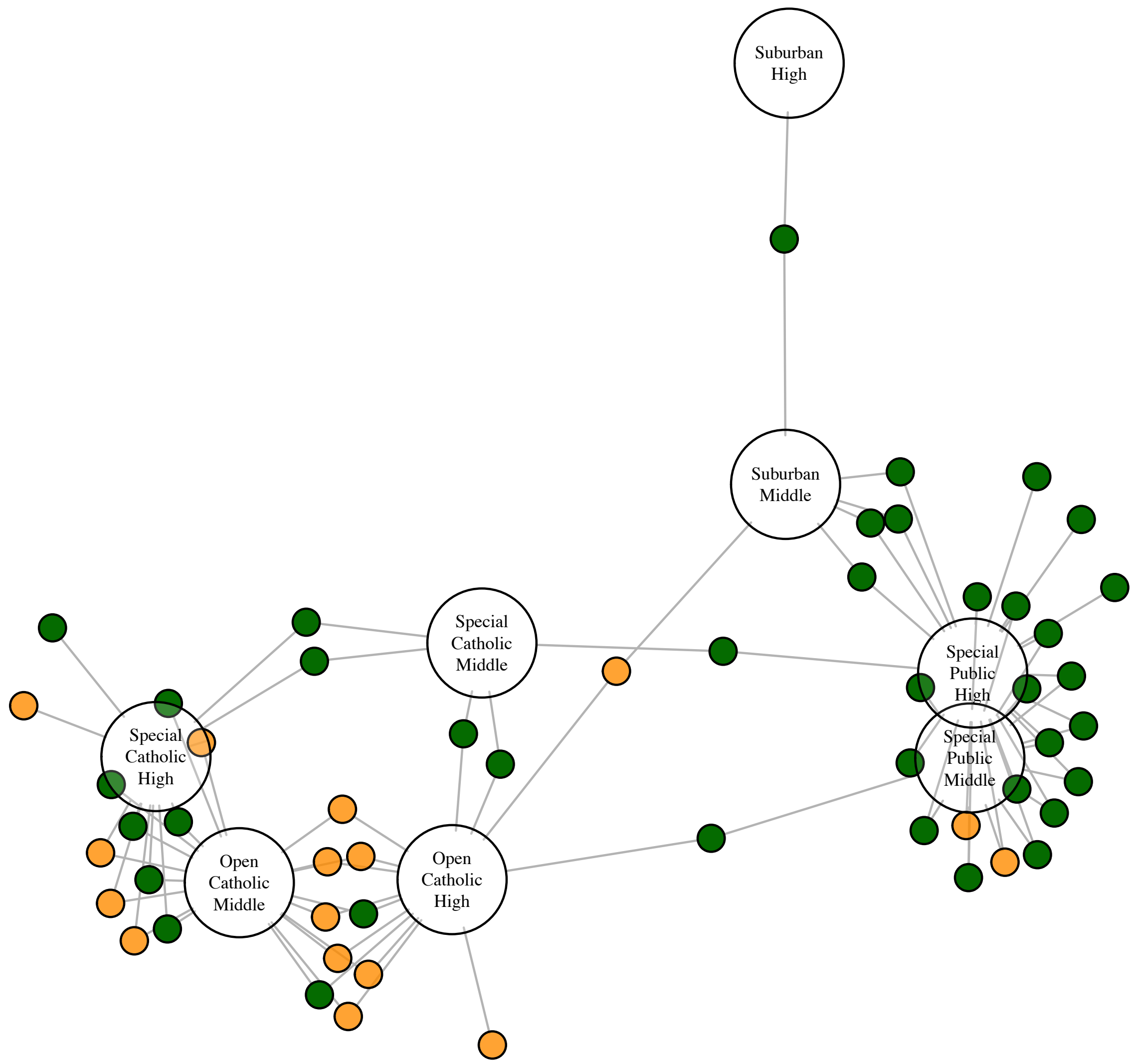


Figure 2.8: Educational paths follwed by IHELP subjects from middle school to high school. Orange = PHL speaker, Green = NAS speaker.

The nodes at the bottom left of the graph show the speakers who attended Open Admissions Catholic schools and Special Admissions Catholic schools. We can see that several speakers in our sample have moved from an Open Admissions Catholic middle school to a Special Admissions Catholic high school; these speakers are more likely to exhibit a NAS-dominant system than their peers who went from an Open Admissions Catholic middle school to an Open Admissions Catholic high school. On the right side of the plot we see the strongly NAS-dominant cluster of speakers who attended Special Admissions middle and high schools. It is worth noting that none of the White speakers analyzed from the IHELP sample attended an Open Admissions public school. This is in large part due to the socioeconomic split in religious schools in Philadelphia whereby working-class Whites use Catholic schools as an alternative to public schools while upper-class Whites turn to elite public schools or expensive private (typically Quaker) schools. In other words, Open Admissions public schools are not a typical choice for White students in general, and are especially underrepresented in this sample of speakers which focuses heavily on speakers who were accepted into regionally and nationally oriented universities. I have included a node for Suburban middle school and Suburban high school; these nodes represent schools that are Open Admissions but are located in a wealthy suburb of Philadelphia. The funding model for these schools, like most American public schools, draws largely on the property taxes of houses in the school's catchment area, meaning that students attending high school in a wealthy suburb of Philadelphia are largely from relatively wealthy or socially elite backgrounds. Perhaps unsurprisingly, speakers who share a connection to the Suburban schools overwhelmingly produce the high prestige NAS system.

Figure 2.8 clearly shows the fragmentation of Philadelphia delineated along school institution type. Students from one type of middle school rarely attend a different type of high school. Perhaps most strikingly, the strongest PHL holdout (Open Admissions Catholic high schools) have almost no connection with the strongest NAS section of the community (Special Admissions Public schools). We can see clearly that the fragmentation of the speech community along the lines of educational institution plays a large role in the diffusion of this linguistic change across the city. As we have noted in Labov et al. (2016), the Catholic school system in Philadelphia serves here as a conservative linguistic force, in which PHL still has a foothold amongst young speakers and NAS may only just be on the way in now. We see also that for the IHELP speakers, the path of linguistic change follows the social fragmentation of the city. In this case, the elite school systems act as filtering devices for young Philadelphians, selecting those that will become the next generation of socially elite and imbuing them with the linguistic capital to signal this social mobility.

2.4 Intergenerational Pattern

In any sound change in progress, intergenerational data provides important insight into the development of the change by tracing its transmission from parent to child. In the course of collecting the IHELP corpus, several of our interviewers obtained data from their family members, which enables us to take a close look into the intergenerational pattern of /æ/. I have previously discussed some of this data in Fisher et al. (2015), which includes a discussion of speakers' productions of THOUGHT as well as speakers' conformity to PHL and NAS as measured by Pillai score. Here, I take a more focused look at the productions of /æ/ for the white speakers reported on in Fisher et al. (2015), using both the overall measure of Pillai score as well as a more in depth look into the production of individual word tokens.

2.4.1 Data from the Family

Here, we have an opportunity to investigate both how children adapt the linguistic system given to them by their parents as well as how those children's peer groups have potentially influenced that system as well. Previous work has found that while children initially acquire the linguistic system of their parents, these early acquired patterns are often lost unless they are reinforced by their peer group. Lacking this reinforcement, children tend to match their peer input by the time they reach adolescence (Labov, 1972; Kerswill and Williams, 2000).

As I've shown in detail above, the educational systems that children attend also have an effect on their language use. In the case of Philadelphia, this is at least partially due to simple population effects – people speak like the people they are around, and as we've seen, the educational system in Philadelphia serves in practice to separate people into distinct subgroupings with relatively little interchange between the subgroups. Education has also been found to play a more directly social role in language use, even after adolescence. In a panel study, De Decker (2006) investigated the production of four young women from a small town in Ontario as they attended college in the larger cities of Toronto and Waterloo. Two of the four women produced a more retracted /æ/ over time, shifting their production to match their more urban-oriented peers. This finding is echoed in Prichard (2016), who found that speaker's local dialect features (in Philadelphia and Raleigh, NC) were affected by the type of college they attended, with local features more likely to be maintained by students attending locally-oriented universities and more likely to be abandoned by students attending nationally-oriented universities.

Lyons Family

I begin this section with a close look at the linguistic production of the Lyons Family. The Lyons are an Irish-Italian family from Northeast Philadelphia. Christine, who was a 20-year-old undergraduate at the University of Pennsylvania at the time of her recording in 2014, was one of the undergraduate interviewers for the IHELP project who interviewed her family members as part of the project. Her father, Antonio, has been referenced above as the speaker in the IHELP data set with the highest conformity to PHL as measured by Pillai scores. Here, we analyze the production of her parents, Antonio and Theresa, Christine, and her younger brother Rocco.

In the figures that follow, each speaker's HAND and CAT class words are plotted in gray with a solid line (for MAN) or a dotted line (for CAT) marking the 95% confidence ellipse. This provides a benchmark for each speaker's tense and lax phonetic targets. Each test token is plotted in text above this, with words in the LAUGH and MAD class plotted in red (as they would be produced as tense under PHL) and words in the MANAGE and HANG class plotted in blue (as they would be produced as lax under PHL). A speaker who fully conforms to PHL should produce red tokens in their HAND cloud and blue tokens in their CAT cloud, while a speaker who fully conforms to NAS will produce blue tokens in their HAND cloud and red tokens in their CAT cloud.

I begin by first analyzing the productions of the parents, Antonio (Figure 2.9) and Theresa (Figure 2.10). Antonio's production fits straightforwardly with a classic PHL system. He produces a phonetically extreme distinction between his tense and lax targets, with almost categorical adherence to PHL. A few exceptional words stick out clearly in Antonio's production: one token of *planet* clearly produced in his tense range, and one token of *asteroid's* clearly produced in his lax range. Overall, however, his production fits with the expected realization of a classic PHL speaker, resulting in a very high PHL Pillai score of 0.74. His NAS Pillai score ranks very low, at only 0.12. Theresa, likewise, produces a classic PHL distribution, with a clear distinction between her tense

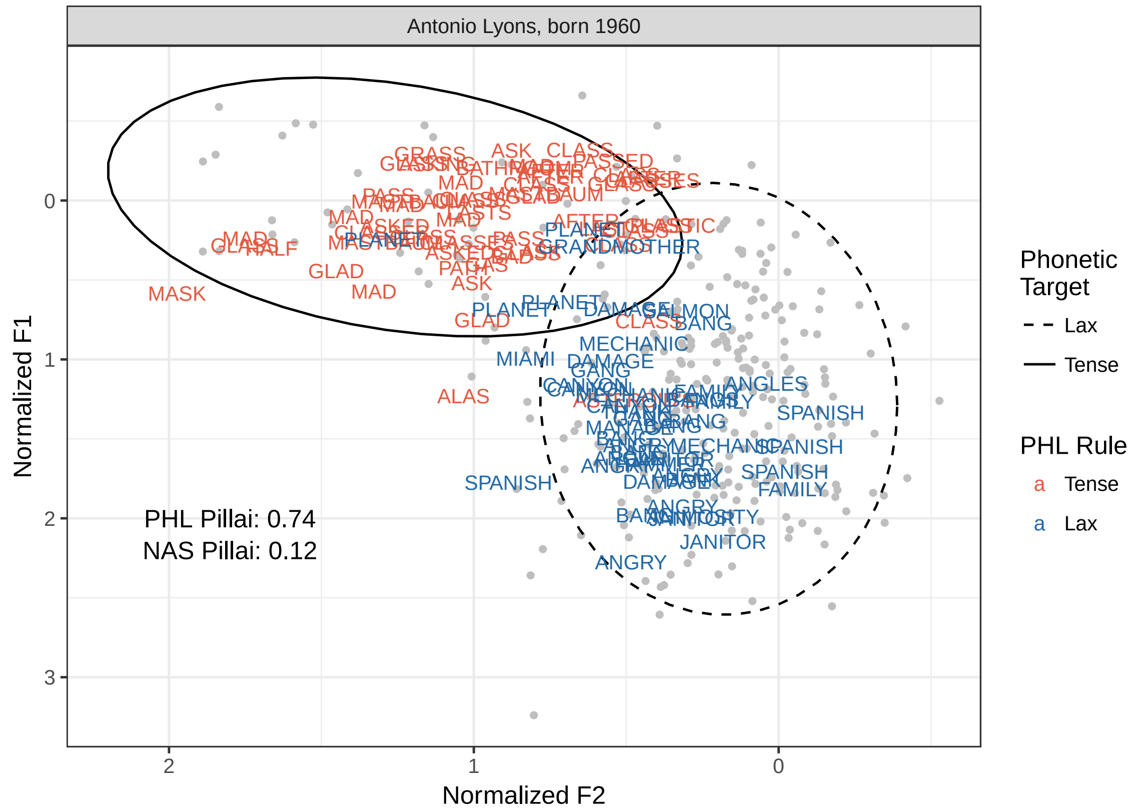


Figure 2.9: Antonio Lyons PHL production.

and lax targets (PHL Pillai: 0.62, NAS Pillai: 0.11). She also produces a token of *planet* as exceptionally tense, as well as a token of *alas* as exceptionally lax. Both of these words are not surprising as lexical exceptions; as discussed above, Brody (2011) found *planet* emerging as a lexical exception to tense for some speakers, and the words *alas* and *asteroid* both classify as "learned words", which are typically produced as exceptionally lax by PHL speakers (Labov, 1989). Overall, the picture from the Lyons parents is that the input to the children would have been a clear PHL system from both parents.

Turning to the children's productions, it becomes possible to see the effect of peer group and the changing community norms on the speech of the children. We begin by examining the speech of Rocco, a 15-year-old high school student at Father Judge, an Open Admissions Catholic school. Based on what we know about his parents' input to him and his demographic data as having

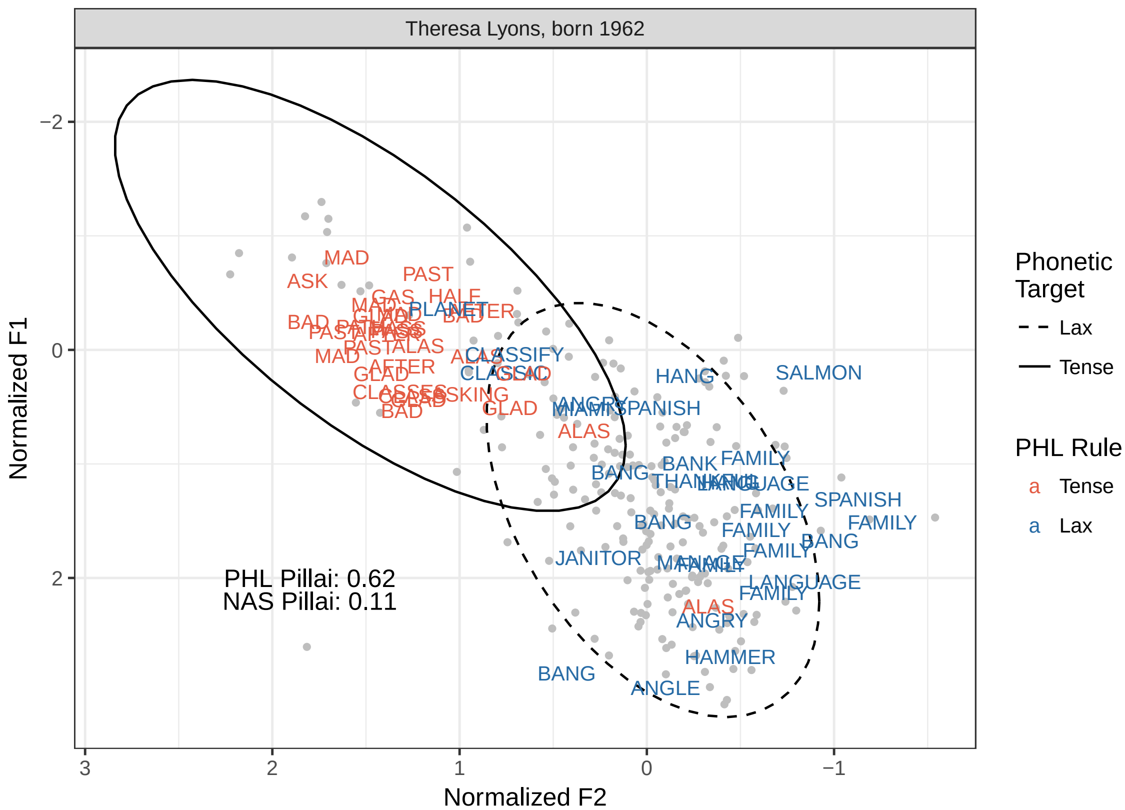


Figure 2.10: Theresa Lyons PHL production.

attended his local diocesan school for middle school and Father Judge for high school, Rocco is a prime candidate for retaining the traditional PHL system. As we can see in Figure 2.11, this is more or less the case. He produces most of his LAUGH and MAD words (in red) relatively in line with his tense target and most of his MANAGE and HANG words (in blue) relatively in line with his lax target. We can see a few exceptional tokens emerging: the unsurprising *planet* as tense, as well as a tense production of *angry* and *hang*. Overall, Rocco produces an output that conforms quite well to the traditional PHL system, even as his Pillai scores appear quite low (PHL: 0.45, NAS: 0.15). This low value for Pillai is partially due to the fact that he was not a verbose speaker, generally providing his sister with very short answers to her interview questions, as perhaps may be expected for a high school boy speaking with his older sister. This low token count increased the standard deviation for each word class, which in turn decreases the Pillai score for both PHL and NAS. Even so, it is clear from his Pillai scores as well as from an examination of his vowel plots that Rocco conforms overall to the expected traditional PHL pattern.

It's in the production of Christine that we begin to see some breakdown of the traditional PHL pattern. Like her parents, she still produces a clear and phonetically distinct tense target and lax target. Her Pillai scores, however, do not reveal a strong conformity to one system over another (PHL: 0.33, NAS: 0.26). In Fisher et al. (2015), using only the Pillai score, we classified Christine as a "weak PHL system speaker". Here, I take a closer look at her actual production to determine the driving force behind her overall Pillai scores. It is clear that Christine produces far more tokens incongruently with PHL than her parents or brother did. We see one tense token of the HANG class (*banker*), and quite a number of lax productions of her LAUGH class. Unlike Jake S., whose production I analyze as a PHL speaker who has phonetically mitigated his LAUGH and MAD class tokens, Christine exhibits clear variation, producing some of her LAUGH tokens as tense and some as lax. This provides a suggestion of the operation of competing PHL and NAS in her linguistic system, which I will return to in Chapter 3. Importantly, Christine's educational background also plays an important role in her linguistic production. Like her brother Rocco, Christine attended her local diocesan middle school followed by an Open Admissions Catholic high school. However, Christine also has gone on to attend the nationally-oriented University of Pennsylvania, which

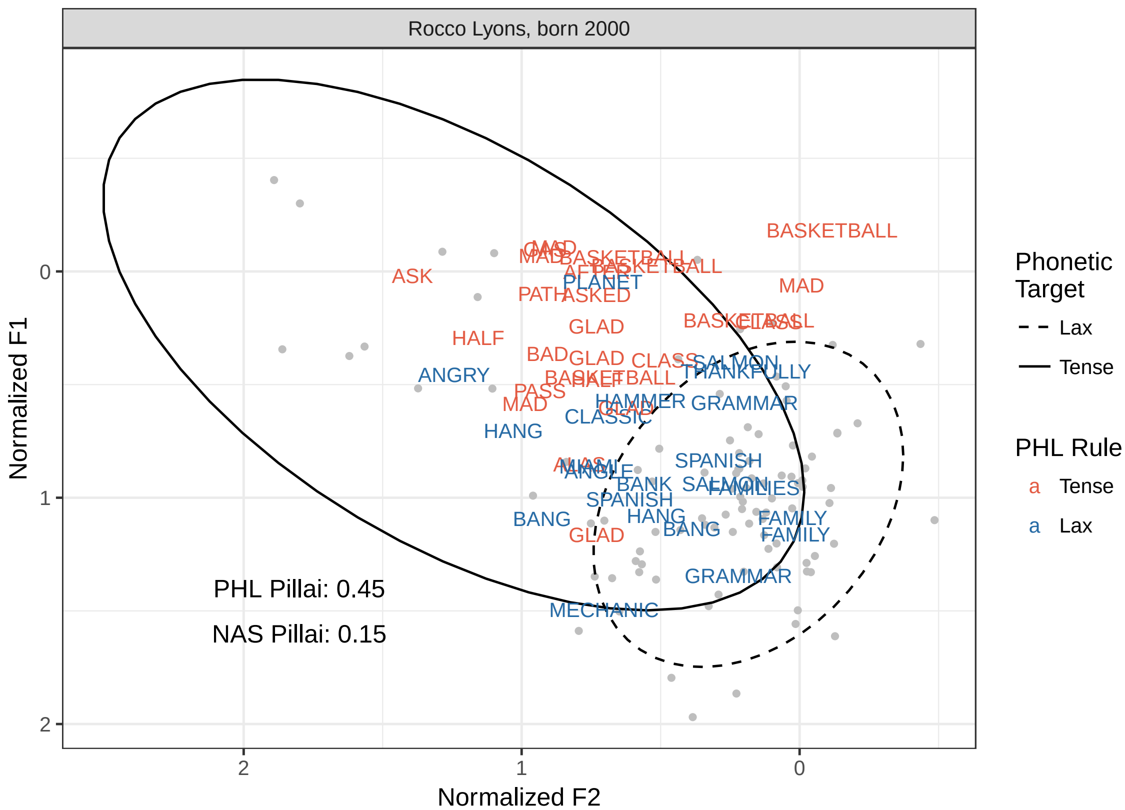


Figure 2.11: Rocco Lyons PHL production.

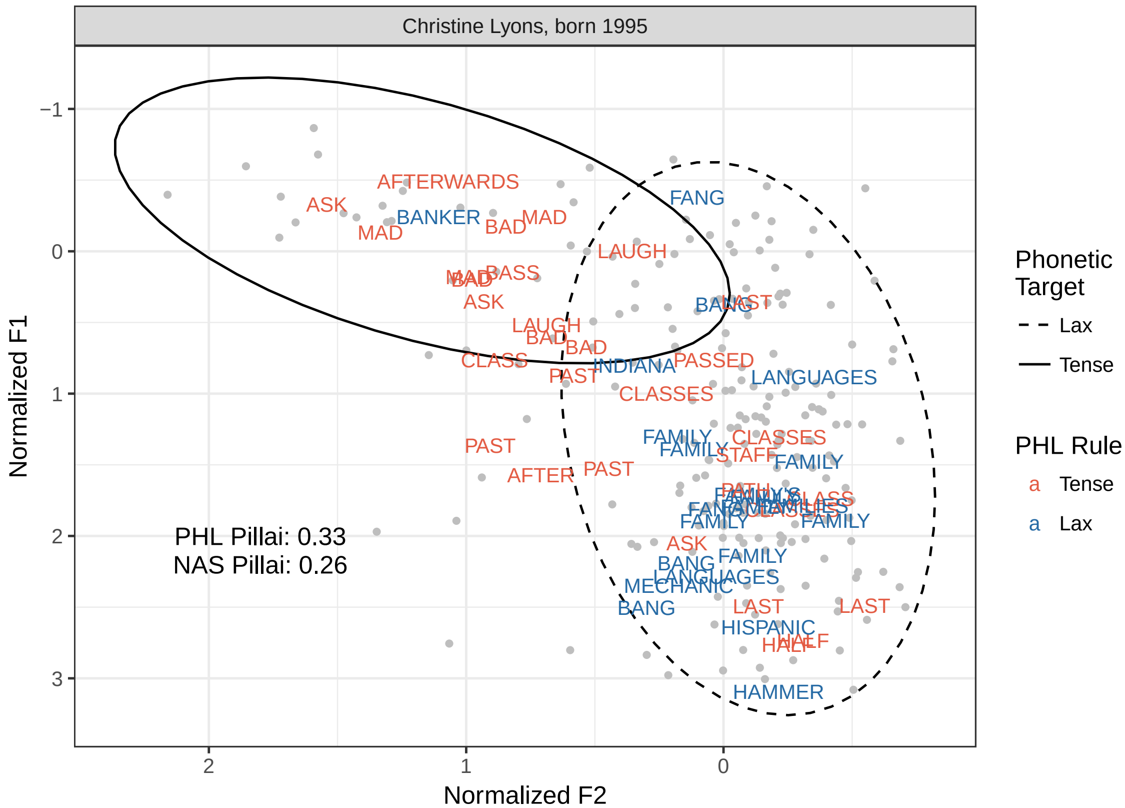


Figure 2.12: Christine Lyons intermediate production between PHL and NAS.

has been found in Prichard (2016) to have an effect on local dialect features. We see this clearly in Christine's production, whereby she produces some lax tokens of traditionally tense PHL words and at least one tense token of a traditionally lax PHL word.

Through close analysis of the Lyons family, we are provided with an in-depth look into how children are adapting the linguistic input of their parents to a intermediate, or potentially mixed-system, production. The Lyons parents provide a classic PHL input to their children. The children in turn, and aligning with their educational history, take that PHL input and either reproduce it (as in Rocco) or take it a step towards NAS (as in Christine). We see clearly the influence of both family and peer education group on the linguistic production of the children, as well as a piece of insight into how PHL becomes NAS intergenerationally.

Vos Family

Just as the Lyons family represents the first step in the transition from PHL to NAS, the Vos family exemplifies the final step the transition to NAS. The Vos family is of Jewish and Persian descent. Data from the mother, a non-native speaker of English, is not presented here, since non-native features are typically disregarded by second-generation children during acquisition (Labov, 2007).

I begin by examining the production of Harry, the Vos family father. Harry's Pillai scores (PHL: 0.29, NAS: 0.2) are immediately reminiscent of Christine's. In Fisher et al. (2015), we similarly classify Harry as a "weak PHL system speaker" based on these overall scores. In Figure 2.13, I take a closer look at how his production of individual tokens has driven this intermediate set of Pillai scores. We can immediately see that, like Christine, Harry produces some of each class of words as both tense and lax. He produces tense forms of MANAGE class words (*Amherst, Miami*) as well as lax forms of these words (*annex, janitor, stammer, planet*). Similarly, he produces instances of HANG as both tense (*hanging, anger, dangle*) and lax (*bank, strangle*). In the word classes that would be produced tense under PHL, we see a similar pattern of variation, with some tense (*after, half, mad*) and some lax (*classes, last, glad, bad*) from each word class. Like Christine's production, Harry's production is suggestive of competing grammars.

That Harry produces an intermediate or mixed-system production is somewhat expected, given his educational history. His parents were also from Philadelphia; while I do not have production data from them to analyze, it is almost certain that Harry would have been given traditional PHL input. Harry attended a prestigious suburban high school outside of Philadelphia. As I have discussed briefly above, suburban schools operate as similarly elite to the Special Admissions non-Catholic schools in Philadelphia. From this, Harry went on to attend a nationally-oriented university (Harvard). This social and educational history aligns with Harry's resultant mixed-system output.

Harry's children, having been given this mixed-system input, take the final step and turn it into a NAS-dominated output. I begin with the production of Nate, who at the time of recording was a 10-year-old Masterman student. As we've seen, Masterman emerges as a stronghold of NAS in Philadelphia. In Figure 2.14 and in Nate's Pillai scores (PHL: 0.06, NAS: 0.73) we see that Nate's

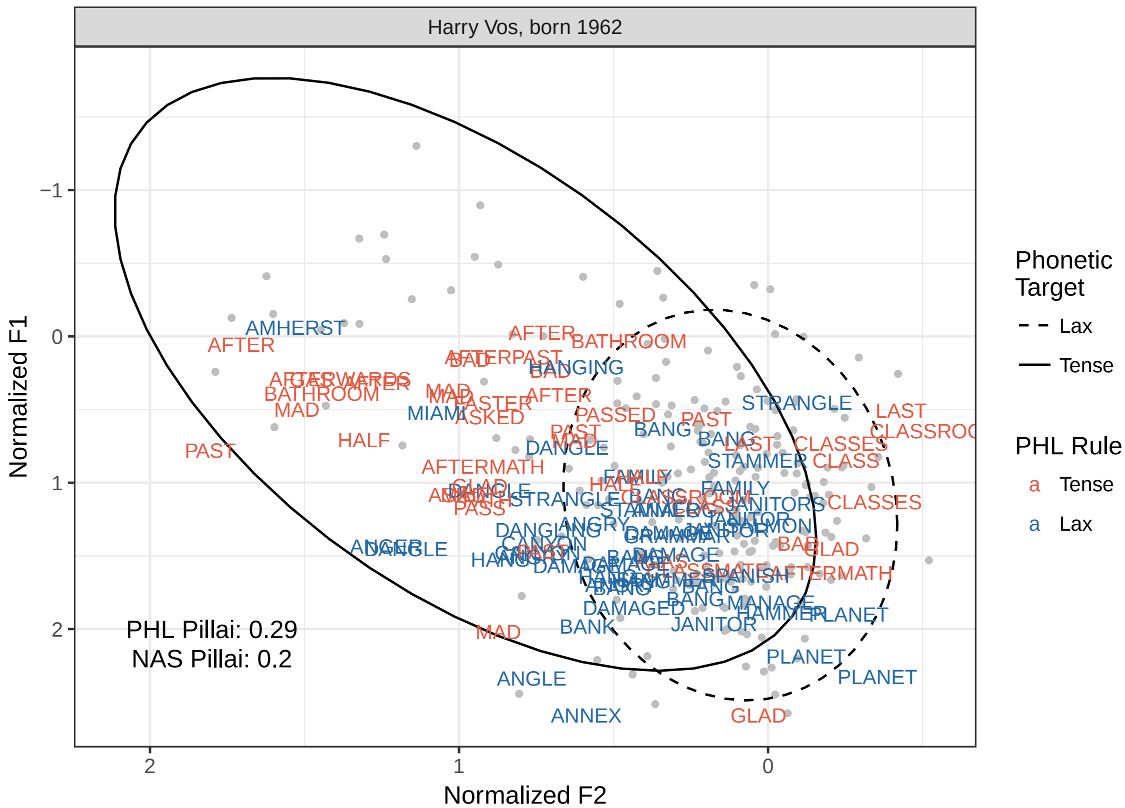


Figure 2.13: Harry Vos intermediate production between PHL and NAS.

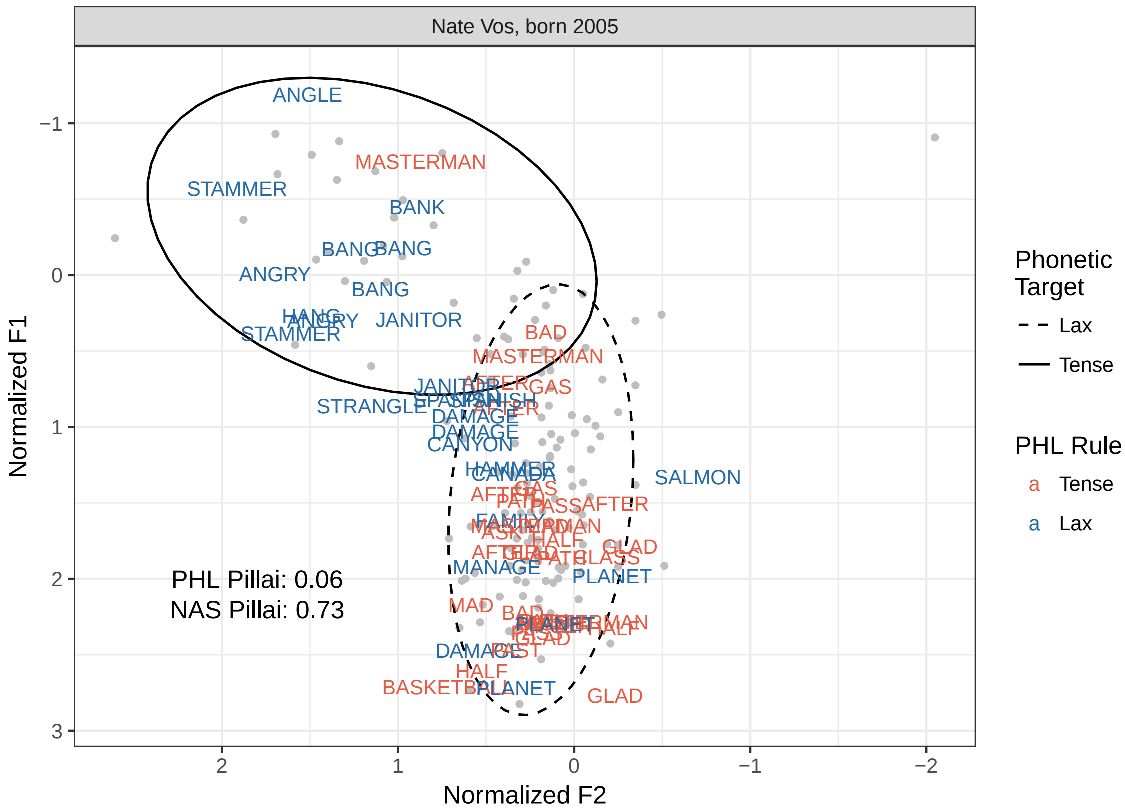


Figure 2.14: Nate Vos NAS production.

production is, overall, dominated by NAS. There are a few exceptional tokens: a tense token of *Masterman*, and a few lax tokens that align with the traditional PHL system (*salmon, planet, manage, family, Canada*). While Nate's attendance at Masterman may predict a stronger NAS system with no lexical exceptions, it is important to note his age at the time of recording. Masterman begins at 5th grade, which 10-year-old Nate had just begun when he was recorded in December of his first semester in Masterman. It is possible, then, that Nate's production represents the beginning of a Masterman influence on his parental input.

His older sister Percia, a 20-year-old undergraduate at the University of Pennsylvania at the time of recording, on the other hand, exhibits the overall expected effect of having attended Masterman through middle and high school as well as the nationally-oriented University of Pennsylvania during college. Her production is shown in Figure 2.15, which clearly exhibits a near-perfect

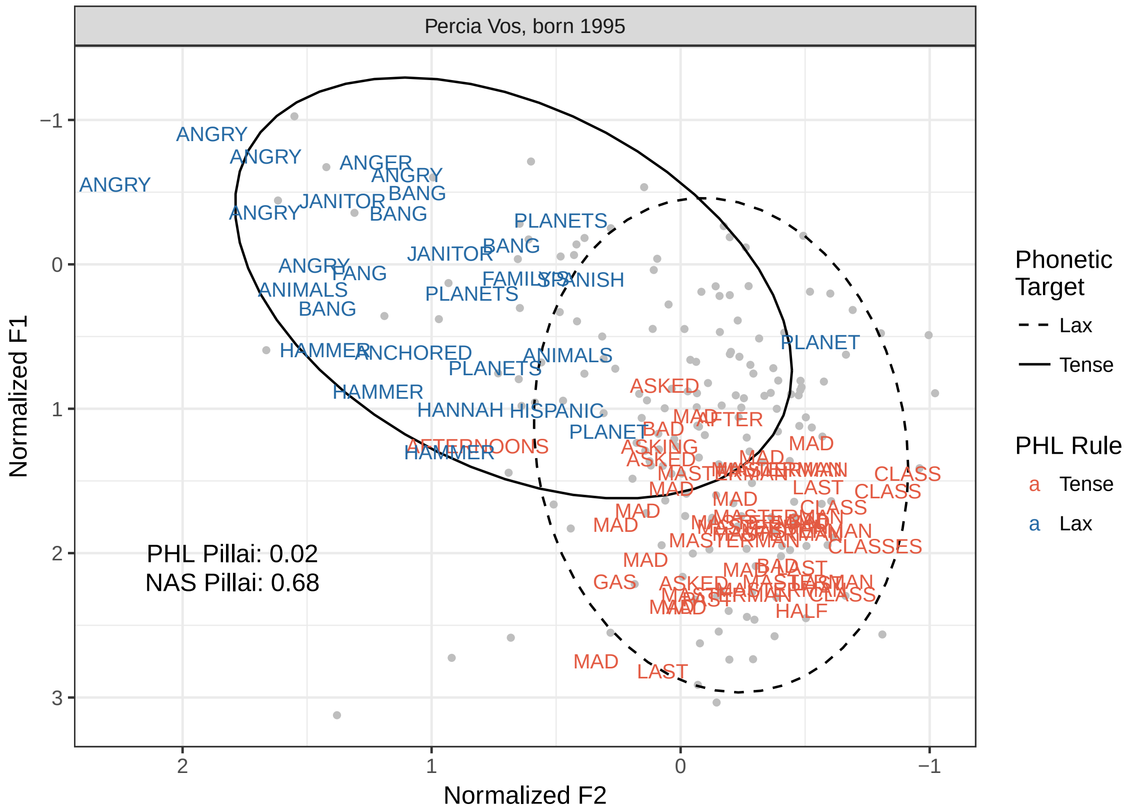


Figure 2.15: Percia Vos NAS production.

conformity to NAS. The only potential exceptions in Percia's production is in a marginally tense form of *afternoons* and a lax form of *planet*. This conformity is also clearly represented in her Pillai scores (PHL: 0.02, NAS: 0.68).

2.4.2 Summary of Intergenerational Change

In both the Lyons and the Vos families, the effect of parental input as well as educational history play an important role in the vowel systems of the children. The data presented here suggest that the transition from PHL to NAS in Philadelphia occurs over a period of three generations, with the first generation (Antonio, Theresa) producing the traditional PHL system as input, the second generation (Christine, Harry) taking that traditional input and, in response to their peer influence, altering it into a mixed-system output, which the third generation (Percia, Nate) take and, in conjunction with their own peer influence, alter this mixed-system input into a fully fledged NAS output. In other words, complete phonological restructuring from PHL to NAS requires the convergence of both parental and peer influence for speakers to take the next step in the change.

Finally, it is also important to note that the Lyons and Vos families can be seen as exemplars from different social subgroupings of Philadelphia. The Lyons send their children to Open Admissions Catholic schools, which we found to be a stronghold for PHL in the community and likewise has a conservative effect on the Lyons' language production. The Vos family, on the other hand, enter into our study having already experienced elite schooling, and continue this trajectory with the children. Here, we see that the fragmentation of Philadelphia along educational system lines has a strong effect not only on the adoption of allophonic restructuring by speakers, but also on the timing of the allophonic restructuring of PHL to NAS. The Vos family, with its educational history of attending elite public schools, exhibits this change a generation ahead of the Lyons family.

2.5 Conclusions

In this chapter, I've gone some depth into the background of /æ/ variation in Philadelphia, providing a detailed look into the community-level pattern of this change as well as the intergenerational pattern of change.

The community-level pattern of this change, as following the fragmentation of the community along educational system lines, provides a detailed sociolinguistic backdrop for the current investigation of phonological change. The sociolinguistic background will emerge in Chapter 4 as a critical component of analyzing the variability within individual speakers. Without an understanding of the community-level pattern, it is impossible to identify the production of an individual as driving phonological change or simply phonetic mitigation as a result of contact with speakers who have already completed the change.

Finally, that this change is found in 2.4 to occur over the course of three generations provides a clear direction to searching for transitional cohort speakers. For younger Philadelphians, particularly those with a Catholic background, transitional cohort speakers are most likely to be those who have attended a combination of Open Admissions Catholic schools and nationally-oriented university. In Chapter 4, this is precisely the demographic of speaker we turn to for an investigation of the mechanism of phonological change in transitional cohort speakers.

Chapter 3

Allophonic Analysis of Traditional Philadelphia /æ/

While Chapter 2 provided empirical support for the traditional Philadelphia /æ/ system as an allophonic rather than phonemic split, here I provide an in-depth theoretical account of the allophonic status of the traditional PHL /æ/ split. I argue that PHL is a productive allophonic rule with a limited set of lexical exceptions. I appeal specifically to the Tolerance Principle (Yang, 2016) to define the upper limit of lexical exceptions; I note, however, that my analysis of PHL as allophonic is compatible with any treatment of productive rules that allow for a precise and limited number of lexical exceptions to that rule.

3.1 Lexical Exceptions in Productive Phonological Processes

Determining whether two sounds in a language hold an allophonic or a phonemic relationship is not always a straightforward task. In generative frameworks (e.g., Chomsky and Halle, 1968; Stampe, 1979; Kiparsky, 1982), defining a phonemic relationship is typically an all-or-nothing undertaking, with segments either considered to be perfectly contrastive or not contrastive at all. Phonologists have traditionally relied on a number of criteria to determine which of these two relationships hold (see, for example Steriade, 2007; Hall, 2013, for an extensive list). The two criteria most commonly appealed to and held up as the most important are *Predictability*, defined as it traditionally has been in 3.1, and *Contrastiveness*, defined in 3.2 (both adapted from Hall 2013).

(3.1) **Predictability:**

Two sounds A and B are considered to be contrastive if, in at least one phonological environment in the language, it is impossible to predict which segment will occur. If in every phonological environment where at least one of these segments can occur, it is possible to predict which of the two segments will occur, then A and B are allophonic.

(3.2) **Contrastiveness:**

Two sounds A and B are contrastive when the substitution of A or B in a given phonological environment causes a change in the lexical identity of the words they appear in. If the use of A as opposed to B causes no change in the identity of the lexical item, A and B are allophonic.

The underlyingly binary approach to phonological classification suggested by the criteria above, in which a phonological relationship is either productive or contrastive and not something in between, holds a great deal of theoretical interest for phonologists who subscribe to a view of phonology in which phonological forms and processes are categorical. There are, however, a number of phonological relationships which are not clearly defined using these criteria or which would even be given contrasting definitions based on these two criteria. The problem of intermediate phonological relationships has been taken up by phonologists for quite some time (e.g., Gleason, 1961; Goldsmith, 1995; Harris, 1990, 1994), with varying degrees of importance given to this problem.

In this chapter, I propose that the primary problem in so-called "intermediate relationships" is not in the resulting classification but rather in the definitions of the criteria used to define phonological relationships. In what follows, I begin by highlighting a synchronic and a diachronic case of lexical specificity in otherwise regular phonological processes. In §3.1.2, I discuss previous solutions to the problem of lexical specificity in regular phonology, and in §3.3 I present my definition of *Predictability* using Yang (2016)'s Tolerance Principle to determine an upper limit to lexical exceptions in productive phonology. In §3.4 I apply this metric to the traditional PHL rule, demonstrating that under all configurations, PHL emerges as sufficiently *Predictable* and therefore as a productive allophonic rule.

3.1.1 Lexical Specificity in Productive Phonological Processes

Here, I outline just a few examples of lexical specificity in otherwise productive phonological processes.

Synchronic Lexical Specificity in the Scottish Vowel Length Rule

The Scottish Vowel Length Rule (SVLR) provides a classic case of lexical specificity (Aitken, 1981). The SVLR is a generally productive phonological process found in Scottish English, whereby vowels are produced as short allophones when they precede voiceless stops, voiceless fricatives, voiced stops, nasals, or /l/. Long allophones of these vowels occur preceding voiced fricatives, /r/, and when in an open syllable. This results in short duration *bead* and *beet* ([bid], [bit]) but long duration *bee* and *beer* ([biː], [biːr]). In addition to this set of conditioning factors triggering the SVLR, the phonological targets of this rule are also somewhat complicated and may vary: the SVLR applies to /i, ʉ/ and /ai/, does not apply to /ɛ, ʌ/ or /ɨ/, and other vowels remain disputed (Scobbie et al., 1999; Ladd, 2005).

In an analysis of the large-scale Glasgow Speech Project, Scobbie and Stuart-Smith (2008) report an additional complication on the SVLR which is most applicable here: lexical specificity in its application, which for some lexical items varies by speaker. Table 3.1 reproduces their findings of young female speakers' production of /ai/ in a word list for words that typically would be produced short under the SVLR. In Table 3.1, each row represents a single speaker, with the top four speakers from a middle-class suburb of Glasgow (Bearsden) and the bottom four from a largely working class area of the city (Maryhill). Cells with a 's' follow the expected pattern, while empty cells represent lexical exceptions to the SVLR. Cells with n/a represent a lack of data due to subject error in reading the word.

Here we see lexical specification within individual speakers, so that Bearsden Speaker 3 has the

	bible	sidle	libel	micro	nitro	hydro	title	tidal	pylon	crisis	miser
Bearsden											
1	s	s	n/a		s		s	s		s	
3	s	s		s	s		s	s			
4	s	s		s			s	s		s	
5	s			s	n/a		s			s	
Maryhill											
1		s	n/a				s	s		s	
2		s	n/a				s	n/a		s	n/a
3	s	n/a					s	s			
4	s	s		s	s		s	s		s	

Table 3.1: Lexical specificity in SVLR for young female subjects. Adapted from Scobbie and Stuart-Smith (2008). Cells with ‘s’ were produced as short (expected pattern), cells with ‘n/a’ were not produced or were errors, and blank cells were produced as long.

following lexical exceptions to the SVLR: *libel, hydro, tidal, miser.* For this individual, who in large part follows the SVLR, there remain some lexical exceptions. Under the strict binary approach to classification presented at the beginning of the chapter, this data raises a problem. Does Speaker 3 now have a phonemic contrast in what is otherwise a productive allophonic process simply because four words are lexically specific? Complicating the picture are speakers like Bearsden Speaker 5, who in addition to six lexical exceptions also produces a marginal minimal pair between *title* [taitl̩] and *tidal* [tai:dl̩]. Under the classic definitions of phonemic classification, Speaker 5’s SVLR is a phonemic relationship in length while Speaker 3’s SVLR is unclear.

Additionally, while there is interspeaker variation in the lexical specificity of the SVLR presented in Table 3.1, a more general community trend also emerges. Across the community as a whole, *bible, sidle, title, tidal* and *crisis* are generally produced as expected, while *libel, nitro, hydro, pylon* and *miser* are exceptionally long. Here we see interspeaker variation aligning overall to produce a larger community-level pattern that may in turn be learned, perhaps with varying degrees of faithfulness in which lexical items are exceptional, by the next cohort of speakers.

The problem of classification for all speakers lies in the overwhelmingly productive nature of the SVLR: while there are a few lexical exceptions for speakers, the pattern is overwhelmingly followed. In following a tradition of analyzing morphological conditioning as a marginal contrast, Scobbie and Stuart-Smith (2008) analyze the SVLR as a *Quasi-Phoneme* with what they term *Fuzzy Contrast* which is morphologically predictable save for a few lexical exceptions. Analyzing the SVLR as a stem-level application, which I do here, accounts straightforwardly for the apparent morphological conditioning; what we are left with is a productive stem-level rule with some lexical specificity.

Diachronic Lexical Specificity in Philadelphia /ay/-Raising

The problem of lexical specificity in phonological processes has also been taken up by scholars of language change, most notably debated by historical linguists in the late 1970s and early 1980s. This debate, dubbed the "Neogrammarian Controversy", debated the relative roles of lexical diffusion and regular sound change in language change. The traditional Neogrammarian position holds that sound change is phonetically gradual and lexically abrupt, affecting all segments in the language that share the same phonological target equally. Lexical diffusionists (e.g., Wang, 1969; Chen and Wang, 1975) hold that sound change is phonetically abrupt but lexically gradual, with segments in only particular words at a time abruptly changing in phonetic output until all words in the language with that segment have changed. Labov (1981) attempts to resolve the Neogrammarian Controversy by proposing two distinct types of changes: Neogrammarian changes, which are lexically abrupt and phonetically gradual, and Lexical Diffusion changes which are lexically gradual but phonetically abrupt. Labov (1981) further proposes that these changes have typical target profiles: that Neogrammarian changes will affect phonological features like raising and fronting (features associated with what I consider to be surface phonological representations), while Lexical Diffusion changes affect the underlying phonological representation, causing a "redistribution of some abstract class into other classes." This predicts that certain changes, like phonemic mergers or secondary allophonic splits, may proceed with lexical diffusion, while other regular sound changes, like /u/-fronting, do not.

While Labov (1981)'s solution carries theoretical appeal, providing both an explanation of seemingly disparate facts and predictions for future sound changes, such a discrete separation between two types of sound changes is not borne out in empirical data on sound change. Take, for instance, Fruehwald (2013)'s analysis of /ay/-raising in Philadelphia, where the nucleus of the PRICE diphthong undergoes raising when it precedes a voiceless segment but remains low elsewhere. On the surface, this appears to be a classic example of Neogrammarian change, with a regular phonological conditioning rule of /ay/ raising before all phonologically voiceless segments and remaining low before all phonologically voiced segments, as shown in Figure 3.1. In the middle of this quite regular sound change, Fruehwald (2013) outlines several lexical items which abruptly change categories from low [aɪ] to raised [ʌɪ]: *Snyder* (a street name in Philadelphia), *cider*, and *spider*. In Figure 3.2, the height of these tokens are plotted against the background of /ay/ raising overall. Each point represents the mean of a single speaker's production of these words, with the size of the point representing the number of tokens per speaker. The baseline community production for /ay/-raising before voiceless segments is plotted in blue, and non-raised tokens before voiced segments is plotted in red.

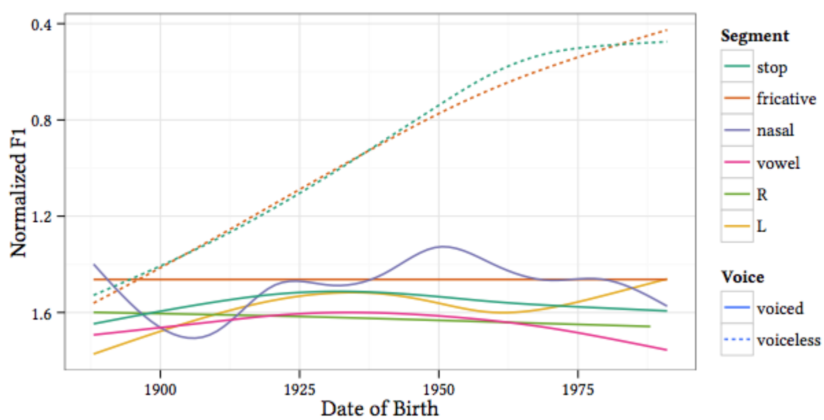


Figure 3.1: PRICE raising by phonological context. From Fruehwald (2013).

Figure 3.2 displays a clear jump in the production of these three words, with most tokens produced low (as predicted by phonological context) near the beginning of the corpus but produced with a raised nucleus near the end of the corpus. The emergence of lexical specificity in the middle of an otherwise regular sound change raises a challenge to the hypotheses laid out in Labov (1981).

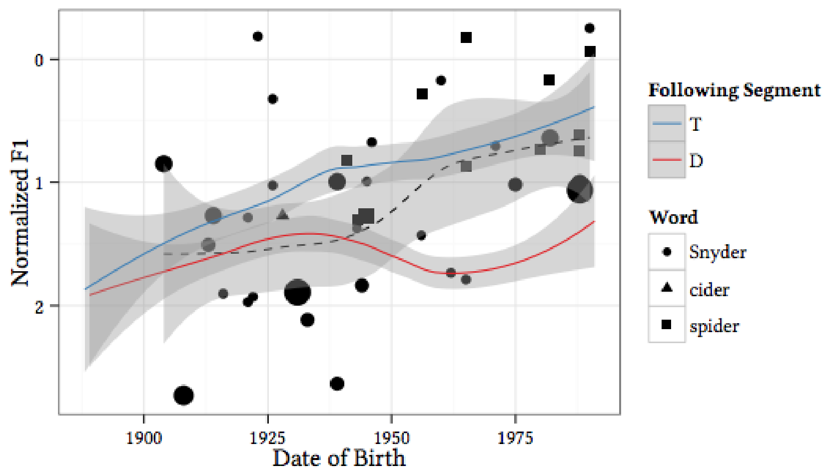


Figure 3.2: Lexical Exceptions in PRICE raising. From Fruehwald (2013).

Here, we see an instance of lexical specificity in the *allophonic* representation of words rather than in the underlying representation. Under Labov (1981), lexical specificity in sound change is hypothesized to occur at the level of underlying specification. This can be potentially resolved by positing that *Snyder, spider* and *cider* did in fact undergo lexical diffusion in their underlying representation, with speakers born after 1940 having re-analyzed the neutralized /ɾ/ as an underlying voiceless /t/. Under this analysis, the lexical specification in the diachronic raising of PRICE is simply an instance of lexical diffusion occurring concurrently with a regular sound change, not an instance of lexical specificity in the allophonic raising rule. However, this solution does not hold for all speakers. Fruehwald also found examples of speakers raising in voiced contexts that do *not* exhibit neutralization between a voiced and a voiceless underlying segment in the output: *tiger* and *cyber*. For these speakers, this lexical specificity cannot be driven by a re-analysis of the underlying form and must instead be accounted for as lexical exceptions to the otherwise regular raising rule.

3.1.2 Solving the Problem of Lexical Specificity

Given that lexical specificity is a well-documented problem in phonology, it is perhaps unsurprising that a number of solutions to these intermediate-type relationships exist. Broadly speaking, these solutions have fallen into one of two main camps. The first camp posits an additional intermediate layer to the phonological architecture to handle these ill-behaved phonological relationships, the idea being that an intermediate relationships is a phonological reality existing between allophonic and phonemic which may be diachronically a step along the way to phonemicization (Kiparsky, 2015). A number of solutions have been brought forward in this vein, with nearly an equal number of distinct labels given to intermediate relationships (e.g. *semi-phonemic, hemi-phoneme, quasi-phoneme, weak contrast, mushy phonemes, marginal contrast* – see Hall 2013, for a robust review). This approach allows for the existence of relationships which would be classified as intermediate under the criteria listed above. There are however, two main critiques to be given to this approach, which fall under an empirical and a theoretical frame. From an empirical perspective, the predictions made by an intermediate phonological category differ from proposal to proposal and often do not make any distinct predictions at all between how an allophonic relationship compared to an intermediate relationship should behave in production (though, see Kiparsky, 2015, for a discussion of quasi-phonemes as a distinct stage in diachronic phonologization).

The second camp takes a gradient view of phonology, arguing that amongst these intermediate relationships, there are some that are more allophonic and some that are more phonemic. This is the view offered in Boulenger et al. (2011), which proposes a Gradient Phonemicity Hypothesis on the basis of gradient responses in an ERP experiment, and in Hall (2013), which redefines the Predictability criterion as a gradient measure of predictability based on the entropy score of a phonological rule. In other words, under both Boulenger et al. (2011) and Hall (2013), the more predictable a pair of sounds is, the more allophonic that pair is and the less predictable a pair of sounds is, the more phonemic that pair is. While this approach provides an overall solution to the problem of intermediate phonological relationships, it introduces fundamental problems to a categorical view of phonology. It predicts, for example, that given two intermediate relationships with different entropy scores, the higher more predictable one will behave more like a productive rule. It is not immediately clear how we may expect this to be borne out in empirical data: perhaps a more predictable intermediate relationship will exhibit more regularity in e.g. nonce word production than a less predictable relationship will. From a theoretical perspective, it is also difficult to incorporate a gradient distinction between allophones and phonemes into a view of phonology that relies on categorical segments and categorical processes, as does any view of phonology consistent with the modular feed-forward approach adopted in this dissertation.

