PERSISTENCE IN THE PRODUCTION OF LINGUISTIC VARIATION

Meredith Johanna Tamminga

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

2014

Supervisor of Dissertation

____

William Labov

Fassitt Professor of Linguistics

Co-Supervisor of Dissertation

____

David Embick

Professor of Linguistics

Graduate Group Chairperson

____

Eugene Buckley

Associate Professor of Linguistics

Dissertation Committee

Anthony Kroch, Professor of Linguistics

PERSISTENCE IN THE PRODUCTION OF LINGUISTIC VARIATION

To my parents, Nancy and Ken Tamminga,

who have endless patience for my academic ramblings.

Acknowledgements

As I wrap up the graduate school chapter of my life, I'm struck by how little inclined I feel to look back and take stock. It is, I think, a testament to the intellectual energy of the Penn Linguistics milieu — in the five years since I started I have been constantly propelled forward, and in my transition from doctoral candidate to junior faculty I seem to be plunging ahead faster than ever. But I can't deposit my dissertation without pausing to thank the many wonderful people who have made it possible.

First, I wish to thank my inimitable co-supervisors. Since before I even started at Penn, Bill Labov has forced me to confront the profound questions and has inculcated in me a veneration for reality over theory. He taught me to look for what is good — in papers and in people — instead of focusing on what is flawed. I suspect that his voice has taken up permanent residence in my academic psyche, and don't doubt that I'll spend the rest of my career discovering over and over again that he already thought of all my best ideas in the 60s (Fruehwald, p.c.). Dave Embick and I have spent hundreds of hours in intense conversation that broadened my linguistic horizons and never failed to keep me on my toes. This dissertation, and my research program, would be immeasurably more limited without his influence. He also has a keen appreciation for the power of small bits of encouragement, and always has a dinosaur sticker, cup of green tea, or exploding fist bump to keep me going.

Many other members of the department have played important roles in my education here. Tony Kroch, my lone committee member, pushes me to think more boldly, argue more forcefully, and write more clearly. His intellectual omnivorism is something I aim to emulate. Gillian Sankoff has quietly been an unofficial mentor throughout my grad school years, and is my role model for how to be a brilliant linguist and unfailingly kind human being at the same time. I've also particularly benefitted from various conversations with Gene Buckley, Robin Clark, and Charles Yang. And I owe a special thanks to a one-time member of the department, Charles Boberg, for not only getting me started in sociolinguistics but also putting me in a position to succeed at Penn before I ever arrived.

As officemates, classmates, collaborators, and friends, my fellow graduate students have taught me as much as our professors. I've been grateful for the ongoing camaraderie and support from my happy little cohort, my fellow regulars at the linguistics lab, and my long-suffering statistics consultants. I would especially like to thank Chris Ahern, Akiva Bacovcin, Aaron Dinkin, Aaron Ecay, Sabriya Fisher, Joe Fruehwald, Duna