Here I submit a third solution to phonological classification, which is to redefine the definition of *Predictability*. This solution will allow phonological relationships to remain categorical, by enabling alternations previously classified as intermediate to be strictly defined as either allophonic or phonemic. In general terms, my point is simple: that productive allophonic rules may allow a limited number of lexical exceptions. In this dissertation, I specifically invoke the Tolerance Principle (Yang, 2016) to define an upper limit to the number of lexical exceptions a productive phonological rule may allow. This principle was derived independently from phonology, as a model of language acquisition. For a detailed description of the derivation of the Tolerance Principle and numerous examples of it working particularly well to explain lexical exceptions in morphology and phonology cross-linguistically, I refer the reader to Yang (2016). In §3.4 I provide a full account of PHL and its lexical exceptions, demonstrating that it falls well below the threshold for excessive exceptions and therefore is a plausible productive rule.

3.2 Philadelphia /æ/

The phonological conditioning of the traditional PHL split is repeated in (9) below. In (9), the Philadelphia /æ/ split is represented as a rule triggered by a disjunctive set of phonological conditions: nasals or voiceless fricatives which are also interior and syllable final. This produces tense *man*, where /æ/ is followed by a syllable final anterior nasal /n/, but lax *manner*, where the following /n/ is syllabified as the onset of the following syllable.

(9) **PHL:** $\text{æ} \rightarrow \text{æh} \,/\, \_\left[\ \text{+anterior}\ \right] \cap (\left[\ \text{+nasal}\ \right] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix})]\ \sigma$

I note briefly that disjunction in the featural representation of segments that trigger a productive phonological rule is necessary for a number of cross-linguistic phonological processes (see Mielke, 2008, for an extensive review); as such, the disjunction in PHL is does not in itself present a challenge to PHL as an allophonic rule. We can conceive of PHL as an example of emergent features, where the segments triggering tensing in PHL become classified as a set of similar features by the speakers of the language, which can be represented as in (10). Here, I employ the stratal aspect of the modular feed-forward approach (Bermúdez-Otero, 2007), in which phonological rules may apply at the stem level, word level, or phrase level. I analyze PHL as a productive rule that applies at the stem level of a word, so that an /æ/ followed by an open syllable in the stem (e.g., *manage*) is produced as lax but an open syllable created by an inflectional morpheme (e.g., *man+ning the ship*) is not relevant to the rule, as it has already been applied at the stem level and is also applied at the word or phrase level.

(10) **PHL:** æ → æh / _{m, n, f, θ, s}] σ

The general PHL rule shown in 9 accounts for much of the Philadelphia /æ/ data. However, there are a number of lexical exceptions to this rule which results in a lack of perfect predictability based on phonological context. For example, while most words with /æ/ followed by a tautosyllabic /d/ (such as *dad* and *fad*) follow the rule and are produced as lax, there are three lexical exceptions which are produced as tense: *mad, bad* and *glad.* There are far more lexical exceptions produced as lax, in which words with an /æ/ followed by a tautosyllabic anterior nasal or voiceless fricative are produced as lax (such as *asterisk, ran, than, carafe*). The total number of lexical exceptions to the general rule is extensive, and includes some words whose status as a lexical exception is dependent on the individual speaking. For example, *planet* follows the traditional rule and is produced as lax by many speakers in Philadelphia, but produced as a lexical exception to tense by a number of speakers born in the 1990s (Brody, 2011). The exact number of lexical exceptions required by a Philadelphia English speaker is the focus of §3.4.

This lack of predictability has made the classification of /æ/ in Philadelphia English historically controversial. Since its first treatment in descriptive dialectology literature by Ferguson (1972), PHL has been sometimes analyzed as phonemic (Ferguson, 1972; Labov, 1989; Dinkin, 2013) and sometimes analyzed as allophonic (Kiparsky, 1995; Labov et al., 2016; Sneller, 2018), with each of these works also acknowledging the controversial nature of the classification of PHL. Proponents of a phonemic analysis have almost categorically appealed to the lack of perfect predictability in the distribution of the two sounds, and to the possible existence of one minimal pair (auxiliary *can* produced as lax and noun *can* produced as tense) as evidence for the phonemic analysis of PHL. Proponents of an allophonic analysis have pointed to the *mostly* predictable distribution of the tense and lax versions and more recently to the community-level competition between PHL and NAS (Labov et al., 2016) as evidence for an allophonic analysis.

In what follows, I demonstrate that applying the Tolerance Principle as a diagnostic of productive phonological processes results in an analysis of PHL as a plausible productive allophonic rule with a number of lexical exceptions.

3.3 Tolerance Principle approach to productive rules

As a model of language acquisition, Yang (2016) outlines a principle that determines the productivity of a rule given a set of input. This principle is shown in (11).

(11) **Tolerance Principle:**

Let R be a rule that is applicable to N items, of which e are exceptions. R is productive iff:

$$e \leq \theta_N \text{ where } \theta_N := \frac{N}{lnN}$$

The Tolerance Principle states that a rule is productive if the number of exceptions to that rule is less than the number of items the rule could potentially apply to divided by the natural log of that number of items. For example, let's assume that a child has 10 verbs in her vocabulary. Some of these verbs take the regular -(e)d suffix to form a past tense (e.g., *walk, smile*), while some of these verbs are exceptions to this regular rule (e.g., *run, fall*). The Tolerance Principle states that the regular past tense -(e)d rule can be productive for this child if her vocabulary has fewer than $10/ln(10)$, or 4.3, exceptions to this rule. In other words, if the child's vocabulary contains 4 or fewer irregular past tense verbs, then the regular past tense -(e)d rule can be a productive rule in her language.

It is important to stress that the Tolerance Principle applies over word *types* rather than tokens. This means that despite evidence that word frequency is an important factor in language processing (Goldinger, 1998; Grainger, 1990; Seguie et al., 1982), it does not play a role in the calculation of the productivity of a rule (modulo the fact that high-frequency words are more likely to be acquired by children and thus more likely to be involved in the calculation of N and e). This predicts that a child would be able to learn a productive rule as long as the word types in her vocabulary fit the Tolerance Principle, regardless of the token frequencies of these words.

Here I highlight a few key features of the Tolerance Principle that are especially relevant for the present dissertation. First, the threshold for exceptions is surprisingly high. Table 3.2 gives a range of values of N and the maximum number of exceptions that a rule defined over N items can tolerate, along with the percentage of total N tolerated as lexical exceptions.

N	e	% exceptions tolerated
10	4	40
20	7	35
50	13	26
100	23	23
200	38	19
500	80	16
1,000	145	14.5

Table 3.2: Number and percent of total lexicon tolerated as exceptions (e) by lexicons of N size.

As shown in Table 3.2, as N increases, the tolerable proportion of exceptions (e) decreases. This suggests that productive rules are relatively *easier* to learn when the learner has a *smaller* vocabulary, a conclusion that may have significant implications for the difference between child and adult language acquisition. Second, the Tolerance Principle has proved effective in accounting for a wide range of problems in language acquisition ranging from phonology and morphology to syntax (see Yang, 2016, , which provides a discussion of over 100 successful applications of the Tolerance Principle). An artificial language learning study (Schuler et al., 2016) found near-categorical support for the Tolerance Principle. In this study, children between the ages of 5 and 6 learned an artificial language comprised of 9 total nouns. According to the Tolerance Principle, such a language can support up to 4 exceptions ($\theta_9 = 4.1$); Schuler et al. (2016) found that children learned a generalized suffix rule when there were only 4 exceptions but failed to learn the rule when the number of exceptions exceeded the tolerance threshold.

In what follows, I will simply assume the correctness of the Tolerance Principle as a diagnostic of productivity and use it to evaluate the viability of PHL as a productive allophonic rule in the face of exceptions.

3.4 Calculating the tolerance threshold for /ae/ in Philadelphia

To analyze the Philadelphia /æ/ split using the Tolerance Principle, we must first determine the value of N. That is, we must determine the total number of lexical items containing /æ/, which will be the total number of lemmas a tensing rule could apply to. In what follows, I outline a number of choices that must be made with regards to calculating N, and provide an analysis of PHL based on both a conservative approach to each of these choices (i.e., bringing PHL closer to not passing the tolerance threshold) as well as what I believe to be a more accurate description of PHL.

Procedurally, once N has been determined, lexical exceptions are then calculated as those words that violate the productive rule. An example is provided Table 3.3, which presents the expected realization (PHL Expectation) and the actual ralization (Traditional Input) for seven lexical items containing /æ/. In the final column, each lexical item is evaluated for whether the actual realization is an exception to the PHL rule or not. Here, we can see that *mad* must be treated as a lexical exception to the regular PHL rule, as its traditional realization does not match the expected output of the regular rule. Once the total number of lexical exceptions (e) has been determined, we can then calculate whether $e \leq$ the tolerance threshold of θ_N. If the lexical items in Table 3.3 were the entirety of a child's /æ/ words, N would be 7, θ_N would be $\frac{7}{ln(7)}$, or 3.59, and e would be 3. Since 3 < 3.59, PHL would emerge as a productive rule in this dummy language.

Here, I calculate the values of N, θ_N, and e under different assumptions about PHL. In all cases, I obtain the total number of lexical items containing /æ/ from the *CHILDES* database (MacWhinney, 2000) to obtain a measure of the total N for a child's vocabulary. This database includes both child and caretaker production data, which gives an approximation of the linguistic input given to a child.

Word	Traditional Input	PHL Expectation	Exception?
hand	tense	tense	no
mad	tense	lax	yes
cat	lax	lax	no
ran	lax	tense	yes
hammer	lax	lax	no
laugh	tense	tense	no
swam	lax	tense	yes

Table 3.3: Input realizations of /æ/ compared to expected realization under PHL.

3.4.1 Productive Morphology

The first major decision that must be made is the role of productive suffixes. Because the Tolerance Principle applies to word types and not tokens, the crucial calculation is over lemmas. This is particularly relevant to calculating lexical exceptions to PHL: while *laugh* [leːᵊf] straightforwardly follows the productive rule, some suffixes (such as -ing) result in resyllabification of the following /f/, producing a surface-level exception to the productive rule: *laughing* [leːᵊ.fɪŋ] is produced with a tense /æ/ despite the /f/ being intervocalic.

Counting pairs like *laugh* and *laughing* as two distinct lemmas has a large impact on the calculations of both the total *N* as well as the total number of exceptions. Because there is robust evidence that children acquire productive suffices for plural, comparative, present and past tense, adjectival -*y* and diminutive fairly early (Brown, 1973), I posit that words with these suffixes are classified as their stem-level lemma. The productive use of suffixes such as -*ify* and those that involve learned vocabulary items generally are not acquired until school age (Jarmulowicz, 2002; Tyler and Nagy, 1989). In other words, I consider *class* and *classes* to belong to a single lemma *class* which is produced with a tense /æ/ following the tensing rule, but *classify* to be a distinct lemma produced with a lax /æ/ following the tensing rule. I note that this formulation of phonology as allowing children to categorize inflected forms under a single lemma fits well with the stratal view of phonology that I adopt throughout this dissertation, in which phonological processes may apply at the stem, word, or phrase level. Under a stratal view of phonology, the discussion above can simply be read as a statement that the PHL rule applies only at stem level.

3.4.2 Status of /ae/ before /l/

A second decision must also be made regarding the status of /æ/ preceding /l/. In the oldest speakers recorded in the Philadelphia Neighborhood Corpus (PNC), we see a noncontroversial production of lax /æl/. However, as noted by Dinkin (2013) and Labov et al. (2013), the production of /æl/ has been increasingly phonetically tensed beginning with speakers born around 1945, in what appears to be a gradual phonetic process rather than a phonological one. In other words, some speakers produce /æl/ in an intermediate phonetic production between their tense and their lax targets, rather than the expected result of lexical diffusion in which some /æl/ tokens would be produced in line with a speaker's tense target. This suggests that /l/ has not simply been added to the PHL rule as an additional tensing environment, since speakers are not producing /æl/ in line with their own tense target. Additional evidence that /l/ has not been added to the set of triggering environments lies in the fact that *all* /æl/ tokens display phonetic raising, not just the tautosyllabic ones. In other words, both *pal* and *palace* display this gradual raising, where only *pal* would be expected to raise if /l/ were part of the PHL rule.

Dinkin (2013) notes further that this raising of /æl/ coincides with the phonetically gradual fronting and raising of /aw/ (as in *owl*) in Philadelphia, and results in a number of misunderstandings between the /awl/ and /æl/ classes: *owl* with *Al*, *vowel* with *Val*, *Powell* with *pal*. Dinkin (2013) argues that the phonetically gradual behavior of raising /æl/, its phonetic realization tracking the realization of awl as it first peripheralizes then retreats in phonetic space, and the large number of misunderstandings between /æl/ and /awl/ suggests that /æl/ has undergone a phonological reanalysis in these younger speakers, in which words traditionally transcribed with /æl/ have been merged phonologically with awl.

The phonological status of /æl/ is important for calculating both *N* and *e*; if /æl/ is phonologically /awl/, then all /æl/ forms should be excluded from both calculations. If /æl/ is still underlyingly part of the /æ/ class, then all /æl/ tokens should be counted as part of *N* as well as part of *e*.

3.4.3 Patterns in the lexical exceptions

Finally, it should be noted that in most treatments of /æ/ in Philadelphia, the lexical exceptions have been noted to follow certain patterns. Setting aside "patterns" that follow straightforwardly from the discussion about productive morphology in §3.4.1 (which serve as the primary evidence in Ferguson 1972 for PHL as a phonemic distinction), the remaining patterns have been described, following Labov (1989), as:

1. Truncations of /æ/ words in originally open-syllable position retain lax /æ/ regardless of surface syllable structure: *math* [mæθ] from *mathematics*, *exam* [ɛɡzæm] from *examination* [3].
2. Function words that can be reduced to schwa are lax: *and, am, than,* auxiliary *can.*
3. Ablaut past tense forms are lax: *ran, swam, began*, the archaic but marginally productive *wan* (past tense of *win*).
4. Rare and late-learned words are lax: *asp, daft, gaffe, carafe* [4].
5. Polysyllabic words with zero onset before voiceless fricatives are lax: *aspirin, Africa* [5].
6. Affective adjectives *mad, bad, glad* are tense [6].

While these patterns can be identified by linguists (though not without their own exceptions, as highlighted in the footnotes), my account here takes on the perspective of the language learner by simply listing all exceptions in a nonhierarchical list. I do this for several reasons. First, this is the more conservative approach: The Tolerance Principle clearly allows for recursive rules, and analyzing these lexical exceptions as the product of additional rules decreases the number of actual lexical exceptions that must be listed. Analyzing them instead as a flat list as I do here makes an allophonic result less likely. Second, this approach takes child language into account: while there is robust evidence that children learn productive derivational suffixes (-ed, -er, -ly, -ing) early on (Brown, 1973), inflectional suffixes (-ify, -ic) and classifications like "Class 3 Strong Verb" are learned quite late, if at all. So for a young child acquiring Philadelphia English, learning

[3] Though note *gas* from *gasoline* does not follow this pattern: [ɡeːᵊs]. Additionally, individual speakers vary with regards to this pattern, with some speakers producing tense *math* [meːᵊθ] and *exam* [ɛɡzeːᵊm]

[4] Though note the "late-learned" effect varies by speaker, with some speakers realizing some of these words as tense

[5] Though note *athlete, afternoon* are tense

[6] Though note the affective adjective *sad* is lax

a phonological pattern based on a classification that that child has not yet acquired is somewhat nonsensical. Finally, it is both unnecessary and inaccurate to consider these patterns as rule: an exception can be found for nearly every pattern described above, and listing recursive exceptional rules as part of the phonological rule unnecessarily complicates the productive phonology.

Instead, I posit that all the lexical exceptions to PHL are simply listed in two lists: L_{tense}\{*mad, bad, glad*\} and L_{lax}\{*math, exam, ran, and*, …\}. This way of listing lexical exceptions means there is no problem in listing some truncations as exceptionally lax (*math, exam*) while leaving other lemmas that were historical truncations to follow the rule (tense *gas*). Additionally, using two lists of lexical exceptions (one for exceptionally tense lemmas and one for exceptionally lax lemmas) easily allows for diachronic additions and subtractions from these lists without expecting changes to affect other words. For example, *planet* is free to join the lexically tense list for the children reported in Brody (2011), then leave it again for speakers reported in Sneller (2018) without any complication to the phonological architecture.

3.4.4 PHL is Productive under All Calculations of N and e

Table 3.4 presents the calculations of N, θ_N and e for all configurations of PHL. As was discussed in §3.3, the tolerance threshold is proportionally higher for smaller vocabularies. This raises the possibility that PHL would be emerge as a productive rule for very young children whose vocabularies are small and therefore more proportionally tolerant of lexical exceptions, but not be productive for older speakers with larger vocabularies. To test this, I calculated N, θ_N and e for different vocabulary sizes. Here, I use the frequency of words defined by the number of instances that word appeared in CHILDES (1, 20, 50, or 100 times) as an approximation of learners' vocabulary at progressive stages of language development. As the frequency value goes up, the total vocabulary goes down; this can be seen in the N values for each row. In each cell, N, θ_N and e are reported, and any cell in which $e \leq \theta_N$ successfully passes the tolerance threshold and is a plausible productive rule.

Table 3.4 reports the results for whether N is calculated with PHL as a stem-level rule (allowing *laugh* and *laughing* to be considered a single lemma) or a surface-level rule under three evaluations

Freq	Surface Rule	Stem Rule	Tense /æl/ Surface Rule	Tense /æl/ Stem Rule	/æl/ as /awl/ Surface Rule	/æl/ as /awl/ Stem Rule
1	$N = 2161$ $\theta_N = 281.4$ $e = 68$	$N = 1412$ $\theta_N = 194.7$ $e = 39$	$N = 2161$ $\theta_N = 281.4$ $e = 165$	$N = 1412$ $\theta_N = 194.7$ $e = 116$	$N = 2064$ $\theta_N = 270.4$ $e = 68$	$N = 1335$ $\theta_N = 185.5$ $e = 39$
20	$N = 660$ $\theta_N = 101.7$ $e = 23$	$N = 487$ $\theta_N = 78.7$ $e = 19$	$N = 660$ $\theta_N = 101.7$ $e = 49$	$N = 487$ $\theta_N = 78.7$ $e = 42$	$N = 634$ $\theta_N = 98.3$ $e = 23$	$N = 464$ $\theta_N = 75.6$ $e = 19$
50	$N = 413$ $\theta_N = 68.6$ $e = 17$	$N = 330$ $\theta_N = 56.9$ $e = 15$	$N = 413$ $\theta_N = 68.6$ $e = 31$	$N = 330$ $\theta_N = 56.9$ $e = 28$	$N = 399$ $\theta_N = 66.6$ $e = 17$	$N = 317$ $\theta_N = 55$ $e = 15$
100	$N = 282$ $\theta_N = 49.9$ $e = 12$	$N = 239$ $\theta_N = 43.6$ $e = 11$	$N = 282$ $\theta_N = 49.9$ $e = 21$	$N = 239$ $\theta_N = 43.6$ $e = 20$	$N = 273$ $\theta_N = 48.7$ $e = 12$	$N = 232$ $\theta_N = 42.6$ $e = 11$

Table 3.4: PHL is productive under all configurations of productive morphology and /æl/ analysis.

of /æl/. The first two columns calculates values based on /æl/ as a lax production of /æ/. The second two columns calculates /æl/ as a tense production of /æ/, and the final two columns calculate values based on /æl/ as no longer belonging to the /æ/ class but rather merged with /awl/. As shown in Table 3.4, there is no configuration of exceptions under which e exceeds θ_N for PHL. In other words, regardless of whether PHL is a stem-level rule or a surface-level rule, and regardless of whether tense forms of /æl/ constitute lexical exceptions for /æ/ or have undergone a secondary split and merged with /awl/, PHL emerges as a plausible productive allophonic rule. A full list of lexical exceptions is provided in Appendix A.

3.4.5 Marginal Contrast in *I can* and *tin can*

While the Tolerance Principle clearly identifies PHL as a plausible productive rule, there is one final sticking point regularly held up as evidence of a phonemic contrast: the marginal contrast between lax auxiliary *can* and tense content *can*.

I have little to say about this contrast, other than to say that whatever mechanism accounts for homophony can be used to account for this contrast. Because the formulation of lexical exceptions relies on a speaker knowing the lexical identity of a word, there is nothing surprising about a speaker being able to distinguish between auxiliary *can* and content *can* underlyingly. In this case, auxiliary *can* is added to the list of exceptions produced as lax, while content *can* remains a regular, unspecified lexical item that is fed straightforwardly through the tensing rule.

3.5 Formulation of PHL as an allophonic rule

As shown in §3.4.4, PHL emerges as a plausible productive rule with some lexical specificity for any configuration of productive morphology and /æl/, using the Tolerance Principle as a measure of productivity. This raises the inevitable question of how to formulate an allophonic rule that has lexical specificity, as well as specifically how I analyze PHL according to the options discussed above.

First, to the problem of representing lexical specificity. Adopting the Tolerance Principle to phonology provides a framework for representing lexical specificity in a productive rule. This principle is formulated as an evaluation metric that "quantifies real time language processing" (Yang, 2016, pg. 40), specifically drawing on the Pāṅinian Elsewhere Condition. To optimize the time-efficiency of representation, the Tolerance Principle argues that speakers list lexical exceptions (w) in order of lexical frequency (w_1 through w_e). When going to process or produce a word containing /æ/, speakers will run through their rules, which are organized first as a rule for each lexical exception ranked by frequency followed by the productive rule and finally the Elsewhere Condition. This is demonstrated in (12), which is adapted from Yang (2016).

(12) IF $w = w_1$ THEN ...
IF $w = w_2$ THEN ...
...
IF $w = w_e$ THEN ...
Apply R

Here, the word w the speaker is processing is evaluated against listed exceptions (w_1 through w_e). If w finds a match, then the relevant exceptional clause is triggered. If the list of exceptions is exhausted without finding a match for w, then rule R applies. The key claim behind this formulation is that the computation of productive rules and their exceptions is a serial rather than an associative process, and that it is the computational search for exceptions that contributes to the cost of real-time processing. While this operation may appear on the surface to be an unwieldy account of processing, Yang (2016) argues that the time cost of adding a lexical exception is minimal and can only be identified in languages where additional processing effects such as neighborhood density and priming do not play a large role. I refer the reader to Yang (2016) for a full derivation and defense of the Tolerance Principle. As for applying the Tolerance Principle to PHL, we can formulate the productive rule as a series of frequency-ranked lexical exceptions followed by the productive rule. This is shown in (13), which applies the computation of frequency ranked lexical exceptions followed by the productive rule R.

(13) **PHL**:

1. IF w = *and* THEN /æ/ → lax
2. IF w = *can* THEN /æ/ → lax

...

39. IF w = *gaffe* THEN /æ/ → lax
40. $\text{æ} \rightarrow \text{æh} \,/\, \_[\ \text{+anterior}\ ] \cap ([\ \text{+nasal}\ ] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix})]\ \sigma$

Following evidence in Chapter 4 that speakers who vary between the productive rules of PHL and NAS also exhibit similar rates of variation in their lexical exceptions, I consider the entire series of computations listed in (13) to be the allophonic rule PHL. This formulation, notably, can accommodate speakers across the speech community having slightly different lexical exceptions and numbers of lexical exceptions, which may be based on differences in exposure to lexical exceptions during acquisition. This would account straightforwardly for the variation that we find between speakers in lexical exceptions. This also allows for diachronic changes to the list of lexical exceptions: when speakers add *planet* to their list of lexical exceptions as a tense production, these speakers simply add a line for *planet* to their lexical exceptions processes. Speakers are only limited by the number of lexical exceptions they may represent, which is capped at θ_N. For my analysis of PHL, $\theta_N = 194.7$.

As to which words qualify as lexical exceptions to PHL, here I take into account the fact that children acquire productive derivational morphology at a relatively young age. This is equivalent to postulating that PHL is a rule that applies at stem-level only, which is the analysis I consider to be accurate. Secondly, while Dinkin (2013) found evidence that /æl/ has merged with /owl/ in the Philadelphia Neighborhood Corpus, the data from speakers in the IHELP corpus (Chapter 2) and the IMPC corpus (Chapter 4) find speakers producing lax tokens of /æl/, in line with their CAT class tokens. For this reason, my analysis of PHL is that it applies at stem level, and includes /æl/ as part of the CAT class of tokens. In other words, I adopt the second column of Table 3.4 as my analysis of PHL. For a full description of my analysis of lexical exceptions, see Appendix A.

3.6 Discussion

Here, I've presented an in-depth analysis of one of the most contested intermediate phonological relationships using the Tolerance Principle to define the upper limit to lexical exceptions. In all formulations, we find that the traditional PHL rule emerges as a productive analysis for language learners. The specific repercussion of this analysis on the dissertation is a confirmation of the position taken by Labov et al. (2016) and Sneller (2018) that PHL is an allophonic rule. The extensions of this approach, however, are far more wide-reaching. This approach provides a solution to phonological relationships previously analyzed as intermediate or problematic, and also brings with it additional empirical predictions.

The first main prediction is that any intermediate relationship classified under the Tolerance Principle as productive should behave like an allophonic relationship rather than a phonemic one. In other words, allophonic rules with lexical exceptions are still expected to be productive: nonce words are expected to follow the regular rule. In any other task that differentiates allophones from phonemes, we would expect allophonic rules with lexical exceptions to also behave like allophones rather than phonemes. One potential additional piece of evidence may come from a phoneme alteration task. It seems to be more difficult for naïve speakers to produce a nonconforming allophonic production of a sound than to produce a different phoneme. Asking a NAS system speaker to produce a lax form of *man* often results in a production more aligned with their /ɑ/ target than their /æ/ target ([mɑːn] rather than [mæn]), but asking a speaker to swap out phonemic productions appears to be easier. We may expect that intermediate relationships classified as phonemic will be easily produced in a production alteration task while those classified as allophonic will be more difficult for speakers to target.

Secondly, this analysis predicts a precise tipping point between an allophonic and phonemic analysis, at the tolerance threshold of θ_N. If a productive rule held enough lexical exceptions to be near this threshold, it is easy to see how phonemicization may differentially affect speakers whose input is comprised of a different set of lexical items. For example, if a speaker of Philadelphia English acquired all the lexical exceptions in their input but through an accident of exposure was not exposed to enough lexical items that conformed to PHL, this speaker would posit a phonemic analysis of PHL while their peers, having been given a more representative vocabulary, would posit an allophonic analysis of PHL. This possibility both reinforces the importance of the individual in phonological change and provides a clear pathway for a productive rule to become phonologized into a phonemic distinction.

Chapter 4

Intraspeaker Variation in æ

An investigation of the mechanism of phonological change as driven by individual speakers relies on an in-depth analysis of the production of individual speakers. In this chapter, I present an analysis of transitional cohort speakers, finding that their data is most consistent with change via competing grammars.

Despite the Philadelphia Neighborhood Corpus providing data spanning the entirety of this phonological change and the Influence of Higher Education on Local Phonology corpus providing key data into the community-level fragmentation of this change, the amount of per-speaker data provided in these corpora do not allow for a robust analysis of change. Here, I create an additional corpus of speech designed specifically to target transitional cohort speakers and to obtain enough test tokens of /æ/ from each speaker to identify the mechanism of phonological change. This corpus, which I refer to as Investigating the Mechanism of Phonological Change (IMPC), is described in §4.3. My method for analyzing individual tokens is outlined in §4.1, and the predictions that each mechanism of change makes for the production of /æ/ by transitional cohort speakers are discussed in §4.2. In §4.5, I analyze all speakers in the IHELP and IMPC data sets that produce enough data to bear on the mechanism of phonological change.

4.1 Analysis of Individual Tokens

To analyze the mechanism of phonological change through the production of individual speakers, it will be necessary to analyze each test token as having been produced by either PHL, NAS, or some intermediate phonetic incrementation of these two systems. In most cases of phonological change, particularly for changes involving phonological mergers or splits, classifying individual tokens as having been produced by the old or the new phonological system is almost always impractical, given the overlapping distributions of tokens in phonetic space and the number of observations collected for typical sociolinguistic data. The allophonic restructuring from PHL to NAS, however, provides a rare opportunity to classify each individual observation according to which underlying system it adheres to. Here, I use the structural similarities and differences between PHL and NAS to my advantage. For PHL, as for NAS, a token of /æ/ preceding a tautosyllabic front nasal (as in *hand* or *ham*) will be produced as lax. Likewise, there are a large number of words that fall into the elsewhere condition for both systems, producing lax /æ/ in words like *cat* and *dad*.

The shared conditioning between PHL and NAS means that we do not need to know whether a speaker has the PHL system or the NAS system in order to characterize that speaker's tense and lax acoustic targets: their tense target will be in the phonetic space of HAND words and their lax target will be in the phonetic space of CAT words. This information about a speaker's phonetic targets can then be used to determine whether each test token aligns best with that speaker's tense target or their lax target.

4.1.1 Classification Methods for Test Tokens

The problem of classifying test tokens for tense, lax, or intermediate realization is not trivial. To determine the optimal classification method, I test different classification methods to actual data from a PHL speaker and a NAS speaker, as well as to simulated data (described in detail in Appendix B), to determine which classification system produces interpretable results for transitional cohort speakers. The classification methods attempted include K-means cluster analysis and Hierarchical cluster analysis run on the F1 and F2 values for tokens as well as these methods run on a Principle Components Analysis (PCA) and a t-Distributed Stochastic Neighbor Embedding (t-SNE) transformation of the data. For the purposes of this dissertation, a glm classifier provided the best fit to the data, and is the method used and described here.

A generalized linear model (glm) is a family of linear regression models, which can be turned into a classification method. As a classification method, a glm model is first fit to training data, which provides coefficients for each term in the model. These coefficients are then used to predict the outcome of test data. Typically, this method is used as a basis for machine-learning: human coders code a random subset of a data set, and the resulting glm model for that training data is then fit to the rest of the test data. As a method for classifying /æ/ production for transitional cohort speakers, this must be slightly modified. We cannot take a random subset of data, precisely because we can not determine *a priori* whether a speaker's test tokens are tense or lax. However, because of the overlapping conditioning factors between PHL and NAS, we can determine the phonetic target of a speaker's HAND and CAT class tokens. Here, I use these tokens as training data for a glm classifier for each speaker, which is then fit to test data to predict whether each test token was produced as phonetically tense or lax.

4.1.2 Applying the Glm Classifier to Speaker Data

The first step in using a glm classifier is to split a speaker's data up into training tokens and test tokens. Here, we use each speaker's HAND class tokens as training tokens for the tense phonetic target and CAT class tokens as training tokens for the lax phonetic target. Figure 4.1 shows the training tokens for Bobby Marx, a Philadelphian born in 1967 whose data is part of the IHELP corpus. 95% confidence ellipses are plotted around Bobby's HAND class tokens (solid line) and his CAT class tokens (dashed line), to show the acoustic characteristics of his tense and lax targets, respectively.

For each speaker, I use these training tokens to train a glm classifier on the acoustic parameters of that speaker's tense and lax targets, using F1 measurement, F2 measurement, duration, and syllable stress as independent variables, shown in (14). The resulting coefficients of this classifier were then used to predict the probability of tense or lax realization for the remaining test /æ/ tokens using the `predict()` function, shown in (15).

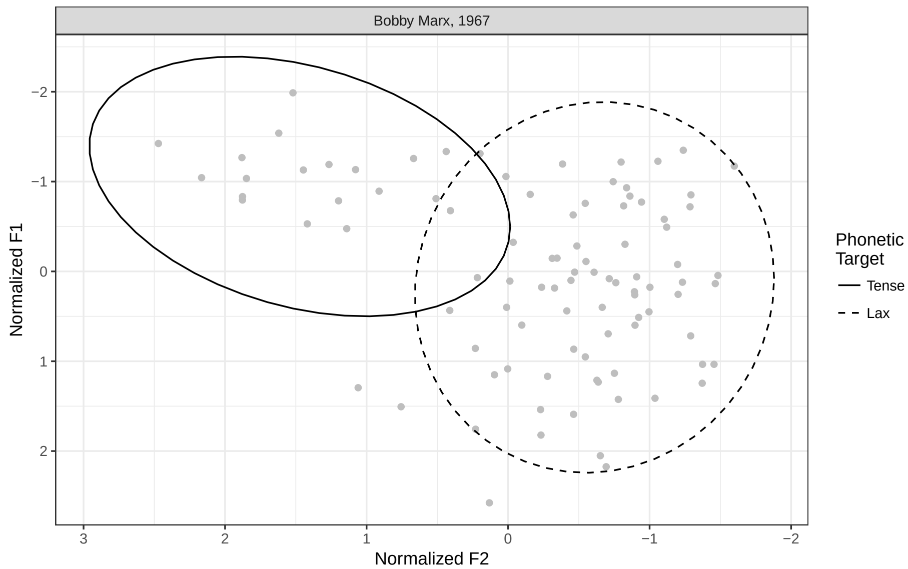


Figure 4.1: Bobby Marx training data.

(14) `predmod <- glm(tense ~ F1*F2*F3*duration*stress)`

(15) `testdata$tenseProb <- predict(predmod)`

Once test tokens have been classified as either tense or lax, each token can then be classified as having been produced by PHL or NAS. Any LAUGH or MAD token that has been classified as tense by the classifier is consistent with PHL but not NAS; likewise, a MANAGE or HANG class token that is classified as tense is consistent with NAS but not PHL. In Figure 4.2, we can see the results of the classifier. Training tokens are again plotted in gray, with 95% confidence ellipses drawn around the tense target (as identified through HAND class tokens) and lax target (as identified through CAT class tokens). Test tokens are plotted over the training data, with tokens classified as PHL in orange and NAS in green. We can see in Figure 4.2 that Bobby overwhelmingly produces tokens consistent with PHL; given his demographic data as a Philadelphian born before 1983, we expect to find predominately PHL data. However, it is also clear that there are a few tokens that are selected by the glm classifier as consistent with NAS. Given that Bobby's overhelming pattern is PHL, I term these tokens *incongruent tokens*, as they are incongruent with his dominant system.

The distribution, number, and lexical identity of these incongruent tokens are an important aspect of identifying whether a speaker's production matches PHL, NAS, or an intermediate system. I come back to this in §4.5.

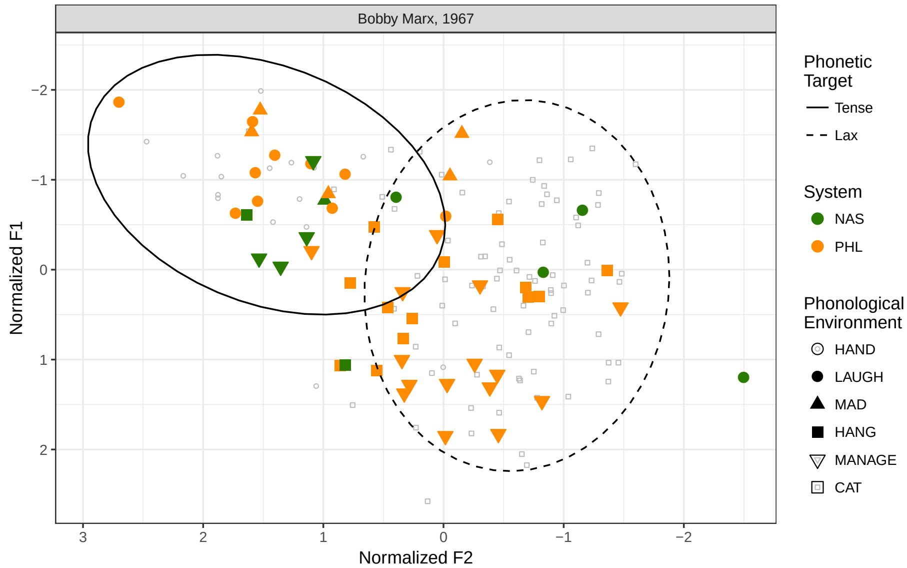


Figure 4.2: Bobby Marx test data.

There is a final point to make about the use of the glm classification system. While the glm classifier produces a probability value for each test token between 0 (lax) and 1 (tense), the break point between which tokens are classified as tense or lax is not 0.5. This is particularly true, given the phonological conditioning of PHL tense and lax tokens. All prenasal tokens in the training data are tense, and all tense training data are prenasal, producing a bias towards classifying prenasal test tokens as tense. Adding to this bias is the acoustic output of PHL tense tokens: Kroch (1996) found, for example, that prenasal tense tokens are acoustically higher and fronter than the LAUGH and MAD class words[7]. To determine the most accurate cutoff thresholds for tense classification, I use the productions of traditional PHL and NAS speakers to obtain a probability threshold that maximizes the accuracy for both types of speakers. This results in a cutoff of 0.22 for prenasal tokens (above which a token will be classified as tense),and a threshold of 0.14 for non-prenasal

[7]While prenasal tokens are realized as acoustically more tense than the rest of the traditional tense class, an ultrasound study (Mielke et al., 2017) finds all tense tokens of the traditional PHL system to be articulatorily identical with regards to tongue position.

tokens.

After being classified as tense or lax, each test token was then categorized as either a PHL token or a NAS token, according to which system it was consistent with. Tokens categorized as PHL are represented in orange, and tokens categorized as NAS are represented in green.

4.2 Profiles of Each Mechanism of Change

Because of the complex set of facts surrounding PHL, it is useful to first run through the predictions that each mechanism of phonological change make. In what follows, I present simulated data to highlight the predicted profile for each mechanism of change. For each simulated change, I create a hypothetical PHL speaker and a hypothetical NAS speaker, then create transitional generation data for three intermediate steps based on the assumptions from each mechanism of phonological change.

The simulated PHL data is generated using F1, F2, covariance matrices, and token count values drawn from an actual PHL speaker (Mary C., whose production of PHL represents a prototypical PHL speaker and who produces one of the highest token counts of /æ/ in the PNC, with $N = 1456$). Simulated NAS data is generated using these same values from Cara G., who is the speaker in the IHELP corpus with the highest token count ($N = 825$). Simulating the productions of PHL and NAS allows me to set the seed for each simulation, resulting in pseudorandom simulated tokens. Setting the seed to the same number for each of these plots enables us to see that any changes between plots is due to differences in the underlying means and covariances of the plots, rather than due to random noise in generated data. Given the PHL speaker and NAS speaker, I then generate transitional generation data following the assumptions of each mechanism of change, which are described in detail below.

Figure 4.3 compares Mary C's actual production data (left) with the simulated plot of her production data (right). In both facets, a 95% confidence ellipse is drawn around the HAND class (solid line) as well as around the CAT class (dotted line), to give a visual representation of her tense and lax phonetic targets. Simulated data was created using the mvtnorm package in R, with F1 and F2 means for each conditioning factor drawn from Mary's data and covariance matrices for F1 and F2 produced with Mary's actual covariance measures. The simulated data contains the same token count for each conditioning factor as Mary's actual data, so that the simulated data contains an accurate snapshot of the relative proportion of /æ/ tokens within each conditioning factor. Mary's simulated data is shown in the right panel.

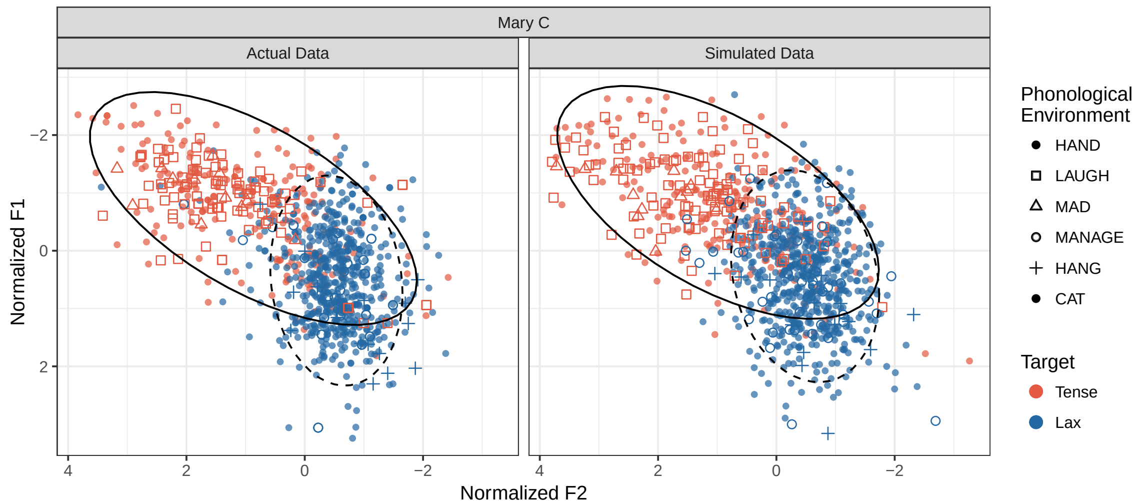


Figure 4.3: Mary C. real data (left) and simulated data (right).

Figure 4.4 displays a similar output for Cara G, with the F1 and F2 means drawn from Cara's actual production of /æ/ in each conditioning factor, and covariance matrices calculated separately for each conditioning factor. The means and covariance for the simulated data matches Cara's production; however, here we have drawn the N values per conditioning factor from Mary's production, so that our simulated speakers are maximally comparable. In other words, the right hand panel depicts the output we would expect if Cara had produced 1456 /æ/ tokens instead of 825.

I use these simulated plots of PHL and NAS so that we may produce a 5-step continuum between PHL and NAS based on the specific predictions from each mechanism of phonological change. Each plot is generated from a pseudorandom gaussian distribution. Using `setseed()` ensures that each plot is generated from the same seed, resulting in psuedorandom rather than fully random data. This allows us to reproduce each plot, changing only the F1, F2, and covariance parameters as predicted by each mechanism of change. In other words, any differences in the position of a particular token between two plots is due to an actual difference in the theoretical predictions

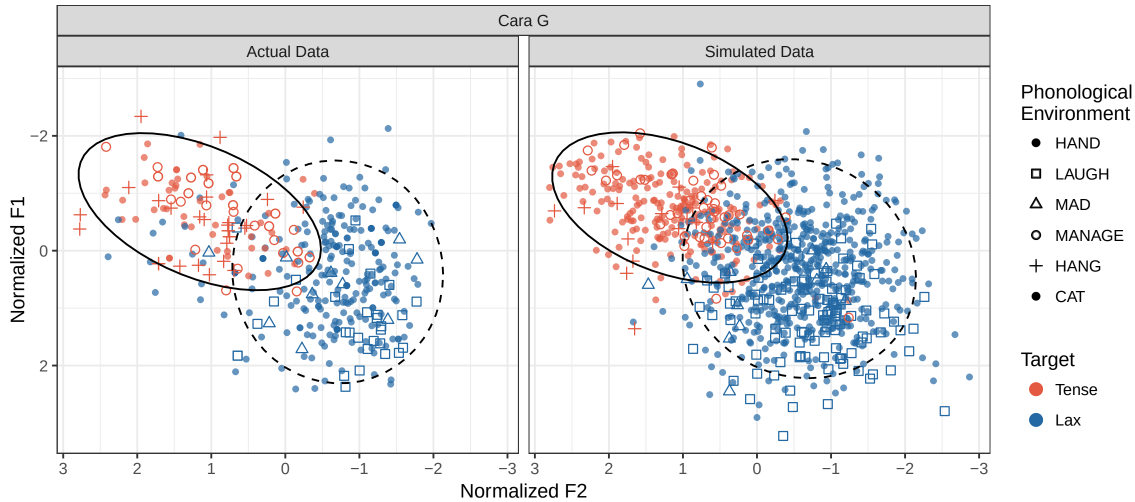


Figure 4.4: Cara Grant actual data (left) and simulated data (right).

rather than to noise in the data. Each step in the 5-step continuum is labelled as a different cohort for the sake of temporal exposition. The actual time difference between "Cohort 1" and "Cohort 2" may only be a short number of years; the main point is that each panel in the following plots represents one speaker who is slightly more advanced in the change from PHL to NAS than the previous panel.

For each simulated speaker, I run the simulated data through our tenseness glm classifier, based on the F1 and F2 values for tense and lax for that simulated speaker's HAND and CAT classes. Each simulated token is then classified as either consistent with PHL (orange) or NAS (green), which provides a profile of what our expected outputs from the transitional cohort speakers will be.

4.2.1 Phonetic Incrementation

There are a few possible profiles for phonetic incrementation to follow in this case, and these depend on the unit that is being incremented.

Tense and Lax Incrementation

The naïve hypothesis for PHL becoming NAS through phonetic incrementation is that phonetic incrementation affects both the tense allophone of /æ/ and the lax allophone simultaneously. The idea here is that due to some unspecified combination of production and perception errors, all conditioning factors contributing to the tense allophone of /æ/ become phonetically laxer while all conditioning factors of the lax allophone becomes phonetically tenser. In the middle of the change, we would expect both allophones to be produced in the same phonetic space, in between canonical PHL lax and canonical PHL tense and appearing merged in phonetic space. After this middle stage of the change, we would expect to see the hint of allophonic restructuring, with the conditioning factors contributing the tense allophone of NAS becoming phonetically raised while the conditioning factors contributing to the lax allophone of NAS become phonetically laxed, leading to a final stage where the production of /æ/ matches our prototypical NAS speaker. The simulated data is created using a 5-step linear interpolation between the F1 and F2 means and covariance matrices of the simulated PHL and NAS data.

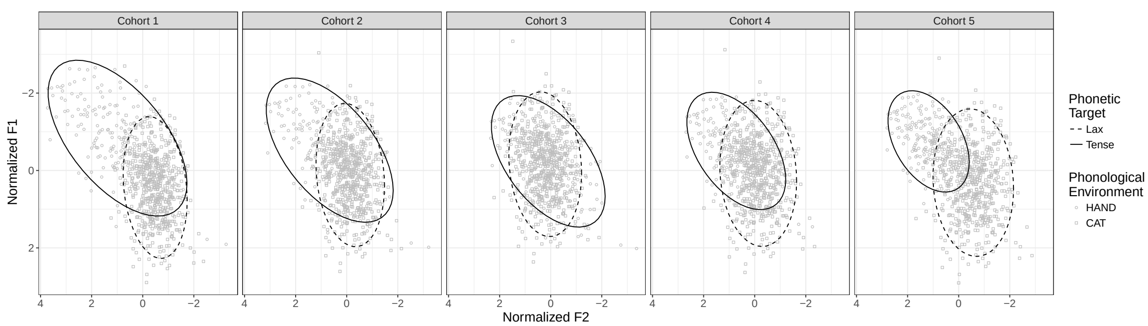


Figure 4.5: Profile of phonological change for HAND and CAT classes under phonetic incrementation of all conditions.

For all the changes that follow, we produce specific predictions about the behavior of HAND and CAT as well as the four conditions that differ between PHL and NAS. For clarity, here we present first the predictions about the shared conditioning factors first, before overlaying the predictions about the test conditions. Figure 4.5 displays the predicted acoustic outputs of HAND and CAT given full phonetic incrementation. As we can see, the two acoustic targets drift together in pho-

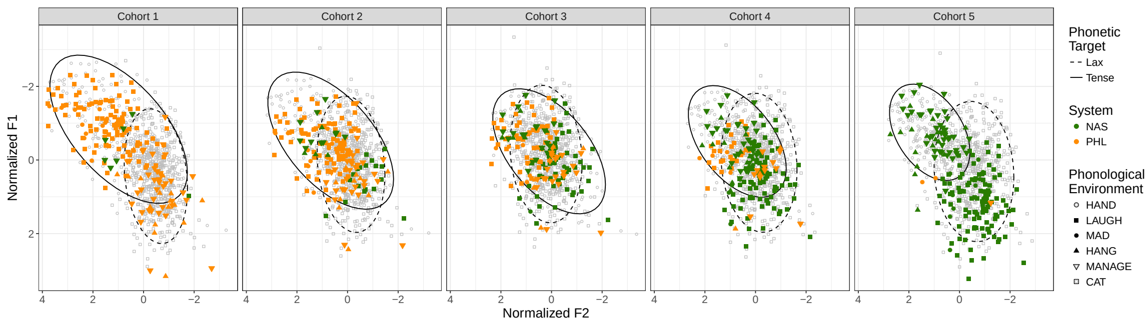


Figure 4.6: Profile of phonological change for test classes under phonetic incrementation of all conditions.

netic space, completely overlapping in Cohort 3, then drift apart again. The test conditions follow suit (4.6); each of the four test conditioning factors (LAUGH, MAD, HANG, MANAGE) increment towards a central position in Cohort 3, then continue on their merry restructuring way to produce full NAS by Cohort 5. I note briefly that like the actual data discussed above, these simulated productions of PHL and NAS contain some test tokens that are classified as incongruent with the rest of the tokens. In analyzing actual speaker data, it is a close analysis of test tokens such as these, that suggest underlying systemic variation, that will provide an account of the mechanism of phonological change.

The main identifying characteristics of phonological change through this type of phonetic incrementation are in the unimodal distribution of all conditioning factors in Cohort 3; the training tokens are merged in acoustic space, as are the test tokens.

Tense Allophone Incrementation

Given the sociolinguistic facts reported for PHL however, we expect phonetic incrementation as shown in §4.2.1 to be unlikely. Most relevantly, Labov (2001) finds Philadelphians who produce PHL to negatively rate only the tense tokens of PHL. This negative evaluation produces a social motivation for phonetically laxing the tense allophone but not the lax allophone. This prediction of phonological change is very similar to the prediction described above; here the only difference is that the lax allophone remains in its lax position throughout the change, while the tense target shifts down to join it in the lax target before moving back to the NAS tense position (Figure 4.7). Here, the transitional cohort values are created first using a 3-step linear interpolation of of F1, F2, and covariance matrices between PHL and Cohort 3 and then a 3-step interpolation between Cohort 3 and NAS. Cohort 3 was created using F1 and F2 for the mean of the lax test class, so Cohorts 1-3 represent a gradual shift of all tense PHL tokens to the CAT target while Cohorts 3-5 represent the gradual shift of tense NAS tokens from the CAT target to the NAS tense target. As with the previous simulation, covariance matrices are a 5-step interpolation between the covariance matrix of PHL and the covariance matrix of NAS.

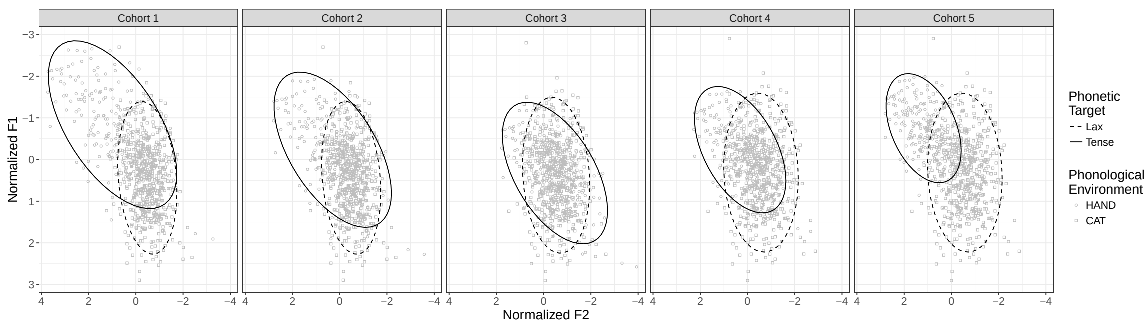


Figure 4.7: Profile of phonological change for HAND and CAT classes under phonetic incrementation of tense allophone.

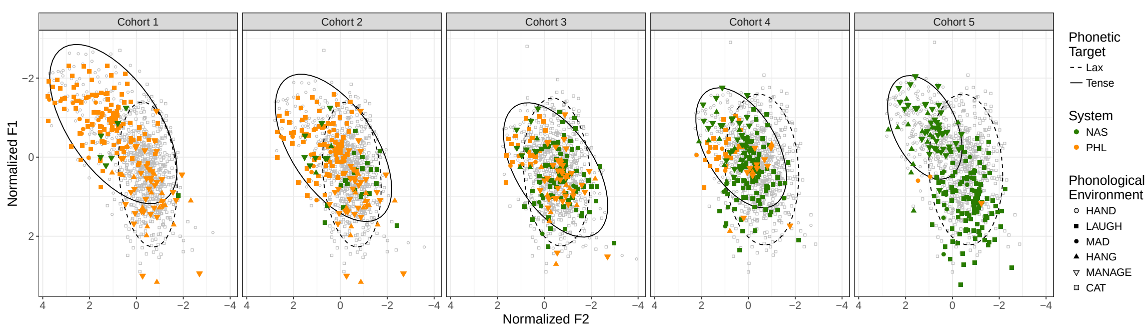


Figure 4.8: Profile of phonological change for test classes under phonetic incrementation of tense allophone.

The characteristics of this change are nearly identical to those laid out above: we find a unimodal distribution of both the training tokens and test tokens in Cohort 3; the only difference here is that we find this unimodal distribution occurring in the acoustic space of speakers' lax targets.

Test Conditioning Factors

The unimodality in all tokens predicted by the phonetic incrementation profiles above are easy to identify as change via phonetic incrementation. Unfortunately, phonetic incrementation could also take on a less clear path. Because phonetic incrementation does not rely on any abstract processes, this mechanism allows for any group of tokens to shift or remain unshifted. This is specifically in contrast to a phonologically based theory of change, in which the target of any phonetic movement is predicted to be a cohesive phonological unit or phonological feature. In other words, there is nothing baked into the theory of phonetic incrementation that predicts that all tokens of an allophone or even all tokens of a phonological conditioning factor will necessarily undergo the same set of errors in production and perception. It is possible, then, that this change from PHL to NAS is the result of phonetic incrementation of only a subset of tokens. Here, we present the most likely version of this, in which the conditioning factors that differ between PHL and NAS phonetically increment but the shared conditioning factors remain stable. In this simulation, the HAND and CAT classes are produced as a 5-step linear interpolation between PHL and NAS F1, F2, and covariance matrices. The test conditioning factors are produced in a 3-step interpolation between PHL and Cohort 3, then a 3-step interpolation between Cohort 3 and NAS. Cohort 3 is produced using the mean F1 and F2 of the HAND and CAT classes and the mean covariance matrices between PHL and NAS for each conditioning factor.

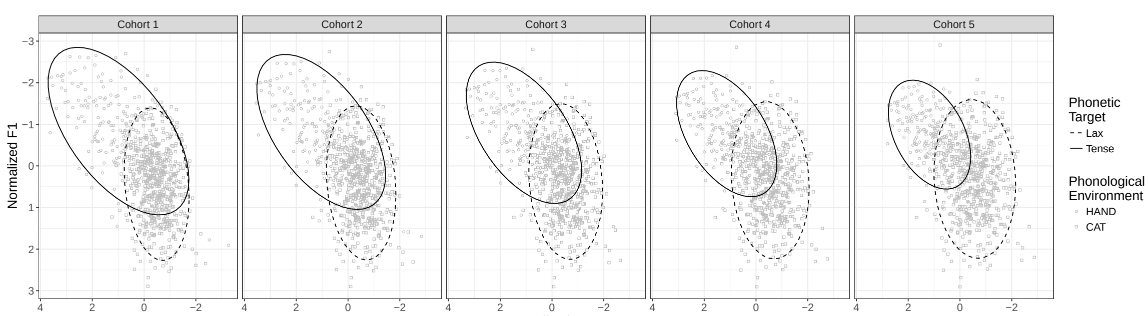


Figure 4.9: Profile of phonological change for HAND and CAT classes under phonetic incrementation of test conditions only.

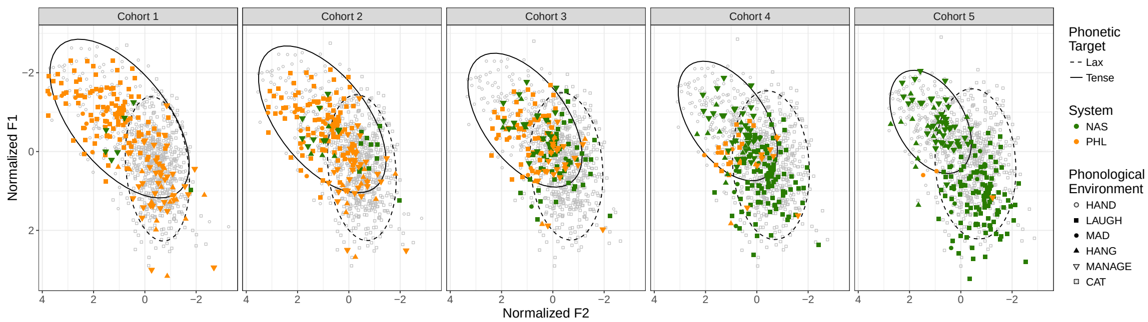


Figure 4.10: Profile of phonological change for test classes under phonetic incrementation of test conditions only.

Figures 4.9 and 4.10 show the predicted outputs for the training and test tokens, respectively. Unlike the previous predictions for change via phonetic incrementation, this type of change predicts that the shared conditioning factors will remain as a tense and lax target; we expect to see the training tokens exhibit bimodal distribution. The test tokens, on the other hand, still go through a period of unimodality in Cohort 3, where they are produced in the intermediate space between the tense target and the the lax target.

As we will see for the predictions made by spontaneous phonologization and grammar competition, it is this unimodality of the test tokens that provides the strongest cue for change via phonetic incrementation.

4.2.2 Spontaneous Phonologization

If the change from PHL to NAS is driven by the transmission mechanism of spontaneous phonologization, this predicts that individual speakers will posit a single allophonic system (either PHL or NAS), and stick to this system in their production. The community-level change, then, will be driven by an increasing number of speakers positing NAS in each successive cohort. This is represented in Figure 4.11, which depicts a representation of four speakers in each cohort. In the first cohort of speakers, representing traditional PHL in the community before any posited change to NAS, every speaker posits and produces PHL. By Cohort 2, one speaker out of four has posited NAS. This increases until Cohort 5, in which every speaker has posited and is producing NAS.

If the change on the community level has been driven by spontaneous phonologization, each

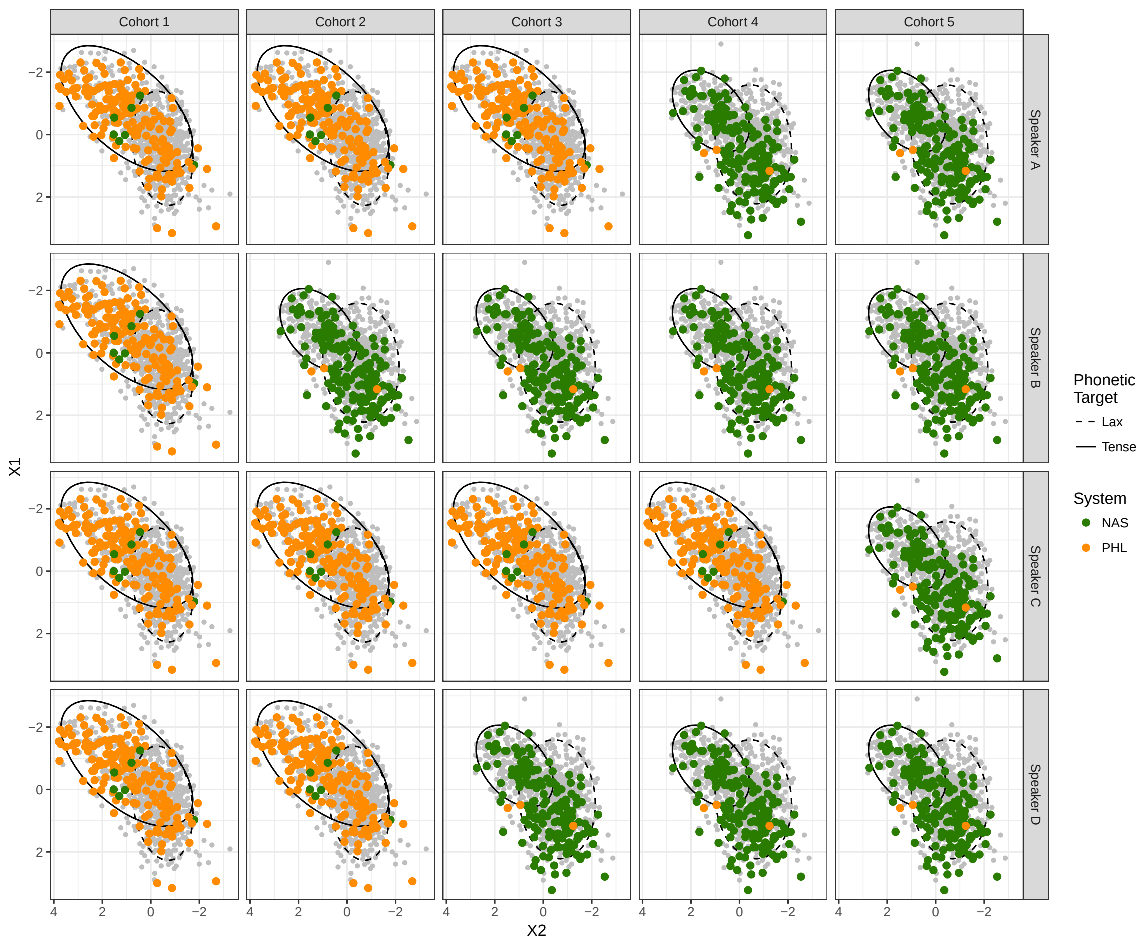


Figure 4.11: Profile of phonological change for change via spontaneous phonologization.

individual speaker should produce strictly PHL or strictly NAS outputs. This mechanism of phonological change requires an analysis of the community as a whole: given a single individual, it will not be possible to determine whether that speaker displays change via spontaneous phonologization or whether that speaker simply is drawn from a subset of the community that has completed the change to NAS or not yet begun the change to NAS. Determining whether change is occurring via spontaneous phonologization will require sociolinguistic data about that speaker's educational peers, as discussed in Chapter 2.

4.2.3 Intraspeaker Grammar Competition

If the change from PHL to NAS is driven through the mechanism of intraspeaker grammar competition, we expect to see a bimodal distribution of all test conditioning factors. This theory of change states that allophonic /æ/ system is a parameter in speakers' grammars, which in this period of change varies between the PHL variant and the NAS variant within a single speaker. In other words, allophonic systems as a whole would act as an abstract level of sociolinguistic variable, with speakers using some proportion of each system. This predicts that speakers in the beginning of the change are using mostly PHL tokens while speakers in the end of the change are using mostly NAS tokens. Here we present only the prediction plot of all tokens, since there is no main difference between the test tokens and the training tokens. Figure 4.12 displays the simulation results for all five steps of this change, beginning with 100% PHL tokens and ending with 100% NAS tokens. Cohort 2 is comprised of 25% NAS tokens, Cohort 3 is comprised of 50% NAS tokens, and Cohort 4 is comprised of 75% NAS tokens.

The predictions for change through grammar competition are clear: in all cases, we expect to see a distinct tense and lax acoustic target from the shared HAND and CAT classes. Test tokens are produced well within the tense and lax targets, with a bimodal distribution, and we expect to see variation between PHL and NAS at roughly equal rates across all test conditions.

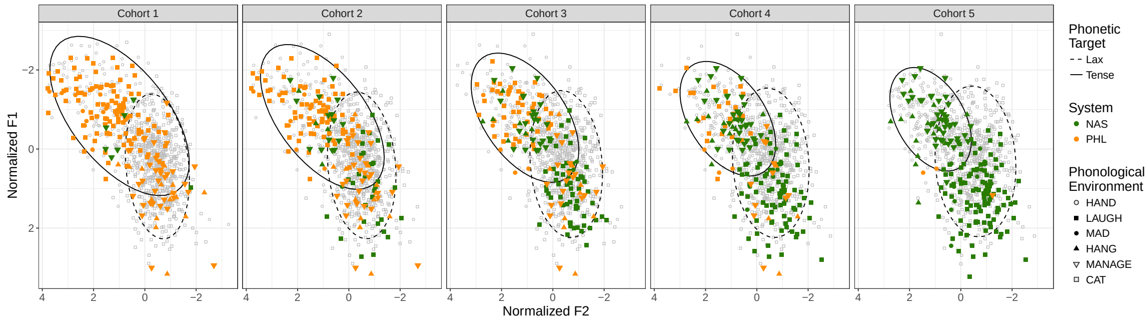


Figure 4.12: Profile of phonological change for change via competing grammars.

4.2.4 Summary of Predictions

Sections 4.2.1 through 4.2.3 provide an outline for predicted outputs based on each mechanism of phonological change. Here, the most crucial point of comparison is between the expected outputs for Cohort 3 from each of the theories of phonological change, since that is where we see the biggest differentiation between the theories. It is the profiles of Cohort 3 simulations that we are particularly looking for in our actual data. In what follows, I will use these profiles of predicted outputs to best characterize transitional generation speakers' productions. If we find speakers producing outputs that match one of the profiles of phonological change, this will serve as evidence that that particular mechanism of phonological change is at play.

4.3 Investigating the Mechanism of Phonological Change Corpus

The data that would bear on our particular question is relatively rare in frequency, both in terms of transitional cohort speakers and the number of tokens each speaker produces. I demonstrated in Chapter 2 that even amongst a single age range in Philadelphia, it is only within particular subsets of the population that PHL and NAS are currently vying for dominance. NAS has won out in elite non-Catholic high schools, and PHL still has a stronghold in the non-elite Catholic high schools. The social networks in which we expect the highest likelihood of mixed system speakers is in the graduates of non-elite non-Catholic high schools and of elite Catholic high schools. In addition to social networks resulting in only a subset of the population likely to acquire competing /æ/ systems, it is also the case that the specific test tokens of /æ/ that would disambiguate between PHL and NAS are rare in conversation. Within the IHELP data, for example, /æ/ tokens comprise only 14% of a speaker's data (with an average of 579 /æ/ tokens and 81 test condition /æ/ tokens per speaker over a one hour sociolinguistic interview).

In Chapter 2, I took a wide sampling approach in an effort to more fully describe the community-level pattern of this change. Here I must take a more targeted approach, with the goal of being able to analyze variation in speakers' test tokens in a way that will determine whether this change is occurring through phonetic incrementation, instant phonologization, or grammar competition. Drawing from the results in Chapter 2, I focus here on the recent graduates of Catholic schools, which is the population currently in flux with regards to this change and therefore the most likely to be transitional cohort speakers. In addition to targeting the population of speakers most likely to be transitional cohort speakers, it is also necessary to increase the number of test tokens obtained per speaker so that I may distinguish between surface-level variation that is driven by grammar competition and surface-level variation driven by another factor such as phonetic incrementation or lexical diffusion. The methods used to target transitional cohort speakers as well as to increase the yield of test tokens are described in some detail below. The data collected under the methods highlighted below is compiled into a single corpus Investigating the Mechanism of Phonological Change (IMPC).

4.3.1 Targeting Transitional Cohort Participants

The IHELP data resulted in some data that suggested that speakers were variable between PHL and NAS; however, there simply were not enough tokens from most of these speakers to rule out any proposed mechanism of phonological change. Here, I target those individuals who are the most likely to be transitional generation speakers. I do this in two ways. First, every participant from the IHELP database whose data suggested variation between PHL and NAS were invited to participate in this as a follow-up study; this resulted in four participants. Second, I extend my reach by targeting Catholic school graduates who were born after 1983, using my existing social networks in Philadelphia. This resulted in sixteen additional participants.

Participants obtained through social networks were first given a screening test, which consisted of a 10 minute Semantic Differential task (see Table 4.1 for the exact questions asked during screening) conducted over the telephone. Participants' productions of each test item were auditorily coded, and any participants who were found to produce variation were invited to participate in a full session. All participants were paid $30 per session they participated in.

4.3.2 Increasing Test Tokens through Interview Methods

Topic Directed Conversations

The first requirement for data collection is that I obtain over an hour of speech per speaker. My aim is to collect at least 10 tokens per test conditioning factor; judging from the rate of test tokens found in the IHELP corpus and the PNC, a classic sociolinguistic interview (Labov, 1984) would need to be roughly two hours in length. To increase the number of test tokens per hour of speech, I introduce the method of *Topic-Directed Conversations*. In this method, participants come into a quiet recording space with a friend, and are recorded as a dyad having a conversation. Following Boyd et al. (2015), the researcher leaves the room and allows the two participants to interact in a naturalistic way. One potential pitfall of using a dyad recording method instead of a traditional sociolinguistic interview is in the expected proportions of participant speech: while a sociolinguistic interviewer is trained to have the interviewee speak most of the time, a more natural conversation between two participants will result in each participant speaking roughly 50% of the time. I find, however, that volume of per-participant speech with participants each speaking roughly 50% of the time in a 1.5 hour Topic-Directed Conversation exceeds the average volume of a participant speaking roughly 80% of the time in a 1 hour sociolinguistic interview (avg. 4855 words vs. 2751 words).

In addition to enabling me to obtain naturalistic data from two participants at once, Topic-Directed Conversations also provide a more important benefit in that they direct participants toward topics with a high likelihood of producing relevant test tokens. For test tokens of /æ/, conversational prompts included the questions in (16)–(18).

(16) When is the **last** time you got really **mad**? Have you two ever gotten into a fight with each other? What do you do when you're **angry**?

(17) What about the **last** time you were embarrassed? Do you remember a time that one of your friends did something really embarrassing?

(18) When's the **last** time that you remember feeling scared? Is there anything that makes you feel like you're going to **panic**?

In (16)–(18), I've bolded the test tokens inherent to the question (these words were not bolded in participants' conversational prompts page). In addition to the questions themselves increasing the use of these words, the answers also typically involved high rates of test tokens for /æ/. Question (16) typically resulted in at least one story from each participant about the last time they were mad, producing multiple stressed tokens per participant of *last, mad, angry, bad.* Question (17), while not containing any test tokens within the question, often resulted in stories from participants that involved *laughing*, additionally, because participants were high school or college friends, many of these stories took place in *class.* Question (18) straightforwardly produced multiple stressed tokens of *panic* per participant. In this way, each of the four test conditions (MAD, LAUGH, HANG, MANAGE) were heightened by this line of questioning.

In this case, the Topic-Directed questions also had the benefit of being thematically related as emotional-state questions, making these inclusion of these questions a natural as a set. Procedurally, participants were told that I was investigating "language and life in Philadelphia," and that I wanted them to chat for about an hour and a half. Participants were told "you may talk about whatever you like, and here is a list of conversational prompts that you're welcome to use." An hour and a half later, I returned to administer the formal methods, outlined in §4.3.2. Because participants were explicitly told that it did not matter whether they followed the prompts or not, not every participant discussed every conversational prompt. However, each dyad did discuss each of the three targeted emotional-state questions, which were found in pilot interviews to be a very productive set of topics that participants were highly engaged in. Overall, this method produced an average of 238 test tokens per participant from the informal conversation section of the interview (contrast with avg. 170 from the traditional sociolinguistic interviews that comprise the

IHELP corpus). The full conversational prompt list is provided in Figure C.1 in Appendix C.

Formal Methods

After an hour and a half of Topic-Directed Conversation, I returned to the recording room to administer the Formal Methods. These were also designed to target the relevant test items for /æ/. Participants were first given a Semantic Differential task, which was slightly modified to enable responses from both participants. Each participant was given a list of different word pairs. Participants took turns reading off a pair, then discussing what they thought the difference between the two words was, followed by their partner responding with their thoughts. In most cases, this resulted in a light debate over the meanings of each pair, producing multiple stressed tokens of the test items per speaker. In the rare case where the partner simply agreed with the first participant or the participant only gave the meaning of one of the words (e.g. "*mad* is more casual"), I prompted further discussion with pointed questions (e.g., "When would use one vs. the other"). The full list of Semantic Differential pairs is given in Table 4.1.

Mad and Angry	Janitor and Handyman	Strangle and Choke	Stammer and Stutter
Sad and Unhappy	Planet and Asteroid	Valley and Canyon	Damage and Destruction
Glad and Happy	Ran and Jogged	Palace and Mansion	Street and Road
Bang and Crash	Swam and Swum	Pal and Buddy	Pollyanna[8] and Secret Santa

Table 4.1: Semantic Differential prompts.

Following the Semantic Differential, participants were asked to read a word list, also provided in Appendix C. The word list included targeted test /æ/ words, as well as several nonce words from the test conditions to help identify the productivity of participant's /æ/ rules. Participants were asked to read the words down rather than across, and were instructed "some of these words aren't real – just say them however you think they should be said."

[8] *Pollyanna* is a term for Secret Santa prevalent in Irish or Italian Philadelphia neighborhoods

4.4 Analysis of individual speakers

Determining which mechanism of phonological change is driving the allophonic restructuring in Philadelphia relies primarily on our ability to determine whether intermediate cohort speakers' productions align with one or more of the predicted outputs highlighted above.

The main distinction in the output between the three mechanisms is in the distribution of the test tokens. In phonetic incrementation, even in the versions that maintain a distinction between HAND and CAT, all test tokens are expected to be produced from a single distribution located intermediately between HAND and CAT. In spontaneous phonologization, the test tokens as a whole will be drawn from two distributions (both HAND and CAT), but each test word class will itself be drawn from a single distribution (either HAND or CAT), following the underlying system that the speaker is adhering to. In competing grammars, each test word class will be drawn from two distributions (HAND and CAT). The basic questions that we seek to answer with statistical evidence are provided in (19)–(20). Each mechanism of change produces a distinct set of answers to these questions, as shown in Table 4.2.

(19) Are the test tokens, overall, bimodal?

(20) Is each conditioning factor bimodal?

	(19)	(20)
Phonetic Incrementation	no	no
Spontaneous Phonologization	yes	no
Competing Grammars	yes	yes

Table 4.2: Profiles of distributions for test tokens and conditions for each of the three mechanisms of change.

As we will see, the statistical methods for analyzing token distribution are not well set up to answer these questions for unsupervised data. A full analysis of individual speakers will rely on bringing statistical, phonological, and sociolinguistic evidence together to bear on the question of the mechanism of phonological change.

4.4.1 Statistical Evidence

The multidimensional and unsupervised nature of this data means that standard statistical methods are not well set up to address (19) and (20). Most tests of bimodality rely on a label given to each group of tokens, and then asking whether each group has been drawn from the same sample. In our case, we cannot give an *a priori* label to any of the test tokens, since the expected distributions under each mechanism of phonological change range from a probability of 0 to 1 for any label that we may give to an individual observation.

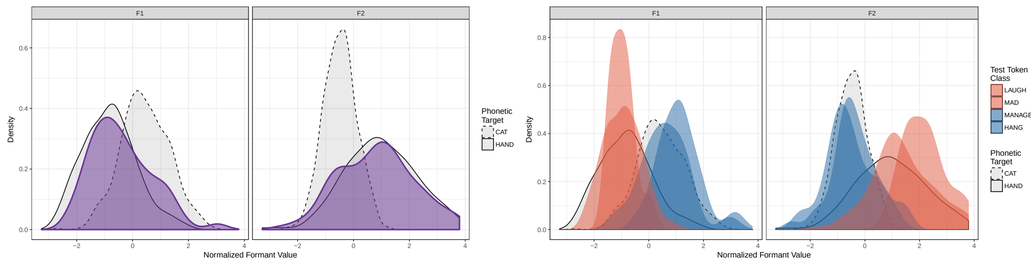
In what follows, I highlight a few statistical methods that in theory would provide some support to our goal of identifying test token observations as being produced by one of the three mechanisms of phonological change. In each case, I test each method on the simulated data (particularly the simulated data represented the expected output of a Cohort 3 speaker) from each mechanism of change outlined above, to test whether each method is able to distinguish data that we know the underlying distribution of.

Multidimensional Kurtosis

As I have discussed in §4.1.1, unsupervised cluster analysis did not provide a useful tool for distinguishing underlying classification of tokens. Turning to the question of testing modality (unimodal vs. bimodal distribution), I first test the usefulness of a multidimensional kurtosis measure (also known as Mardia's test). Mardia's test in the MVN package in R provides Pearson's adjusted kurtosis values, which are generally interpreted as unimodal (or normal) between the range of -2 and 2, and bimodal outside of this range. Unfortunately, Mardia's test does not provide a reliable distinction between the simulated data for phonetic incrementation (which should produce a unimodal distribution) and competing grammars (which should produce a bimodal distribution), so cannot be used as a reliable measure of bimodality for the data from transitional cohort speakers.

Kernel Density

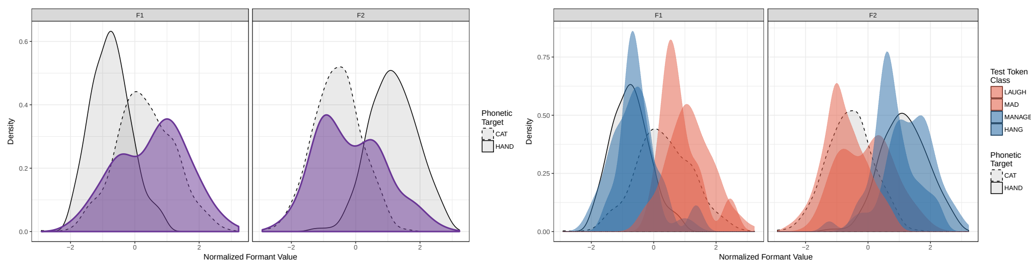
Kernel density estimation provides an estimate of the probability density function of a variable, essentially providing an output that can be interpreted visually. Using the ks package in R, I fit a



(a) Bimodal distribution of test tokens for PHL speaker.

(b) Unimodal distribution of each test conditioning factor for PHL speaker.

Figure 4.13: PHL speaker kernel density plots.



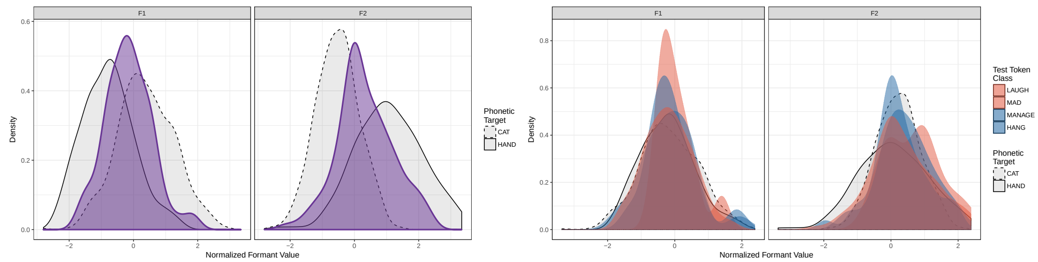
(a) Bimodal distribution of test tokens for NAS speaker.

(b) Unimodal distribution of each test conditioning factor for NAS speaker.

Figure 4.14: NAS speaker kernel density plots.

multidimensional kernel density estimate to each class of test tokens. These estimates can then be compared to each other to test whether the distribution of each class is from the same sample or not. In principle, this would provide a useful tool for testing whether each conditioning factor was equally participating in any changes (as expected for a competing grammars analysis and most phonetic incrementation predictions) or if each conditioning factor was separately affected by the change from PHL to NAS. In practice, the number of observations per conditioning factor for most speakers makes the kernel density plots difficult to lean on as analysis tools. However, the predictions that each mechanism of phonological change make with regards to kernel density are worth briefly discussing. In the final analysis of each individual speaker, I use sociophonological as well as kernel density evidence to classify each speaker as consistent with PHL, NAS, competing grammars, or phonetic incrementation.

Here, a visual plot of the predicted kernel density outcomes for each mechanism of change is



(a) Unimodal distribution of test tokens for phonetic incrementation.

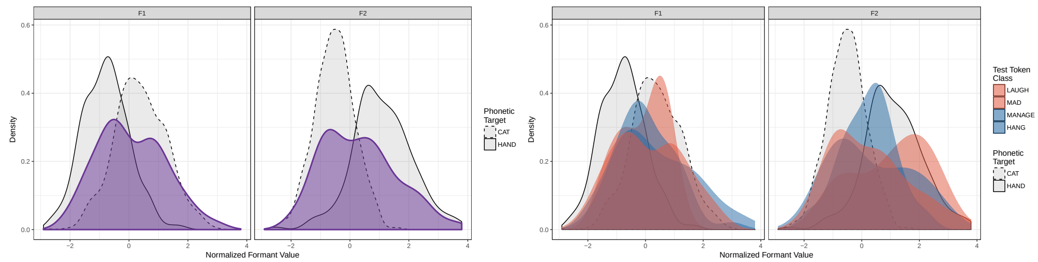
(b) Unimodal distribution of each test conditioning factor for phonetic incrementation.

Figure 4.15: Kernel density plots of predicted output for phonetic incrementation of test tokens.

useful. Figures 4.13 and 4.14 present kernel density plots for the simulated PHL speaker's data and the simulated NAS speaker's data, respectively. Each of these plots comprise a kernel density plot of the test tokens as a whole (a) as well as a kernel density plot of each test conditioning factor (b). The target HAND class and CAT class tokens are also presented in each plot, as a benchmark for each simulated speaker's tense and lax targets. For maximal clarity, each conditioning factor in (b) is colored according to its expected realization under PHL: red for the LAUGH and MAD class, and blue for the MANAGE and HANG class. Figures 4.13a and 4.14a clearly display a bimodal distribution in the test tokens overall. Figures 4.13b and 4.14b display unimodal distributions for each conditioning factor, as either within the tense target or the lax target.

Compare this to Figure 4.15, which presents the kernel density plots for F1 and F2 for the simulated data of phonetic incrementation of the test tokens. Here, I use the variation of phonetic incrementation that results in only the test tokens incrementing, because the stability of the HAND and CAT classes in this variation makes it the most difficult to disambiguate this variation from the outputs of competing grammars and spontaneous phonologization. In Figure 4.15a, the unimodal distribution of all test tokens, produced intermediately between the tense and the lax phonetic targets, is clear. This unimodality also holds for each test condition (Figure 4.15b), in which each test condition also displays a unimodal distribution centered between the tense and the lax targets.

Finally, in Figure 4.16, I present the kernel density results for the predicted output from a Cohort 3 competing grammars speaker. This data comprises 50% PHL-consistent tokens and 50%



(a) Bimodal distribution of test tokens for competing grammars.

(b) Bimodal distribution of each test conditioning factor for competing grammars.

Figure 4.16: Kernel density plots of predicted output for competing grammars speaker.

NAS-consistent tokens. In Figure 4.16a, the bimodality of the test tokens as a whole is clear and reminiscent of the test kernel density plots for the PHL speaker in Figure 4.13a and the NAS speaker in Figure 4.14a. The test conditioning factors, on the other hand, provide a stark contrast to the unimodal distributions for each test condition found in the PHL speaker, the NAS speaker, and the phonetic incrementation speaker. Here, each test condition is bimodal, with 50% of each condition produced as tense and 50% produced as lax.

Kernel density plots for test conditions overall and each test conditioning factor are provided in Appendix D for each transitional cohort speaker analyzed in this chapter, along with that speaker's production plots.

4.4.2 Sociophonological Evidence

Where the statistical methods for determining underlying distribution of test tokens fall short of our goal, the sociophonological evidence provides additional important cues.

Central to a sociophonological account /æ/ is found in the details of speaker production of PHL described Labov (1989). In Labov (1989), PHL-dominant speakers were found to occasionally produce lax forms of traditionally tense PHL tokens, which I've termed *incongruent tokens*. Speakers' rates of inongruent tokens increased during the more formal components of the interview, with the highest rate (15%) found during the reading list. This behavior is in line with the finding of Labov (2001) that PHL speakers downgrade tense PHL tokens but not lax PHL tokens. Taken together, these two findings produce a prediction that speakers occassionally phonetically mitigate up to 15% of their tense tokens, particularly in more formal methods.

From the viewpoint of understanding speakers' systematic production of their language, both the style shifting and the targets of the style shifting provides some important pieces of information: first, that speakers are consistent within their casual style of speech, and second, that they phonetically mitigate a stigmatized tense form to lax in more formal settings, but do not produce traditionally lax forms as tense. From a sociophonological point of view, this provides an important backdrop to the data that follow. We may posit that any speaker producing a clearly tense token of LAUGH or MAD (the tense classes that could only be produced by an underlying PHL system) as well as a clearly tense token of MANAGE or HANG (the tense classes that could only be produced by an underlying NAS system) is exhibiting the operation of two systems within their speech. If such variation is found within the casual portion of the sociolinguistic interview, it can be given more analytic weight, since it is during the most casual speech that speakers behave most systematically. We can add to this the clear prediction drawn from a Competing Grammars framework: not only would we expect tense tokens from PHL and NAS, but we would also expect lax tokens from each system as well.

Here, I take a conservative approach and require a speaker to produce more than 15% incongruent tokens, either produced as lax or tense, in order to be classified as a speaker exhibiting competing grammars. The modified Magnitude Estimation task presented in Chapter 5 finds Philadelphians do not rate tense productions of NAS poorly; therefore, a tense NAS token cannot be used as evidence for an underlying NAS system in the same way that a tense PHL token can be used as evidence for an underlying PHL system. As a result, I simply set a limit of 15% incongruent tokens as a defining limit for a competing grammars speaker. I use this along with the kernel density plots to determine (1) whether the production of test tokens is unimodal and (2) if not, whether the production of test tokens is consistent with a competing grammars hypothesis.

4.5 Results

Here, I will first demonstrate my analyis of two individual speakers before turning to the community-wide pattern. In §4.5.1, I return to the data from Bobby Marx, who is analyzed as an underlyingly PHL speaker who produces fewer than 15% incongruent tokens. In §4.5.2, I analyze the production of "Orange Juice", a 25-year-old transitional cohort speaker from the IMPC data set who produces the highest token count for test tokens of /æ/ and whose production is most compatible with a competing grammars analysis of change. In §4.5.3, I present the results of analyzing every transitional cohort speaker (speakers from the IHELP and IMPC data sets born after 1983) who produce enough data (at least 5 tokens per test condition) to bear on the mechanism of phonological change.

4.5.1 PHL speaker: Bobby Marx

First, I return to the data from Bobby Marx, whose data is ultimately classified as a PHL-conforming speaker. As we've noted above, Bobby produces some incongruent tokens, both as PHL and as NAS. From a social evaluation perspective, the existence of tense tokens from both systems suggests the operation of both PHL and NAS. However, here I take a conservative approach to classifying speakers as competing systems speakers, and use instead the benchmark of whether more than 15% of his test tokens overall are incongruent.

In Figure 4.17, Bobby's test tokens are plotted as text, to enable a sociophonological analysis. His kernel density plots are provided in Figure 4.18a (the kernel density of his test tokens as a whole) and Figure 4.18b (the kernel density of each test conditioning factor). Bobby's kernel density plots for his HAND class (solid line) and CAT class (dotted line) are provided as benchmarks for his tense and lax targets as well. The relative sparsity of Bobby's data results in kernel density plots that are difficult to read, but the main takeaway from these plots is the apparent separation between the red conditions (LAUGH, MAD classes) and the blue conditions (MANAGE, HANG classes)

Figure 4.17 provides more insight into the incongruent tokens in Bobby's production. Several of the lax tokens (*alas, asterisk*) are straightforwardly lexical exceptions to lax for most PHL speakers. Likewise, we see that Bobby produces all tokens of *planet* as exceptionally tense forms (in

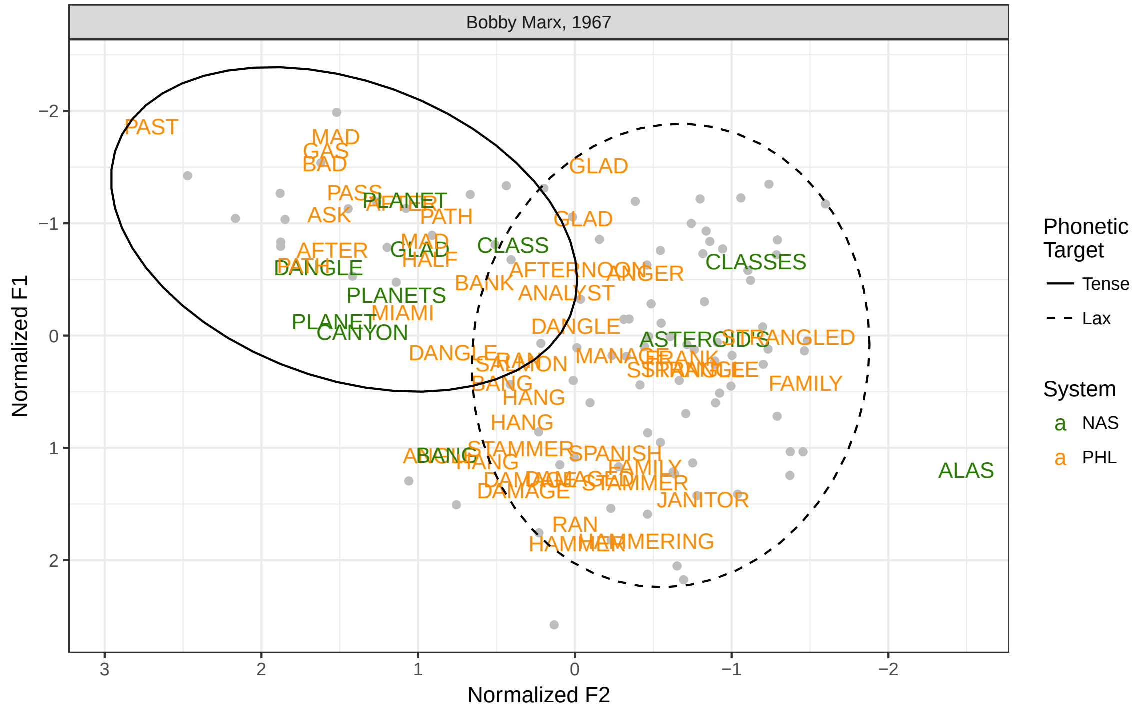
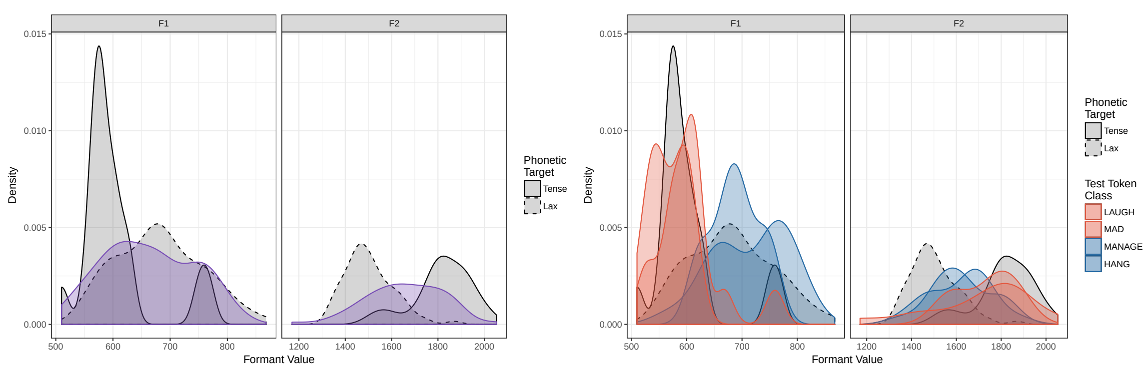


Figure 4.17: Bobby Marx test tokens by word identity.



(a) Bimodal kernel density distribution of test tokens overall.

(b) Kernel density distributions of each conditioning factor.

Figure 4.18: Bobby Marx kernel density plots.

contrast to the rest of his MANAGE class words, which he produces well within his lax target). Bobby represents one of the earliest speakers in our dataset who has adopted *planet* as an exceptionally tense form, in line with Brody (2011)'s findings. Because these tokens are canonical lexical exceptions, I do not count them towards Bobby's overall incongruent count.

Aside from these expected exceptions, Bobby also produces several more interesting tokens consistent with NAS. Both of his tokens of *class* are classified as lax. This appears to be a lexical exception that has been added to his lax list, or is perhaps a single lexical item that has been singled out for phonetic mitigation to lax. This leaves four tokens unaccounted for: *glad, canyon, dangle*, and *bang*. There are three possible explanations for this production. The first is that Bobby has added *dangle* and *canyon* to his exceptional tense list and *glad* and *bang* to his exceptional lax list. The second is that he is exhibiting an early stage of grammar competition between PHL and NAS, with NAS only produced a small percentage of the time. The third is that the classifier misclassified tokens or that these tokens were speech errors. In order to truly disambiguate between these three options, we would need more data from Bobby, including multiple tokens from each lemma.

For the purposes of classifying the data that exists, I count these tokens as incongruent. By token count, this means that Bobby produces six incongruent tokens *dangle, canyon, glad, class, classes, bang* out of 56 total test tokens, meaning that 10.7 % of his tokens as incongruent. As this falls below the threshold of 15%, I classify Bobby as a PHL speaker.

4.5.2 Competing Grammars speaker: Orange Juice

I turn next to the speaker in the IMPC data set with the highest token count, "Orange Juice". Orange Juice is a 25 year old women who graduated from an Open Admissions Catholic high school and the regionally-oriented Drexel University, produces the highest number of /æ/ tokens (894) and the highest number of test /æ/ tokens (331) of all the IMPC speakers.

On the surface, Orange Juice's data (Figure 4.19) is clearly most consistent with a competing grammars profile. Her kernel density plots (shown in Figure 4.20) provide support for this conclusion, particularly in the apparent bimodality in the test token conditions shown in 4.20a. Given what appears on the surface to be clear variation between PHL and NAS in a competing grammars

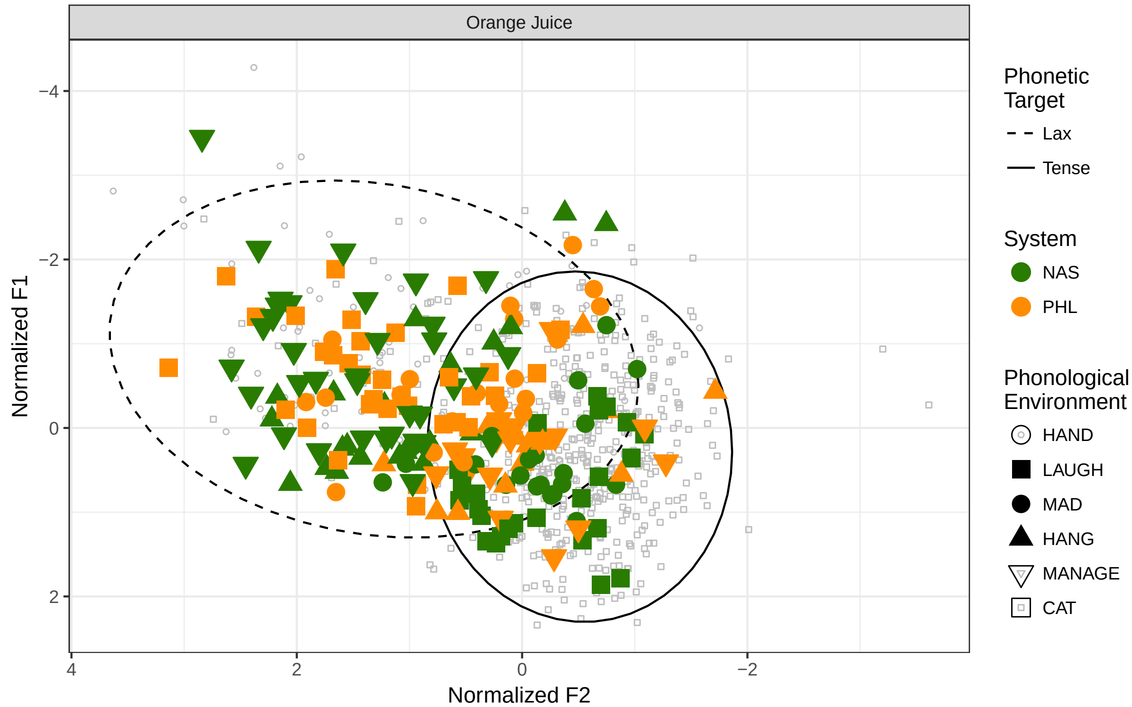
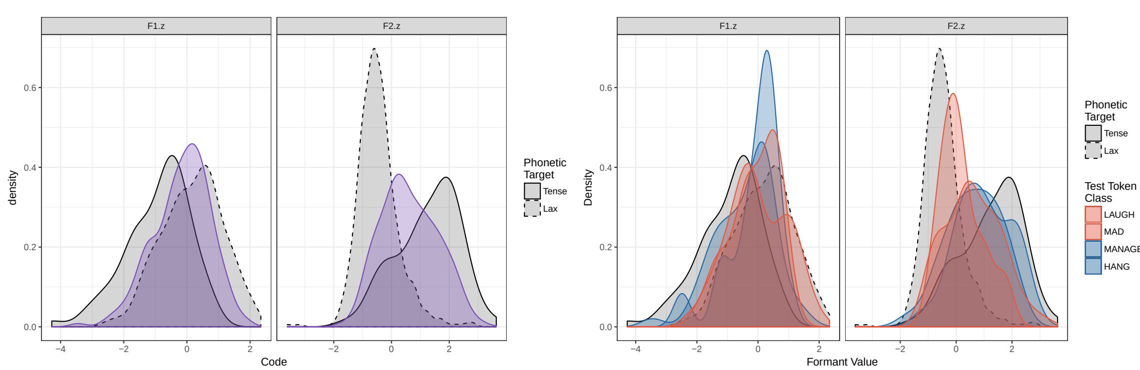


Figure 4.19: Orange Juice production of /æ/.



(a) Kernel density of Orange Juice's test tokens overall.

(b) Kernel density of Orange Juice's test conditioning factors.

Figure 4.20: Orange Juice kernel density plots.

type output, there are two alternative explanations for Orange Juice's distribution that must be ruled out before classifying her as a competing grammars speaker.

Disjunctive rules

Recall that PHL is comprised of a disjoint set of phonological triggers. While I have represented PHL as a single rule with disjoint triggers, it would also be possible to represent the traditional /æ/ system as two separate rules, shown in (21).

(21) a. **PHL$_1$:** $\text{æ} \rightarrow \text{æh}$ / $\_\left[\ \text{+anterior}\ \right] \cap \left[\ \text{+nasal}\ \right]]\ \sigma$

b. **PHL$_2$:** $\text{æ} \rightarrow \text{æh}$ / $\_\left[\ \text{+anterior}\ \right] \cap \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix}]\ \sigma$

If speakers represent the traditional input as two distinct rules rather than a single parameter as I have hypothesized, it is possible that the surface level variation found in Figure 4.19 is simply the result of Orange Juice discarding one of the two rules. If, for example, Orange Juice rejected PHL$_2$, she would produce tense /æ/ preceding anterior tautosyllabic nasals (HAND) and lax tokens elsewhere. This means that tokens preceding intervocalic nasals (MANAGE) and velar nasals (HANG) would be produced lax, appearing as surface-level PHL tokens. Orange Juice would also produce lax tokens preceding voiceless fricatives (LAUGH), which would appear as surface-level NAS tokens. This scenario can fairly quickly be ruled out by taking a closer look at Figure 4.19, in which we see instances of each test conditioning factor in both the tense and lax targets, as depicted by token shape.

Lexical diffusion

A final possibility that must be falsified before concluding that we have found competition between PHL and NAS is whether the surface-level variation is simply the result of lexical diffusion. Traditional PHL input requires speakers to memorize a fairy extensive list of lexical exceptions both as tense and as lax. We've also seen evidence that the specific lexical entries are subject to diachronic change, with *planet* joining the exceptionally tense class for many speakers born around 1990 (Brody, 2011) and various words leaving the exceptionally lax class (e.g. *ran, swam, began, and* for speakers born around 1985). This raises the distinct possibility that the variation within conditioning factors found in speakers may actually the result of lexical diffusion into or out of the list of lexical exceptions. For example, if Orange Juice produced PHL but added *janitor* to her list of exceptionally tense tokens, she would produce tense *janitor* and lax *manage*, resulting in what appears on the surface to be variation within the intervocalic nasal conditioning factor. If this she then also added *hang* to their exceptionally tense list and *class* to their exceptionally lax list, she would appear on the surface to produce variation within all conditioning factors that distinguish between PHL and NAS. If, on the other hand, Orange Juice's is the result of competing PHL and NAS, she is expected to produce variation between PHL and NAS within a each lemma.

Competing Grammars in Orange Juice

With her high token count, Orange Juice's data provides the best opportunity to test whether what looks like variation between PHL and NAS is the result of a selective adherence to only one of the PHL constraints, the result of lexical diffusion, or the result of competition between PHL and NAS.

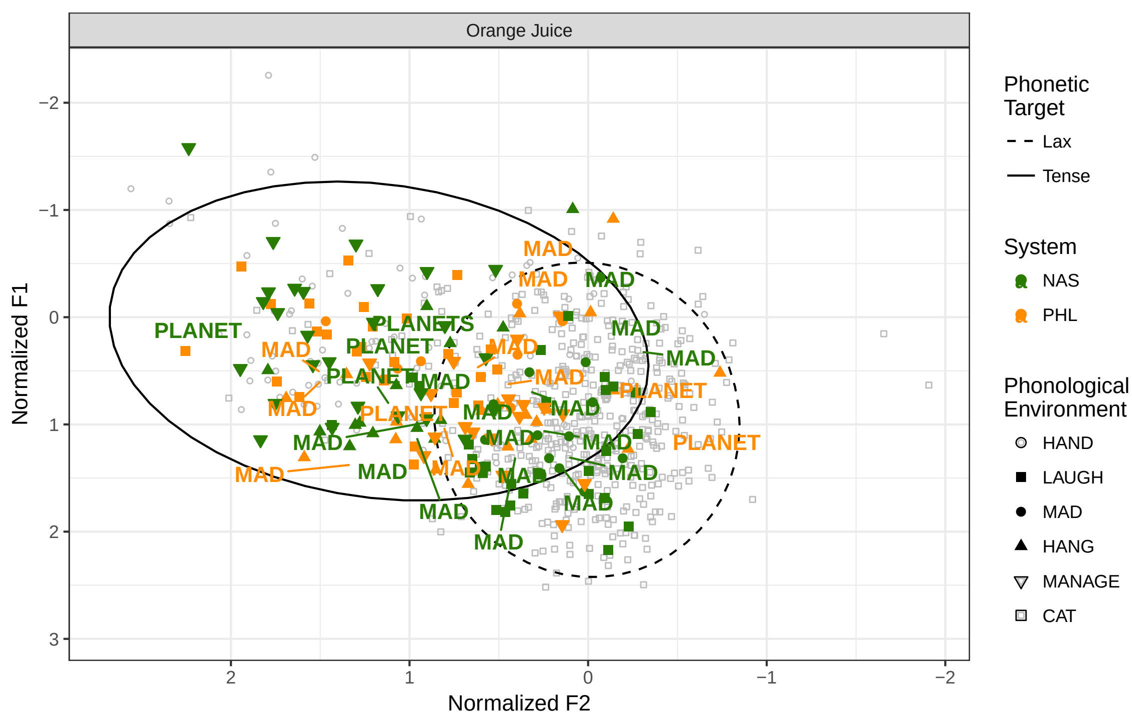


Figure 4.21: Within-lemma variation in the production of Orange Juice.

First, to the disjointed rules. We have already seen both tense and lax forms from each test conditioning factor in Figure 4.19 (see Appendix D for a full presentation of Orange Juice's data). The next possibility to rule out is variation through lexical diffusion. Figure 4.21 presents the wide variation in Orange Juice's production of two lexical items (*mad* and *planet*). Because Orange Juice exhibits variation in all test conditions, and because this variation is not driven by lexical diffusion, her data is best classified as an example of competing PHL and NAS within a single speaker.

Having found a clear example of competing grammars within a single speaker, I now turn to the data from the community as a whole to see whether this mechanism of change is the primary driving force for phonological change across the community.

4.5.3 Analysis of the community

Using the methods described above for Bobby Marx and Orange Juice, I classify each transitional cohort speaker in the IHELP and IMPC data sets who produce at least 5 tokens per test condition according to how their production matches up with the profiles of phonological change outlined above (47 speakers in total). Each speaker's data is also provided in Appendix D, which includes word identity and kernel density plots for each speaker. The majority of transitional cohort speakers conform either to PHL (7 speakers, Figure 4.22) or NAS (19 speakers, Figure 4.23).

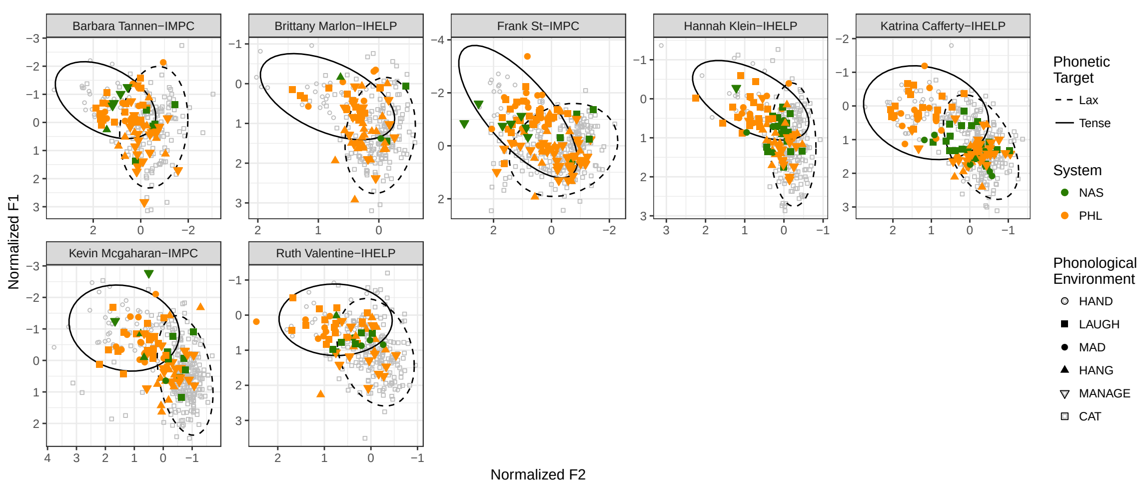


Figure 4.22: PHL speakers.

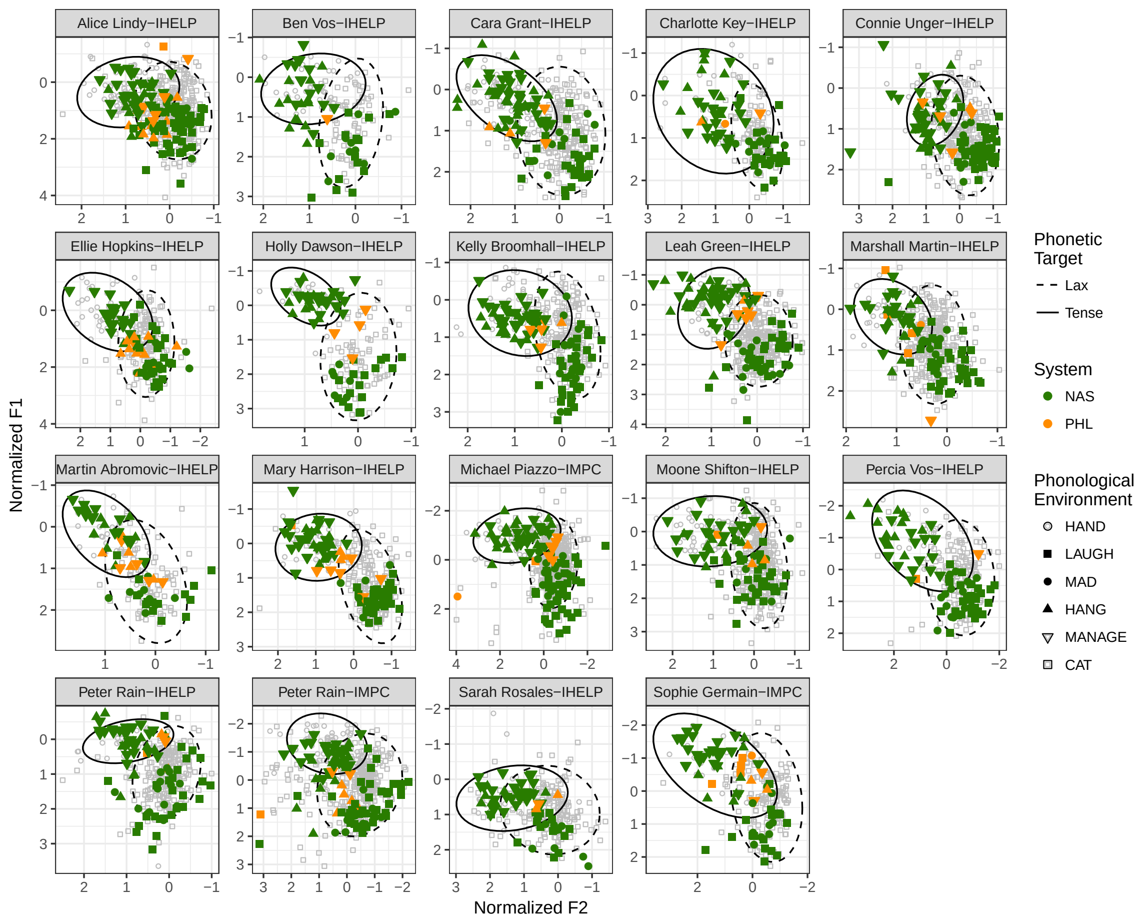


Figure 4.23: NAS speakers.

As discussed briefly in Chapter 1, finding speakers conforming primarily to PHL or NAS does not in itself provide support for any of the proposed mechanisms of phonological change. This data could be consistent with an analysis of some speakers driving phonological change via spontaneous phonologization, but could also simply be the result of the fragmentation of the speech community. A closer look at the educational history of each of these speakers reveals that their production is most likely the result of their school either not yet having undergone the change to NAS or having already completed the change. With the exception of Kevin, all of the speakers in Figure 4.22 attended an Open Admissions Catholic school, which as we have seen in Chapter 2 are conservative strongholds for PHL. Kevin emerges as somewhat of an outlier: he produces a PHL system despite having attended a Special Admissions non-Catholic high school. I note, however, that the neighborhood he grew up in (in South Philadelphia) is a stronghold for traditional PHL, and that he attended an Open Admissions Catholic middle school, providing an avenue for his acquisition and maintenance of PHL.

Similarly, most of the speakers classified as NAS speakers in Figure 4.23 were graduates of Special Admissions non-Catholic schools, with six exceptions. Five of these exceptions (Alice Lindy, Michael Piazzo, Moone Shifton, Peter Rain, and Sophie Germain) attended Special Admissions Catholic schools, which were identified in Chapter 2 as a segment of the schooling system that produces both NAS and PHL speakers. Here, these five speakers represent graduates of Special Admissions Catholic schools who have adopted the new NAS system. The sixth exception, Marshall Martin, attended an Open Admissions suburban public school, which we have previously identified in Chapter 2 as socially similar to the Special Admissions non-Catholic schools within the city. The educational histories of speakers classified as PHL or NAS in Figures 4.22 and 4.23 reveal that these speakers are not driving phonological change, but rather represent a segment of the population that either has not yet undergone the change from PHL to NAS or has already completed this change.

There are an additional fourteen speakers (Figure 4.24) who produce outputs most consistent with a competing grammars analysis of change. Seven of these speakers (Jacob Ambrose, Elizabeth Rina, Steve Rina, Ariana Tocci, Orange Juice, Speedy Racer, Wendy Juice) graduated from

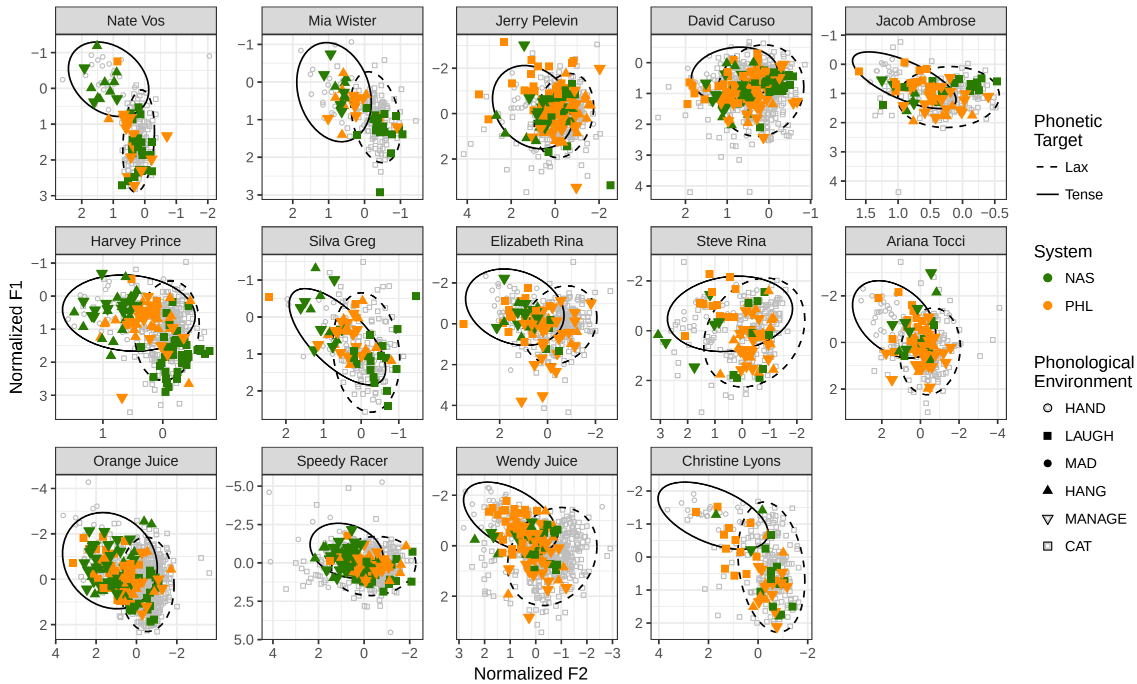


Figure 4.24: Competing grammars speakers.

Special Admissions Catholic schools, which is that section of the school system that has been already identified as the most vigorously changing in Philadelphia. Several of these speakers, Nate Vos and Christine Lyons, has already been discussed in some detail in Chapter 2. While Nate appears exceptions, as he attended Masterman middle school at the time of his interview, he had only been in that environment for a few months and had received mixed-system input from his father (Harry). Nate's production of a competing grammars output suggests that it is the result of his not yet settling on a NAS output. Christine is exceptional for the opposite reason: she is the graduate of an Open Admissions Catholic high school who nevertheless produces a competing grammars output. Recall, however, that Christine also attended Penn – a nationally oriented Ivy-league university – which is likely to have had an impact on her production. There are a few other speakers whose productions are exceptionally conservative or innovative given their educational background. Mia Wister, David Caruso, Jerry Pelevan, Silva Greg, and Harvey Prince were graduates of Special Admissions non-Catholic high schools. Jerry and Harvey, however, also attended Open Admissions Catholic middle schools. Likewise, Jerry and Silva attended the locally-oriented Temple University for college. The influence of traditional middle schools and universities for these speakers may provide some insight into their ability to retain a competing grammars system. Mia Wister and David Caruso emerge as surprising conservative exceptions. Both Mia and David attended Special Admissions non-Catholic middle and high schools, before attending Penn for college.

In the detailed examination of speakers' educational histories provided above, we have found that Open Admissions Catholic schools largely produce PHL-conforming speakers while Special Admissions non-Catholic schools largely produce NAS-conforming speakers. Special Admissions Catholic schools are the main schools that produce competing grammars speakers. This is unsurprising, given the results in Chapter 2: this is the segment of the broader speech community that bridges the Special Admissions non-Catholic schools and the Open Admissions Catholic schools and is most likely to have contact with enough NAS and PHL speakers to adopt both systems (I argue in Chapter 6 that a competing grammars speaker requires between roughly 46% and 54% input from both systems during acquisition in order to acquire both as plausible productive rules). We see a few speakers (Mia, David) who trail their peers in the adoption of NAS, as well as one speaker (Christine) who is on the vanguard of this change given her educational history. Overall, the educational histories and production of speakers in Figures 4.22 through 4.24 provide a compelling argument that the change from PHL to NAS is occurring via the mechanism of competing grammars.

4.5.4 Unclassified Speakers

There are a few speakers, all found in the IHELP data set, whose data is not easily classified by the combination of statistical and sociophonological methods that I employ here. Six of these speakers, shown in Figure 4.25, display overall conformity to either NAS or PHL in the non-overlapping parts of their tense and lax targets. For these speakers, the proportion of incongruent tokens exceeds the 15% threshold while they also produce bimodal distributions of their test tokens. Under the classification system I outline above, these speakers would be classified as competing grammars speakers. However, due to the distribution of their tokens, I remain skeptical of analyzing them as competing grammars speakers.

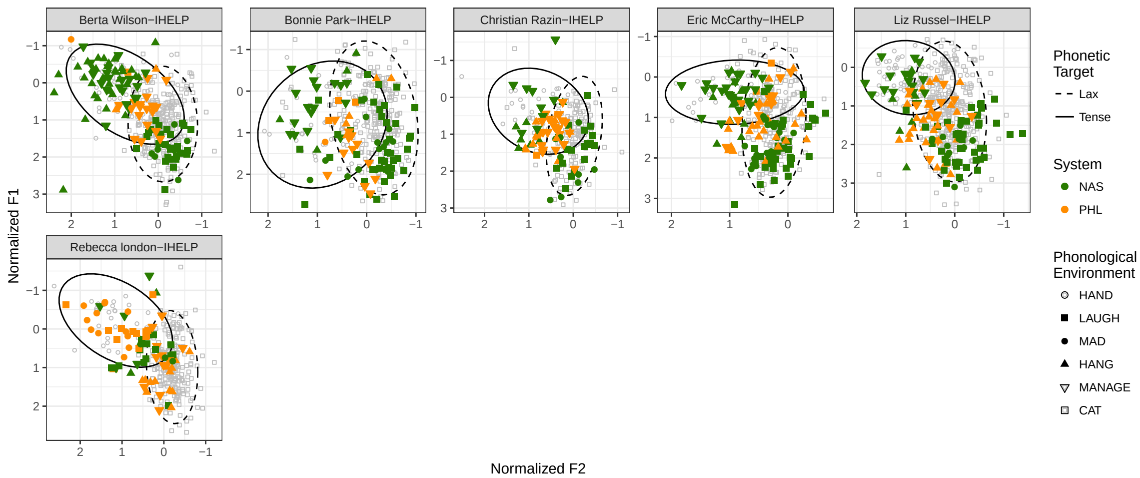


Figure 4.25: Largely PHL or NAS speakers who nevertheless produce more than 15% incongruent tokens.

In addition to the six speakers analyzed above, we find one speakers whose production is consistent with the profile of phonetic incrementation of the test tokens: Jake Stone. Jake's production was already discussed as a community outlier in Chapter 2. As previously discussed, Jake's sociolinguistic background suggests that he is not a transitional cohort speaker producing phonetic incrementation, but rather is simply phonetically mitigating the stigmatized tense production of his underlying PHL system. Jake attended Masterman high school at a time when his peers already demonstrate conformity to NAS, meaning that his phonetically mitigated production of tense PHL tokens is not a likely driver of this phonological change. As in Chapter 2, I analyze Jake as having an underlying PHL system, but phonetically mitigating his traditionally tense tokens to lax, in a sociolinguistic avoidance of a stigmatized form.

Finally, I turn to the last unclassified speaker, Carlos Santana. Carlos' data largely conforms to the NAS system, though he has a few PHL-congruent tokens. He shows a few non-extreme tokens of traditional tense PHL forms, *path, gas, mad, after.* The rest of his PHL-consistent tokens are lax productions of MANAGE and HANG class words. Carlos' production is not consistent with any of

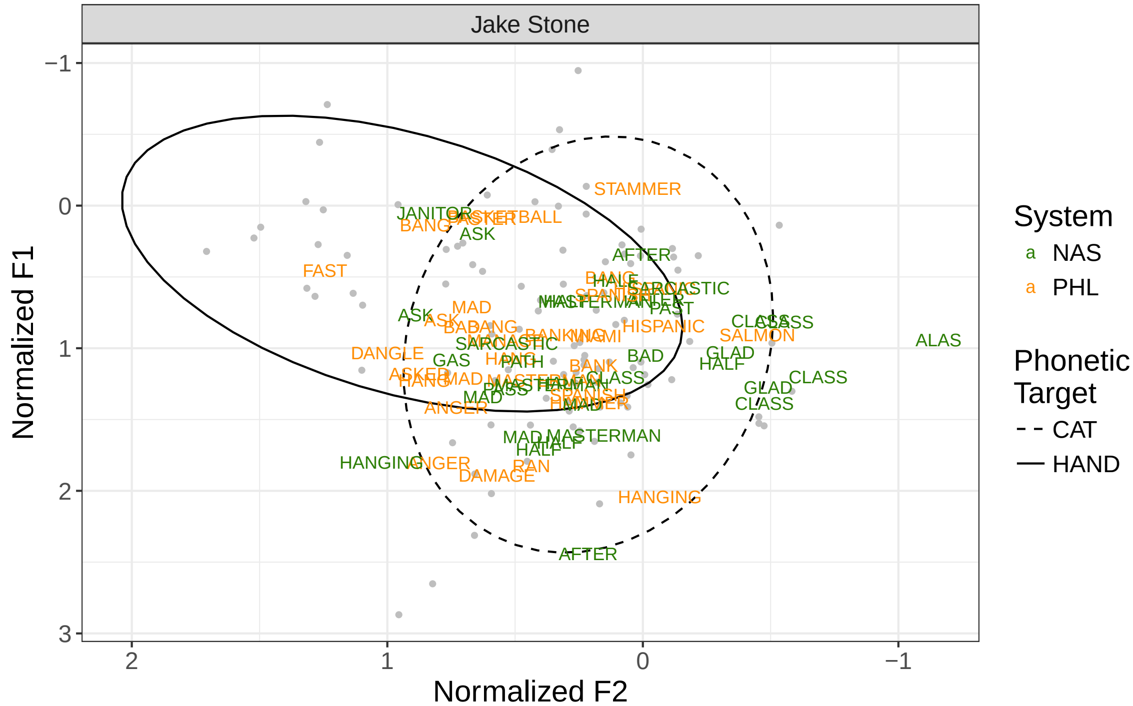


Figure 4.26: Phonetic mitigation in Jake S.

the profiles of phonological change outlined above. He does not adhere to NAS or PHL entirely, as a mechanism of spontaneous phonologization predicts. He also does not produce clearly tense and lax tokens of each conditioning factor, as a mechanism of competing grammars predicts. Carlos' production also does not fit a model of phonetic incrementation, with tokens produced clearly within and even more extreme than his underlying phonetic tense and lax targets.

So what can be said about Carlos and his unexpected production? One potential answer to Carlos' data is that he produces a combination of grammar competition and phonetic mitigation. As I will argue in Chapter 5, younger speakers in Philadelphia acquire two evaluation systems for /æ/ tokens. Any token consistent with NAS is rated highly, as is any lax token of PHL. PHL-specific tense tokens, on the other hand, are rated poorly by younger speakers. If Carlos underlying produced competition between PHL and NAS, then added a filter of phonetically mitigating his PHL-tense tokens (LAUGH and MAD class tokens), the expected output would be as in Figure 4.27.

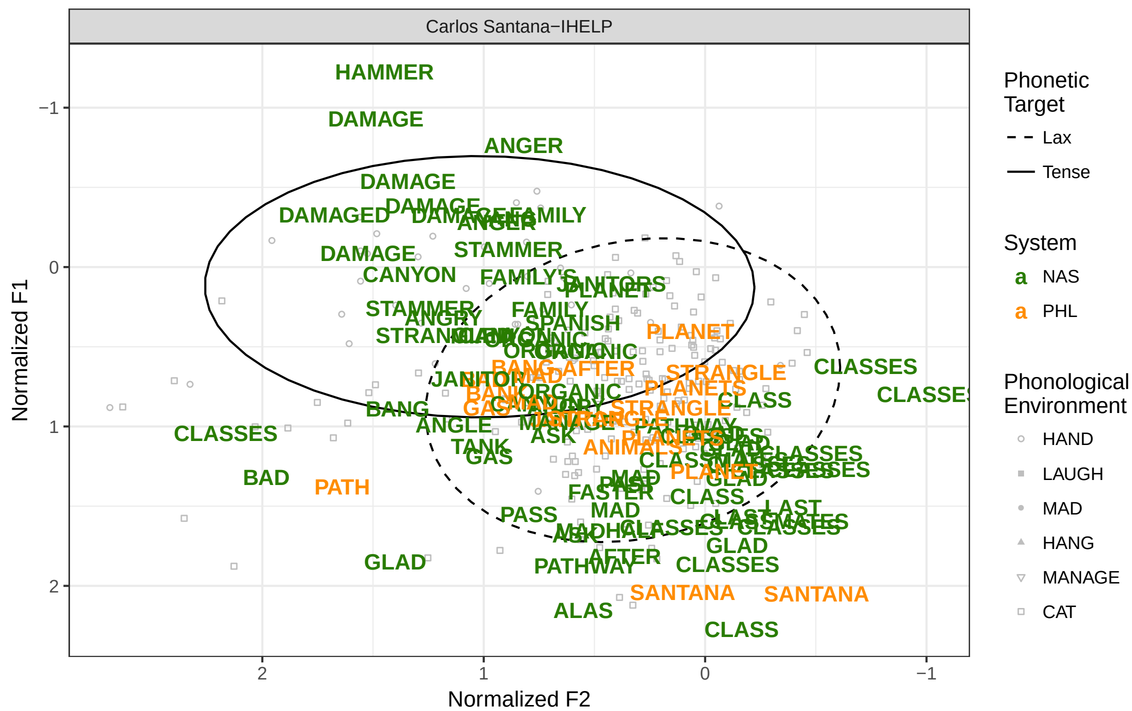


Figure 4.27: Phonetic mitigation of tense PHL tokens in a mixed-system speaker (Carlos Santana).

4.6 Discussion

Throughout this chapter, I've appealed to a number of pieces of evidence to identify what the individual speakers in the IMPC and IHELP data sets are doing. I find that most speakers adhere to either PHL or NAS, but that a sizeable number of speakers also produce variation between the two systems. By closely analyzing the variable speakers with the highest token count, we find that all test conditioning factors exhibit both PHL and NAS productions, and furthermore that variation is even found within the lexical exception lists. This suggests two important outcomes.

First, that grammar competition between PHL and NAS is the mechanism by which this change is occurring, at least in the transitional cohort speakers analyzed here. This is competition between two outcomes of a single "allophonic system" parameter that selects between PHL and NAS each time the speaker goes to produce a token of /æ/. This parameter governs a number of conditioning factors, which all exhibit variability. If our data included longitudinal data from transitional cohort speakers, we would expect these distinct parameters to exhibit a constant rate of change, in accordance with the Constant Rate Hypothesis (Kroch, 1989), also termed the Unity Principle when used to refer to phonological constant rate (Fruehwald, 2013).

Second, we find that the lexical exceptions exhibit variability at roughly the same rates as the phonological conditioning factors. This suggests that the sets of lexical exceptions are cognitively stored under the same single parameter rather than externally to that parameter. It appears that speakers are not producing lexical exceptions and then applying a variable rule which selects whether the non-exceptional forms are produced under PHL or NAS. Instead, we find that the variable rule selects each system wholesale including its lexical exceptions. This finding has important consequences for the structure of the grammar, meaning that any lexical specificity to a phonological process is stored as a component of the phonology itself. This is, in fact, in line with the predictions of the Tolerance Principle (Yang, 2016). The concept behind the Tolerance Principle is that it is a calculation of whether it is more efficient to memorize all lexical items or some lexical items as well as a rule. This is based on the premise that under a rule scenario, all lexical exceptions are processed *before* the rule, in order of lexical frequency. In the strongest formulation of this theory, this is proposed to be an actual model of word production, whereby speakers run through first the list of lexical exceptions then the productive rule any time they go to produce a word.

The data presented in this chapter provide surprising support for this model. If it is the case that lexical exceptions must be serially processed before a regular rule, it follows that that speaker's cognitive representation of that rule includes the lexical exceptions as well. This means that, phonologically, however we represent productive processes must include the possibility of storing lexical exceptions as part of that process as well. Here, I have focused on the most frequent lexical exceptions *mad, bad, glad*, which have also been amplified by the interview methods. As outlined in Chapter 3, I interpret all lexical exceptions to the productive rule in PHL to be stored as a series of exceptions. Given the results reported in this chapter, I would expect *all* instances of lexical exceptions to the productive PHL rule for a given speaker to exhibit the same proportions of variability as do *mad, bad, glad*, since these additional exceptions would also be stored in a similar list preceding the productive rule.

(22) **PHL:**

1. IF w = *and* THEN /æ/ → lax

2. IF w = *can* THEN /æ/ → lax

...

39. IF w = *gaffe* THEN /æ/ → lax

40. $\text{æ} \rightarrow \text{æh} \;/\; \_[\ +\text{anterior}\ ] \cap ([\ +\text{nasal}\ ] \cup \begin{bmatrix} \text{-voice} \\ +\text{fricative} \end{bmatrix})]\ \sigma$

(23) **NAS:**

1. $\text{æ} \rightarrow \text{æh} \;/\; \_[\ +\text{nasal}\ ]$

The addition of lexical exceptions to the rule itself requires a slight modification of the notation for PHL, which is argued for in Chapter 3 and reproduced in (22). Here, PHL is comprised first of the lexical exceptions, followed by the regular productive rule. NAS, then, as any other productive phonological rule, also carries the potential for its own list of lexical exceptions, as shown in (23). For the majority of NAS speakers, these potential lexical exceptions list will remain empty, though we do find some speakers in the IHELP corpus whose primary production is NAS but who retain a lexical exception (e.g., in the speaker's own name *Hannah*). In such a case, this speaker's formulation of NAS would be as in (24).

(24) **NAS:**

1. IF w = *Hannah* THEN /æ/ → lax

2. $\text{æ} \rightarrow \text{æh} \;/\; \_[\ +\text{nasal}\ ]$

In other words, our *Hannah* lexical exception speaker is not producing variation between NAS and PHL (using PHL only and every time she says *Hannah*), but rather producing a single system NAS that has a lexical exception. When PHL and NAS are in competition within a single speaker, as found in the competing grammars speakers in Figure 4.24, it is the entirety of (22) and (23) that is in variation. This means that a competing grammars speaker produces tense *mad* ([meːᵊd]) due to the lexical exceptions listed in (22) when PHL is selected, but lax *mad* ([mæd]) when NAS is selected, accounting for the variation found in Orange Juice's production of *mad*.

Here, we have found data suggesting that not only is the change from PHL to NAS in Philadelphia driven by the mechanism of competing grammars, but also that what we consider to be an allophonic rule or an allophonic system is best represented as a single unit containing the possibility for lexical exceptions, as in (22). Finally, it is worth reiterating the major point drawn on in this chapter that an identification of individual speaker's production is not fully possible without an understanding of the sociolinguistic facts of the speech community as a whole.

Chapter 5

The Social Evaluation of Abstract Phonological Structure

Given the robust evidence that speakers are producing variation between the abstract parameter of PHL and the abstract parameter of NAS demonstrated in Chapter 4 and the community-wide social patterning of this change outlined in Chapter 2, it follows that the allophonic restructuring of /æ/ in Philadelphia may also attract social evaluation. While social evaluation and the social motivation for sound change have been at the heart of sociolinguistic inquiry since Labov (1963), the ability of abstract phonological structure to be the target of social evaluation has been contested (see, e.g., Labov, 1993; Eckert and Labov, 2017). Because the allophonic restructuring of /æ/ in Philadelphia is a socially stratified abstract phonological change, it provides an important opportunity to test the social evaluation of phonological structure.

In this chapter, I present two experiments conducted prior to the projects reported in Chapters 2 and 4, which were designed to test the social evaluation of abstract structure. In §5.2, I present a matched-guise experiment designed to test the overall implicit social evaluation of PHL and NAS, finding that Philadelphian participants can in fact identify PHL and NAS along a scale of *accentedness.* In §5.3, this is followed by a magnitude estimation task which obtains participants' explicit evaluation of the six primary conditioning factors between PHL and NAS. I find that participants produce surprisingly systematic evaluations of these allophonic systems, with younger speakers rating NAS highly and tense PHL tokens poorly and older speakers evaluating the conditioning factors rather than their phonetic realizations. These findings suggest that abstract phonological structure is targeted for social evaluation in this change from PHL to NAS in Philadelphia English.

5.1 The Unnobservability of Structure

Speakers' ability to identify and furthermore evaluate structural variables such as a phonological rule is not well established in the literature. Labov (1993) argues that linguistic structure is unobservable, and that it is instead the phonetic output that is subject to social evaluation by listeners. This is not conceived of as *purely* phonetic output, but rather as the phonetic implementation of a surface phonological form, as in the tense production of an /æ/ allophone. Eckert and Labov (2017) point out, for example, that a production of [eː^{ə}] is not negatively evaluated when it appears in the word *yeah*, but is negatively evaluated as the phonetic output of the tense PHL rule. Additional work (Campbell-Kibler, 2007; Dinkin, 2015) carries this argument further with evidence that listeners attach social meaning to a variant itself (such as the use of "like" across the different variables of quotatives and discourse markers), regardless of the structural composition of the variable.

Eckert and Labov (2017) make the question of the evaluation of phonological structure explicit: "what kinds of phonological structures take on social meaning?" Eckert and Labov (2017) argue that while phonological variables are well suited for relaying social meaning, given that phonological variation rarely has referential meaning and is therefore maximally available for indexical meaning, the abstract structures governing relations between phonological entities is not well suited for this task. They go on to examine the case of phonological mergers, which occasionally attract social meaning, as in the case of the PIN-PEN merger in Northern California which is associated in production with an 'outdoorsy' lifestyle (Geenberg, 2014). Despite clear social associations being given to structural mergers, (Eckert and Labov, 2017, pg. 482) go on to discuss the lack of structural *commentary* about structural changes: "the merger of /i/ and /e/ before nasals is more likely to be noted as 'He says *pin* for *pen*' than 'He says *pin* and *pen*' the same." This focus on lexical items or specific pronunciation of lexical items is taken as evidence that the structure of the merger is invisible to speakers.

While there is clear evidence from nearly every speaker interviewed in the PNC who provides metalinguistic commentary about language that their evaluation is attached to the phonetic form rather than the phonological structure, it does not necessarily follow that the phonological structure does not attract implicit social evaluation. The evaluation given to the PIN-PEN merger in California is one example of a case where listeners do provide social evaluation of a structural feature, even if they are not themselves aware of the structural component to their evaluation.

The phonological restructuring of /æ/ in Philadelphia provides a useful case study for investigating the observability of structure. As we have seen in Chapters 2 and 4, the tense and lax phonetic targets of a PHL speaker and a NAS speaker are almost identical. If listeners evaluate only the phonetic form of an allophone rather than the abstract structure of it, this predicts that listeners will provide a similar evaluation to a PHL and NAS speaker whose phonetic targets are similar. In this chapter, I present two experiments designed to test different aspects of the social evaluation of PHL and NAS. In Experiment 1 (§5.2), I employ a Matched Guise technique to test for the overall social evaluation of PHL and NAS, finding that listeners do identify a PHL speaker as more *accented* than a NAS speaker. In Experiment 2 (§5.3), I take a closer look into how listeners evaluate the different conditioning factors that make up PHL and NAS, finding that listeners' explicit acceptability scores are best described along structural, rather than phonetic, lines.

5.2 Experiment 1: Matched Guise

Since its development by Lambert et al. (1960) (see also, Anisfeld et al., 1962; Lambert et al., 1965; Lambert, 1967), the Matched Guise technique has been a widely used tool for obtaining implicit attitudes towards language. The basic concept of a Matched Guise experiment is to provide participants with two (or more) recordings. The participants do not know that the two samples of speech are from the same person, and are asked to judge the speaker of each recording along a number of social dimensions. As outlined in Gaies and Beebe (1991), Matched Guise tasks have two main purposes:

1. to elicit reactions to particular features indirectly, rather than having participants express

opinions about the features themselves

2. to control all variables other than the features in question

The Matched Guise technique has been applied to a wide range of sociolinguistic features, including obtaining participant attitudes toward specific languages in multilingual settings (see, e.g. Edwards, 1983; Lambert et al., 1965; Wölck, 1973; Gibbons, 1983), dialectal differences (Strongman and Woosley, 1967; Giles et al., 1992a; Elwell et al., 1984; Ohama et al., 2000; Arthur et al., 1974; Cargile, 1997), and has been particularly useful in obtaining attitude reactions to raciolinguistic dialects (Purnell et al., 1999). In addition to linguistics, social scientists have used the Matched Guise approach to investigate participant evaluation of visual cues (Elwell et al., 1984), including race (Dixon et al., 2002; Rubin and Smith, 1990), and age (Giles et al., 1992a).

Sociolinguists have also used the Matched Guise technique to investigate the social evaluation of more fine-grained linguistic features, such as speech rate and pitch variation (Addington, 1968; Brown et al., 1985; Ray et al., 1991; Giles et al., 1992b; Apple et al., 1979; Ray and Zahn, 1999). The ability to synthetically manipulate a recording has also made it possible to investigate listener attitudes towards specific features: these features can be manipulated within a single recording, mitigating the potential effect of phonetic differences in instances recorded.

As a first step towards investigating whether listeners evaluate the abstract organization of PHL distinctly from the abstract organization of NAS, a Matched Guise task provides a controlled way to elicit listeners' implicit evaluations. It is particularly important to investigate implicit social evaluation, given that the evidence drawn on in Eckert and Labov (2017) is primarily explicit in nature. Here, instead of asking whether participants comment on abstract structure, we rely on differences in social evaluations of a matched guise experiment as evidence of listeners' ability to evaluate abstract structure.

5.2.1 Participants

Participants were recruited through social media. Demographic data, including age, gender, race, and childhood zip code was collected. Only participants who reported living in a Philadelphia-area zip code between the ages of 1-18 were considered, resulting in a total of 52 participants.

Because the change in /æ/ occurred around 1983 in the community, participants born before this year were considered "older" and participants born after this year were considered "younger". The data consisted of responses from 17 older and 35 younger participants.

5.2.2 Methods

Stimuli

Previous treatments of the Matched Guise technique have highlighted that task effects may play an important role in participants' responses. Specifically, read passages differ from spontaneous speech in their prosody (Fowler, 1988; Blaauw, 1994), speech rate (Kowal et al., 1975), pause quantity and quality (Kowal et al., 1975; Guaitella, 1999), and tone boundaries (Howell and Kadi-Hanifi, 1991). These linguistic differences translate into differences in participant behavior: Smith and Baley (1980) demonstrate that the difference in speech activity (whether it was read or spoken spontaneously) influences speaker perceptions. Furthermore, recent research on the effects on nonstandard speech in experimental settings (e.g., Perry et al., 2017) reveal that nonstandard speech may be processed differently based on participant expectations. Because reading is a task associated with education, providing participants with one supraregional standard guise (in the form of NAS) and one local nonstandard guise (in the form of PHL) in read form is likely to introduce a potential task mismatch effect. In other words, participants may rate the PHL guise more harshly because it is seen as an unacceptable way to *read* rather than an unacceptable way to *speak*. Furthermore, the primary interest at hand is whether PHL and NAS receive distinct social evaluations in everyday interactions (not in read speech).

However, as any researcher who has attempted to use natural sociolinguistic interview data in an experiment can attest, finding passages from naturalistic sociolinguistic interviews that can be used for experimental purposes is a difficult feat. Many interviews are conducted in noisy settings, making acoustic manipulation very difficult and unnatural sounding. In addition, the researcher must find a section of the recording that contains the appropriate number and phonological conditions of the variable under investigation. For very frequent features, such as *ing-in* variation or t/d deletion, this may be possible. As highlighted in Chapter 4, however, test tokens of /æ/ occur relatively infrequently in natural speech.

With a goal of including natural-sounding oral narrative stimuli that can be easily acoustically manipulated and also includes the right proportions of test /æ/ tokens, I adapt an oral narrative found in a sociolinguistic interview from the IHELP corpus. Because its baseline was an oral narrative, the story maintains a cadence of spoken – not read – speech. The narrative was modified to include more test /æ/ tokens, with special care towards ensuring that the PHL guise and the NAS guise each contained 9 tense tokens and 15 lax tokens. A trained phonetician read the story twice: once with all tense /æ/ tokens and once with all lax /æ/ tokens. To ensure that listeners did not obtain external social cues independent of /æ/ realization, both the PHL guise and the NAS guise used the same baseline recording of the story. All /æ/ tokens were spliced into this baseline story, meaning that all test tokens for both guises were comprised of spliced /æ/. The text for both guises is provided below. Tokens that would be tense under PHL are in bold, tokens that would be lax under PHL are in italics, and tokens that would be tense under NAS are underlined.

> I got in a lot of trouble that night. And I didn't do anything wrong! Okay.
>
> There was a big blizzard, and we didn't have ***class***, so we all went down to Jake's to *hang* out there and play in the snow.
>
> My mom was like "Don't bring your phone out", because I had just gotten a ***brand*** new phone. So she was like "Don't bring it, because if you *manage* to ruin it, your dad's not gonna be happy."
>
> So I left it at Jake's house because I didn't wanna *damage* it.
>
> So we were *hanging* out in the snow all day. He has like a little *canyon* behind his house that we were sledding in and stuff. So this ***lasted*** for like hours.
>
> We got back to Jake's house ***after*** that, changed because our ***pants*** were all snowy, and went out again.
>
> I get home that night, and I find out that my parents had called my cell phone like a hundred times, and it was this whole big thing. So I called her back and she started

going <u>*bananas*</u> on me. I started ***laughing***, like "You told me not to bring my phone out!"

And then she got really <u>*angry*</u> that she hadn't heard from me all day. It was pretty ***bad***. And then supposedly I was grounded, but that ***lasted*** like a day because she doesn't stay ***mad*** at me for very long.

Task

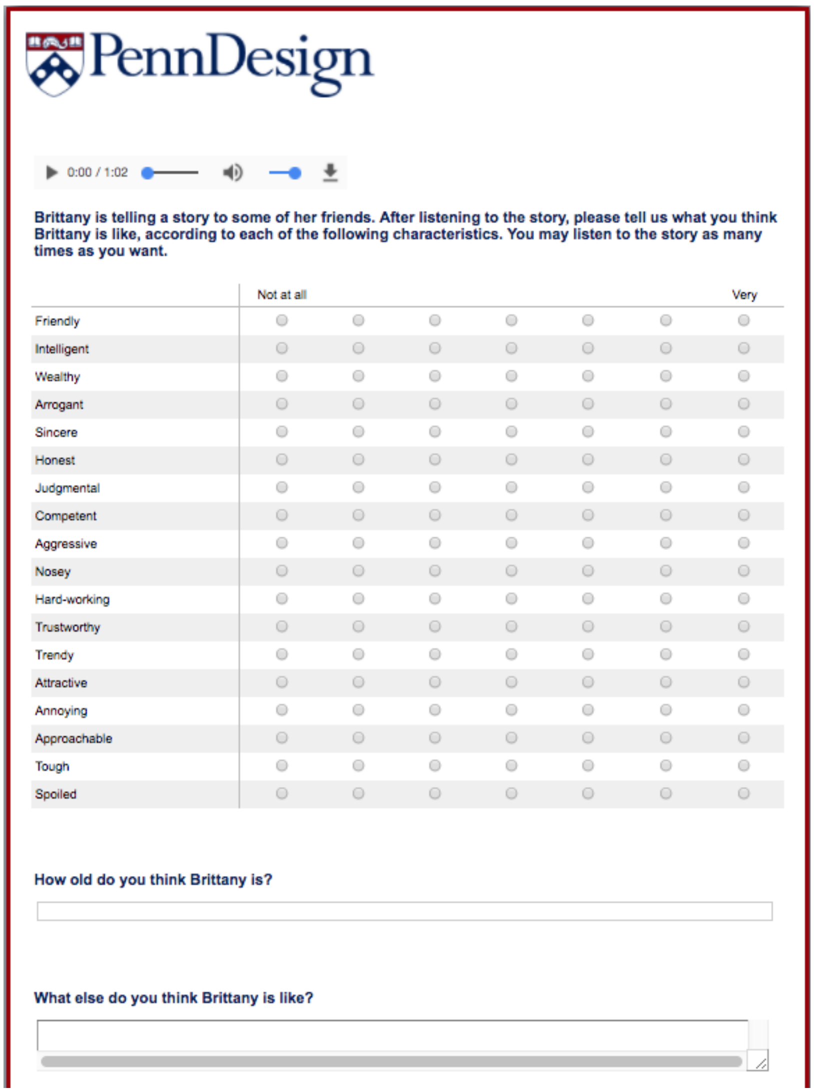
Participants each heard only one guise (either PHL or NAS), and were asked to rate the speaker on a number of social dimensions based on what they had heard using a Likert scale, as shown in Figure 5.1.

Participants were able to listen to the story as many times as they liked. The social attributes selected for the Likert scale were chosen to match the broad social characteristics reported Campbell-Kibler (2007). While Campbell-Kibler (2007) ran several pilot studies to determine the most relevant social characteristics for her subjects, here I adopt the reported list of social characteristics as a broad insight into the social evaluation of the phonological structure of PHL vs. NAS. Future work may investigate a more nuanced set of social characteristics, but this is beyond the scope of the current dissertation.

In addition to the Likert scale ratings for social characteristics, participants were also provided with a free-form response box asking "How old do you think Brittany is" and a second free-form response prompting participants for additional reactions (see Figure 5.1).

5.2.3 Analysis and Results

Participant ratings were analyzed using ANOVA, with story guise as the first independent variable. Because 21 attributes were tested for, resulting p-values were Bonferroni corrected. Because the changing /æ/ system in Philadelphia is a change in progress, and because NAS is most prevalent in elite circles for younger speakers, we anticipate that participant age will be an important factor in participant ratings. Specifically, a speaker growing up before the advent of NAS in elite schools will be expected to have a different overall rating of the NAS guise than a younger speaker, for whom



PennDesign

0:00 / 1:02

Brittany is telling a story to some of her friends. After listening to the story, please tell us what you think Brittany is like, according to each of the following characteristics. You may listen to the story as many times as you want.

	Not at all						Very
Friendly							
Intelligent							
Wealthy							
Arrogant							
Sincere							
Honest							
Judgmental							
Competent							
Aggressive							
Nosey							
Hard-working							
Trustworthy							
Trendy							
Attractive							
Annoying							
Approachable							
Tough							
Spoiled							

How old do you think Brittany is?

What else do you think Brittany is like?

Figure 5.1: Screen shot of Matched Guise Task.

NAS may serve as a strong indicator of social class or social mobility. In the figures presented here, responses are binned by age of participant, with a date of birth of 1983 as the break point. Because 1983 was the changepoint in the speech community (Chapter 2), where NAS began to emerge in the production of Philadelphian speakers, this presents a sociophonological argument for binning participant age by this date. It is expected that on average, speakers born before this date acquired language in a PHL-only environment while speakers born after this date acquired language in a radically different environment which included two allophonic /æ/ systems as the input.

5.2.4 Results

For the majority of attributes, /æ/ system did not have a significant effect. I include a brief plot of these non-significant attributes in Figure 5.2, which provides some insight into the overall social evaluation of the speaker (regardless of /æ/ guise). Immediately apparent is the effect of story context: this is a narrative about a speakers' parents not grounding her, and we see she is somewhat unsurprisingly rated high on *spoiled.* This young sounding female voice also is rated as *approachable, friendly, sincere, trendy,* and *wealthy*. She is not considered *hard working, aggressive* or *shy*.

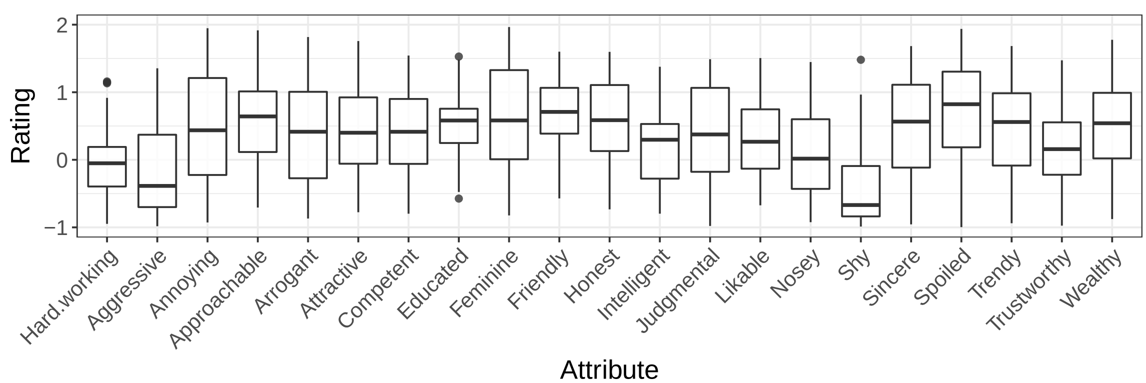


Figure 5.2: Non-significant attributes from the Matched Guise task.

Against the overall social characteristics attributed to the speaker, there is a single trait that is affected by story guise: *accented.* This result is shown in Table 5.1.

		Estimate	Std. Error	t value	Pr(> |t|)
Accented					
	Story (PHL)	0.88	0.26	3.38	**0.02***
	Age (younger)	-0.52	0.39	-1.3	.99
	Story:Age	-0.52	0.57	-0.89	.99

Table 5.1: ANOVA results for *Accented*; $p-$value presents Bonferroni correction.

5.2.5 PHL is rated as more Accented than NAS

As shown in Table 5.1, a PHL guise has a strongest effect on the standardized coefficient for *accented* ratings, with an estimate of 0.88. This serves as an important sanity check on the sociolinguistic awareness of the participants. Unlike a supraregional standard like the NAS system, Philadelphia English is a nonstandard regional dialect, and is interpreted and often maligned by the general public as an "accent". Philadelphia English was included as a contestant in Gawker's 2014 "America's Ugliest Accent" competition (Evans, 2014), and dozens of sociolinguistic interviews in the Philadelphia Neighborhod Corpus contain metalinguistic commentary by Philadelphians about the Philadelphia accent. It is not surprising, therefore, that Philadelphian participants from both age groups rate the PHL guise as more *accented*.

That Philadelphians of both age groups rate the PHL guise as more *accented* speaks to their ability to detect linguistic variation. However, it does not necessarily follow that an identification of linguistic variation equates to social evaluation of that variation. We may expect, for instance, that a Philadelphian aware of the social patterning of PHL and NAS across school systems may rate a NAS guise as more *wealthy* or more *educated*, and a PHL guise as adjectives that align with social evaluation of working class speakers, such as *aggressive* or *hard working*. The lack of social adjectives assigned to the PHL or NAS guise suggests that this change has not attracted overall social meaning. However, as we have seen, listeners are still able to identify PHL as sounding more *accented*; we may then turn to the question of how listeners rate the six main conditioning factors governing the allophony of /æ/. For this, we turn to a Magnitude Estimation task.

5.3 Experiment 2: Magnitude Estimation

While Experiment 1 demonstrated that PHL and NAS are identifiably different, the next step is to ask exactly how the six main conditioning factors governing the allophony of /æ/ contribute to listerner evaluation. We can conceive of several levels to the phonological architecture which may be the target for acceptability judgments. First, we've seen in Chapter 4 that PHL and NAS behave as two variants of an overall /æ/ parameter, with these two systems competing wholesale in production. The uniformity of these systems in production might lead a reader to expect similar uniformity in acceptability: in other words, we might expect to see all the PHL phonological contexts rated alike and all the NAS phonological contexts rated alike.

However, as we have seen in the results from previous Matched Guise experiments, there is additional evidence that phonetic variation such as speech rate (Brown et al., 1985) or F0 (Levon, 2014) may also be the target of evaluation. Adding to this, we have seen that allophones have also been found to be the target of evaluation: Labov (2001) found Philadelphia speakers negatively rating only the tense forms of /æ/, rather than the system as a whole. This has been taken (e.g., Eckert and Labov, 2017) as evidence that social evaluation targets a surface form (i.e., the phonetics of hyper-tense *bad* differently from the phonetics of phonetically mitigated lax *bad*) rather than the underlying grammar; the evidence provided in §5.2 suggests instead that social evaluation may target any number of levels of phonology: the abstract parameters governing an allophonic split, as we have seen in §5.2, an allophone (Labov, 2001), and the phonetics (Brown et al., 1985; Levon, 2014).

Here, I investigate how the phonological conditioning factors differentiating PHL and NAS are rated, using a modified version of a Magnitude Estimation task (Sprouse, 2007, 2011; Bard et al., 1996; Cowart, 1997; Featherston, 2005). Magnitude Estimation is a task quite widely used in experimental syntax (Sprouse, 2007), in which participants are encouraged to rate items in comparison to a reference item. For example, participants may be told a reference line is length 100, and asked to rate subsequent lines by comparing them to the reference, as shown in Figure 5.3.

The goal of a magnitude estimation task is to capture a perceptual scale, rather than a physical scale. For instance, while doubling the lumens of a light will double its physical brightness, par-

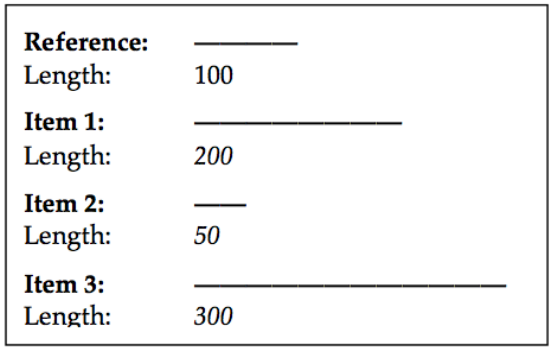


Figure 5.3: Magnitude estimation of the length of a line. From Sprouse (2007)

ticipants do not react in a linear way to this increase; such a light is rated as brighter but not by double. Bard et al. (1996) adapted this task to acceptability judgment data, allowing participants to rate sentences with marginal acceptability along a gradient and non limited scale. Here, I adapt this method to acquire acceptability judgments of phonetic realizations. I present participants with auditory stimuli and ask them to rate each stimulus in comparison to a reference stimulus. The task and stimuli are reported in more detail below.

5.3.1 Participants

Participants were the same as in Experiment 1; participants completed the Matched Guise task first, then went on to complete the Magnitude Estimation task.

5.3.2 Methods

Stimuli

Stimuli consisted of 96 tokens total, comprised of 50% test tokens containing a target /æ/ word and 50% filler tokens that did not contain /æ/. Of the test words, each participant heard a tense and a lax form of each word. Lists were presented in four blocks, and were prerandomized so that a participant did not hear a tense and a lax token of the same token within a single block. Likewise, each list contained no more than three test tokens in a row. Stimuli were recorded in a sound-attenuated sound booth. A tense and a lax form were recorded for each /æ/ word, meaning that no stimuli had to be acoustically manipulated.

Task

The experiment consisted of a training and a test phase. During the training phase, participants were introduced to the concept of magnitude estimation with the line task presented in Figure 5.3. After this training phase, participants entered the phonological ratings phase. They were presented with a reference stimulus (*chocolate*) and told that it received a rating of 100 for being "well pronounced."

Participants were then asked to rate stimuli for how "well pronounced" they sounded, using the reference stimulus rated 100 as a reference. An example is provided in Figure 5.4. Each page of the experiment contained 24 tokens, and the reference stimulus was repeated at the beginning of each page. This task included one important modification from the classic Magnitude Estimation paradigm: rather than allowing participants to input any unbounded value, they were asked to slide a slider somewhere between 0 and 150 for pronunciation value[9]. The experiment was run through Qualtrix and results were analyzed using R.

5.3.3 Analysis and Results

The results of the Magnitude Estimation task suggest a somewhat complicated social evaluation of /æ/ conditioning factors, which differ between the older participants and the younger participants. Here, I split participants into age groups based on the community-wide sociolinguistic patterns found in Chapter 2. Older speakers are defined as any speaker born before 1983, which was selected as the best changepoint in the community-wide data from the PNC and IHELP corpora. Older speakers would have largely acquired their language in a PHL-only environment, while younger speakers would have acquired language in a mixed environment consisting of both PHL and NAS.

[9] A pilot study giving participants a blank line for response resulted in a majority of '99' answers, presumably because participants wanted to finish the experiment as quickly as possible, and typing '99' provides a quick response. Changing to a slider bar resulted in a much wider range of responses.

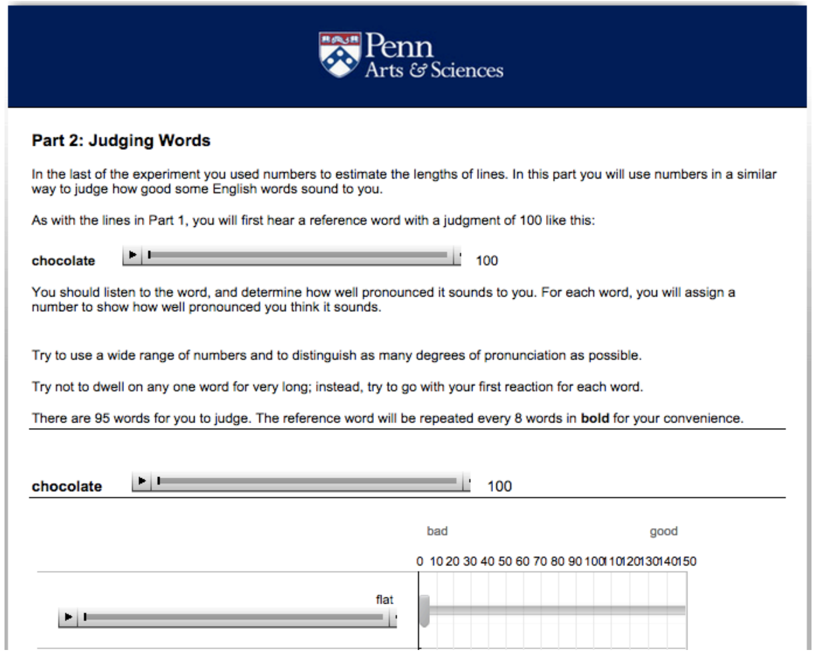


Figure 5.4: Modified Magnitude Estimation task rating the "well pronouncedness" of words against a reference word with score 100.

Older participants downgrade tense PHL

I begin by analyzing the results of the older speakers rating PHL-consistent tokens. We expect this data to align with the findings of (Labov, 2001), who found Philadelphian listeners negatively rating the tense allophone of /æ/ but not the lax allophone of /æ/. We see in Figure 5.5 a direct replication of these findings, with these older listeners downgrading tense /æ/ tokens and rating lax /æ/ tokens quite highly.

	Estimate	**Std. Error**	**t value**
(Intercept)	0.55	0.28	**2.03***
Realization (tense)	-0.69	0.18	**-3.85****
Gender (male)	0.41	0.54	-.76
Realization(tense):Gender(m)	-0.61	0.35	-1.75

Table 5.2: Tense PHL tokens downgraded by older speakers.

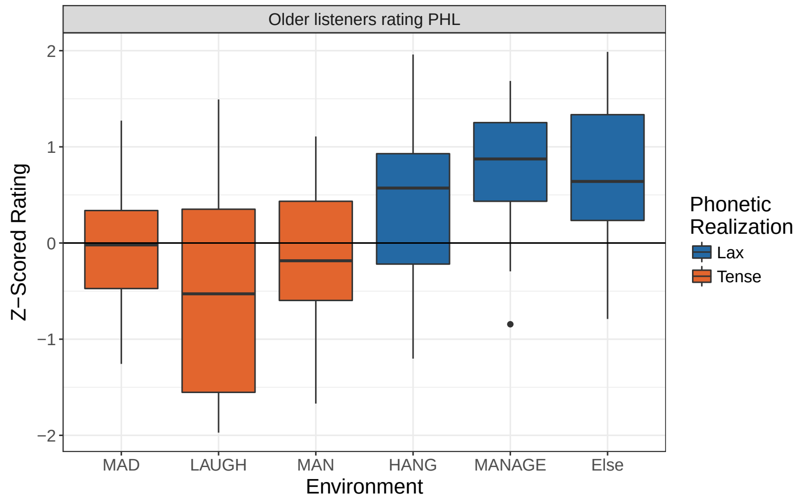


Figure 5.5: Older listeners downgrade tense PHL tokens.

A mixed effects model of this data with main effects of Realization (tense or lax) and Gender (male or female) and random intercept for participant is presented in Table 5.2, which finds a significant effect of tense realization on the evaluation of these tokens. This data serves primarily as a validation of the experiment: we find that the older participants rate PHL tokens consistently with the data reported in Labov (2001). In §5.3.3, I explore the systemic properties of this evaluation in more detail.

Younger participants learn two evaluation systems

I turn next next to the results from younger participants, meaning any participant born after 1983. While we do not have production data from participants, we can reasonably expect that these younger participants would have been exposed to both PHL and NAS in the community. The results from Chapter 2 demonstrating the social stratification of NAS in the elite non-Catholic schools in Philadelphia combined with the different social evaluations of PHL and NAS found in the Matched Guise experiment in §5.2 furthermore suggest that we might see a different pattern of overt ratings for PHL-consistent and NAS-consistent tokens from younger participants than from participants born before 1983. In other words, as the production of the community is in flux, younger participants may in turn adjust their overt ratings of pronunciations in line with the changing community norms.

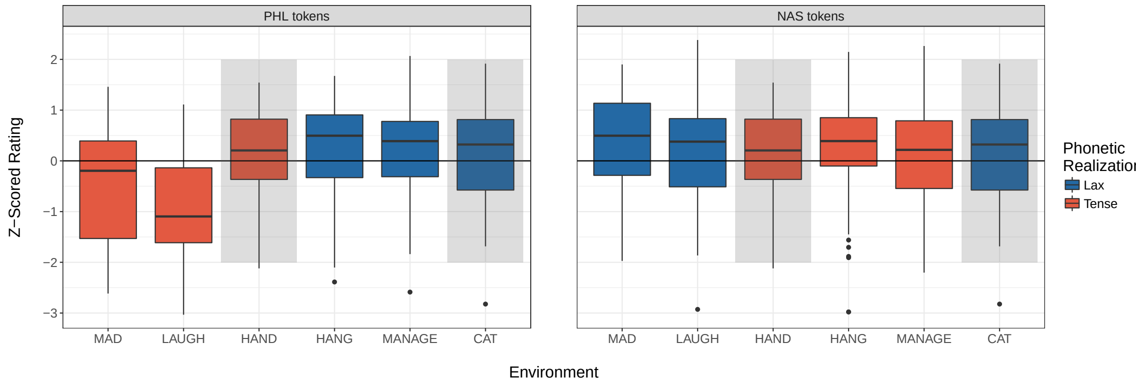


Figure 5.6: NAS rated highly by younger speakers (right); tense PHL downgraded (left).

Figure 5.6 shows the results from younger participants rating PHL-consistent tokens (left) and NAS-consistent tokens (right). Note that the HAND condition and the CAT conditions are the same in both facets, since HAND is produced as tense by both systems and CAT is produced lax by both systems; these boxplots have been grayed out as a visual aid to this fact. Let's first address the NAS-consistent tokens (right panel). Younger participants rate all NAS-consistent tokens highly, regardless of phonetic realization. This suggests that younger speakers have adopted a systemic evaluation of NAS: namely, that NAS-consistent tokens are all positively evaluated. Turning to the PHL-consistent tokens, we find that the younger participants have also learned the traditional community evaluations of PHL-consistent tokens, with tense realizations downgraded and lax realizations rated highly. Note that the only violation of this generalization is in the high ratings young speakers give to the MAN class, which I analyze as interference from participants' positive NAS evaluations.

These results suggest that younger participants are applying two evaluation systems. As an evaluation system, this means that younger speakers first apply a positive rating to any tokens that are NAS-consistent. This is relatively unsurprising, given the overt nature of this task: we have seen in Chapter 2 that the social patterning of NAS in Philadelphia resembles a change from above, in which the incoming NAS system is expected to be evaluated positively. That NAS is rated positively may predict all PHL-consistent tokens to be downgraded. However, this is not what we see. After applying a NAS-positive evaluation, participants then also apply a PHL evaluation system to any remaining tokens. That is, any tense tokens of MAD or LAUGH are rated low, in accordance with the PHL evaluation system. Tense tokens of HAND remain high, as they have already been highly evaluated using the NAS evaluation. Finally, lax tokens of HANG and MANAGE get rated highly, also in accordance with the PHL evaluation system. In other words, participants have learned a NAS evaluation as well as the traditional community norms for evaluation of PHL-consistent tokens, which results in a high rating for lax PHL tokens and a low rating for tense PHL tokens. These results are confirmed by a parsimonious mixed-effects model (Bates et al., 2015), which I describe in detail below.

Older speakers evaluation of conditioning factors

I turn finally to the older speakers' ratings of NAS, comparing these ratings to their ratings of PHL. Again, the HAND and CAT class words are grayed out, as a visual reminder that these two classes share conditioning between PHL and NAS, and are therefore given the same ratings. Here, a somewhat surprising picture emerges (Figure 5.7. We see here that participants are rating tokens according to their conditioning factor, rather than according to their phonetic realization or the system they are consistent with. In other words, older speakers rate tokens MAD, LAUGH, and HAND negatively regardless of whether they were produced as tense or lax. Likewise, older participants rate tokens of HANG, MANAGE, and CAT positively regardless of phonetic realization.

Mixed effects modelling

Here, I present the results of a parsimonious mixed-effects model fit and optimized separately for the younger participants and the older participants. In both models, I begin with a maximal model with the following fixed effects.

Realization Realization was treatment coded as a binary factor (Tense or Lax), with Lax as the reference level. This was chosen as a reference level due to the evidence that PHL speakers treat lax realizations as a default and tense as a negative value (Labov, 2001).

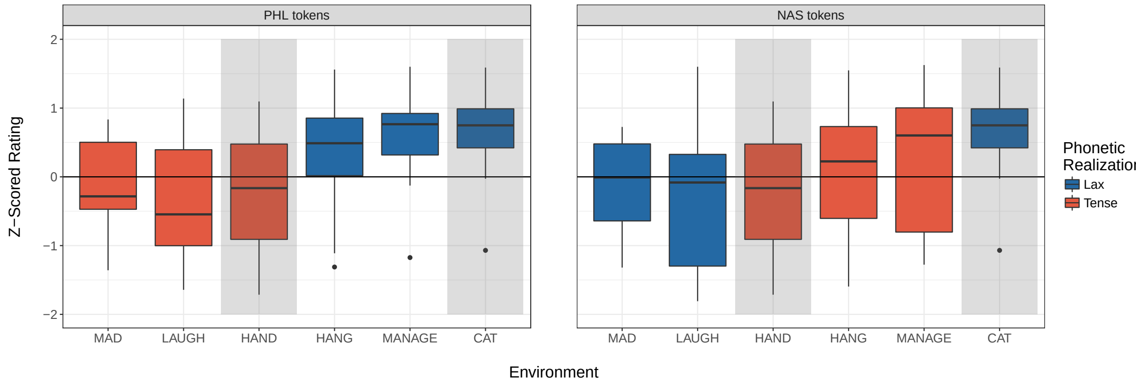


Figure 5.7: Older speakers rate MAD, LAUGH, MAN tokens low and HANG, MANAGE, CAT tokens high.

System Conformity The overlapping conditioning factors between PHL and NAS require some thought for the model, because they could potentially be analyzed as either PHL-consistent or NAS-consistent but not both. I resolve this by splitting the "system" parameter into two fixed effects: Conformity to PHL and Conformity to NAS. Conformity to PHL was coded as a (1) for tense tokens of HAND, LAUGH, MAD, and lax tokens of MANAGE, HANG, CAT, and a (0) elsewhere. Conformity to NAS was coded as (1) for tense tokens of HAND, MANAGE, HANG and lax tokens of LAUGH, MAD, CAT, and a (0) elsewhere. This enables us to test ratings of tense HAND and lax CAT as members of PHL as well as NAS.

Conditioning Environment Conditioning Environment was treatment coded, with six levels (HAND, LAUGH, MAD, MANAGE, HANG, CAT). Here, CAT was selected as the reference level because its lax production can be considered the default, unmarked variant.

PHL Conditioning This effect was included to test the effect suggested by the results in Figure 5.7 that older speakers downgrade the PHL tense-producing conditioning environments as a whole rather than the tense realization of those environments. PHL Conditioning, unlike Conformity to PHL, represents the conditioning environments only and not the realization of those environments. For PHL Conditioning, HAND, LAUGH, and MAD received a (1) and MANAGE, HANG, and CAT received a (0).

Gender Models were tested with three different methods of coding participant self-reported gender[10]. The first method of coding gender was to sum code, given an assumption that males and females may produce different evaluations but that neither gender should be considered the reference level. However, in cases of language evaluation, it is not clear that sum coding gender is the best approach. There is an argument to be made that because the change from PHL to NAS has been described as a Change from Above, in which we expect women to lead in production, women may also lead in evaluation. For this reason, a second version of each model was also run with treatment coded gender with female as the reference level. Finally, because PHL is associated with an "accented" local dialect feature, there is the possibility that PHL-consistent tokens may be rated by participants as carrying covert prestige (Trudgill, 1974), which may predict that males positively evaluate PHL-consistent tokens. In all three versions of coding Gender, Gender did not improve model fit and was subsequently removed.

There is a large redundancy in including terms for Realization, Conformity to PHL, Conformity to NAS, Conditioning Environment, and PHL Conditioning in the same model. Conditioning Environment is colinear with PHL Conditioning, and the interaction of Conditioning Environment and Realization is colinear with Conformity to PHL and Conformity to NAS. For this reason, the maximal model and several of the near-maximal models were rank-deficient. All terms were tested for model fit using AIC and BIC comparison.

The results of the parsimonious mixed effects models for the younger listeners are consistent with the analysis provided above. Younger listeners have learned to downgrade tense tokens, but positively evaluate tense tokens that are consistent with NAS. In other words, younger listeners exhibit the operation of two evaluation systems: one in which NAS tokens are positively rated, and a second in which tense tokens that are inconsistent with NAS are negatively rated. It is worth pointing out that this is a slight break from the traditional rating pattern for PHL, since

[10]Participants were given a free-form response box for gender, to allow for queer and non-binary participants to self-identity. Participant responses fell categorically into a 'male' ('m', 'M', 'man', 'male') or 'female' ('f', 'F', 'female', 'femail', 'woman') response.

	Estimate	Std. Error	t value
(Intercept)	0.27	0.08	**3.36****
Realization (Tense)	-1.00	0.12	**-8.56*****
PHL (true)	-0.06	0.07	-0.87
NAS (true)	-0.03	0.09	-0.34
Conditioning(PHL)	-0.02	0.07	-0.25
Realization (Tense) : NAS (true)	0.94	0.13	**7.024*****

Table 5.3: Younger speakers downgrade Tense but positively rate Tense NAS tokens.

	Estimate	Std. Error	t value
(Intercept)	0.20	0.25	0.81
Realization (Tense)	-0.17	0.32	-0.54
PHL (true)	0.33	0.21	1.58
NAS (true)	0.06	0.27	0.22
Conditioning(PHL)	-0.66	0.21	**-3.13****
Realization (Tense) : NAS (true)	0.01	0.41	0.03

Table 5.4: Older speakers downgrade PHL conditioning factors, regardless of phonetic realization.

the traditional evaluation is to negatively evaluate all tense PHLtokens (including HAND tokens), while the results from the younger listeners demonstrate that tense HAND tokes are considered to be part of a NAS system and subsequently rated positively.

The results from the older listeners are somewhat more complicated. While we see the expected pattern of downgrading tense PHL tokens and upgrading lax PHL tokens, it is not clear how to interpret their evaluation of NAS-consistent tokens. Rather than rating all lax tokens of /æ/ positively and all tense tokens negatively, as would be expected if it is the phonetic production listeners evaluation rather than the phonological context, we in fact see older speakers not rating NAS tokens by their phonetic output. Instead, older speakers rate all conditioning factors that would be tense under PHL (MAD, LAUGH, HAND) as negative regardless of the phonetic production of the tokens, and all conditioning factors that would be lax under PHL (MANAGE, HANG, CAT) as positive regardless of the phonetic production of the tokens. There are two possible explanations for these results.

The first explanation is that older listeners' evaluation is tied to the phonological conditioning factors rather than to the phonetic production. In other words, listeners learn that the conditioning factors MAD, LAUGH, and HAND are negative while MANAGE, HANG and CAT are positive. Whether these tokens are produced as tense or lax does not matter all that much, as it is the underlying phonological environment that is evaluated rather than the phonetic production of that phonology. This would suggest that what seemed on the surface in Labov (2001) to be a straightforward case of participants negatively evaluating a tense production of an /æ/ allophone may instead have been participants evaluating the underlying conditioning of the allophone. This interpretation finds listeners evaluating the phonological system in a systematic way, contra the expectations in Eckert and Labov (2017).

A second explanation may be that older participants have several competing social evaluations available. First, any tense PHL token gets negatively evaluated while lax tokens are taken to be neutral or positive. Second, any tense token that conforms to NAS only may either be unnoticed or may be associated with a positive accent and so receives a high rating. This accounts for the positive ratings of HANG and MANAGE regardless of phonetic output. Finally, listeners would also need to apply an additional socially-motivated negative evaluation for lax productions of traditionally tense PHL tokens (MAD and LAUGH class), perhaps as a negative response to tokens that sound out-group. So any tense tokens of MAD and LAUGH are negatively evaluated because of the traditional evaluation, but lax tokens of these classes are also negatively evaluated because they don't sound Philadelphian enough.

5.4 Discussion

In this chapter, I have attempted to shed some light on the phonological target of social evaluation. In §5.2, Philadelphian participants were found to identify a PHL guise as distinctly more *accented* than NAS, using a Matched Guise paradigm. With the addition of the Magnitude Estimation results, I find that not only are Philadelphians at least implicitly aware of their sociolinguistic environment, but also that their explicit evaluations of "well pronouncedness" fall out from a structural rather than phonetic evaluation. Young Philadelphians exhibit the operation of two evaluation standards in their responses: the new NAS system is rated positively overall while the older PHL system tokens receive the expected downgrading of the tense forms. The responses from older Philadelphians provide what is potentially the biggest surprise: here, we find that the target of listener evaluation may be the abstract conditioning factors, rather than the phonetic output of those conditioning factors. These findings reveal two important points: First, it suggests that abstract phonological structure may act as the target of social evaluation. Secondly, it reinforces the importance of diachronic work: what appeared synchronically to be participants rating the phonetic output of an allophone is revealed diachronically to be a potential case of participants rating the underlying phonological structure rather than the phonetic realization of that structure.

Chapter 6

The Inevitability of Phonological Change

Throughout this dissertation, my main objective has been to identify the mechanism of phonological change for the allophonic restructuring of /æ/ in Philadelphia. I've argued that this phonological change occurs via intraspeaker grammar competition between the abstract parameters of PHL and NAS, and furthermore that these abstract parameters are the subject of social evaluation. In this chapter, I turn to the question of how inevitable this change is. Many frameworks of phonology take articulatory and cognitive simplification to be a motivating factor for sound change. The allophonic restructuring from the phonologically complex PHL system to a simple surface-true NAS system appears on the surface to be a confirmation of the inevitability of phonological change via simplification. In this chapter, I conduct a computational simulation based on the Tolerance Principle to investigate whether this change was the result of an inevitable simplification. The work presented in this chapter is a slightly modified version of a collaboration with Josef Fruehwald and Charles Yang, which is currently under review.

6.1 The Role of Simplification in Sound Change

Simplification, whether cognitive, phonological or articulatory, is often appealed to as a major motivation for sound change. This notion can be found in a number of different theoretical frameworks, from European structuralism to generative phonology. While an appeal to simplicity is often considered intuitive, a definition of simplicity depends on the framework and what the target of simplicity is.

First, to phonological simplicity. The specifics of simplicity depend on the framework under consideration, but the primary cohesive factor is the idea that marked or dispreferred forms and systems are more "cognitively complex" (Givon, 1991) and therefore more susceptible to change. Cognitive complexity is, in itself, a somewhat slippery term to define. Writing in the functionalist tradition, Martinet (1952) appeals to the notion of structural harmony as a motivating factor in the history of a language. Here, structural harmony refers specifically to a linguistic, or phonemic, inventory that is maximally symmetrical and makes use of a limited number of active features, resulting in a cognitively efficient system. This idea is echoed in Feature Economy (Clements, 2003), in which the simpler systems are those that maximize the ratio of sounds to features. Under a featural phonological framework, a simpler system with simpler forms would be defined as a system needing fewer features to encode it than a complex system. The specifics of feature simplicity depend further on the framework involved, with Feature Geometry (Clements, 1985) and Contrastive Hierarchy (Dresher, 2011) providing a hierarchical account of active features, Government Phonology (Kaye et al., 1985) relying on nonlinear representations and classical Generative Phonology Chomsky and Halle (1968) calculating simplicity through binary feature bundles, just to name a few. Regardless of specific definition of feature simplicity, however, there is a shared notion across these frameworks that simplicity is a driving force in phonology.

If simplicity as measured by cognitive complexity is a main driving force in language change, we may be tempted to echo the question articulated by Martinet (1952): "How is it that after so many millennia of uninterrupted speech practice, patterns should still be in need of structural integration?" In other words, why, after so many thousands of years of speaking, have languages not settled on a maximally cognitively efficient system? Why do they continue to change?

One potential answer to this lies in the physical facts of using language. A cognitively perfect system must still pass through human articulators, whether oral or manual. This interface introduces a second type of simplicity which has been thought to have an effect on language change, namely, articulatory ease. The role of articulatory ease can be found hand-in-hand with cognitive complexity in nearly every framework: Martinet (1952) refers to the strain of physiology as a "germ of instability" within a linguistic system. While some markedness constraints in Optimality Theory refer to cognitive complexity, other markedness constraints refer to articulatory ease (see Haspelmath, 2005), whereby processes like coarticulation and consonant cluster reduction which may initially occur due to articulatory ease become phonologically encoded into the underlying system. An Exemplar Theoretic account simultaneously appeals to ease of articulation and ease of cognitive recall: developed from the observation that high frequency words exhibit reductive processes in production (Bybee, 1999) as well as faster recall (Segui et al., 1982; Grainger, 1990); many proponents of Exemplar Theory suggest that high frequency tokens will likewise exhibit distinct profiles of change (e.g. Hay et al., 2015). Blevins (2006) exemplifies this view of language, arguing that human perceptual and articulatory biases are the source of many of the phonological patterns found in languages today. The proliferation of framework-specific considerations outlined here highlight how the specific predictions of simplification-motivated sound change will depend on the framework used to define simplicity. Regardless of framework, however, the implicit notion is that simplicity in form and system will be preferable to speakers, and that given the choice between two plausible representations, speakers will select the simpler choice.

The change from PHL to NAS in Philadelphia English seems, on the surface, to be a case study in support of simplification as a driving factor in sound change. While the specific definition of complexity is framework-dependent, it is uncontroversial under any framework to state that PHL, with its disjoint set of phonological triggers and syllable structure references and lexical specificity, is simpler than NAS, a surface-true allophonic rule with little complexity. In this chapter, I delve into this question in detail, asking whether this change from PHL to NAS was the inevitable result of simplification.

Here, we again make use of the Tolerance Principle (Yang, 2016) as a method of diagnosing whether a proposed phonological rule would be plausible for a language learner. We apply the Tolerance Principle to the allophonic restructuring in /æ/, to investigate the likely route by which NAS is supplanting PHL in the community. In §6.2.1, we find that a child receiving entirely traditional input could not plausibly posit NAS as a productive rule. In §6.2.2, we further find that the change from PHL to NAS is unlikely to be the result of children positing incrementally simpler changes in PHL, removing a conditioning factor at a time until the speech community is left with NAS. Finally, in §6.3 we turn to the possibility that Philadelphian children have acquired NAS through receiving mixed input from both PHL and NAS speakers, in a situation of dialect contact.

6.2 Could Children have Endogenously Postulated NAS?

Recall that under a featural rule-based framework, PHL is described as in (25)[11]: tense before tautosyllabic anterior nasals and voiceless fricatives. In addition to a rule with relatively complex conditioning, PHL also requires speakers to memorize a list of lexical exceptions, as outlined in Chapter 3. In contrast to this, NAS (shown in 26) is a simple allophonic rule comprised of a single conditioning factor which typically requires no lexical exceptions.

(25) **PHL:** $\text{æ} \rightarrow \text{æh} \;/\; \_[\ \text{+anterior}\ ] \cap ([\ \text{+nasal}\ ] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix})]\ \sigma$

(26) **NAS:** $\text{æ} \rightarrow \text{æh} \;/\; \_[\ \text{+nasal}\ ]$

Given that NAS is a surface-true generalized rule where PHL produces surface exceptions, one possibility to be addressed is whether Philadelphian children are spontaneously simplifying their traditional input into the new NAS system. In other words, a Philadelphia child, perhaps at some transient stage of language acquisition, may have postulated a NAS system despite receiving consistent PHL input: as we have seen, a statistical majority of the lexical items produced under the PHL system is in fact compatible with the NAS system, and children's tendency of regularizing inconsistent input to form a majority rule is well documented in naturalistic acquisition (Singh et al., 2004) and in artificial language learning experiments (Hudson Kam and Newport, 2009, 2005). The

[11] Because this chapter deals primarily with counting lexical exceptions under different versions of the regular rule, here I exclude a full list of lexical exceptions as part of PHL or NAS.

NAS system, once postulated, would of course encounter exceptions, but as outlined in Chapter 3, linguistic systems – including PHL – that have lexical specificity can still be stably acquired (Scobbie and Stuart-Smith, 2008; Payne, 1980; Roberts and Labov, 1995). Here, we are interested in whether the NAS system can become a viable endogenous response to the PHL system; if so, it would provide an example of simplification by children as a source of language change.

6.2.1 Can a NAS Postulation Tolerate PHL Input?

We begin first by asking the question "can a child who has posited NAS tolerate traditional PHL input?" To apply the Tolerance Principle to short-a systems in Philadelphia, assume a child is receiving input generated only by the traditional PHL system, with its disjunctive featural specification, syllabic sensitivity, and lexical exceptions. This learner could possibly hypothesize that their target grammar is simply 6, tense before nasals, producing tense æ in *ham, man*, etc. If they do so, they must somehow account for words they acquire that violates this generalization, such as lax æ in *bang*, or tense æ in *last.* If they maintain the generalization in 6, they must treat these and all other words that violate the "tense before nasals" generalization as stored lexical exceptions. If the number of such exceptions (e) is less than the tolerance threshold for that child's vocabulary size, then it is plausible that learners in Philadelhpia could endogenously hypothesize a NAS grammar given only PHL input. However, if the number of exceptions exceeds the tolerance threshold, then some other source of the NAS grammar in Philadelphia must be sought. As described in §3, N will be the entire set of æ words in a child's vocabulary, and e will be the list of words that violate R, where R = NAS.

We begin by using the CHILDES database (MacWhinney, 2000) to obtain a measure of the total N for a child's vocabulary. Each word type was coded for its realization under traditional Philadelphian input, under R = PHL, and under R = NAS. An example is shown in Table 6.1. Note that the mismatch between traditional input and PHL for *bad* reflects the fact that *bad* must be treated as a lexical exception, while PHL captures the regular phonological generalization.

This coding system allows us to measure the total number of exceptions produced by positing either PHL or NAS as a rule. Using Table 6.1 as a dummy lexicon with N = 5 words, we can see that a child positing R = PHL would have to list $e = 1$ exception to that rule, because the realization of *bad* under R = PHL does not match the child's input. Because 1 (fix), PHL emerges as a plausible rule for this dummy language. By contrast, a child positing R = NAS would have to list $e = 4$ exceptions, which does not pass the tolerance threshold of 3.11, rendering NAS an unproductive rule for the dummy language in Table 6.1.

Word	Traditional input	PHL	NAS
bad	Tense	Lax	Lax
hammer	Lax	Lax	Lax
cat	Lax	Lax	Lax
fast	Tense	Tense	Lax
bang	Lax	Lax	Tense

Table 6.1: Input realizations of /æ/ compared to expected /æ/ realizations for PHL and NAS. Mismatches between actual input and expected input (in gray) result in an exception.

Using the full list of /æ/ word types in CHILDES, we calculated whether the number of exceptions a child would need to list under R =PHL and R =NAS would pass the tolerance threshold of e. We find that given the traditional Philadelphian input distribution, a child positing R =PHL would have to store e =39 lexical exceptions (mostly *mad, bad, glad*, strong verbs and function words), well under the tolerance threshold of 194.7. This, of course, is expected: children have been successfully learning PHL and its listed exceptions for well over 100 years (Labov et al., 2016, 2013). Turning to the question of whether NAS can be a productive rule given traditional input, we find that positing R =NAS requires listing a total of 324 exceptions (e.g. all tense /æ/ before anterior voiceless tautosyllabic fricatives), well over the tolerance threshold.

Thus, despite being a formally simpler rule, and in fact a featural subset of PHL, NAS is not a plausible innovation for Philadelphian children on the basis of only traditional Philadelphian /æ/ input. Positing NAS requires storing too many lexical exceptions for it to be productive.

6.2.2 Can NAS replace PHL incrementally over time?

It remains, however, that NAS is rapidly replacing PHL as the dominant allophonic rule for /æ/ in Philadelphia. Given the finding the r =NAS is not a plausible re-analysis of the traditional input, we can now turn to the question of incremental re-analysis. In other words, we ask whether it is possible that a child might posit an intermediate rule given traditional input, which might then be re-analyzed as a productive NAS rule by the subsequent generation of language learners. We take PHL, reproduced in 25, and break it down into its four constituent aspects. R =PHL can be spelled out as /æ/ is tensed when it precedes a (a) tautosyllabic (b) anterior (c) nasal or (d) voiceless fricative.

Using these four components of PHL, we construct six intermediate grammars between full PHL and NAS, beginning with excluding only one aspect of PHL at a time and ending with excluding two aspects of PHL. We do not analyze intermediate forms of PHL which consist of excluding the nasal trigger, since that would not produce an intermediate form between PHL and NAS; NAS being the result of excluding every component of PHL except the nasal constraint. In Table 6.2, these rules are described as PHL minus the components that have been excluded. We note that some rule exclusions result in the expansion of the set of triggering forms (as in PHL-ant). The set of triggering phonological contexts resulting from each intermediate rule is shown in the third column of Table 6.2. We note finally that NAS is the same as PHL minus the tautosyllabic, anterior, and voiceless fricative components.

Name	**Rule**	**Triggering Segments**
PHL-ant	$\text{æ} \rightarrow \text{æh} \;/\; \_[\;\text{+nasal}\;] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix}]\,\sigma$	m, n, ŋ, f, θ, s, $ʃ]_\sigma$
PHL-taut	$\text{æ} \rightarrow \text{æh} \;/\; \_[\;\text{+anterior}\;] \cap ([\;\text{+nasal}\;] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix})]\,\sigma$	m, n, f, θ, s
PHL-fric	$\text{æ} \rightarrow \text{æh} \;/\; \_[\;\text{+anterior}\;] \cap [\;\text{+nasal}\;]]\,\sigma$	$\text{m, n}]_\sigma$
PHL-ant-taut	$\text{æ} \rightarrow \text{æh} \;/\; \_[\;\text{+nasal}\;] \cup \begin{bmatrix} \text{-voice} \\ \text{+fricative} \end{bmatrix}$	m, n, ŋ, f, θ, s, ʃ
PHL-ant-fric	$\text{æ} \rightarrow \text{æh} \;/\; \_[\;\text{+nasal}\;]]\,\sigma$	$\text{m, n, ŋ}]_\sigma$
PHL-taut-fric	$\text{æ} \rightarrow \text{æh} \;/\; \_[\;\text{+anterior}\;] \cap [\;\text{+nasal}\;]$	m, n

Table 6.2: Intermediate grammars between PHL and NAS.

In addition to testing the intermediate rules shown in Table 6.2, we also consider the effects of a smaller vocabulary. As mentioned in 3, smaller vocabularies are able to tolerate a higher proportion of exceptions. This is particularly relevant to the question at hand: perhaps children with smaller vocabularies would be able to plausibly posit NAS as a productive rule for their traditional input. To test this, we also test the plausibility of NAS and intermediate PHL forms on several subsets of the most frequent words in CHILDES, with at least 20, 50, and 100 mentions in the corpus, so as to provide a rough approximation of learners' vocabulary composition at progressive stages of language development. The results are shown in Table 6.3.

Rule	1 Mention N =1412 T =194.7	20 Mentions N =498 T =80.2	50 Mentions N =334 T =57.5	100 Mentions N=239 T =43.6
PHL	39	19	15	11
PHL-ant	244	60	42	31
PHL-taut	155	55	36	25
PHL-fric	155	64	48	38
PHL-ant-taut	273	94	63	45
PHL-ant-fric	237	93	67	51
PHL-taut-fric	240	92	65	50
NAS	324	121	84	63

Table 6.3: Exceptions required for each intermediate rule for vocabularies consisting of words with 1, 20, 50, and 100 mentions in CHILDES. Plausible grammars shaded.

As shown in Table 6.3, NAS does not emerge as a plausible analysis of traditional input, even with a limited vocabulary. However, we see that traditional input can be plausibly re-analyzed as any of the three intermediate rules that result from deleting one component of PHL. For example, a child could plausibly posit a phonological rule tensing /æ/ before all nasals and voiceless fricatives, including ŋ and ʃ (PHL-ant) without having to list more exceptions than the threshold. Given the plausibility of *at least some* children positing these intermediate grammars, we must now turn to the question of whether these intermediate children could plausibly contribute enough examples to the linguistic environment that in turn favors NAS, resulting in wholesale change for all subsequent language learners. To do so, we introduce the model of rule learning under heterogeneous input.

6.2.3 Rule learning under a mixture of PHL and Intermediate Grammar input

As stressed in 3, the Tolerance Principle applies on individual learner's lexicon composition; even if a representative sample of words (e.g., the 498 that appear at least 20 times per million) can be expected to support an intermediate grammar (e.g., PHL-fric, which is expected to have 64 exceptions an thereby falls below the tolerance threshold of 80), it remains possible that some learners may learn from a somewhat skewed sample, whose lexicon fails to support an intermediate grammar. Thus, if the endogenous emergence of NAS is achieved through successive generations of learners, we must consider the situation in which learners are exposed to a mixed input: some produced by speakers who happened to successfully acquire an intermediate grammar and some produced by speakers who have retained the traditional PHL grammar. The question of whether NAS is a plausible reanalysis must then be reframed as "what proportion of intermediate input does a child need in order to plausibly posit NAS?"

To answer this question, we simulate a child's acquisition given dialect contact between PHL and each intermediate rule, in the following way. First, we let m represent the proportion of input from the intermediate grammar that a child receives during acquisition, and 1-m the proportion of traditional input. We then construct a simulation of the plausibility of positing NAS, for values of m between 0 and 1 in steps of .01 for each of the three intermediate rules. We begin with the assumption that a child will store one form for each word type. For each run of the simulation, we generate a full mixed lexicon according to m. Each word is assigned lax or tense /æ/ on the basis of an intermediate rule or traditional input, according to m. For example, if m =24, each word in the lexicon will have a 24% chance of its /æ/ allophone being determined by an intermediate rule. This assumption is motivated by empirical studies of how children deal with mixed input where each lexical item is subject to probabilistic variation at the level of token frequency. In a series of studies (Hudson Kam and Newport, 2005, 2009), children were found to regularize mixed input to the statistically dominant variant. In the present case of mixed input with the level of m, we assume that each word type has an m probability of beign internalized in the child learner's vocabulary as an example of the intermediate grammar, and a $1-m$ probability of being internalized as an example of PHL. That is, the child regularizes a *probilistic* mixture of word tokens in the input as a *discrete* mixture of word types representing the two variant grammars: this is implemented by stochastically assigning each word type into one of two grammars with the associate probabilities. We then evaluate the viability of the two grammars on the basis of the resulting lexicon.

It is worth stressing several important features of the learning model. First, it is crucial to note that this is an acquisition model of how a single learner evaluates rules given variable input. This is clear from the description of the model, where the sample lexicon for the learner is stochastically drawn from the mixture distribution in the environment. By running the model many times, we can understand the outcome of learning for the speech community at large. Second, the model is agnostic as to the real-world source of the variable input. Both dialect contact scenarios and endogenous innovation scenarios are treated identically by the model. An individual learner evaluates rules on the basis of the lexicon they acquire from the mixed environment, and it is immaterial how such a mixture is introduced in the environment in the first place; see Yang (2000) for additional discussion and applications to syntactic change. Third, the model also does not imply any particular time course for change. For a given mixture of input data, it estimates the probability that PHL or NAS may be a plausible grammar for a speaker, but does not predict what the rate of use of either grammar would be for a speaker who has successfully acquired both systems. In other words, this model does not predict *m* for the next generation of learners. Fourth, we stress that this model does not address how a child may *generate* a possible rule, it is simply a model of how a child evaluates possible rules that have already been generated.

We calculate whether an input lexicon comprised of a mixture of PHL and intermediate grammars would allow NAS to be a productive rule for each trial. 1000 trials were run for each value of *m* between 0 and 1 in steps of .01, for each intermediate grammar.

Figure 6.1 presents the results of this simulation, with rates of *m* plotted along the x-axis and the proportion of trials that pass the tolerance threshold along the y-axis. It is important to note that the y-axis represents only the predicted proportion of children whose input would allow them to evaluate PHL (in stars) or NAS (in circles) as a plausible grammar for each value of *m*; it does not represent the predicted production of these children. Each intermediate rule was tested for whether NAS passed the tolerance threshold for each value of *m* (circles) and for whether PHL

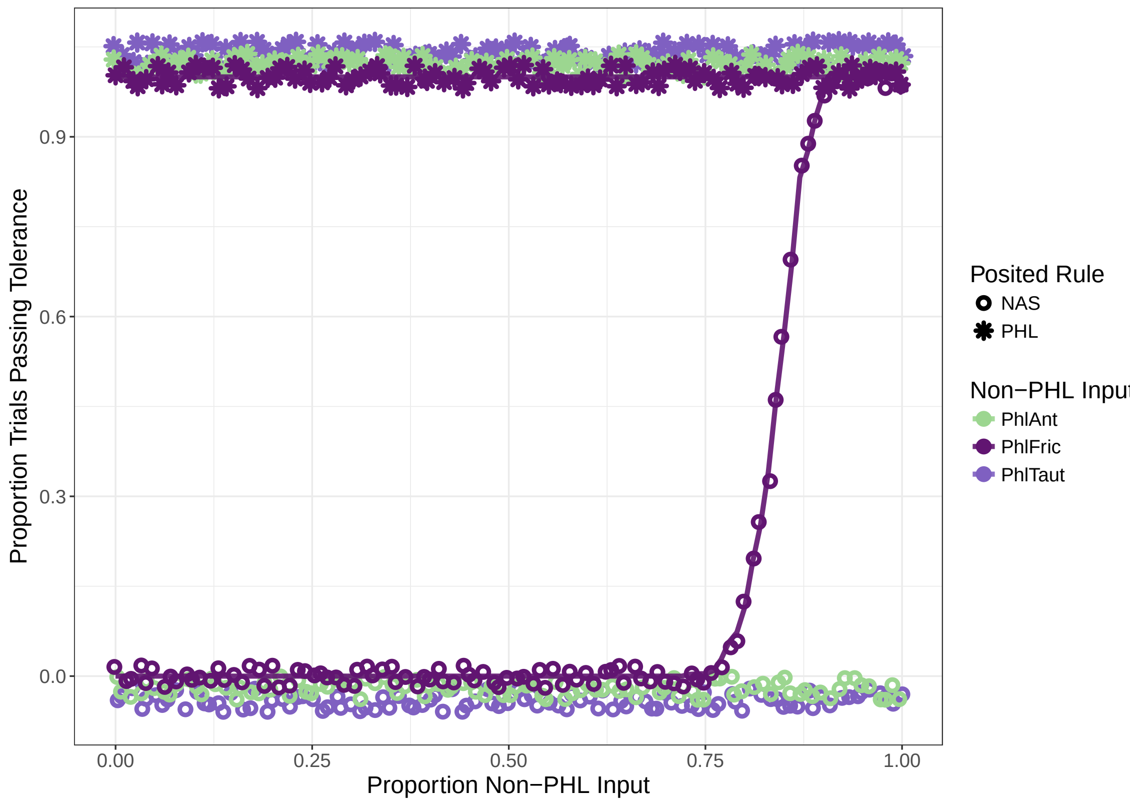


Figure 6.1: Proportion trials which pass the tolerance threshold for each proportion of intermediate rule input for positing NAS or PHL.

passed the tolerance threshold (stars).

We find two striking results. First, PHL is a plausible reanalysis of every intermediate rule, for all proportions of intermediate input, including 100% intermediate input. This speaks to the history of stability of PHL in Philadelphia; even if speakers have been spontaneously positing intermediate rules throughout the history of the /æ/ split in Philadelphia, these intermediate rules can still be reanalyzed as PHL by the next generation of speakers. Second, of the three intermediate rules that are plausible given traditional input, it is only the PHL-fricative rule that will allow NAS to be a plausible reanalysis of the intermediate rule. And this is only possible when children are receiving approximately 73% PHL-fricative input, which is the point at which the probability of accepting NAS becomes non-zero. That is, if at least 73% of Philadelphian children lost the voiceless fricative conditioning, then NAS can endogenously emerge as a consequence. We note that this possibility mirrors the argument in Ash (2002), who models the change from PHL to NAS in central New Jersey as occurring via an intermediate step of PHL-fricative.

However, we find this route of change to be highly implausible for Philadelphia, given the results of an empirical search for speakers exhibiting a PHL-fricative type grammar. Only 1 speaker out of 184 who had enough data to allow such an investigation was found[12]: Jake S, our outlier from Chapter 2. As I have argued in Chapter 2, Jake's social profile suggests he developed a PHL-fricative grammar as a *result* of NAS contact, rather than as an endogenous modification of the PHL system. Jake was born in 1992, and attended the elite Masterman middle and high school, then went on to graduate from the University of Pennsylvania. Most of Jake's peers – speakers born around 1992 who attended Masterman – produced NAS. Given the data, this suggests that language learners positing PHL-fricative was not the route by which NAS came into Philadelphia. In addition to a social profile that renders Jake's production of PHL-fricative an unlikely step in the change to NAS for his own subset of the speech community, it is also noteworthy that finding only one speaker out of 184 falls well short of the 73% PHL-fricative speakers necessary for NAS to be plausible for the following cohort of speakers.

[12]Using data from the PNC and the IHELP corpus , we analyzed every white speaker who produced at least 5 /æ/ tokens in both the fricative environment and lax nasal environment. The search was restricted to white speakers, as African American and Hispanic speakers in Philadelphia traditionally produce a neutral /æ/ system, produced as a raised lax form [ɛː] for all phonological categories (Fisher et al., 2015; Labov and Fisher, 2015).

To summarize the theoretical results so far, we have found that it is impossible for NAS to directly arise from a PHL system. It is conceivable that an intermediate grammar, specifically PHL-fricative, may eventually lead to NAS, but only if the vast majority of learners all converge onto that grammar under homogeneous PHL input. This, however, proves to be highly unlikely. Finally, although our simulation is intended to model the terminal state of language acquisition, it can also be used to understand the developmental time course of language acquisition in a single child/generation. It is clear that unless a child is nearly completely surrounded by PHL-fricative input (as indicated by the value m), it is virtually impossible for the grammar to survive until the stabilization of language acquisition (e.g., pre-teen years; Herold (1997); Johnson (2010); Johnson and Newport (1997)).

6.3 Acquiring NAS through dialect contact

Given the unlikelihood of and lack of empirical support for NAS emerging endogenously in Philadelphia, either through direct reanalysis of the original system or via a sequence of reanalyses, we now turn to the possibility of NAS emerging as the result of dialect contact between NAS and PHL.

6.3.1 Sociolinguistic background

The idea that Philadelphian children may be exposed to NAS speaking outsiders is not altogether unlikely. According to the Atlas of North American English (Labov et al., 2006), NAS has been found in the geographic area surrounding Philadelphia; it is not unlikely that some of these speakers may have access to and influence within Philadelphia. Ash (2002) also provides clear evidence of NAS gaining ground over both PHL and the New York split-/æ/ system in the Mid-Atlantic region in the region between Philadelphia and New York City. Furthermore, as outlined in Chapter 2, NAS is more likely to be found in graduates of elite non-Catholic high schools such as Masterman and Friends Central than in graduates of local diocesan schools. This pattern fits with an analysis of NAS as a change from above: the wealthier, more nationally-oriented schools adopt NAS early (perhaps via the influence of externally raised teachers), while the more locally-oriented neighborhood schools act as conservative forces holding on to PHL.

6.3.2 Theoretical analysis and predictions

Given that dialect contact with NAS speakers is a likely situation given the geographic and social patterns around Philadelphia, we now turn to the question of how much contact with NAS speakers is necessary for a Philadelphian child to accept NAS as a plausible system. Using the same simulation procedure described in §6.2.3, with NAS as the non-PHL input at the proportion of m, we tested which proportion of NAS input is necessary for a child to plausibly posit NAS. Figure 6.2 presents the results of this simulation, plotting the proportion of trials in which NAS emerged as a plausible rule (in circles) and PHL emerged as a plausible rule (in stars). Simulations were run for different sized lexicons, from words with one mention to words with 100 mentions in CHILDES, in order to capture the potential effect of differently sized lexicons. The full results are displayed in Table 6.4, which displays the proportion NAS input necessary for NAS and PHL to be viable at all as well as viable for 100% of trials.

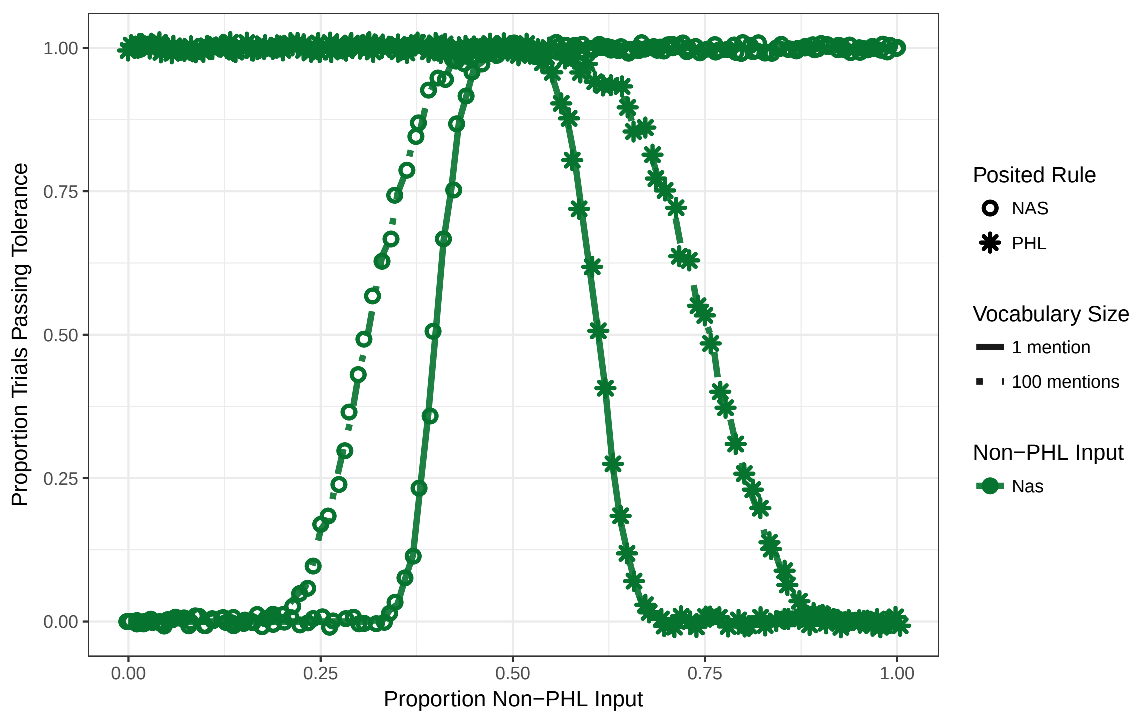


Figure 6.2: Proportion trials that pass the tolerance threshold for NAS (circles) and PHL (stars) for different proportions of NAS input.

As expected, higher word frequency cutoffs produce shallower slopes; this is a reflection of the fact that these lexicons are smaller and therefore more proportionally tolerant of exceptions, resulting in a slightly higher proportion of trials that pass the tolerance threshold for each value of m. In contrast to the endogenously posited intermediate rules simulated in the previous section, we find that dialect contact between traditional input and NAS makes positing NAS a highly plausible solution for a child receiving both inputs. In other words, NAS becomes a plausible analysis of a child's input if that child is receiving at least 32% NAS input.

Vocabulary size	**NAS leaves 0% viable**	**NAS reaches 100% viable**	**PHL leaves 100% viable**	**PHL reaches 0% viable**
1 mention	.32	.48	.53	.7
20 mentions	.25	.46	.52	.82
50 mentions	.2	.47	.54	.86
100 mentions	.17	.48	.54	.9

Table 6.4: Proportion NAS input at which NAS and PHL become variable viable and categorically viable.

Again, although our model has been used to study contact-induced change, it is also applicable to children's developmental time course, and the sociolinguistic conditions of language acquisition. For example, as documented in detail (Johnson, 2010), young children may initially acquire the grammar of the parental input and then adopt a new grammar once immersed in their peer group under certain conditions. In the current study, the viability of PHL and NAS as a function of contact can be understood as follows. If there is a relatively weak presence of NAS in the environmet (e.g., $m < .2$), even if a child were to acquire NAS at home they will still end up adopting PHL. Likewise, if NAS is already quite dominant in a child's peer group (e.g., $m > 0.7$), then the home PHL system will be abandoned in favor of NAS. In the region where m assumes an intermediate value, both systems are predicted to be viable. In other words, whichever system a child acquires at home, the linguistic environment of their peer group is sufficiently heterogeneous for these intermediate m values that either system will be sufficiently supported (i.e., neither will encounter an intolerable number of exceptions).

The above discussion is particularly applicable when the community network structure is taken into account. For instance, while m may be quite low over the entire speech community of Philadelphia, there may be local networks which may be geographically or socially defined, in which the concentration of NAS speakers is quite high, which may lead to the rise of NAS in specific groups before diffusing it to the wider dialect region. This is precisely the situation found in Labov et al. (2016) and reported in Chapter 2, which finds the highest concentration of NAS speakers amongst the graduates of elite public high schools, with other school networks lagging behind in the change to NAS.

6.4 Stability, Change, and Variation

So far, we have focused exclusively on what kind of input is necessary in what mixture for children to acquire a NAS grammar. However, the conclusion for the acquisition modelling is that across a broad range of mixtures, both PHL and NAS grammars are plausible. This raises two clear questions. First, is it possible that some learners acquire both PHL and NAS as a result of dialect contact? Second, once both grammars are in use within the speech community, is it inevitable that one should replace the other, as is being observed in Philadelphia?

Let's first consider the question of co-existing variation as the outcome of learning. There is considerable evidence that even for fully native bilingual speakers, one of the phonemic systems appears dominant (Cutler et al., 1989; Bosch and Sebastián-Gallés, 2003). The acquisition of the low-back merger system at the dialect boundary appears to be a case in point. At the beginning of this change, despite the presence of the merged system in their peer group, children retained the traditional distinct system. Once the merged system reached a certain level of prominence – above 23% – children acquired it en masse, resulting in the dramatic contrast in the vowel systems used by siblings separated by a few years as documented by Johnson (2010). However, the evidence provided in Chapter 4 suggests that for this allophonic restructuring in Philadelphia English, transitional cohort speakers do in fact acquire both PHL and NAS, and produce variation between the two systems as a whole. That Figure 6.2 finds both systems fully viable for such a wide overlap of NAS input (between 46% and 54% NAS input) provides a suggestion of the input data provided to the competing grammars speakers found in Chapter 4.

We now turn to the second question: if both NAS and PHL are viable, and speakers evidently acquire them, what is the long-term prognosis of this competition? Will one system necessarily replace the other, as it appears to be doing in Philadelphia? This question is quite different from the issues discussed so far in this chapter. We have mainly been concerned with the viability of a *single* system given a mixed environment. The Tolerance Principle based model has identified numerical conditions under which one grammar will replace the other as the terminal stage of language acquisition. It is a separate question entirely whether, having posited and acquired *two* competing systems, one will prevail. For the intermediate values of m, the learner can – and apparently does – acquire both systems, assigning a probabilistic distribution over them. Here, we have the more familiar sociolinguistic situation of variable rules, in which a speaker sometimes uses one variant of the allophonic system parameter and sometimes the other. The suitable mathematical model to study the dynamics of change is the variational model (Yang, 2000, 2002), where the terminal stage of language acquisition is a statistical distribution over two (or multiple) grammars. Language change is characterized by the change in this simulation over time, as governed by the differential utilities ("fitness") of the grammars in competition. Unlike the Tolerance Principle, which operates over *type* frequencies for rules and exceptions in the learner's lexicon and has a discrete outcome (whether a rule is tenable or not), the variational model presupposes the productivity of both rules and evaluates them on the basis of *token* frequencies.

The adaption of the variational model to a case of allophonic restructuring is not entirely straightforward. By the traditional formulation, this model evaluates the proportion of input produced by the each grammar that can be parsed by the other. The inevitable "winner" will be that grammar that can parse more of the other grammar's production (i.e., receives a lower penalty probability). The idea here is that many utterances will be compatible with either underlying grammar that is in competition, but that the few utterances that are *not* compatible with one of the two possible grammars generates a penalty probability for that grammar. Whichever grammar receives the lowest penalty probability will eventually win. This is visually represented in Figure 6.3, which displays the overlapping production of two mutually incompatible grammars (G_1 and G_2). In this visualization, G_2 will eventually win out over G_1, because it can analyze a higher proportion of the other's output.

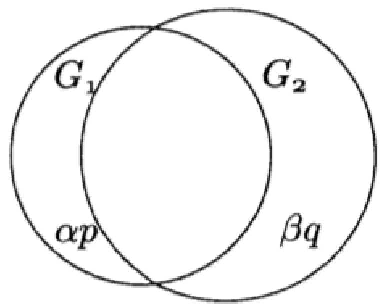


Figure 6.3: Two mutually incompatible grammars produce some proportion of ambiguous utterances. From Yang (2000).

This model has been successfully applied to syntactic parameters like the acquisition of a V2 grammar or pro-drop (Yang, 2000) as well as to phonological parameters like the LOT-THOUGHT merger (Yang, 2009), which produce assymetrical α and β values, resulting in an inevitable winning grammar. The problem of applying the variational model to the competition between PHL and NAS is that because any test token incompatible with PHL will be compatible with NAS and vice versa, here the penalty probabilities between the two grammars will be identical. While confusability can not be used here as a penalty probability, a potential direction for future research may lie in the social evaluation metrics reported in Chapter 5. PHL and NAS may be able to parse the exact same proportion of output, but they do not receive identical social evaluation scores. That structural sound change may be socially motivated has been a longstanding aspect of sociolinguistics (Labov, 1963); while future work may fruitfully apply the magnitude estimation scores of Chapter 5 to the variational model for the competition between PHL and NAS, this remains beyond the scope of the current dissertation.

6.5 Conclusion

The formulation of precise theoretical formulations such as the Tolerance Principle enables specific predictions, which hopefully lead in turn to theoretical advancement. In this chapter, we've demonstrated that applying the quantitative precision of the Tolerance Principle to the question of phonological change through language acquisition has allowed us to articulate a clearer model of the allophonic restructuring of /æ/ in Philadelphia in a way which would not be possible otherwise. Given a number of prima facie plausible hypotheses for the source of NAS innovation (grammar simplification, endogenous reanalysis, and dialect contact), we have been able to determine that only dialect contact is the likely source of this change.

We've investigated the possibility of NAS emerging in Philadelphia through regular transmission, finding that not only is NAS an implausible reanalysis of PHL input, but that it is also unlikely for NAS to have emerged through successive transmission simplifications of PHL. We've furthermore demonstrated that dialect contact is a far more likely source of NAS in Philadelphia, with the finding that NAS becomes a plausible analysis of mixed-environment input if that input is comprised of only 46% NAS. Importantly, it is not necessary for the *entire* speech community to be using NAS 46% of the time in order for NAS to make inroads into the speech community. Rather, it is only necessary for *some* learners to receive 46% NAS input.

This point bolsters the claim in Labov et al. (2016) that the shift from PHL to NAS is a change from above through dialect contact with NAS speakers who are unevenly distributed across social networks. Chapter 2 provides insight into the way educational systems in Philadelphia produce this uneven distribution, as well as the community level social characteristics that fit a classic change from above. Chapter 5 provides further evidence for this change as a change from above, with younger speakers in the Magnitude Estimation task rating all NAS-conforming tokens positively but rating tense PHL-conforming tokens negatively.

Finally, we've also found a relatively wide overlap in the tenability of PHL and NAS, with both systems completely viable when the input comprises between 46% and 54% NAS input. These findings are quickly turned into their own empirical predictions. We expect a child who is receiving less than 32% NAS input to posit PHL, and a child who is receiving more than 70% NAS input to posit NAS. A child receiving roughly 50% NAS input is expected to learn both systems and produce variation between the two. This predicts that a child with one NAS-speaking caregiver and one PHL-speaking caregiver who receives roughly equivalent input from both will emerge as a variable speaker, at least before they enter school and receive input from their peer group. We note that this prediction aligns neatly with the empirical results of Payne (1980), who found children with one PHL parent producing some /æ/ tokens that were inconsistent with PHL.

Chapter 7

Discussion and Conclusions

In this dissertation, I've taken a sociophonological approach to identifying the mechanism of phonological change for the allophonic restructuring of /æ/ in Philadelphia. Through a combination of community-wide corpus analysis, targeted interviews with individual speakers, experimental evaluation techniques, and computational simulations, I have presented a robust analysis of the sociolinguistic and phonological mechanisms and ramifications of a phonological change in Philadelphia English. I find that this phonological change is occurring through intraspeaker competing grammars. I argue, furthermore, that the systemic behavior of speakers

In Chapter 2, I provide background into the community-level variation in /æ/ systems. In this chapter, I showed that overall, each of the six primary conditioning factors of PHL and NAS participate in the change to NAS at the same time point in the community. Rather than change proceeding from context to context – as may have been predicted by coarticulatory effects of a following velar nasal, for example – we find instead that the changing conditioning factors take an abrupt turn together in the community, heading towards NAS beginning with speakers born around 1983. I demonstrate that among younger speakers, the educational system in Philadelphia creates social fragmentation of the community that has linguistic consequences. Special Admissions non-Catholic schools are at the forefront of this change to NAS, while Open Admissions Catholic schools are the conservative stronghold of PHL. I show, furthermore, that this change occurs in three steps intergenerationally. From a child's parental input, given sufficient peer influence, children can take a PHL input and produce a mixed PHL-and-NAS-system output or take a mixed input and produce a fully NAS output.

In Chapter 3, I provided a theoretical account for PHL as a productive allophonic rule, using the Tolerance Principle (Yang, 2016) as a diagnostic of productivity. I use the Tolerance Principle to define a clear definition to the phonological classification criterion of *Predictability*, and demonstrate that under this definition of predictability, a number of phonological relationships that have previously been analyzed as "intermediate" may be straightforwardly classified as productive. I show that under all considerations, PHL emerges as a plausible productive allophonic rule.

Chapter 4 provides an in-depth investigation into intraspeaker variation between PHL and NAS for the transitional cohort speakers. Using a combination of statistical, sociolinguistic, and phonological methods, I find that the mechanism of phonological change for the allophonic restructuring of /æ/ in Philadelphia is best analyzed as competing phonological grammars. Most of the speakers analyzed produce a PHL system (for those who have not yet undergone the change), a NAS system (for those who have completed the change), or grammar competition between the two systems as a whole. The few exceptional speakers are clearly analyzed as producing phonetic mitigation of their tense tokens as a reaction to the phonological change surrounding them rather than as a driver of that phonological change.

Given the finding that speakers produce variation between the abstract parameter governing PHL and the abstract parameter governing NAS, Chapter 5 turns to the question of whether these abstract parameters may be the target of social evaluation. In a Matched Guise task containing the same number of tense and lax tokens for the PHL and NAS guise, participants rated the PHL guise more *accented* than NAS. In a modified Magnitude Estimation task, I demonstrate that participants also produce systematic overt ratings of tokens of PHL and NAS. Younger participants born after the introduction of NAS into the community learn the traditional PHL evaluation system of downgrading tense PHL-consistent tokens, as found in Labov (2001), but have also adopted a second evaluation system of upgrading any tokens consistent with PHL. Older participants born before the introduction of NAS into the community also reproduce the findings in Labov (2001) by downgrading tense PHL-consistent tokens. However, when asked to rate NAS-consistent tokens, older participants' ratings are best analyzed as following abstract conditioning factors rather than phonetic production of those conditioning factors. In other words, older participants are found to produce negative evaluations of the *conditioning factors* rather than the phonetic realization of those conditioning factors. These results reveal a surprisingly abstract evaluation by older participants which only emerges when testing evaluations of a changing phonology. This surprising result demonstrates that sociolinguistic inquiry of a phonological change may reveal sociolinguistic facts that would otherwise be obfuscated by analyzing that variable during a period of stability.

Finally, in Chapter 6, I take on the question of whether the change from complex PHL to simple NAS was an inevitable outcome. Using the Tolerance Principle (Yang, 2016) as a measure of the plausibility of a phonological rule given a child's input, I demonstrate that NAS is not a plausible reanalysis of PHL, despite being a subset of the featural specification of PHL. Using a computational simulation, I furthermore demonstrate that successive simplifications of PHL leading to NAS is not a likely route by which this change occurred. Finally, I turn to dialect contact with NAS as a source for this phonological change, finding that NAS becomes a plausible reanalysis of the input when a child is receiving roughly 40% NAS input and 60% PHL input. Furthermore, both PHL and NAS emerge as fully viable when a child is receiving between 46 and 54% NAS input, suggesting an input profile that may account for the systemic variability between PHL and NAS found in the transitional cohort speakers. We find that the allophonic restructuring of Philadelphia English /æ/ is not an inevitable simplification of a complex rule, but more likely the result of a relatively high degree of dialect contact with NAS speakers for a particular subset of the community.

7.1 Similarity to Syntactic Change

In this dissertation, I've found evidence drawn from targeted recordings of transitional cohort speakers to support a competing grammars hypothesis for phonological change. The particular change investigated here, from a PHL parameter governing a complex set of conditioning factors to a NAS parameter, provides strong evidence for competing grammars in phonology, given that variation is found in all conditioning factors that differentiate PHL from NAS. As a mechanism of phonological change, this provides a challenge to the conventional wisdom of change driven by accruing phonetic errors. The parallels with Kroch (1989) are fairly striking, as at the time of writing, competing grammars provided a challenge to the widely accepted idea “that language change proceeds context by context, with new forms appearing first in a narrowly restricted context and spreading to others only later” (Kroch, 1989).

Here, I find similar evidence for competing grammars as the mechanism by which phonological change occurs, where change proceeds not context by context, but rather by intraspeaker competition of the two outputs of a single allophonic system parameter. This produces a probabilistic variation between PHL-consistent tokens and NAS-consistent tokens. As with syntactic change, this variation manifests within a single speaker and even within a single speech style. My findings here echo the results in Fruehwald (2013), who finds evidence for Kroch (1989)’s Constant Rate Hypothesis applying to phonological change, as well as the results in Fruehwald et al. (2013), who argue for competing grammars as the mechanism of change in Middle High German stop fortition.

That phonological change is found here to proceed by the same mechanism that syntactic change proceeds raises a number of additional questions. The first is whether the competing grammars found here is a mechanism of phonological change more generally or whether it is the mechanism by which change via dialect contact proceeds. This is a question that may only be answered with more investigations into community endogenous changes, which will require extensive corpora of speech in order to capture the community norms spanning the entire change. As more large-scale speech corpora are being built, this emerges as a possibility for future research. The second point is a more general theoretical one. There is no clear reason for phonological change to proceed by the same mechanism as syntactic change, yet I find evidence here that it does, which suggests that competing grammars may be the mechanism by which language in general changes.

7.2 Directions for Future Research

This dissertation represents one step towards an overall research program in phonological change. In it, I have demonstrated that an analysis of phonological change requires a robust understanding of the full set of sociolinguistic and phonological facts both on the macro-social (community) and micro-social (individuals and subsets of the community) level. I've also demonstrated the benefit that the study of sound change in progress provides to a larger understanding of phonological processes more generally as well as to our understanding of the sociolinguistic evaluation: by investigating changing norms in linguistic production and linguistic evaluation, we gain a deeper insight into the target of linguistic variation as well as the target of linguistic evaluation. While the work presented here provides a thorough sociophonological account of the allophonic restructuring in Philadelphia /æ/, it also paves the way for a number of future research directions.

First, in Chapter 6, we have presented specific numerical predictions about the acquisition of PHL and NAS under a mixed input. Namely, we have argued that a child receiving between 46% and 54% NAS input will be able to posit both systems. This predicts that a child with one NAS speaking parent and one PHL speaking parent who receives roughly equivalent input from both parents will acquire both systems, at least before their peers become a strong influence on acquisition. This prediction may also extend to school peer input – a child whose peers produce between 46% and 54% NAS input may be expected to acquire both systems, but a school environment that is tipped more strongly towards NAS or PHL predicts that child will only acquire one. Testing these predictions would provide important empirical support for the models presented in Chapter 6, though I note that because this change is rapidly coming to completion in the community, such an investigation must be conducted relatively soon.

The phonological representation of allophonic rules that I have argued for in Chapter 3 also generates predictions that may be tested empirically. That Chapter 4 finds speakers producing variation in the lexical exceptions provides one piece of support for such a representation. Future work may additionally benefit from more experimental approaches. For instance, Schuler et al. (2016) found experimental support for the claim that morphological rules follow the Tolerance Principle for productivity in an acquisition-like period. Schuler et al. (2016) found that in an artificial language experiment, children (aged 5-6) given a rule with greater than θ_N exceptions do not form a productive rule while children given a rule with fewer than θ_N exceptions do form a productive rule. This work could be extended to test the limit of lexical exceptions for phonological rules as well. If the Tolerance Principle should replace the traditional definition of *Predictability*, as I have argued in Chapter 3, then an artificial language experiment conducted on phonological processes should exhibit the same patterns found for morphological productivity.

Additionally, community-level language change is not the only locus of phonological variation that may be investigated for an individual speaker. As very young children acquire language, articulatory constraints result in different stages of child phonology. During acquisition, children must acquire both the abstract phonological features of their adult phonology as well as the articulatory capabilities of producing that phonology. As children mature from child phonology to a more adult-like phonology, it may be the case that the transition between the two occurs via grammar competition as well. Becker and Tessier (2011) provide some support for this idea, finding that during acquisition, Trevor (Comppton and Streeter, 1977) produces variation between consonant harmony and faithful productions for non-harmonious lexical items (e.g., *goat, cat, duck*). Becker and Tessier (2011) analyze this as variation that occurs when Trevor acquires a new constraint in his phonology, though they name it as the effect of stored lexical productions rather than variable grammars. If competing grammars is the mechanism by which longitudinal phonological change occurs, it follows that competing grammars may also be the mechanism by which children develop their adult-like competencies. If this is the case, it predicts a Constant Rate of development across all contexts affected by the child phonology in question, as well as a bimodal distribution of production between the child phonology and the more adult phonology parameter.

Finally, it is my hope that this dissertation may serve as an example of a return to the study of variables as a structural unit. Labov's original formulation of the linguistic variable, as outlined in Labov (1966) "The Linguistic Variable as a Structural Unit", conceives of the linguistic variable as a systemic property. In discussing variable non-rhoticity in New York English, (Labov, 1966, pg. 6) describes the variability not as variation between two segments /ɹ/ and /ø/, but rather as "the oscillation of entire phonemic categories: the set of ingliding phonemes appears and disappears as a whole." In other words, Labov analyzed speakers as varying between one phonemic *system* which includes ingliding phonemes (the vocalized variants of /r/ nuclei) and a second system that does not include ingliding phonemes, capturing the vocalic variation that accompanies /r/-vocalization in

New York English as well as the /r/-vocalization itself. In this dissertation, I approach the variation in /æ/ as a systemic variable as well. Analyzing the variation between PHL and NAS as grammar competition between a single parameter that governs /æ/ allophony, both on the community level as well as the individual level, provides the best explanatory account for the data, and produces additional testable hypotheses for both sociolinguistic variation and phonological architecture.

Appendix A

Lexical Exceptions for Traditional PHL

My full formulation for lexical exceptions in PHL is provided in (27), which orders all lexical exception according to word frequency as measure in the SUBTLEX-US corpus. Words that vary from speaker to speaker as to whether they are exceptional are denoted with an asterisk. We find, for example, wide variation in production of diminutive names (e.g. *Danny*, *Annie*), which I have listed here as an exception to lax because as noted in Chapter 3, children acquire a productive diminutive suffix -y relatively early.

(27) **PHL:**

1. IF w = *and* THEN /æ/ → lax
2. IF w = *can* THEN /æ/ → lax
3. IF w = *an* THEN /æ/ → lax
4. IF w = *am* THEN /æ/ → lax
5. IF w = *than* THEN /æ/ → lax
6. IF w = *bad* THEN /æ/ → tense
7. IF w = *glad* THEN /æ/ → tense
8. IF w = *mad* THEN /æ/ → tense
9. IF w = *ran* THEN /æ/ → lax
10. IF w = *Danny** THEN /æ/ → lax
11. IF w = *program** THEN /æ/ → lax
12. IF w = *planet** THEN /æ/ → tense

13. IF w = *Annie** THEN /æ/ → lax
14. IF w = *began* THEN /æ/ → lax
15. IF w = *Africa* THEN /æ/ → lax
16. IF w = *math** THEN /æ/ → lax
17. IF w = *Sammy** THEN /æ/ → lax
18. IF w = *exam* THEN /æ/ → lax
19. IF w = *nanny* THEN /æ/ → lax
20. IF w = *candidate** THEN /æ/ → lax
21. IF w = *Granny* THEN /æ/ → tense
22. IF w = *aspirin* THEN /æ/ → lax
23. IF w = *Fanny** THEN /æ/ → lax
24. IF w = *astronaut** THEN /æ/ → lax
25. IF w = *Nana* THEN /æ/ → tense
26. IF w = *alas** THEN /æ/ → lax
27. IF w = *aft* THEN /æ/ → lax
28. IF w = *swam* THEN /æ/ → lax
29. IF w = *asteroid* THEN /æ/ → lax
30. IF w = *Daffy** THEN /æ/ → lax
31. IF w = *Grammie* THEN /æ/ → tense
32. IF w = *afro* THEN /æ/ → lax
33. IF w = *asphalt* THEN /æ/ → lax
34. IF w = *affirmation* THEN /æ/ → lax
35. IF w = *asterisk* THEN /æ/ → lax
36. IF w = *badminton** THEN /æ/ → tense
37. IF w = *aspirate* THEN /æ/ → lax
38. IF w = *carafe* THEN /æ/ → lax
39. IF w = *gaffe* THEN /æ/ → lax
40. $\text{æ} \rightarrow \text{æh} \;/\; \_[\;\text{+anterior}\;] \cap ([\;\text{+nasal}\;] \cup \begin{bmatrix}\text{-voice}\\ \text{+fricative}\end{bmatrix})]\,\sigma$

All /æl/ words found in CHILDES are listed below. While I do not count these words as lexical exceptions to PHL for reasons discussed in Chapter 3, I include them here for completeness.

Word	N
Al	1518
ala	7
Albert	46
album	10
albums	4
Albuquerque	31
alcohol	10
Aleck	3
Alex	238
alfalfa	2
Alfred	8
algae	3
Alice	59
Alison	7
alkaseltzer	2
Allen	254
allergy	3
alley	10
alligator	335
alligators	54
ally	13
alphabet	116
alphabets	32
alphabits	4
Al's	15
alto	4
Alvin	17
balance	92
balanced	9
balances	5

Table A.1: /æl/ words.

Word	N
balancing	18
balcony	16
balla	2
calculator	20
calendar	62
callous	2
calorie	2
calories	8
calvary	5
Calvin	9
challenge	5
Dallas	10
falcon	36
gal	2
galaxy	5
gallery	4
galley	2
gallon	3
gallop	16
galloping	3
immortality	2
Italian	37
Hal	9
hallo	5
malapropism	2
Malcolm	3
Malik	24
mallard	30
mallards	2
mallet	4

Table A.2: /æl/ words.

Word	N
medallion	2
pal	43
palace	63
palaces	2
Palo	5
pals	8
personality	3
rally	2
Ralph	16
reality	2
Sal	4
salad	230
salads	3
Salazar	2
Sally	446
scalps	2
shall	1734
shallots	2
shallow	12
talent	6
talented	6
talon	3
Val	10
Valentine	225
Valentine's	40
Vallerie	22
valley	22
valuable	2
valve	2

Table A.3: /æl/ words.

Appendix B

Alternative Methods for Token Classification

B.1 Clustering algorithms

K-means Clustering K-means clustering is a simple unsupervised learning algorithm that clusters observations into *k* clusters, through minimizing the within-cluster sum of squares and maximizing the between-cluster sum of squares. With the data analyzed here, k-means clustering could potentially identify underlying clusters of tokens, enabling us to then identify (1) the tenseness value of tokens and (2) whether each cluster contains only PHL-tense tokens (in the case of spontaneous phonologization) or tokens from each test condition (in the case of competing grammars). One downfall of k-means clustering is that *k* must be set *a priori*, and there is not a statistical method for determining the optimal number of clusters. This is often done visually through an "elbow plot" method, which plots the decrease in variance captured by the clusters as the number of clusters increases. K-means was tested here on tokens using just F1 and F2 values (as these are the primary perceptual indicators of tenseness), as well as on the output of PCA and t-SNE analysis (as these methods enable the incorporation of all measurements).

Algorithmically, a k-means algorithm assigns each observation to the cluster whose mean has the least squared Euclidean distance (B.1), then updates the new means to be the centroids of the observations in the new clusters (B.2).

$$S_i^{(t)} = \{x_p : \| x_p - m_i^{(t)} \|^2 \leq \| x_p - m_j^{(t)} \|^2 \; \forall j, 1 \leq j \leq k\} \tag{B.1}$$

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \tag{B.2}$$

Hierarchical Clustering Hierarchical clustering is a method of cluster analysis that builds a hierarchy of clusters. Here, I use the `hclust()` function in R, which applies an agglomerative hierarchical clustering. Each observation is first counted as its own cluster, then pairs of similar clusters are merged as the hierarchy is built up. For clustering /æ/, tokens were merged according to similarity as measured by complete linkage, shown in (B.3).

$$\max\{d(a,b) : a \in A, b \in B\} \tag{B.3}$$

B.1.1 Applying Clustering Algorithms to F1 and F2 values

Figures B.1 and B.2 display the results of K-means clustering (right panel) and Hierarchical clustering (left panel) for our simulated PHL speaker and NAS speaker. Because the simulated data is constructed using known underlying phonological values, this enables us to identify where the clustering algorithms have assigned specific tokens to the wrong cluster, shown in red.

Figure B.3 displays the results of the clustering algorthims on F1 and F2 for the Cohort 3 Competing Grammars speaker. Again, we can identify the inaccurately classified tokens, because the simulated data contains information about whether any individual token was drawn from the PHL sample or the NAS sample. Inaccurate tokens are displayed in red.

For analyzing the production of phonological change via phonetic incrementation, we turn to the Cohort 3 production of the test-token phonetic incrementation speaker. We use this version of phonetic incrementation because it is the most difficult to distinguish from a competing grammars analysis of sound change, so it is crucial to obtain a classification method that can distinguish between these two mechanisms of sound change in the actual data. Because change via phonetic

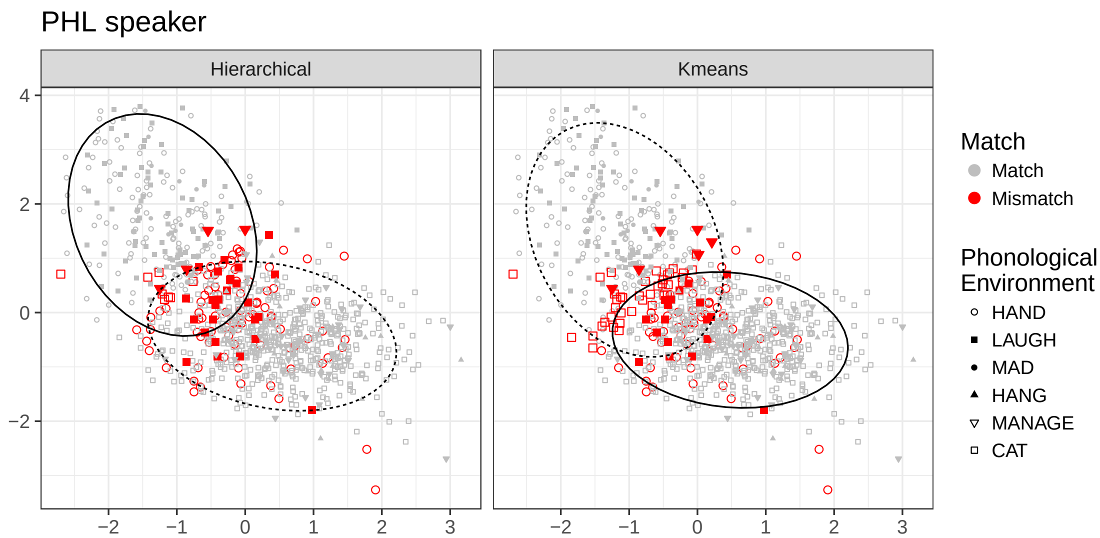


Figure B.1: Accuracy of clustering algorithms for PHL speaker F1 and F2. Red tokens display inaccurately classified tokens.

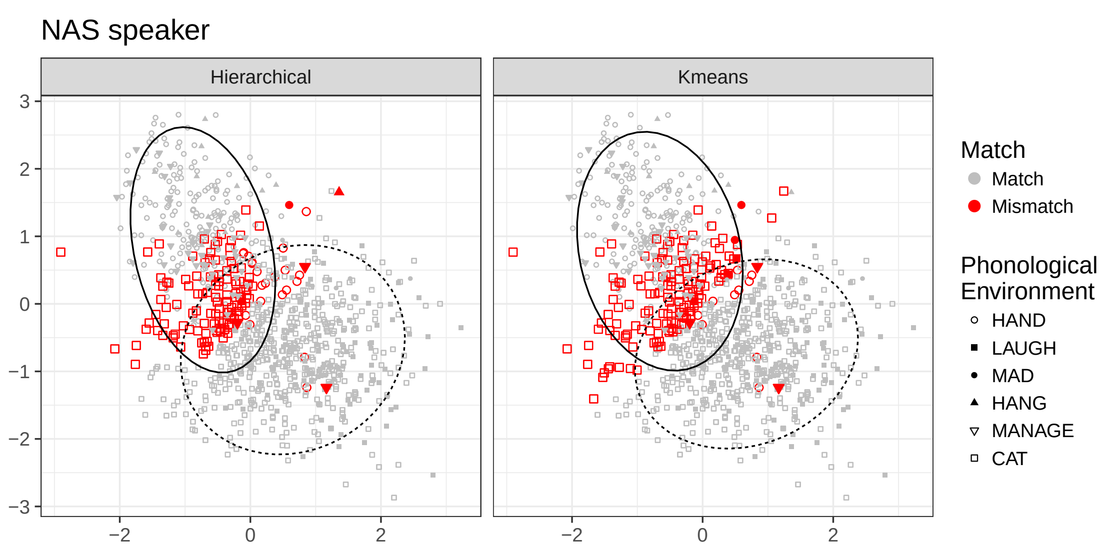


Figure B.2: Accuracy of clustering algorithms for NAS speaker F1 and F2. Red tokens display inaccurately classified tokens.

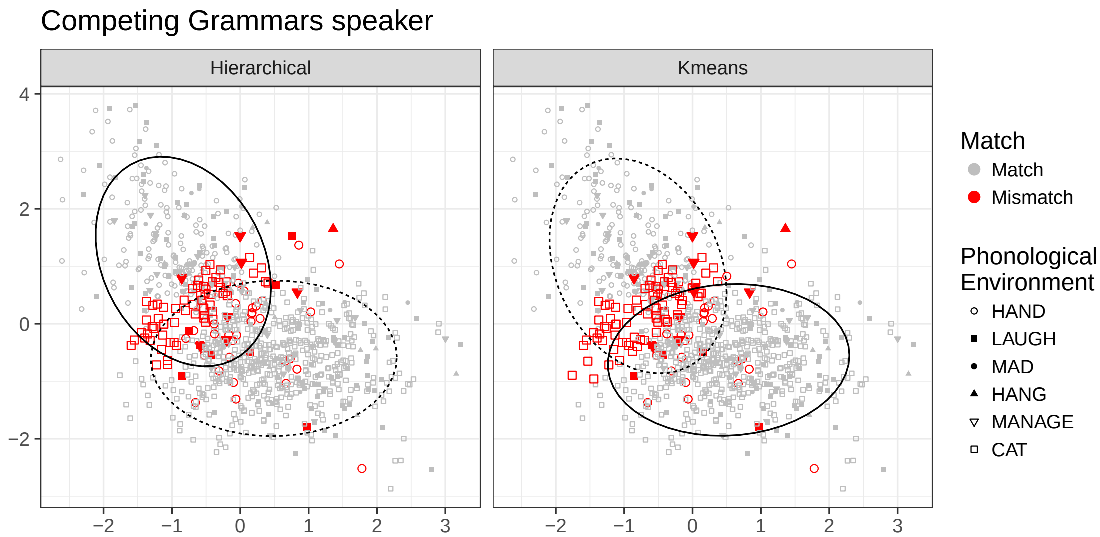


Figure B.3: Accuracy of clustering algorithms for Competing Grammars speaker F1 and F2. Red tokens display inaccurately classified tokens.

incrementation does not assume any underlying phonological reasons for a change, it is not possible to identify whether tokens classified by either K-means or Hierarchical algorithms are accurate (given that there is no underlying phonological classification to the tokens in the first place).

The one expectation that can be made for a phonetic incrementation transitional cohort speaker is that they would produce test tokens as a distinct phonetic target from their HAND and CAT classes. In other words, any clustering algorithm set to find three clusters should identify HAND as one cluster, CAT as a second cluster, and all test tokens as a third cluster. Figure B.4 displays the results of a K-means (right) and Hierarchical (left) clustering model set at $k = 3$ for the simulated phonetic incrementation speaker.

As we can see in Figure B.4, neither the Hierarchical model nor the K-means model selects test tokens accurately as belonging to an intermediate cluster of tokens. Similarly, the clustering algorithms for the PHL and NAS speakers produce a fairly high rate of misanalyzed tokens near the overlapping space between PHL and NAS, where the glm classifier produced between 3 (for the NAS data) and 7 (for the PHL data) inaccurately classified tokens.

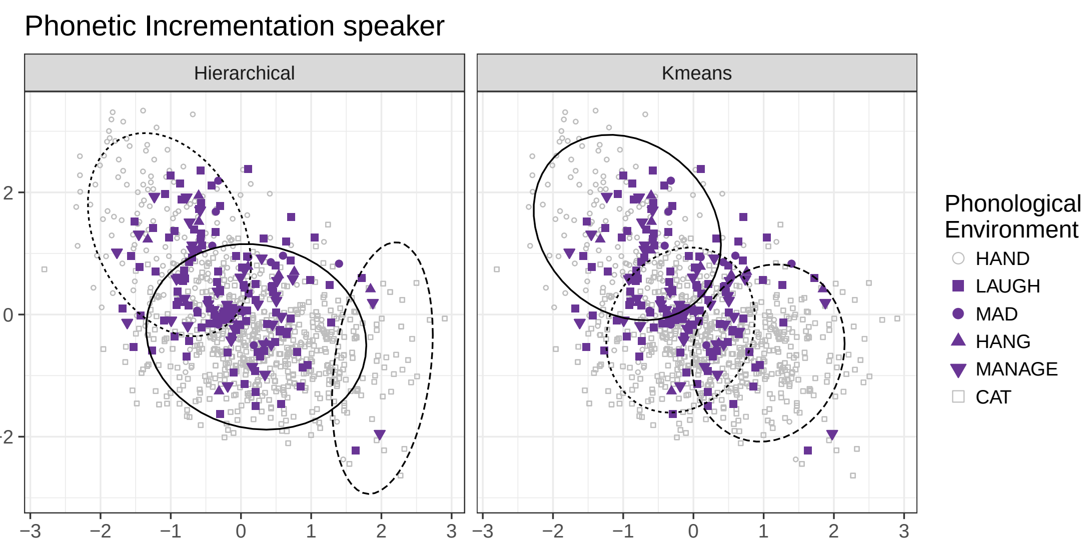


Figure B.4: Accuracy of clustering algorithms for Phonetic Incrementation speaker F1 and F2. Purple tokens display test tokens, which should fall within a single cluster.

B.2 Dimensionality Reduction

While F1 and F2 serve as the primary acoustic cues for tenseness, we must also consider the possibility that additional parameters contribute to a token's identity as tense or lax, and that the misanalyses presented above are the result of taking only two dimensions of tenseness into account. Indeed, in the glm classifier that emerges as the best classifier of token tenseness, F3, duration and stress all factor into the classification of tokens as tense or lax. In addition to the glm classifier described in Chapter 4, I also tested whether a K-means and Hierarchical clustering algorithm accurately classified the data when it had been submitted to a dimensionality reduction algorithm. For this, simulated data for a competing grammars speaker and a phonetic incrementation speaker were created that included a simulated duration for each token, calculated using the covariance matrices for F1, F2, and duration.

Principal Components Analysis Principal Components Analysis (PCA) is an unsupervised dimensionality reduction algorithm that reduces a set of observations into linearly uncorrelated variables, or *principal components*. The first principal component accounts for the highest variance. After this, the second component accounts for the highest of the remaining variance. In theory, a PCA analysis would be able to determine the similarity of test tokens and training tokens along all relevant measurement dimensions and produce groupings that cluster test tokens either as part of the HAND or CAT class underlyingly or as phonetically distinct. The resulting data can then be submitted to a K-means and a Hierarchical cluster algorithm.

T-distributed stochastic neighbor embedding T-distributed stochastic neighbor embedding (t-SNE) provides a type of dimensionality reduction similar to PCA. t-SNE creates a probability distribution over pairs of observations in high-dimensional space. This results in a set of probabilities $p_{i,j}$ that represent the similarity of observations x_i and x_j. Based on these probabilities, t-SNE produces a k-dimensional map of clusters (typically set to $k = 2$), which can then be either visually distinguished or clustered by K-means or Hierarchical analysis.

B.2.1 Applying Clustering Algorithms to Reduced Dimension Data

In what follows, I present accuracy plots for clustering algorithms run on PCA and t-SNE transformed data. The resulting plots display high levels of inaccuracy for clustering dimensionality reduced data for all simulated speakers.

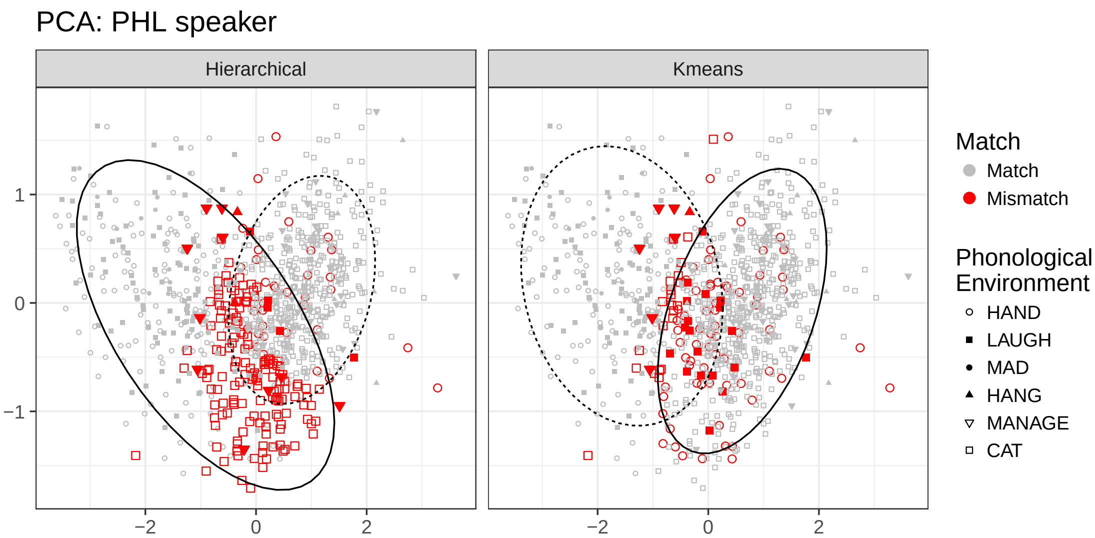


Figure B.5: Accuracy of clustering algorithms for PHL speaker PCA data. Red tokens display inaccurately classified tokens.

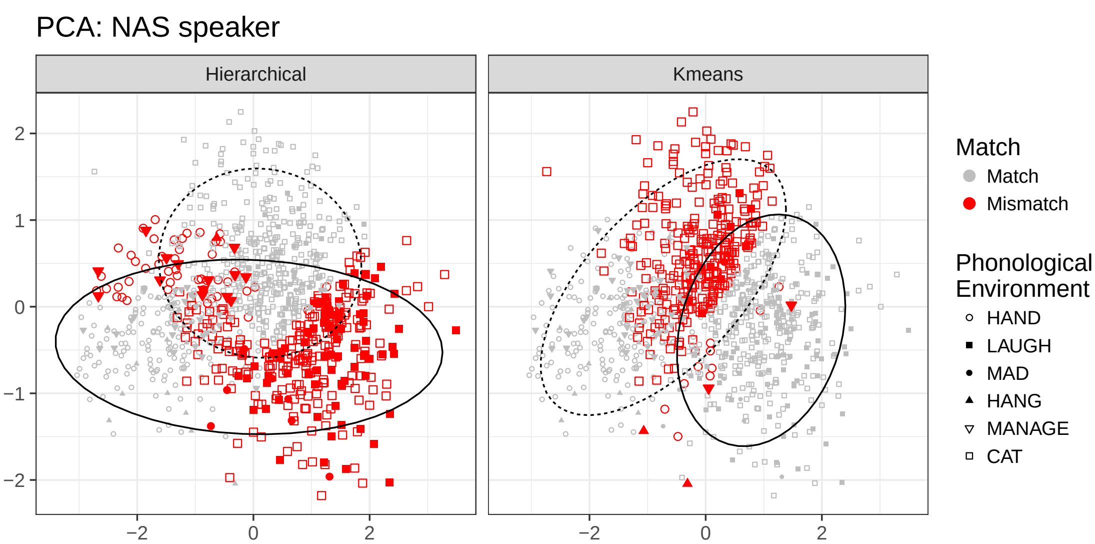


Figure B.6: Accuracy of clustering algorithms for NAS speaker PCA data. Red tokens display inaccurately classified tokens.

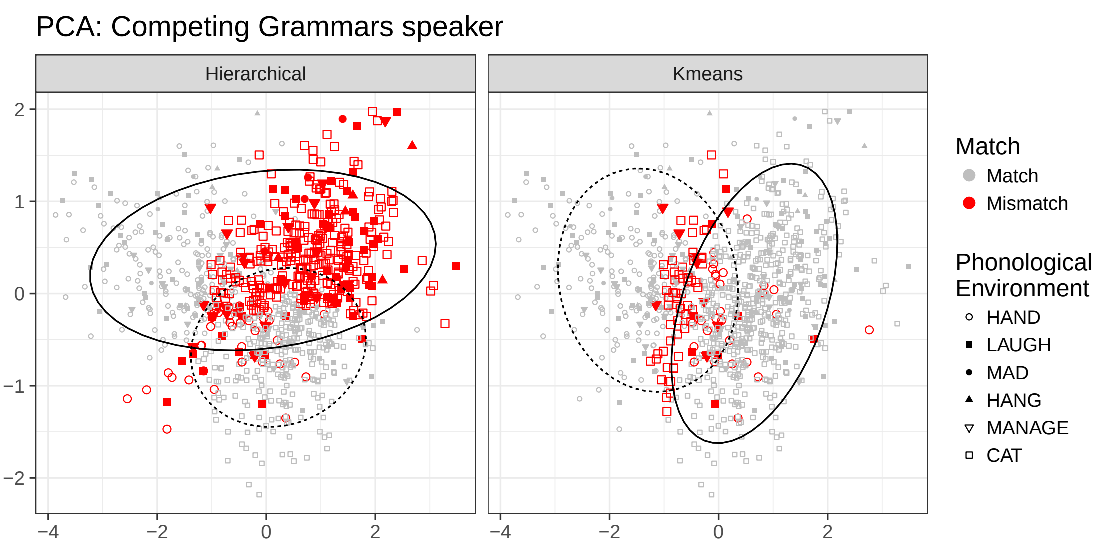


Figure B.7: Accuracy of clustering algorithms for Competing Grammars speaker PCA data. Red tokens display inaccurately classified tokens.

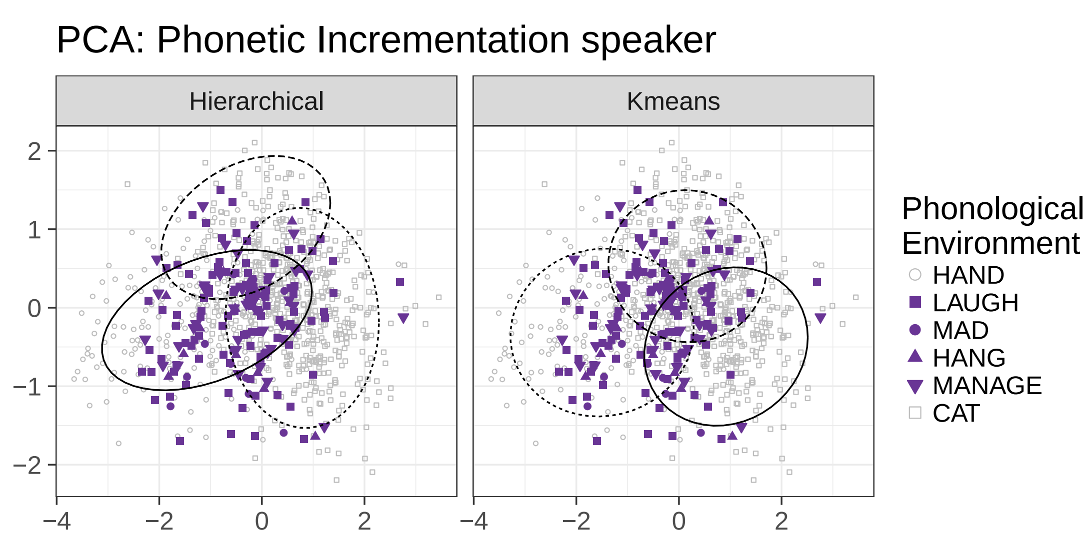


Figure B.8: Accuracy of clustering algorithms for Phonetic Incrementation speaker PCA data. Purple tokens display test tokens, which should fall within a single cluster.

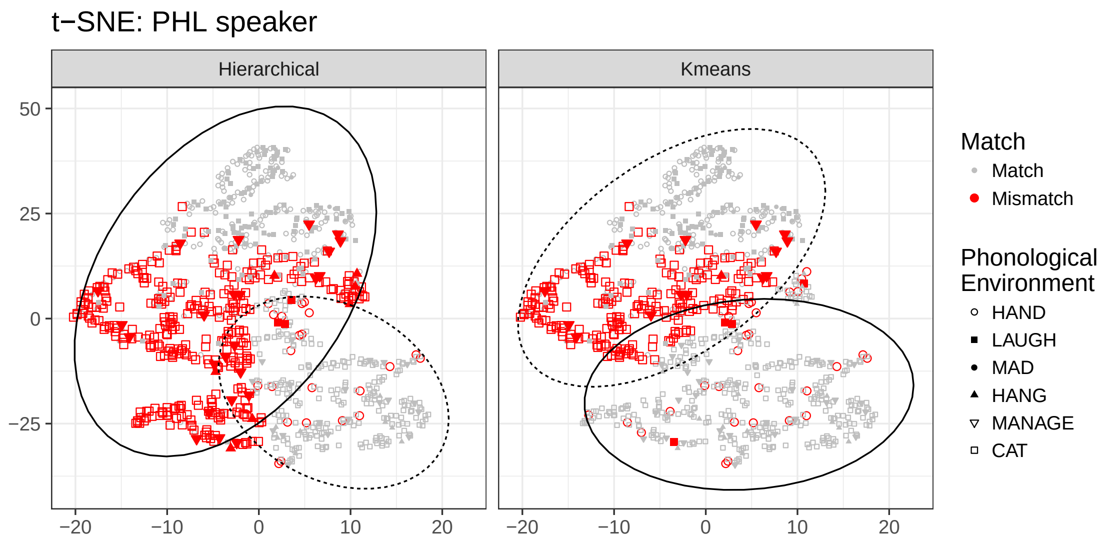


Figure B.9: Accuracy of clustering algorithms for PHL speaker t-SNE data. Red tokens display inaccurately classified tokens.

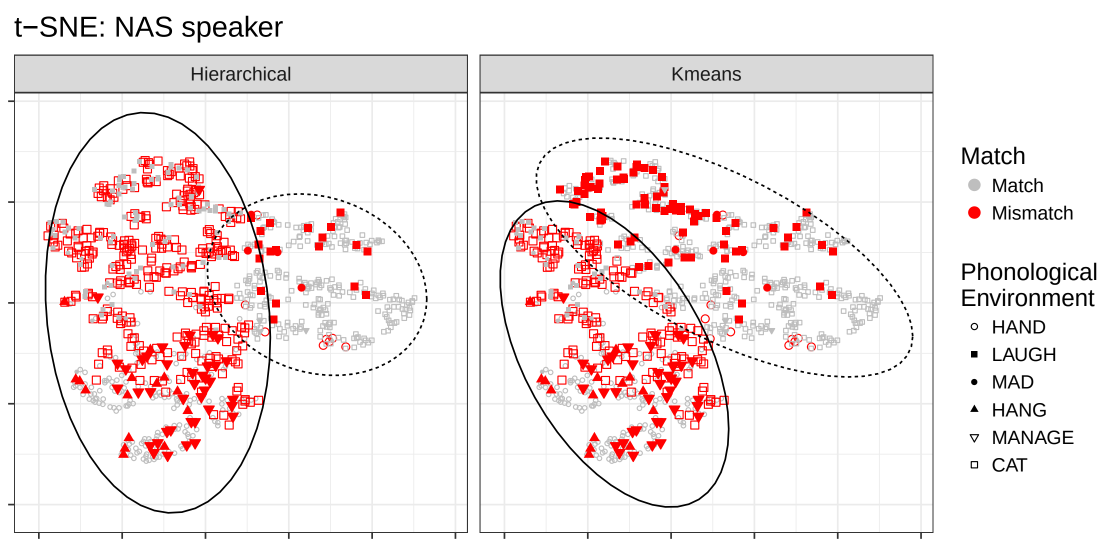


Figure B.10: Accuracy of clustering algorithms for NAS speaker t-SNE data. Red tokens display inaccurately classified tokens.

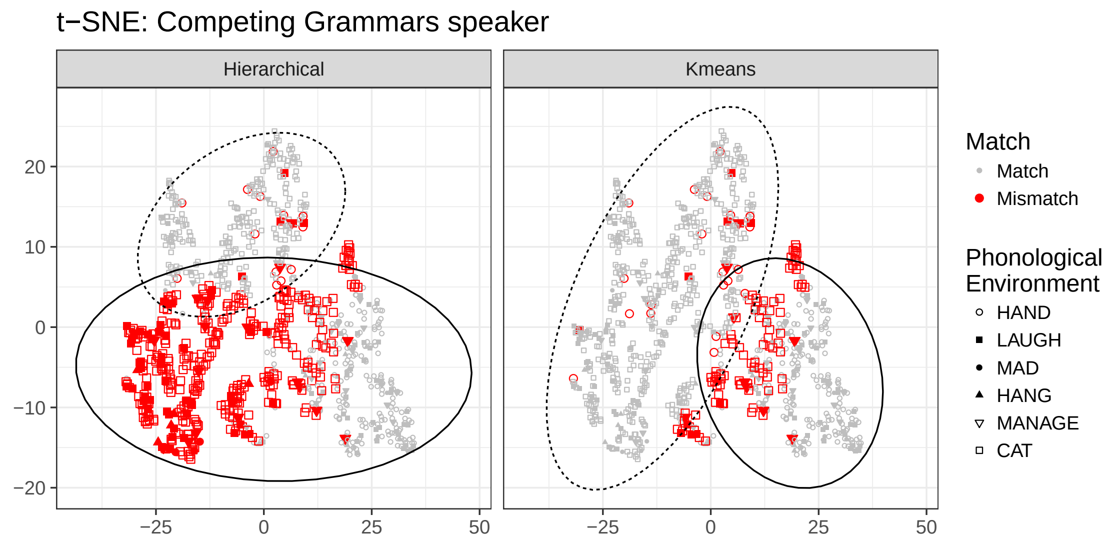


Figure B.11: Accuracy of clustering algorithms for Competing Grammars speaker t-SNE data. Red tokens display inaccurately classified tokens.

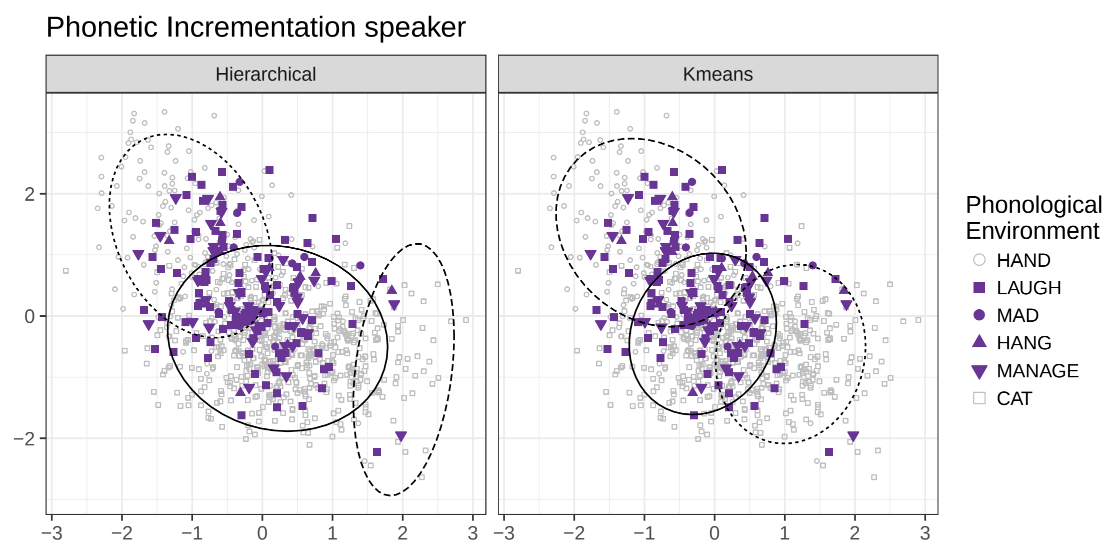


Figure B.12: Accuracy of clustering algorithms for Phonetic Incrementation speaker t-SNE data. Purple tokens display test tokens, which should fall within a single cluster.

Appendix C

IMPC Methods

marry	dog	bother	stock
hassit	hand	Miami	planet
stalk	personality	sauce	ask
father	dad	corner	Don
collar	Murray	awed	big
pacify	calm	alas	tiny
am	trannel	prass	Alice
merry	spider	ham	chocolate
Snyder	log	path	have
I can	ice	class	tiger
wide	very	bank	baff
nath	Friday	classic	palm
gas	Girard	white	sad
rider	tot	cash	right
dawn	league	Spanish	law
bang	croth	furry	Mary
and	athlete	pal	tin can
odd	down	bad	aspirin
valley	taught	asterisk	glad
angle	mouth	ferry	crown
classify	manage	man	south
caller	ride	lang	pass
groll	eyes	toss	crayon
Charlie	half	math	nearer
salve	mad	hammer	salmon

Table C.1: IMPC wordlist.

Conversational prompts

How did you two meet? Did you become friends right away?

What did you do for fun when you were a kid? Were your parents strict? Did you play with the kids on your street? Were there games you liked to play with your friends/neighbors?

Does school/work stress you out? What about the election? What do you do for fun and to de-stress?

When is the last time you got really mad? Have you two ever gotten into a fight with each other? What do you do when you're angry?

What about the last time you were embarrassed? Do you remember a time that one of your friends did something really embarrassing?

What's the last time that you remember feeling scared? Is there anything that makes you feel like you're going to panic?

Do you remember the 90s? What about the 2000s? What kind of trends do you associate with being a child of that decade? (Clothes? Music?)

Have you maintained strong relationships with the people who you grew up with? Or are you making new friends as an adult? Is it harder to make new friends as an adult?

Do you like the idea of traveling or would you rather stay home? Have you been anywhere cool? Where would you go if you could go anywhere in the world for free?

Figure C.1: IMPC Conversational Prompts

Appendix D

Production Plots for IMPC and IHELP Speakers

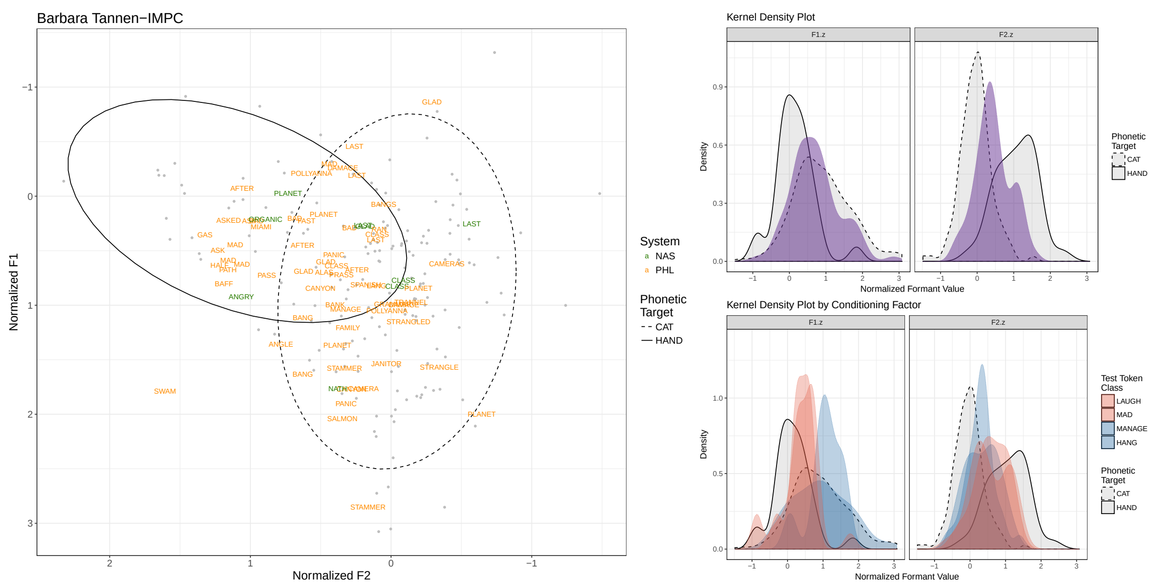


Figure D.1: Barbara Tannen, PHL

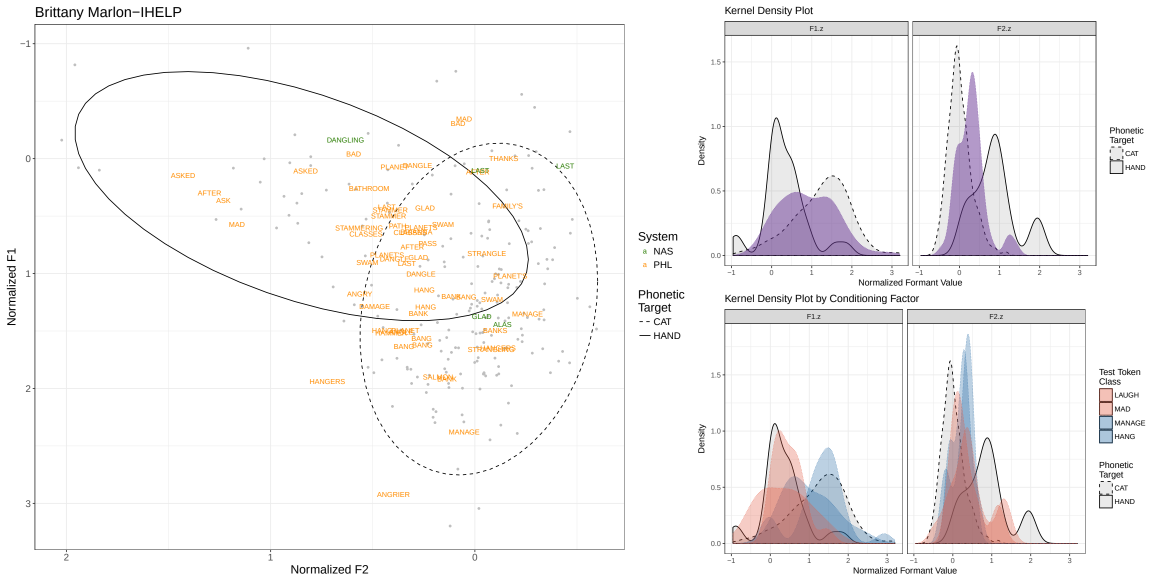


Figure D.2: Brittany Marlon, PHL

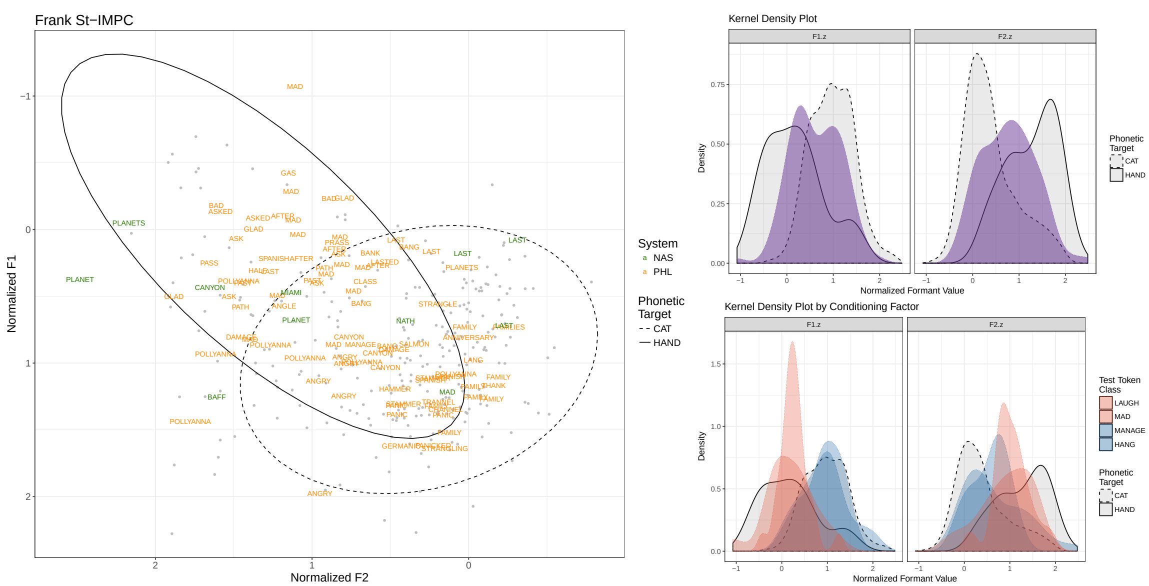


Figure D.3: Frank St, PHL

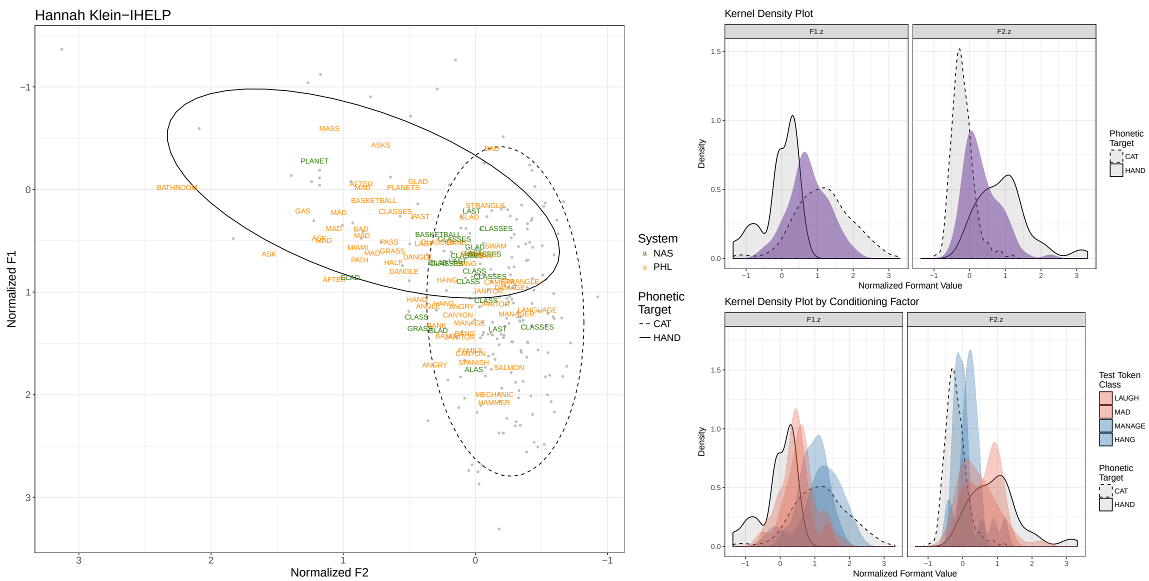


Figure D.4: Hannah Klein, PHL

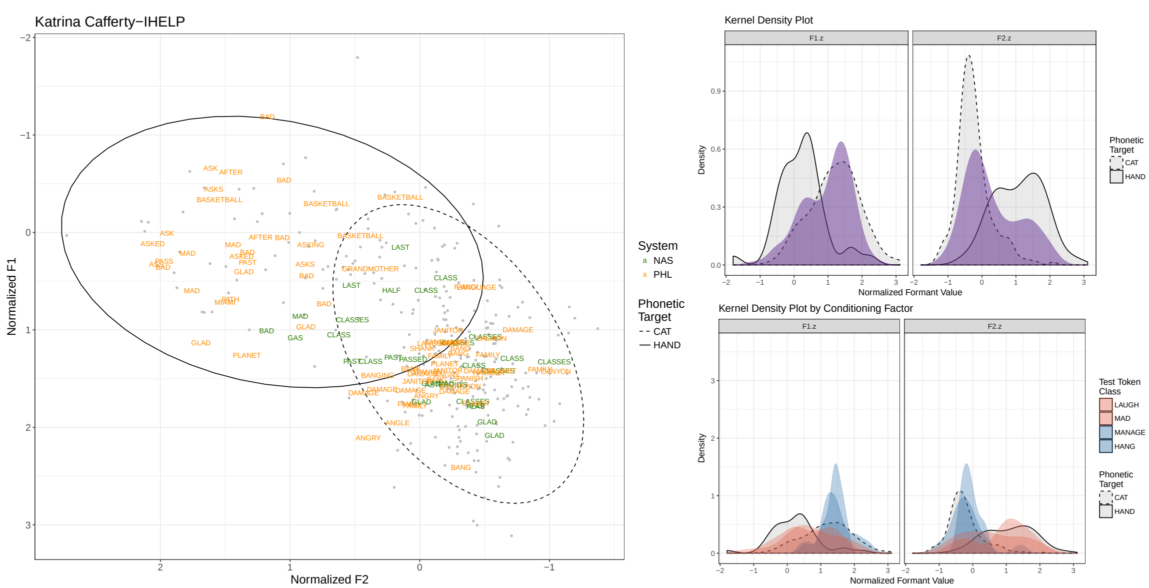


Figure D.5: Katrina Cafferty, PHL

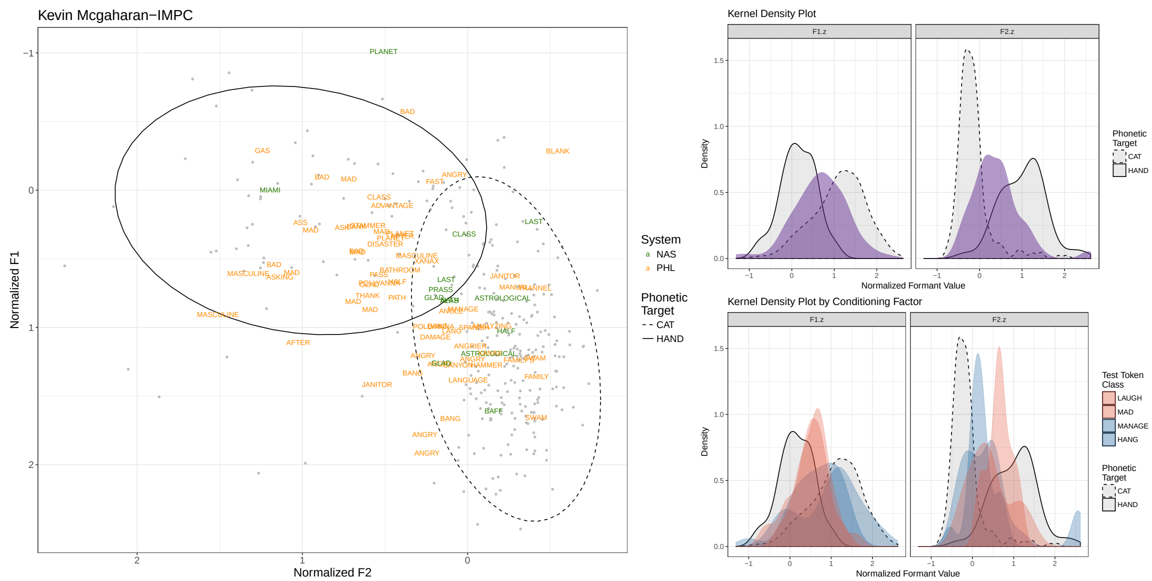


Figure D.6: Kevin Mcgaharan, PHL

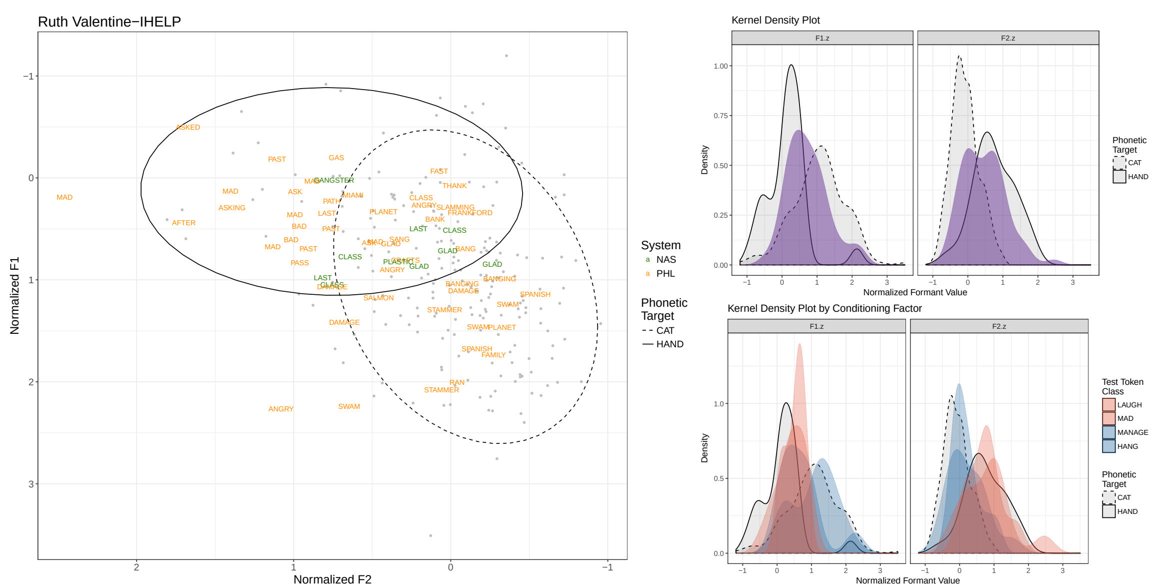


Figure D.7: Ruth Valentine, PHL

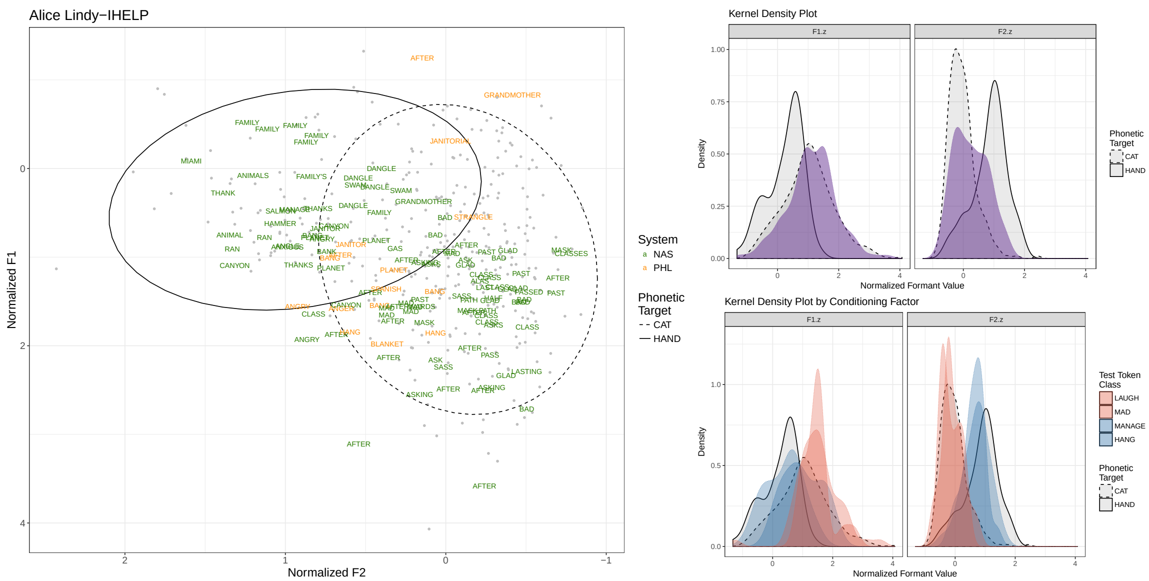


Figure D.8: Alice Lindy, NAS

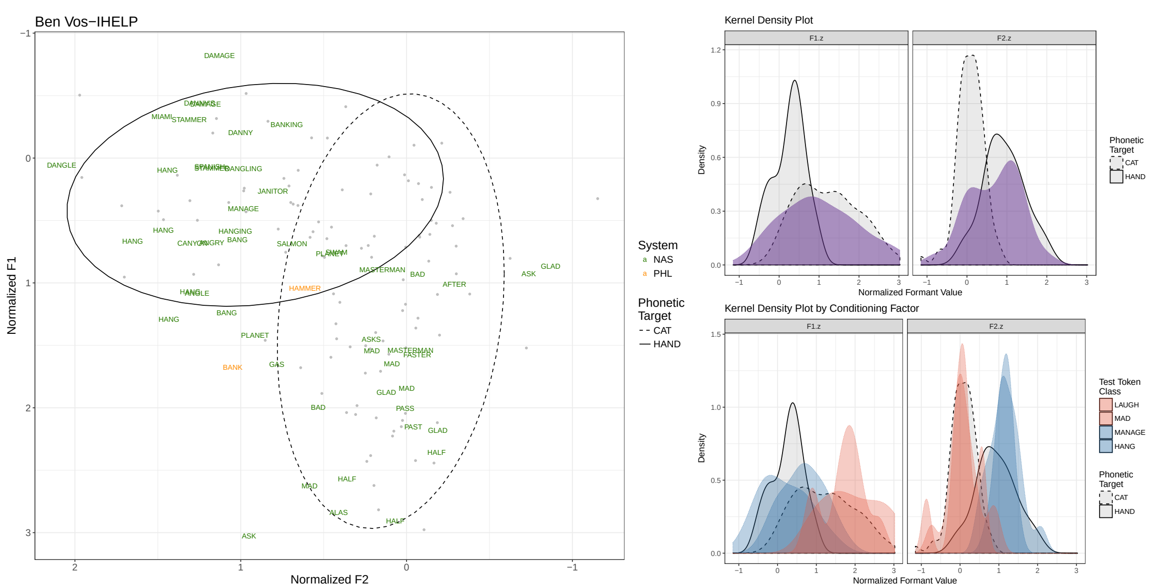


Figure D.9: Ben Vos, NAS

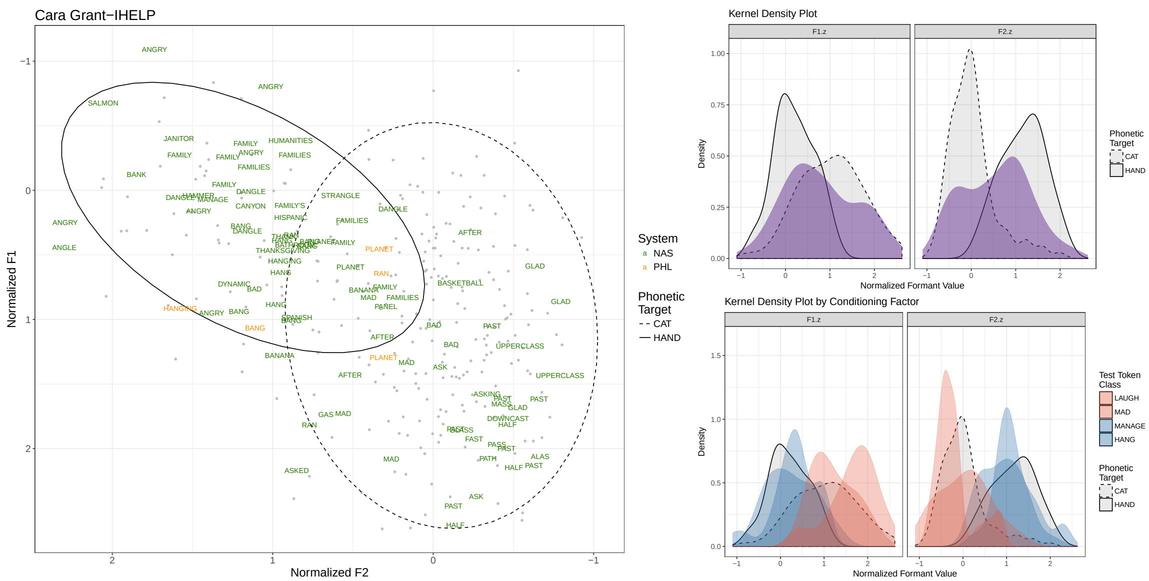


Figure D.10: Cara Grant, NAS

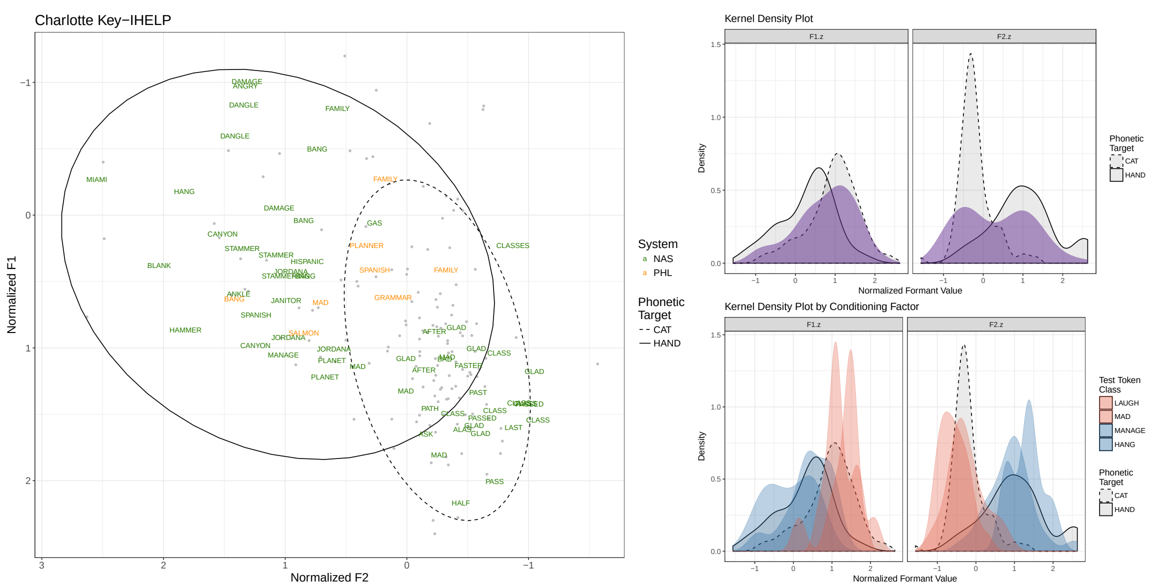


Figure D.11: Charlotte Key, NAS

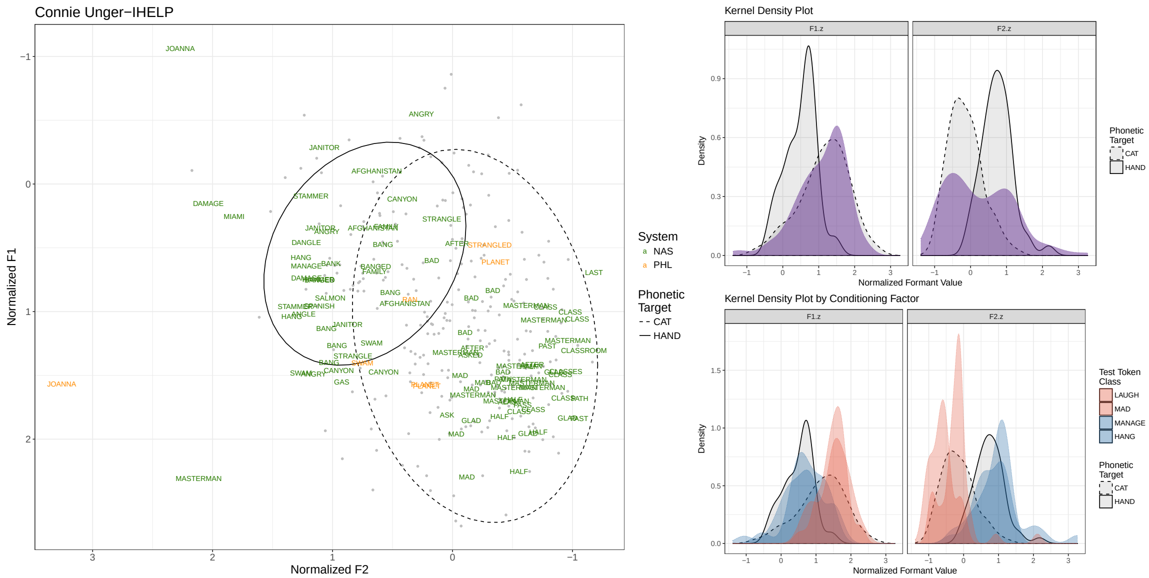


Figure D.12: Connie Unger, NAS

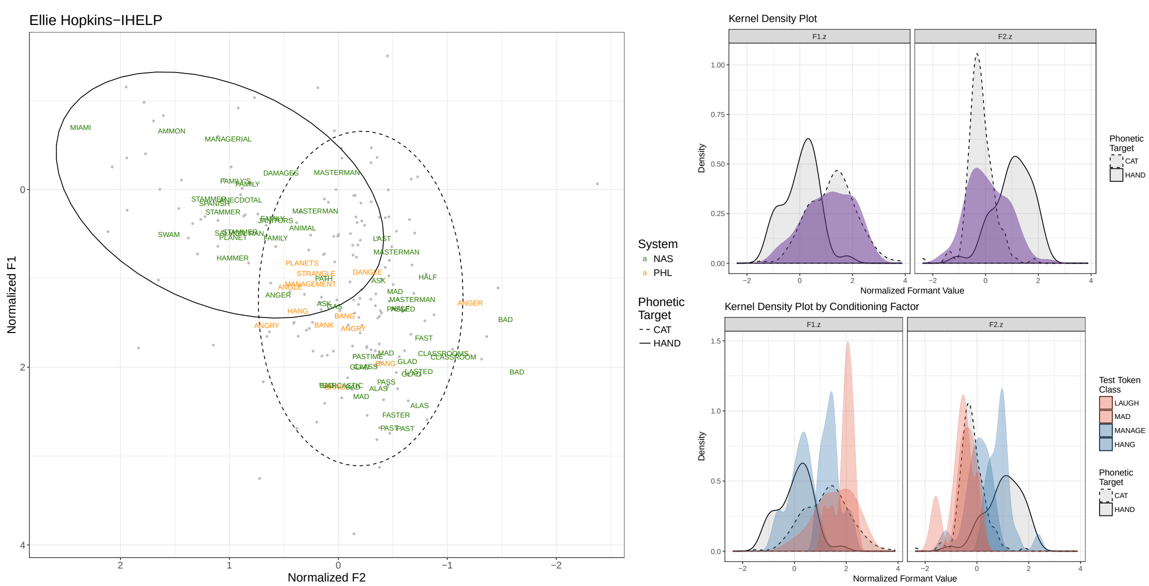


Figure D.13: Ellie Hopkins, NAS

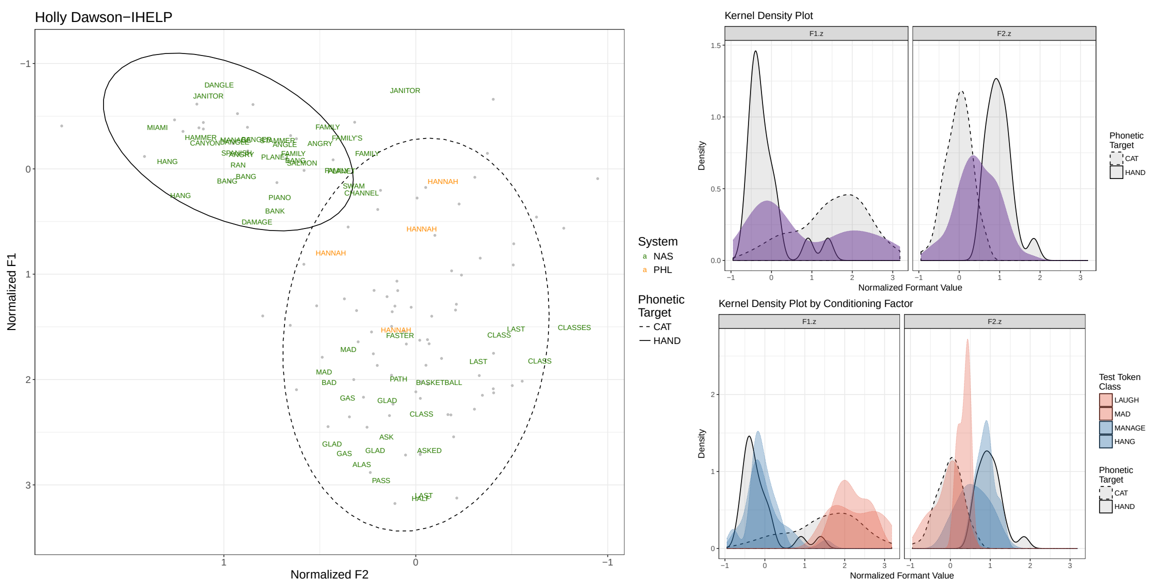


Figure D.14: Holly Dawson, NAS

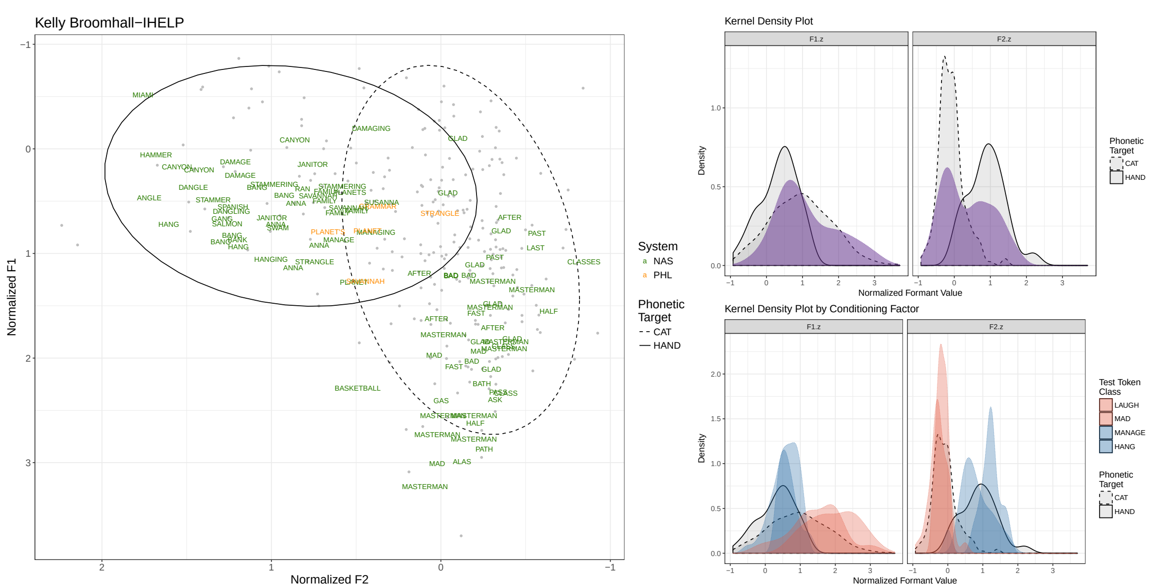


Figure D.15: Kelly Broomhall, NAS

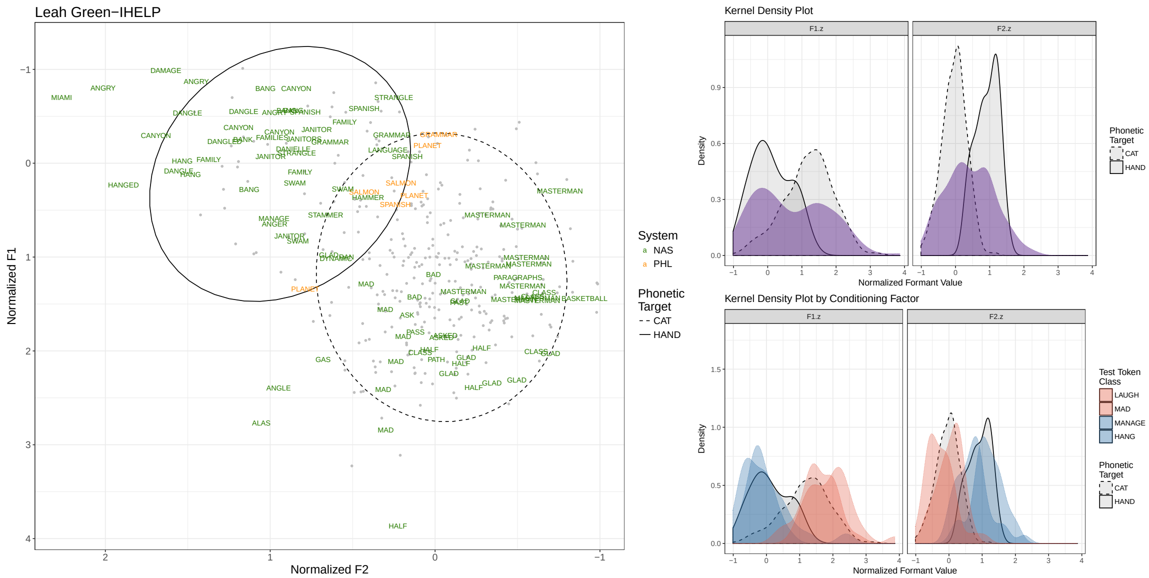


Figure D.16: Leah Green, NAS

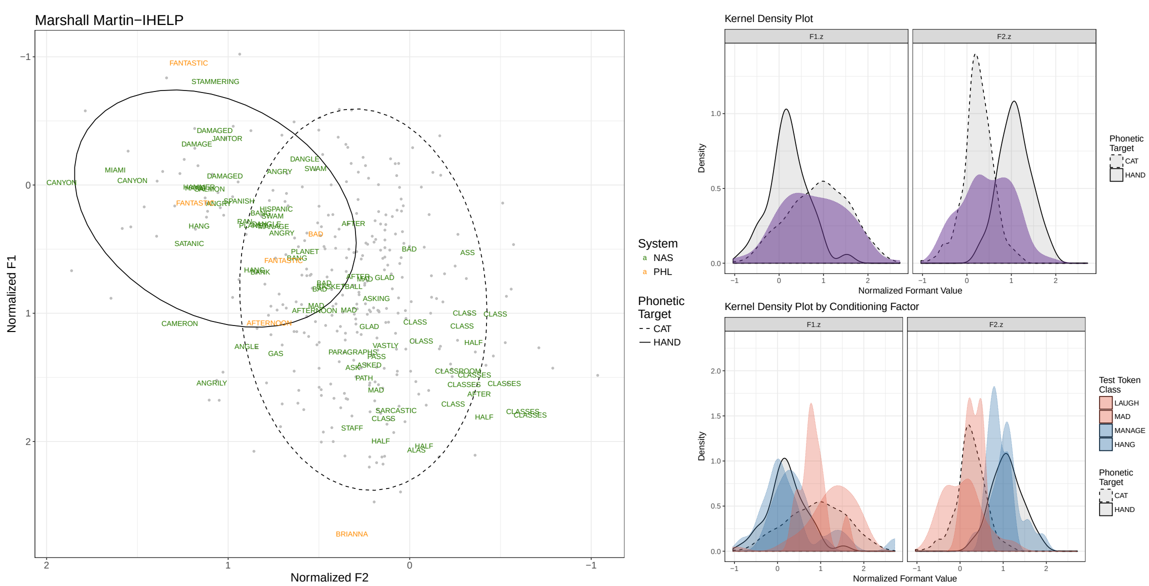


Figure D.17: Marshall Martin, NAS

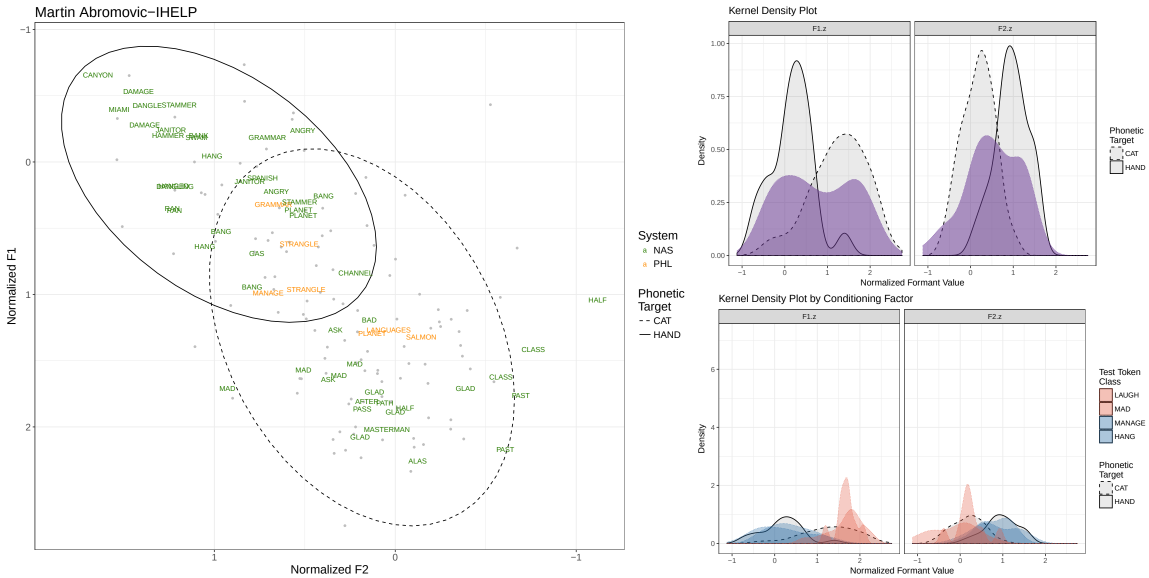


Figure D.18: Martin Abromovic, NAS

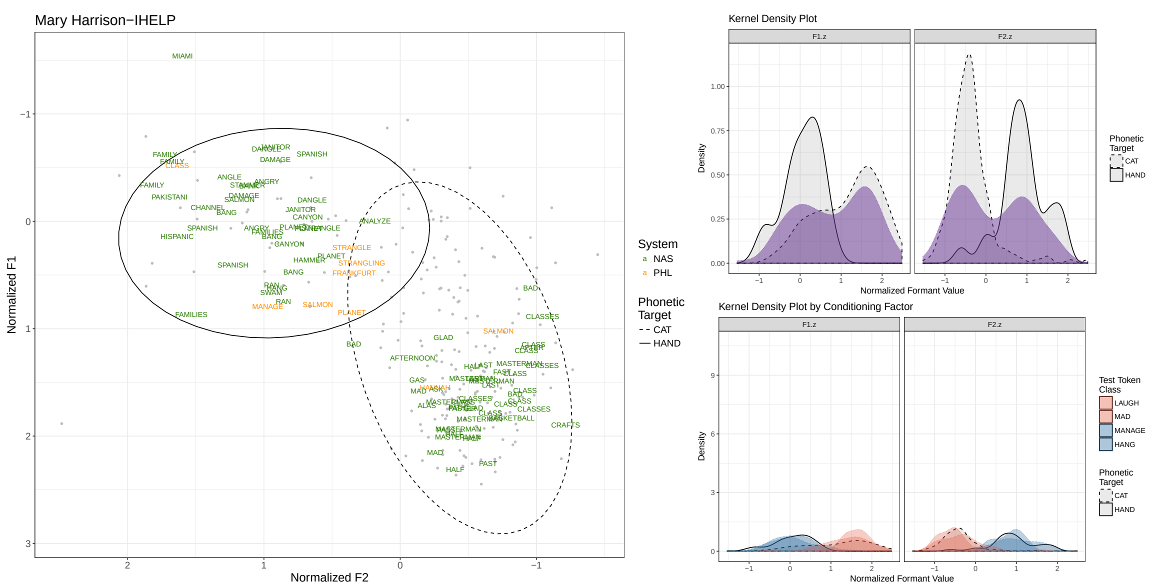


Figure D.19: Mary Harrison, NAS

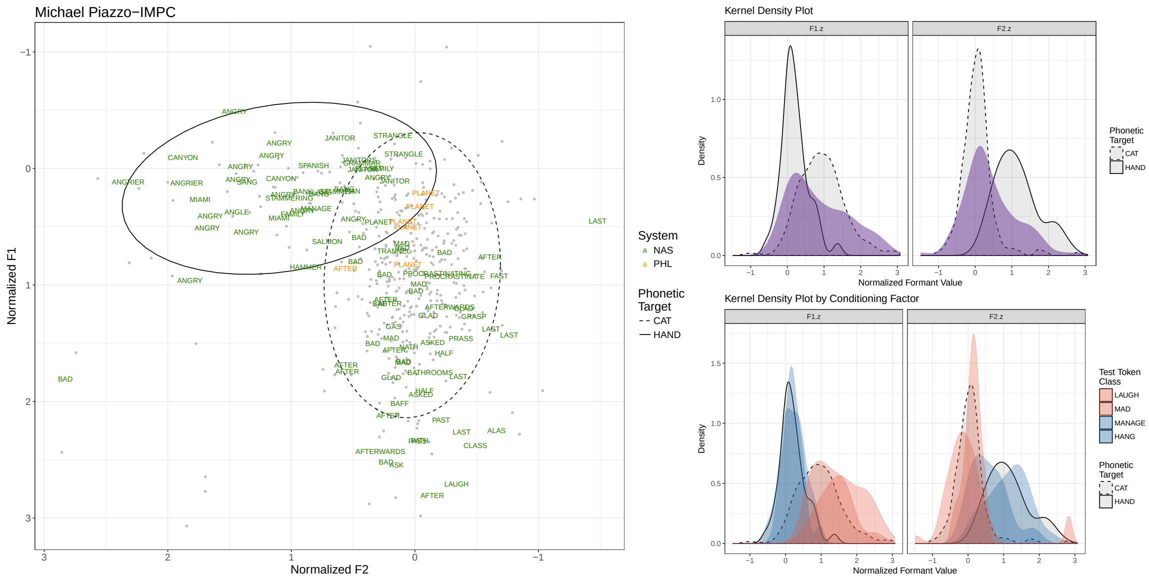


Figure D.20: Michael Piazzo, NAS

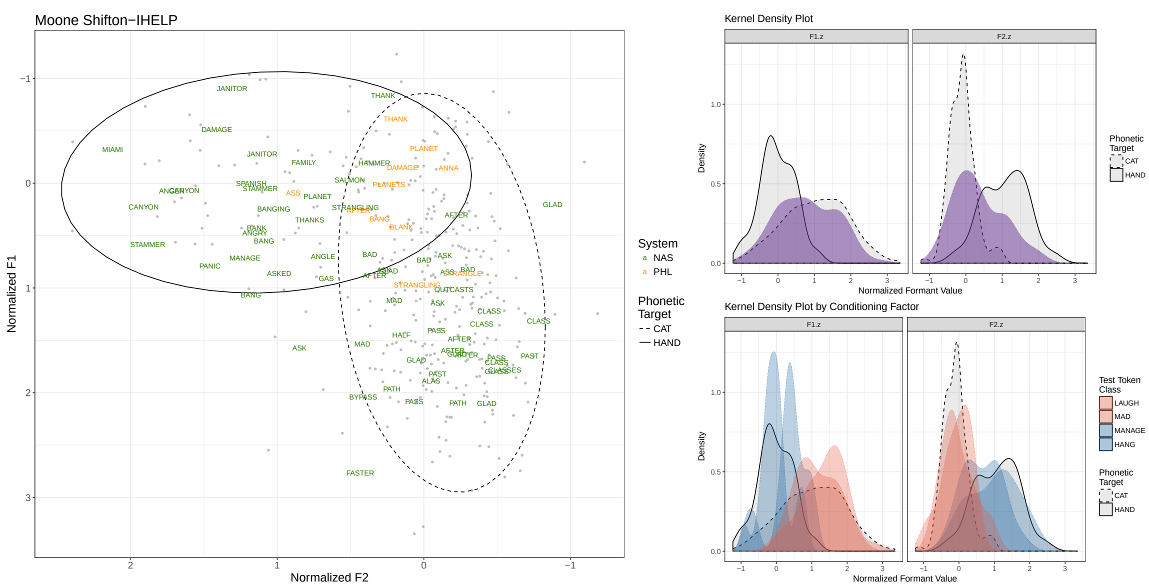


Figure D.21: Moone Shifton, NAS

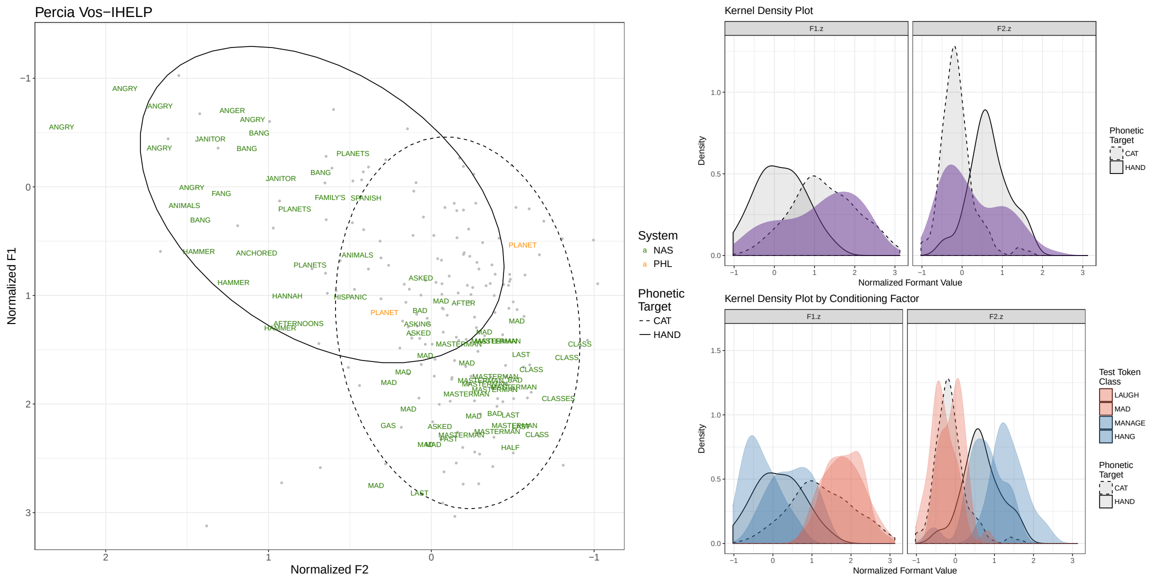


Figure D.22: Percia Vos, NAS

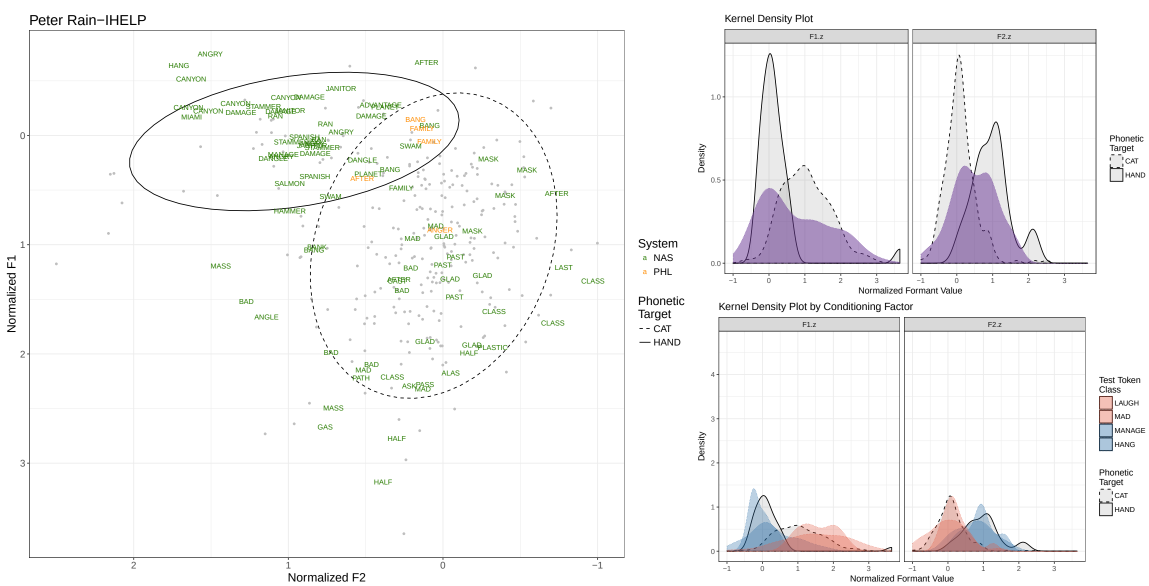


Figure D.23: Peter Rain, NAS

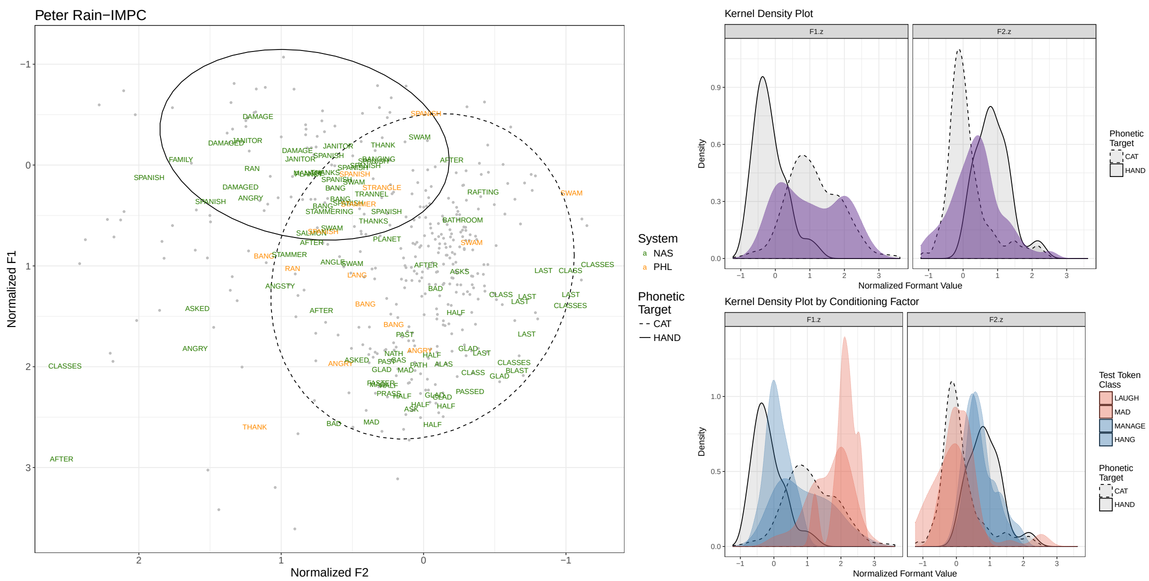


Figure D.24: Peter Rain, NAS

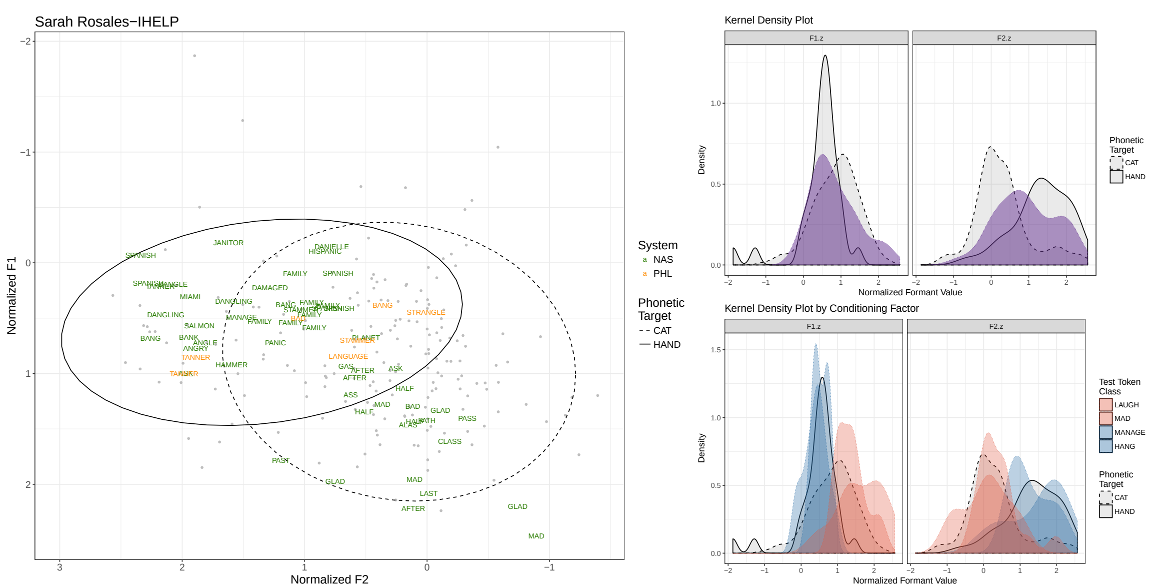


Figure D.25: Sarah Rosales, NAS

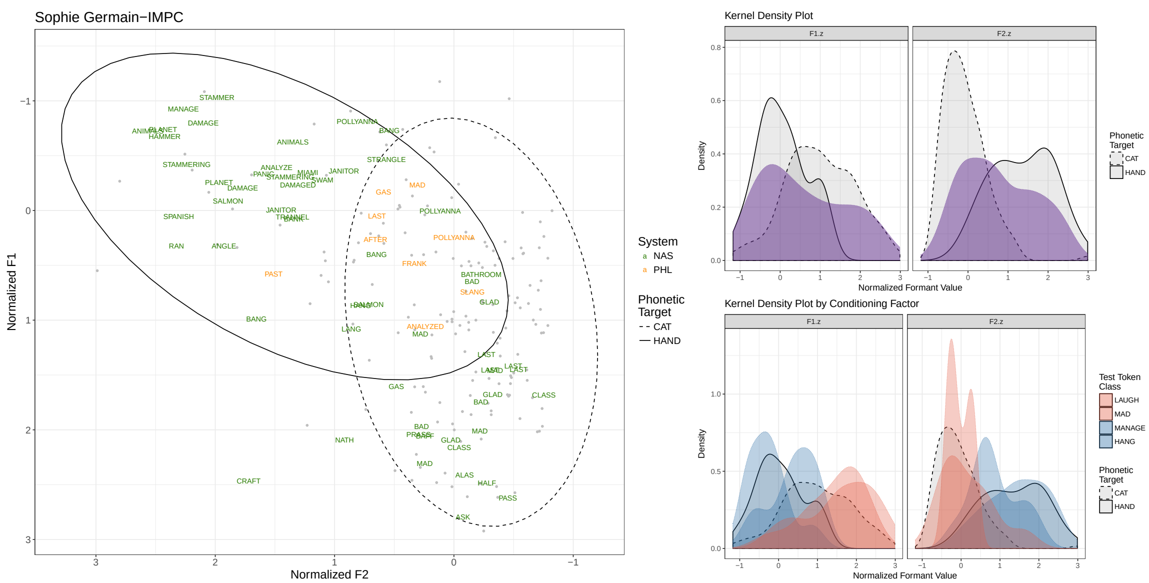


Figure D.26: Sophie Germain, NAS

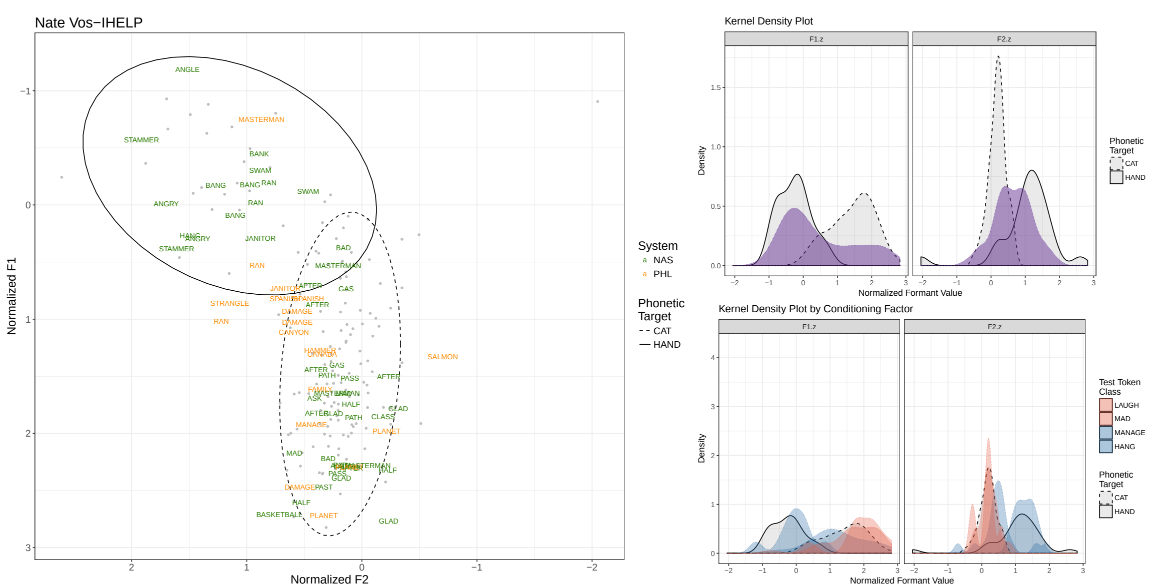


Figure D.27: Nate Vos, competing grammars

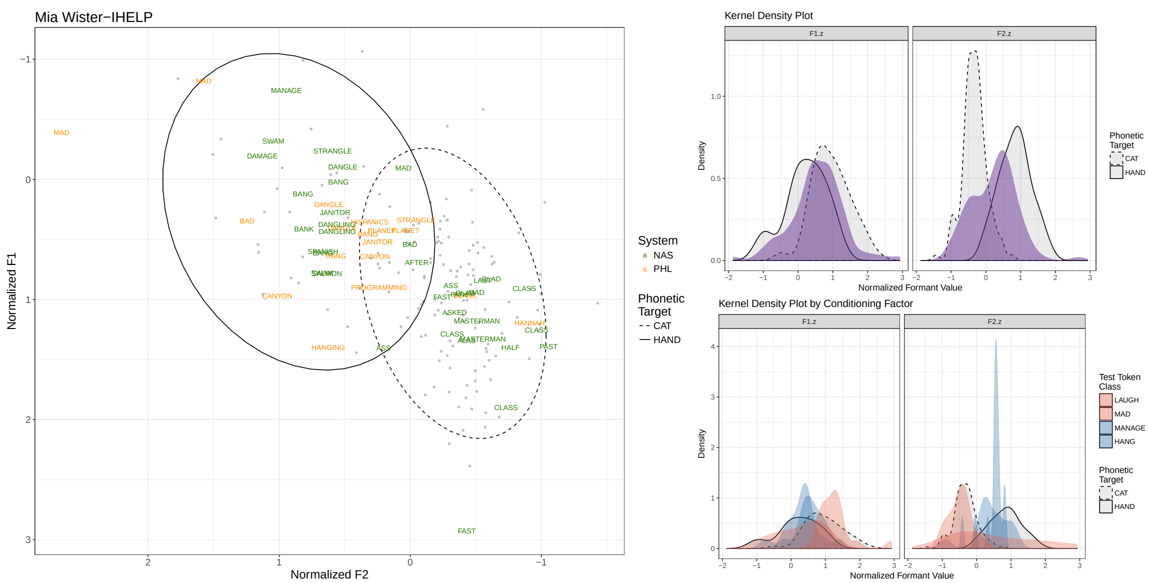


Figure D.28: Mia Wister, competing grammars

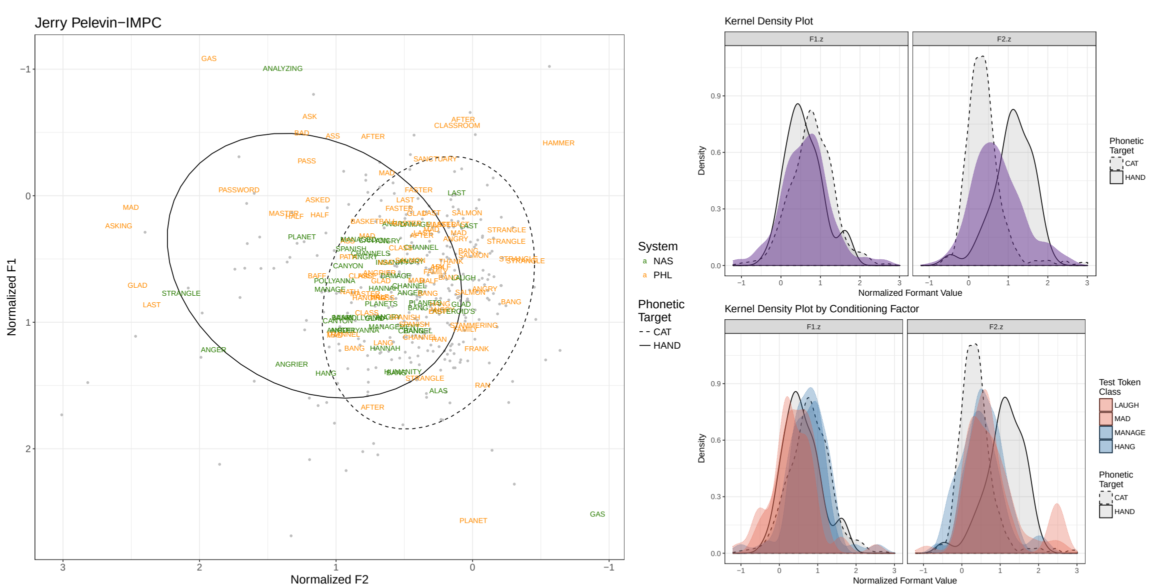


Figure D.29: Jerry Pelevin, competing grammars

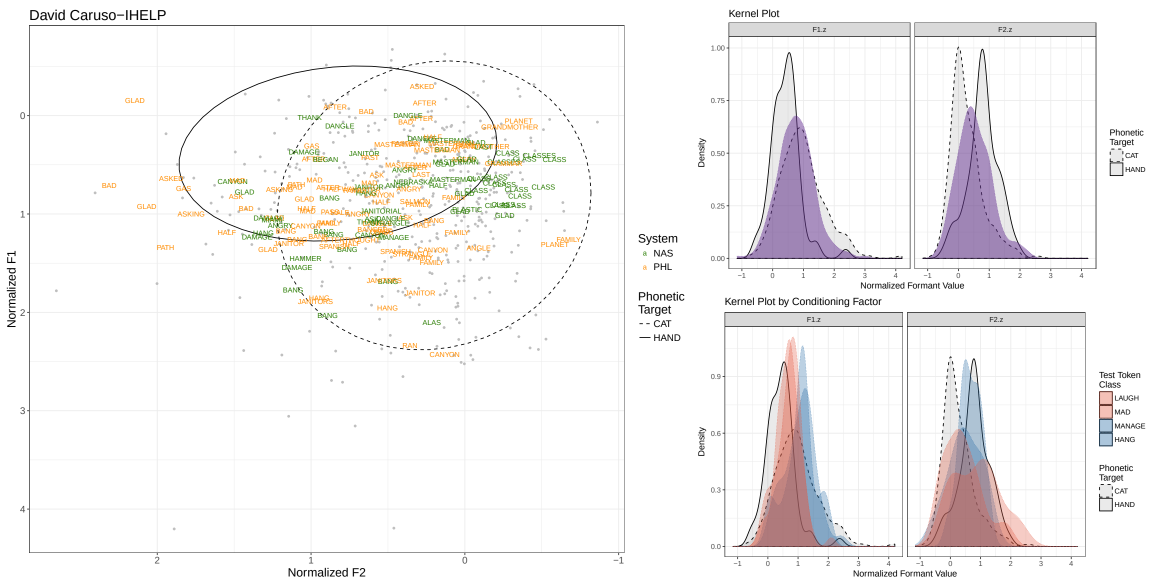


Figure D.30: David Caruso, competing grammars

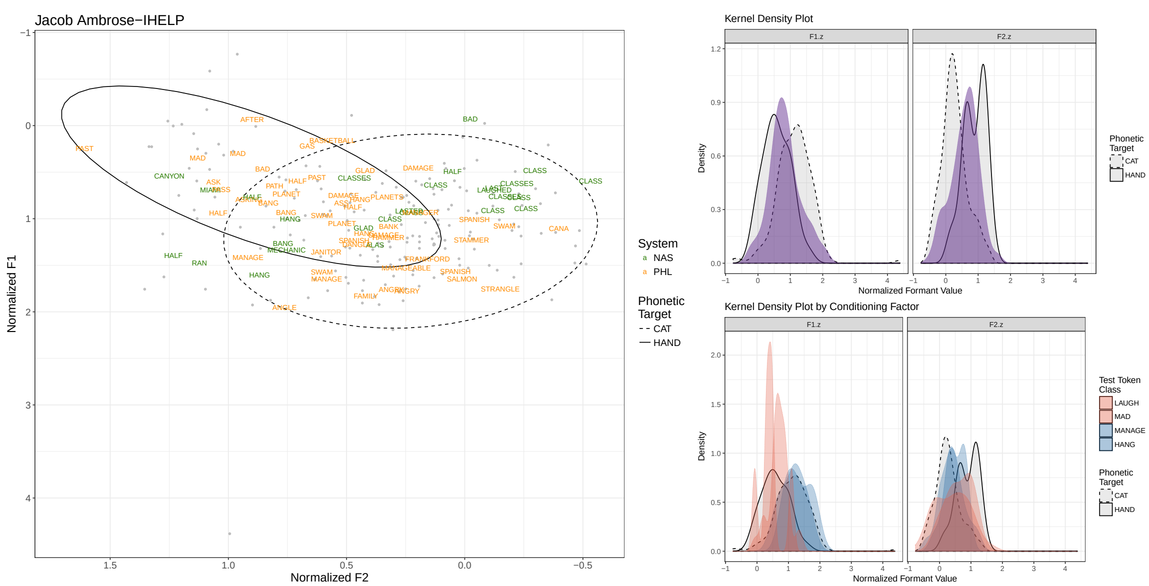


Figure D.31: Jacob Ambrose, competing grammars

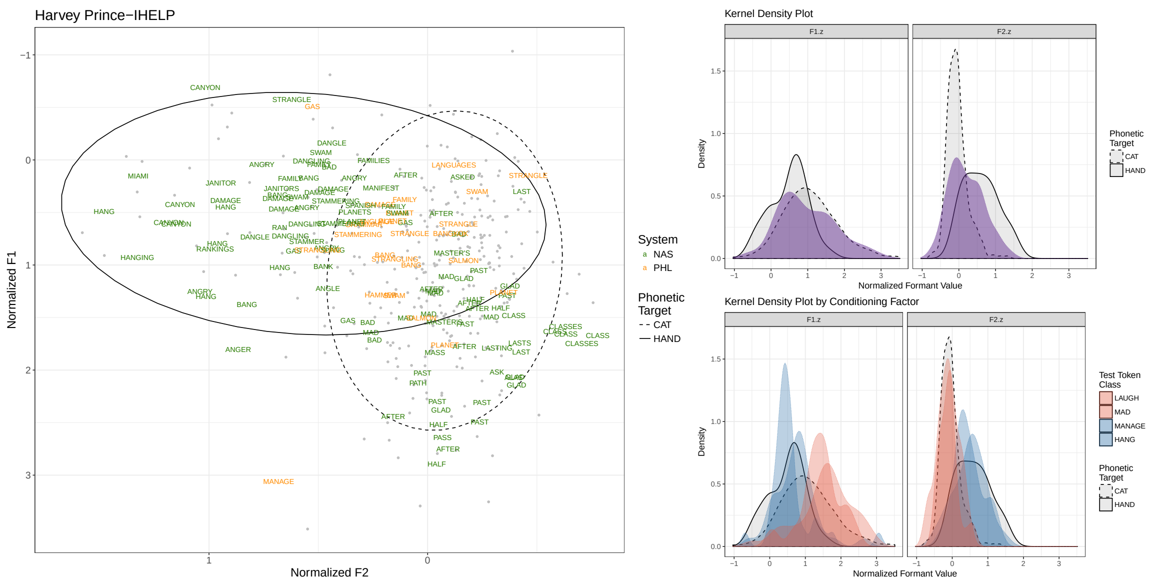


Figure D.32: Harvey Prince, competing grammars

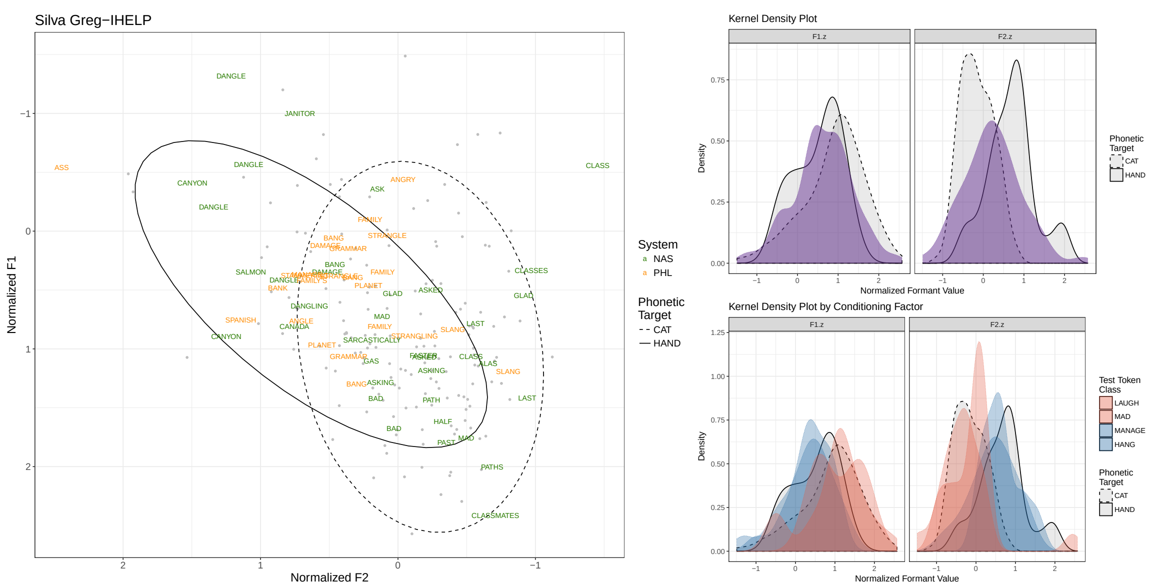


Figure D.33: Silva Greg, competing grammars

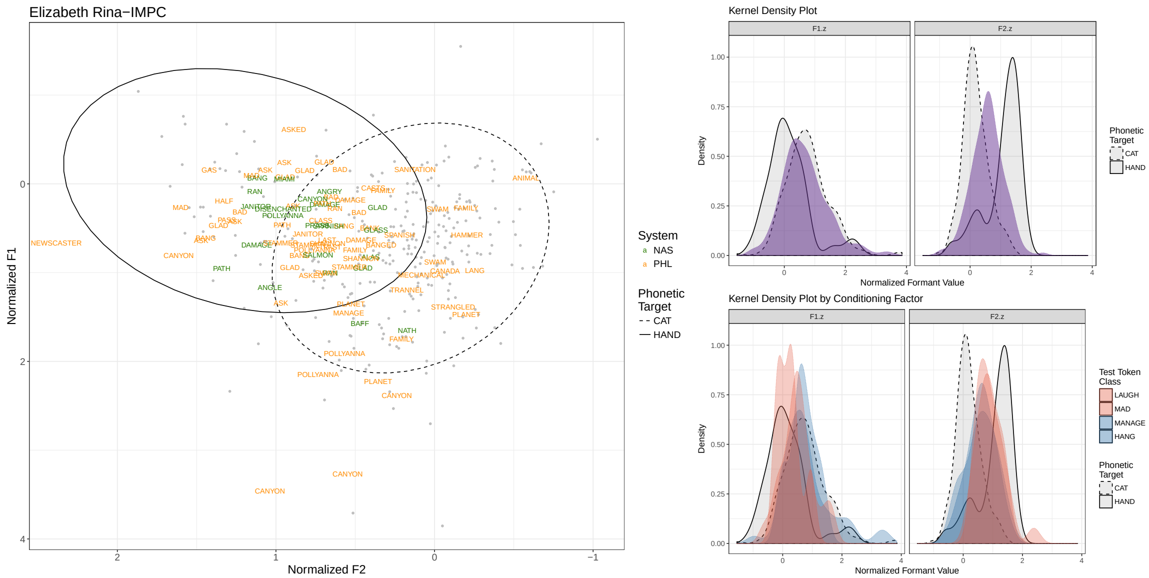


Figure D.34: Elizabeth Rina, competing grammars

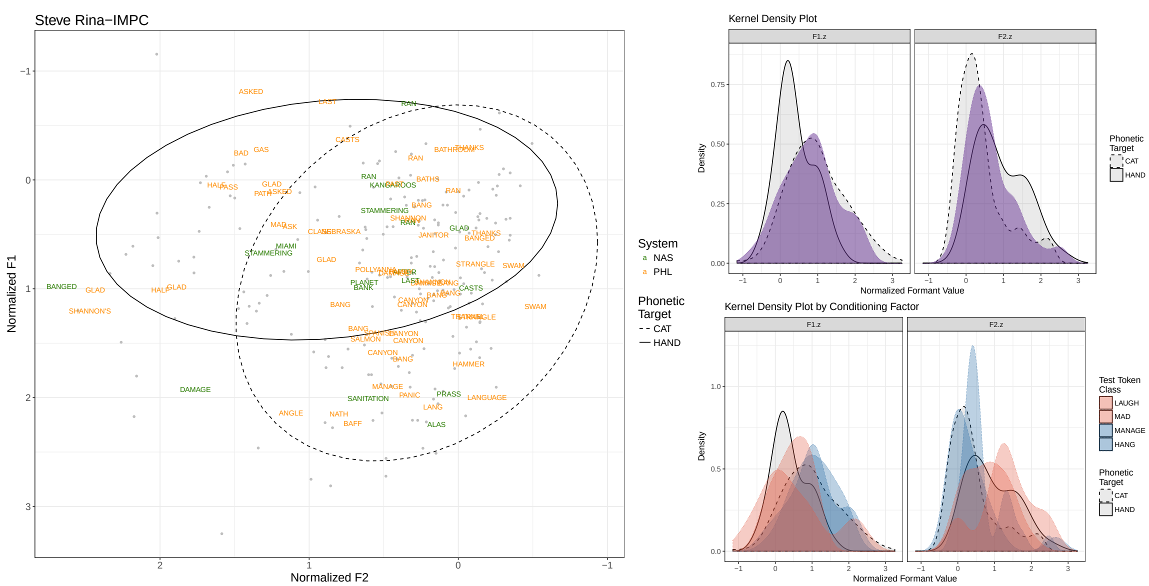


Figure D.35: Steve Rina, competing grammars

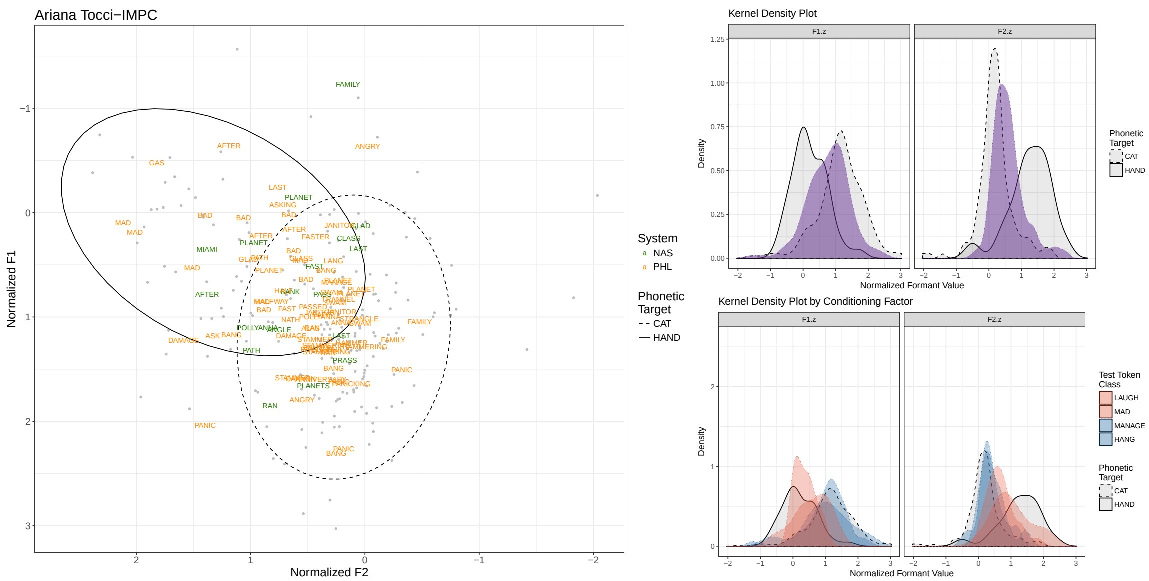


Figure D.36: Ariana Tocci, competing grammars

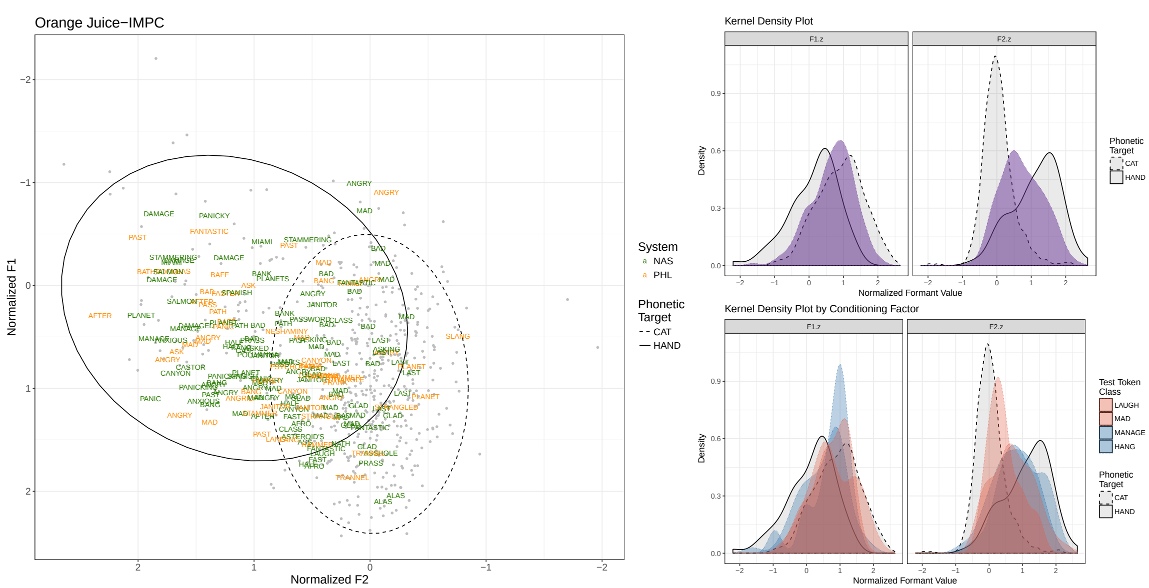


Figure D.37: Orange Juice, competing grammars

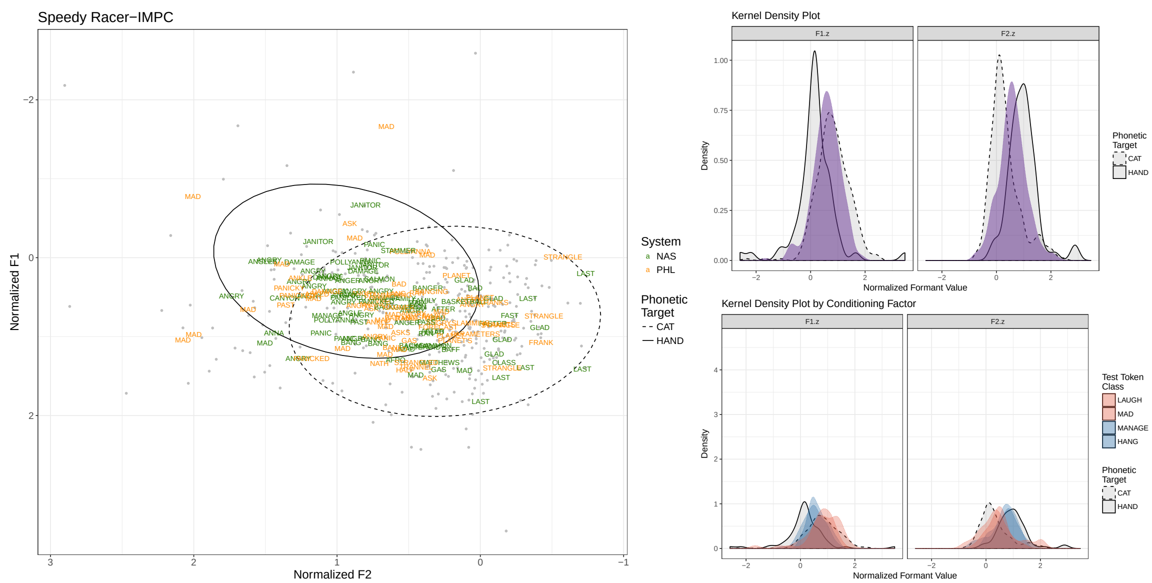


Figure D.38: Speedy Racer, competing grammars

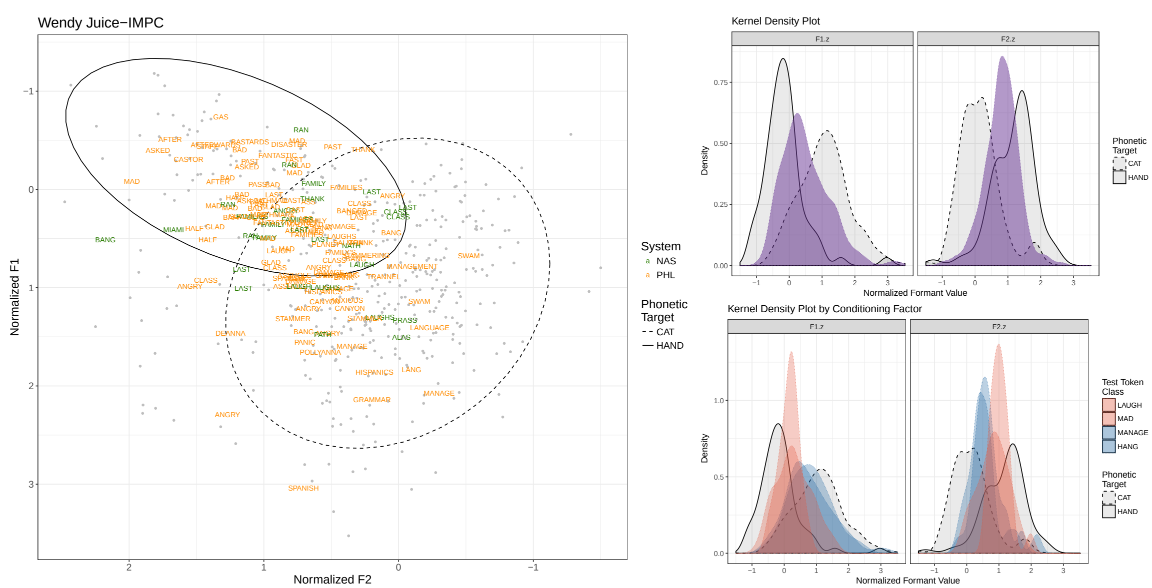


Figure D.39: Wendy Juice, competing grammars

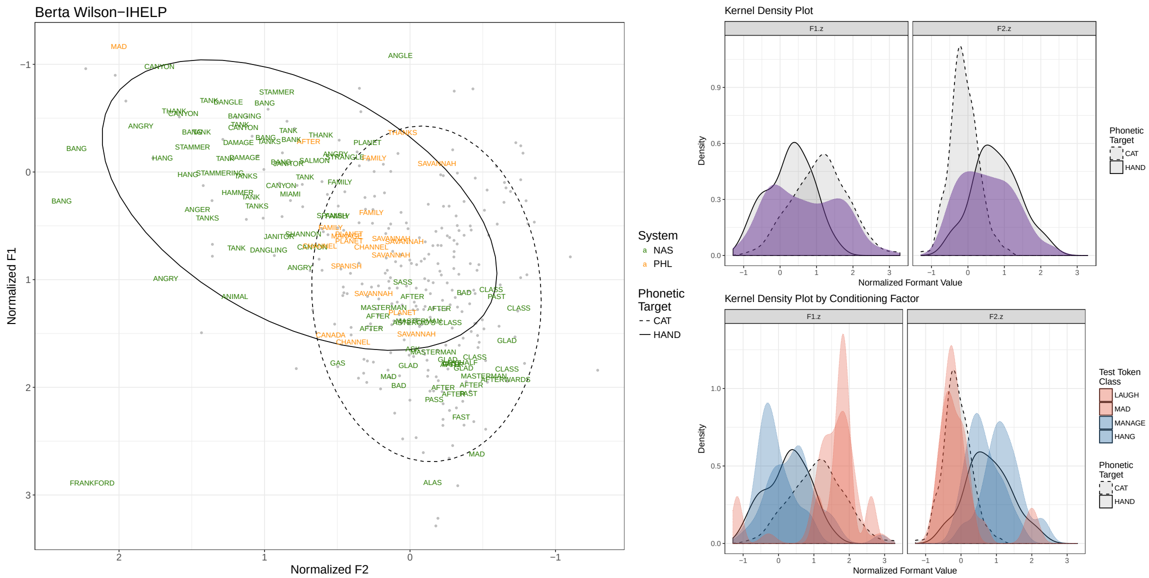


Figure D.40: Berta Wilson, possible NAS

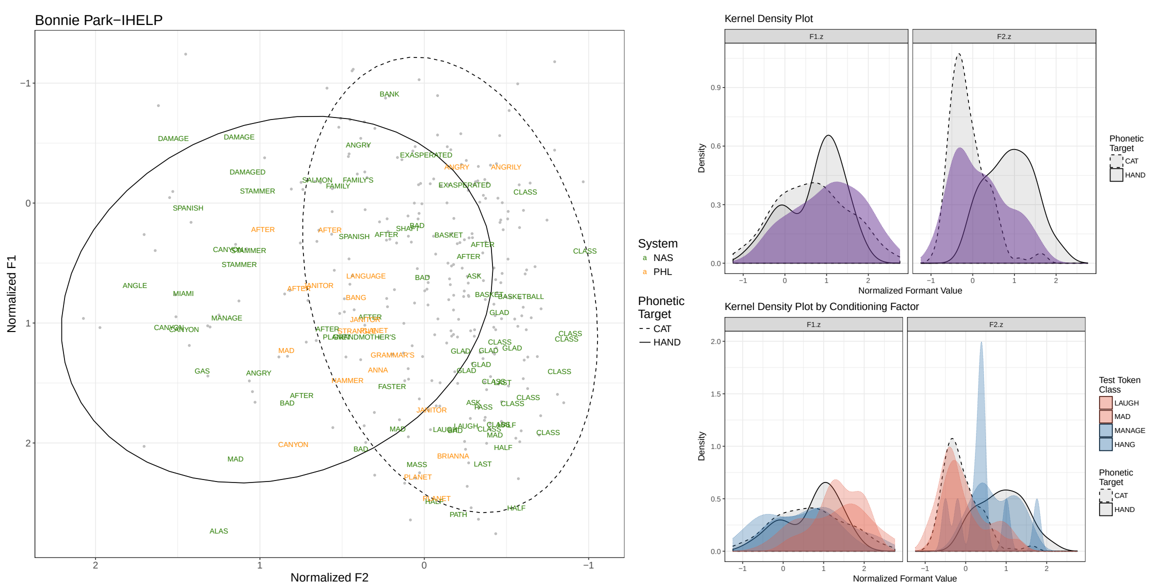


Figure D.41: Bonnie Park, possible NAS

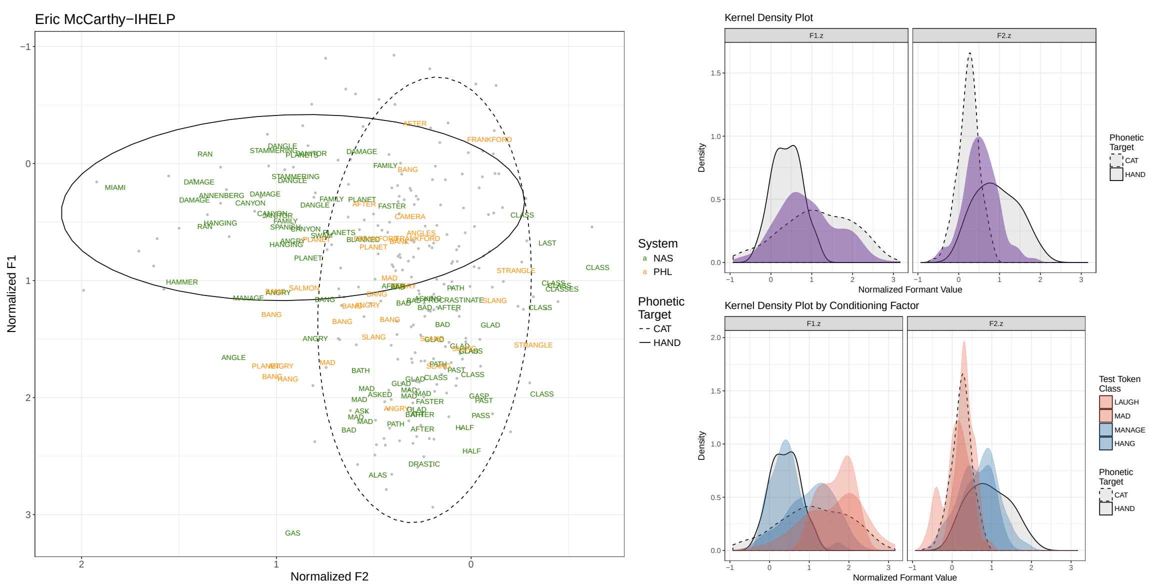


Figure D.42: Eric McCarthy, possible NAS

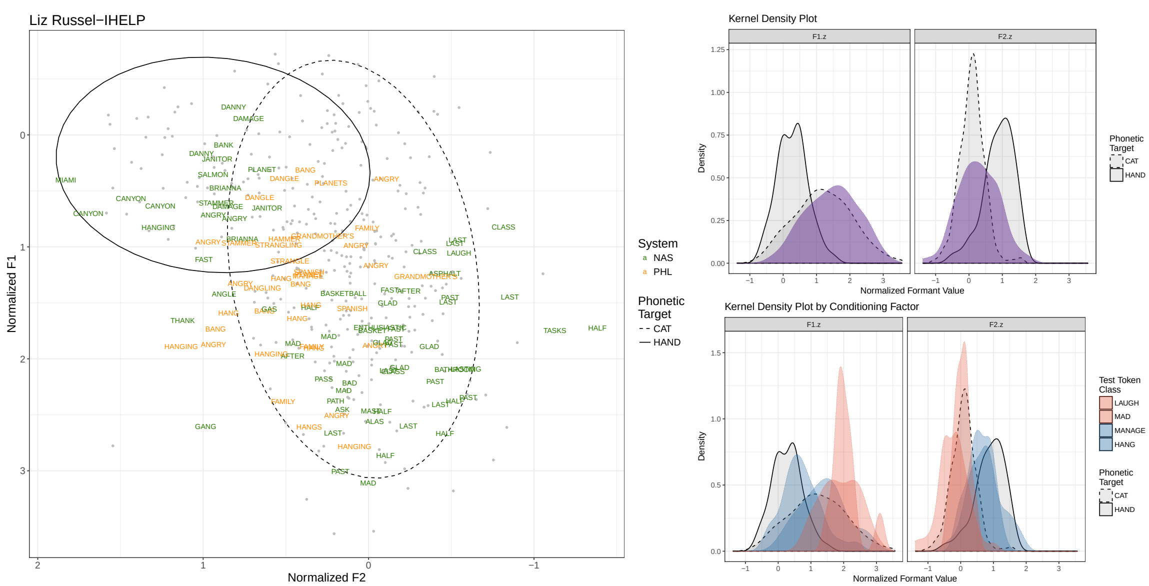


Figure D.43: Liz Russel, possible NAS

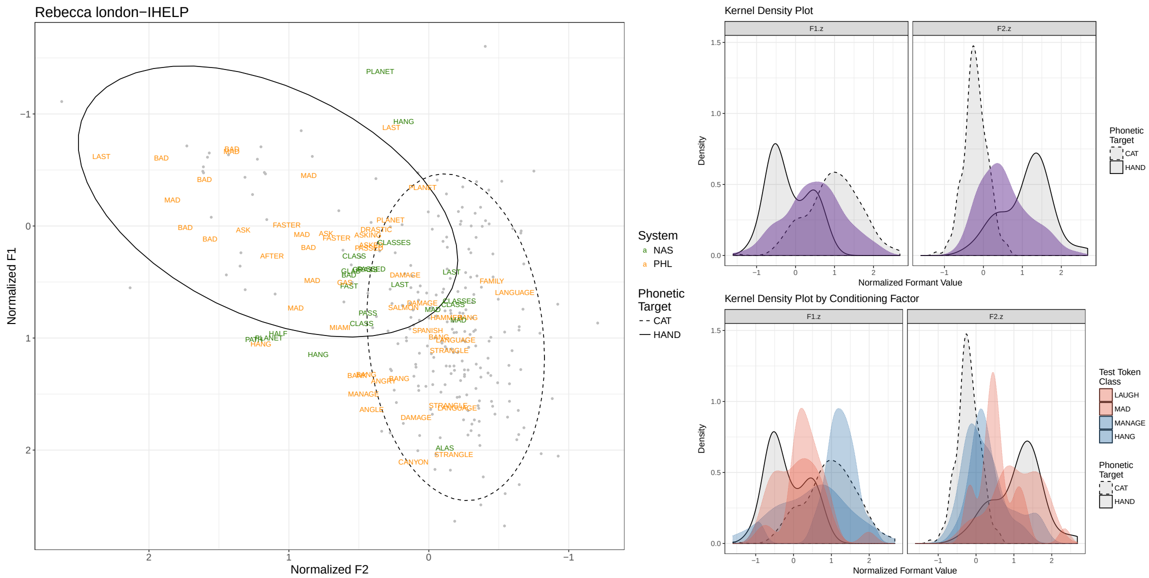


Figure D.44: Rebecca London, possible PHL

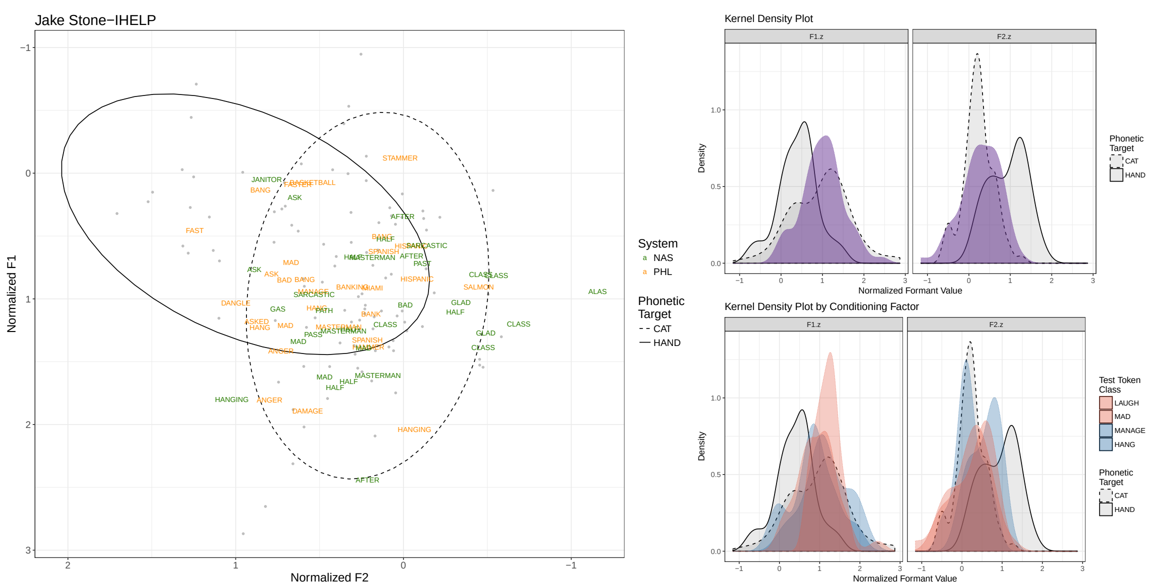


Figure D.45: Jake Stone, phonetic mitigation of tense PHL

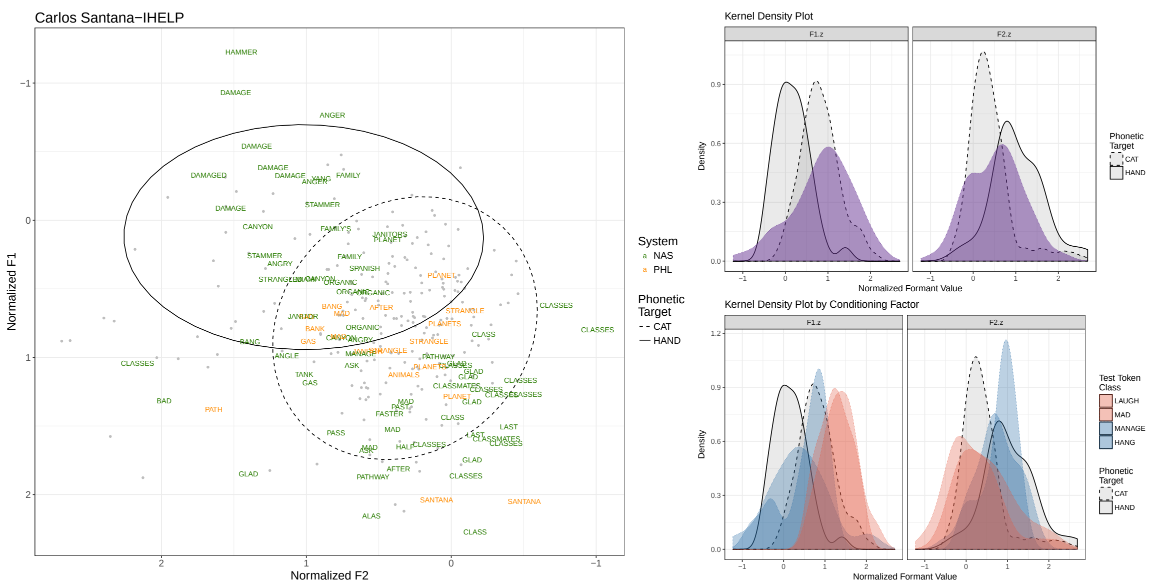


Figure D.46: Carlos Santana, competing grammars plus phonetic mitigation of tense PHL

Bibliography

Addington, D. (1968). The power of language in computer-mediated groups. *Speech Monographs*, (35):492–503.

Aitken, A. (1981). *So meny People Longages and Tonges*, chapter The Scottish Vowel-Length Rule, pages 131–157. The Middle English Dialect Project, Edinburgh.

Anisfeld, M., Bogo, N., and Lambert, W. (1962). Evaluational reactions to accented english speech. *Journal of Abnormal and Social Psychology*, 65:223–269.

Apple, W., Streeter, L., and Krauss, R. (1979). Effects of pitch and speech rate on personal attributes. *Journal of Personality and Social Psychology*, (37):715–727.

Arthur, B., Farrar, D., and Bradford, G. (1974). Evaluation reactions of college students to dialect differences in the english of mexican-americans. *Language and Speech*, (17):255–270.

Ash, S. (2002). The distribution of a phonemic split in the mid-atlantic region: yet more on short-a. *University of Pennsylvania Working Papers in Linguistics*, 8.

Bailey, G. (2017). Velar nasal plus in the north of (ing)land. In *UKLVC 11*, Cardiff.

Bard, E., Robertson, D., and Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72:32–68.

Bates, D., Kleigl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *Journal of Memory and Language*. Preprint.

Becker, K. and Wong, A. (2009). The short-a system of New York City English: An update. In Gorman, K. and Mackenzie, L., editors, *Penn Working Papers in Linguistics: Selected papers from NWAV 37*, volume 15, pages 9–20.

Becker, M. and Tessier, A.-M. (2011). Trajectories of faithfulness in child-specific phonology. *Phonology*, 28:163–196.

Bermúdez-Otero, R. (2007). Diachronic phonology. In de Lacy, P., editor, *The Cambridge handbook of phonology*, pages 497–517. Cambridge University Press.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, (14):359–375.

Blevins, J. (2006). A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics*, 32(2):117–166.

Boberg, C. and Strassel, S. M. (2000). Short-a in cincinnati: A change in progress. *Journal of English Linguistics*, 28:108–126.

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

Boulenger, V., Ferrange, E., Bedoin, N., and Pellegrino, F. c. (2011). Derived contrasts in scottish english: An eeg study. *Proceedings of ICPhS XVII*, pages 352–355.

Boyd, Z., Elliot, Z., Fruehwald, J., Hall-Lew, L., and Lawrence, D. (2015). An evaluation of sociolinguistic elicitation methods. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Brody, S. (2011). Transmissions of philadelphia short-a: the changing gravity of "planets". Master's thesis, University of Pennsylvania.

Brown, B., Giles, H., and Thakerar, J. (1985). Speaker evaluations as a function of speech rate, accent and context. *Language and Communication*, (5):207–220.

Brown, R. (1973). *A first language: The early stages.* George Allen & Unwin, London.

Bybee, J. (1999). *Functionalism and formalism in linguistics*, volume 1, chapter Usage-based phonology. Cambridge University Press, Cambridge.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14:261–290.

Bybee, J. and McClelland, J. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22:381–410.

Campbell-Kibler, K. (2007). Accent, (ing), and the social logic of listener perceptions. *American Speech*, 82(1).

Cargile, A. (1997). Attitudes toward Chinese-accented speech: An investigation in two contexts. *Journal of Language and Social Psychology*, (16):434–443.

Chen, M. and Wang, W. (1975). Sound change: actuation and implementation. *Language*, 51:255–281.

Chomsky, N. (1995). *The Minimalist Program.* MIT Press.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English.* Harper and Row, New York.

Chomsky, N. and Lasnik, H. (2008). *Syntax: an International Handbook of Contemporary Research*, volume 1. de Gruyter, Berlin.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2:225–252.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, 20(3):287–333.

Cohn, A. (1990). Phonetic and phonological rules of nasalization. *UCLA Working Papers in Phonetics*, (76).

Cohn, A. (1993). Nasalisation in English: Phonology or phonetics. *Phonology*, 10(1):43–81.

Comppton, A. and Streeter, M. (1977). Child phonology: data collection and preliminary analyses. *Papers and Reports on Child Language Development*, (13):99–109.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments.* Sage, Thousand Oaks, CA.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1989). Limits on bilingualism. *Nature*, 340:229–230.

De Decker, P. (2006). A real-time investigation of social and phonetic changes in post-adolescence. In Ravindranath, M., editor, *Penn Working Papers in Linguistics: Selected papers from NWAV 34*, volume 12.2.

Dinkin, A. (2015). The like conspiracy: Avoiding accountability. In *Presented at 39th meeting of the Penn Linguistics Conference*, Philadelphia.

Dinkin, A. J. (2013). What's really happening to short-a before l in philadelphia? *American Speech*, 88(1):7–31.

Dinkin, A. J. and Dodsworth, R. (2017). Gradience, allophony, and chain shifts. *Language Variation and Change*, 29:101–127.

Dixon, J., Mahoney, B., and Cocks, R. (2002). Accents of guilt?: Effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology*, (21):162–168.

Dodsworth, R. (2014). Network embeddedness and the retreat from southern vowels in raleigh. In Gylfadottir, D., editor, *Penn Working Papers in Linguistics: Selected papers from NWAV 42*, volume 20.2.

Dresher, B. E. (2011). *The Blackwell Companion to Phonology*, volume 1, chapter The Phoneme, pages 241–266. Wiley-Blackwell.

Durian, D. (2012). *A new perspective on vowel variation across the 19th and 20th centuries in Columbus, OH.* PhD thesis, The Ohio State University.

Eckert, P. and Labov, W. (2017). Phonetics, phonology and social meaning. *Journal of Sociolinguistics*, 21(4):467–496.

Eckert, P. and McConnell-Ginet, S. (1992). Locating power: Proceedings of the second berkeley women and language conference. pages 89–99.

Edwards, J. (1983). Language attitudes in multilingual settings: A general assessment. *Journal of Multilingual and Multicultural Development*, (4):225–236.

Elwell, C., Brown, R., and Rutter, D. (1984). Effects of accent and visual information on impression formation. *Journal of Language and Social Psychology*, (3):297–299.

Evans, D. (2014). America's ugliest accent. `http://gawker.com/tag/americas-ugliest-accent-tournament`. [Online; accessed 14-March-2018].

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in german. *Lingua*, 115:1525–1550.

Ferguson, C. (1972). "short a" in philadelphia english. In Smith, M. E., editor, *Studies in Linguistics in Honor of George L. Trager*, pages 259–274. Mouton.

Fisher, S., Prichard, H., and Sneller, B. (2015). The apple doesn't fall far from the tree: incremental change in philadelphia families. In Fisher, S., editor, *University of Pennsylvania Working Papers in Linguistics*, volume 21.

Fogle, D. (2008). Expansino or approximation: The low-back merger in indianapolis. In *Presented at the 37th meeting of the New Ways of Analyzing Variation NWAV Conference*, Houston.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, (31):307–319.

Friedman, L. (2014). *The St. Louis corridor: mixing, competing, and retreating dialects.* PhD thesis, University of Pennsylvania.

Fruehwald, J. personal communication.

Fruehwald, J. (2013). *Phonological Involvement in Phonetic Change.* PhD thesis, University of Pennsylvania.

Fruehwald, J., Gress-Wright, J., and Wallenberg, J. C. (2013). Phonological rule change: The constant rate effect. In Kan, S., Moore-Cantwell, C., and Staubs, R., editors, *Proceedings of the 40th Annual Meeting of the North East Linguistic Society.*

Gaies, S. and Beebe, J. (1991). The matched guise technique for measuring attitudes and their implications for language education: A critical assessment. ERIC Report.

Geenberg, K. (2014). *What it means to be Norcal Country: Variation and marginalization in rural California.* PhD thesis, Stanford University.

Gibbons, J. (1983). Attitudes towards languages and code-mixing in hong kong. *Journal of Multilingual and Multicultural Development*, (4):129–147.

Giles, H., Coupland, N., Henwood, K., Harriman, J., and Coupland, J. (1992a). *Studies in the Pronunciation of English: A Commemorative Volume in Honor of A. C. Gimson*, chapter The social meaning of RP: An intergenerational perspective, pages 191–211. Routledge, New York.

Giles, H., Henwood, K., Coupland, N., Harriman, J., and Coupland, J. (1992b). Language attitudes and cognitive mediation. *Human Communication Research*, (18):500–527.

Givon, T. (1991). Markedness in grammar: distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language*, (2):335–370.

Gleason, H. (1961). Reviewed work(s): African language studies 1 by malcolm guthrie; the role of tone in the structure of sukuma by i. richardson. *Language*, 37(2):294–308.

Goldinger, S. (1998). Echoes of echoes? an episodic theory of lexical access). *Psychological Review*, 105(2):251–279.

Goldsmith, J. (1995). *Phonological Theory*, pages 1–23. Blackwell, Oxford.

Gordon, E., Maclagan, M., and Hay, J. The onze corpus.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29(2):228–244.

Guaitella, I. (1999). Rhythm in speech: What rhythmic organizations reveal about cognitive processes in spontaneous speech production versus reading aloud. *Journal of Pragmatics*, 31:509–523.

Hall, K. C. (2013). A typology of indeterminate phonological relationships. *The Linguistic Review*, 30(2):215–275.

Hall-Lew, L. (2010). Improved representation of variance in measures of vowel merger. In *Proceedings of Meetings on Acoustics*, volume 9.

Hall-Lew, L., Cardoso, A., Kemenchedjieva, Y., Wilson, K., Purse, R., and Saigusa, J. (2010). San francisco english and the california vowel shift. In *Proceedings of Meetings on Acoustics*, volume 9, pages 1–10.

Harris, J. (1990). *Derived phonological contrasts*, pages 87–105. Routledge, London.

Harris, J. (1994). *English sound structure.* Blackwell, Oxford.

Haspelmath, M. (2005). Against markedness (and what to replace it with). *Journal of Linguistics*, 42:25–70.

Hay, J., Pierrehumbert, J., Walker, A., and LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, 139:83–91.

Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34:458–484.

Herold, R. (1997). Solving the actuation problem: Merger and immigration in eastern pennsylvania. *Language Variation and Change*, 9(2):165–189.

Hockett, C. F. (1958). *A course in modern linguistics.* Macmillan, New York.

Howell, P. and Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, (10):163–169.

Hudson Kam, C. and Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, pages 151–195.

Hudson Kam, C. and Newport, E. (2009). Getting it right by getting it wrong: when learners change languages. *Cognitive Psychology*, 59:30–66.

Janda, R. D. and Joseph, B. D. (2003). Reconsidering the Canons of Sound-Change: Towards a "Big Bang" Theory. In Blake, B. and Burridge, K., editors, *Historical Linguistics 2001. Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne*, pages 205–219. John Benjamins.

Jarmulowicz, L. D. (2002). English derivational suffix frequency and children's stress judgments. *Brain and Language*, 81:192–204.

Johnson, D. E. (1998). The tensing and laxing of short-a in new haven, connecticut. Online manuscript.

Johnson, D. E. (2010). *Stability and change across a dialect boundary: the low vowels of Southeastern New England.* Duke University Press, Durham, NC.

Johnson, J. and Newport, E. (1997). Critical period effects in second language learning: The influence of maturational state on the acquisition of english as a second language. *Cognitive Psychology*, 21:60–99.

Kaye, J., Lowenstamm, J., and Vergnaud, J. (1985). The internal structure of phonological representations. *Phonology Yearbook*, 2:305–328.

Keating, P. (1985). *Phonetic Linguitics*, chapter Universal phonetics and the organization of grammars, pages 115–132. Academic Press.

Keating, P. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, 18:321–334.

Kerswill, P. and Williams, A. (2000). Creating a new town koine: Children and language change in milton keynes. *Language in Society*, 29(1):65–115.

Kiparsky, P. (1982). *The structure of phonological representations*, chapter From Cyclic to Lexical Phonology, pages 131–175. Foris Publications, Dordrecht, The Netherlands.

Kiparsky, P. (1995). The phonological basis of sound change. In Goldsmith, J., editor, *Handbook of Phonological Theory*. Blackwell.

Kiparsky, P. (2015). *The Oxford Handbook of Historical Phonology*, chapter Phonologization. Oxford University Press.

Kowal, S., O'Connell, D., O'Brien, E., and Bryant, E. (1975). Temporal aspects of reading aloud and speaking. *American Journal of Psychology*, (88):549–569.

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

Kroch, A. (1996). *Dialect and Style in the Speech of Upper Class Philadelphia.* John Benjamins.

Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.

Labov, W. (1966). The linguistic variable as a structual unit. ERIC Report.

Labov, W. (1972). *Sociolinguistic Patterns.* Blackwell Publishers.

Labov, W. (1981). Resolving the neogrammarian controversy. *Language*, 57(2):267–308.

Labov, W. (1984). *Field methods on the Project of Linguistic Change and Variation*, pages 28–53. Prentice Hall.

Labov, W. (1989). The exact description of a speech community. In Fasold, R. and Schiffrin, D., editors, *Language Change and Variation*, chapter The exact description of a speech community, pages 1–57. John Benjamins.

Labov, W. (1993). The unobservability of structure and its linguistic consequences. In *Presented at the 22nd meeting of the New Ways of Analyzing Variation conference*, Ottawa.

Labov, W. (2001). *Principles of Linguistic Change*, volume 2: Social Factors. Blackwell, Oxford.

Labov, W. (2007). Transmission and diffusion. *Language*, 83:344–387.

Labov, W. (2009). Enigmas of uniformity. In *Presented at the 38th meeting of the New Ways of Analyzing Variation NWAV Conference*, Ottawa.

Labov, W. (2012–2015). Influence of Higher Education on Local Phonology. With support of NSF contract number 1251437.

Labov, W., Ash, S., and Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Mouton de Gruyter.

Labov, W. and Fisher, S. (2015). *The Oxford Handbook of African American Language*, chapter African American Phonology in a Philadelphia Community. Oxford University Press.

Labov, W., Fisher, S., Gylfadottir, D., Henderson, A., and Sneller, B. (2016). Competing systems in Philadelphia phonology. *Language Variation and Change*, 28(3):273–305.

Labov, W., Rosenfelder, I., and Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: linear incrementation, reversal and reanalysis. *Language*, 89:30–65.

Labov, W., Yager, M., and Steiner, R. (1972). A quantitative study of sound change in progress. Report on National Science Foundation Contract NSF-GS-3287.

Ladd, B. (2005). Vowel length in scottish english: new data based on the alignment of pitch accent peaks. In *13th Manchester Phonology Meeting*.

Lambert, W. (1967). A social psychology of bilingualism. *Journal of Social Issues*, (23):91–109.

Lambert, W., Anisfeld, M., and Yeni-Komshian, G. (1965). Evaluational reactions of jewish and arab adolescents to dialect and language variations. *Journal of Personality and Social Psychology*, (2):84–90.

Lambert, W., Hodgson, R., Gardner, R., and Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, (60).

Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society*, (5):539–566.

MacWhinney, B. (2000). The childes project: Tools for analyzing talk.

Martinet, A. (1952). Function, structure, and sound change. *Word*, (1):1–32.

Mielke, J. (2008). *The Emergence of Distinctive Features*. Oxford University Press.

Mielke, J., Carignan, C., and Thomas, E. R. (2017). The articulatory dynamics of pre-velar and pre-nasal /ae/-raising in english: an ultrasound study. *Journal of the Acoustical Society of America*, 142(1):332–349.

Ohala, J. (1981). The listener as a source of language change. In Masek, C., Hendrick, R., and Miller, F., editors, *Papers from the parasession on language and behavior*.

Ohama, M. L. F., Gotay, C., Pagano, I., Boles, L., and Craven, D. (2000). Evaluations of hawaii creole english and standard english. *Journal of Language and Social Psychology*, (19):357–377.

Payne, A. C. (1980). Factors controlling the acquisition of the philadelphia dialect by out-of-state children. In Labov, W., editor, *Locating Language in Time and Space*, chapter Factors Controlling the Acquisition of the Philadelphia Dialect by Out-of-State Children. Academic Press, New York.

Perry, L., Mech, E., MacDonald, M., and Seidenberg, M. (2017). Influences of speaker-listener similarity on shadowing and comprehension. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Philadelphia Neighborhood Corpus of LING560 Studies (1972–2013). With support of NSF contract 921643 to W. Labov.

Pierrehumbert, J. (2001). *Frequency effects and the emergence of linguistic structure*, chapter Exemplar dynamics: Word frequency, lenition and contrast. John Benjamins.

Pierrehumbert, J. (2002). *Laboratory Phonology*, volume 7, chapter Word-specific phonetics. de Gruyter Mouton.

Pintzuk, S. (1996). Old english verb-complement word order and the change from ov to vo. *York Papers in Linguistics*, 17:241–264.

Pitt, M.A., D. L. J. K. K. S. R. W. H. E. F.-L. E. (2007). Buckeye corpus of conversational speech (2nd release).

Prichard, H. (2016). *The role of higher education in linguistic change*. PhD thesis, University of Pennsylvania.

Prichard, H. and Tamminga, M. (2012). The impact of higher education on philadelphia vowels. In Prichard, H., editor, *University of Pennsylvania Working Papers in Linguistics*, volume 18.

Prince, A. and Smolensky, P. (1993). Optimality theory. Manuscript.

Purnell, T., Idsardi, W., and Baugh, J. (1999). Perceptual and phonetic experiments on american english dialect identification. *Journal of Language and Social Psychology*, (18):10–30.

Ray, G., Berlin Ray, E., and Zahn, C. (1991). Speech behavior and social evaluation: An examination of medical messages. *Communication Quarterly*, (39):119–129.

Ray, G. and Zahn, C. (1999). Language attitudes and speech behavior: New Zealand English and Standard American English. *Journal of Language and Social Psychology*, (18):310–319.

Ringe, D. and Eska, J. F. (2013). *Historical Linguistics: Toward a Twenty-First Century Reintegration*. Cambridge University Press, New York.

Roberts, J. and Labov, W. (1995). Learning to talk philadelphian. *Lanuage Variation and Change*, 7(1):101–112.

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., and Yuan, J. (2014). Fave (forced alignment and vowel extraction) program suite v1.2.1. doi:10.5281/zenodo.15530.

Rubin, D. and Smith, K. (1990). Effects of accent, ethnicity and lecture topic on undergraduates' perception of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, (14):337–353.

Schuler, K., Yang, C., and Newport, E. (2016). Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. In Papafragou, A., Grodner, D., Mirman, D., and Trueswell, J., editors, *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.

Scobbie, J., Hewlett, N., and Turke, A. (1999). *Standard English in Edinburgh and Glasgow: the Scottish vowel length rule revealed*, pages 230–245. Arnold, London.

Scobbie, J. and Stuart-Smith, J. (2008). *Contrast in phonology: Perception and acquisition*, chapter Quasi-phonemic contrast and the indeterminacy of the segmental inventory: Examples from Scottish English, pages 87–113. Mouton de Gruyter.

Segui, J., Mehler, J., Frauenfelder, U., and Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia*, 20(6):615–627.

Seguie, J., Mehler, J., and Frauenfelder, U. (1982). The word frewuency effect and lexical access. *Neuropsychologia*, 20(6):615–627.

Singh, L., Morgan, J., and White, K. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 52(2):172–189.

Smith, M. K. and Baley, G. H. (1980). *Language: Social Psychological Perspectives*, chapter Attitude and activity: Contextual constraints on subjective judgments, pages 209–226. Pergamon Press, Oxford.

Sneller, B. (2018). Allophonic systems as a variable within individual speakers. In Lightfoot, D. and Havenhill, J., editors, *Variable properties in language: Their nature and acquisition*, Washington, D.C.

Sprouse, J. (2007). *A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge*. PhD thesis, University of Maryland.

Sprouse, J. (2011). A test of the cognitive assumptions of magnitude estimation. *Language*, 87(2):274–288.

Stampe, D. (1973, 1979). *A dissertation on natural phonology*. PhD thesis, University of Chicago.

Steriade, D. (2007). *Contrast*, pages 139–157. Cambridge University Press.

Strongman, K. and Woosley, J. (1967). Stereotyped reactions to regional accents. *British Journal of Social and Clinical Psychology*, (6):164–167.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich.* Cambridge University Press.

Tyler, A. and Nagy, W. (1989). The acquisition of english derivational morphology. *Journal of Memory and Language*, 28(6):649–667.

Wagner, S. E., Mason, A., Nesbitt, M., Pevan, E., and Savage, M. (2015). Reversal and re-organization of the northern cities shift in michigan. In *New Ways of Analyzing Variation 44*, Toronto.

Wang, W. (1969). Competing changes as a cause of residue. *Language*, 57(2):9–25.

Weinriech, U., Labov, W., and Herzog, M. (1968). *Empirical foundations for a theory of language change.* University of Texas Press.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity.* Cambridge University Press.

Wölck, W. (1973). *Language Attitude Studies: Current Trends and Prospects*, chapter Attitudes towards Spanish and Quechua in bilingual Peru, pages 129–147. Georgetown University Press, Washington, D.C.

Yang, C. (2000). *Knowledge and learning in natural language.* PhD thesis, Massachusetts Institute of Technology, Cambridge.

Yang, C. (2002). *Knowledge and Learning in Natural Language.* Oxford University Press, New York.

Yang, C. (2009). Population structure and language change. Online manuscript.

Yang, C. (2016). *Price of Productivity: How children learn and break rules of language.* MIT Press.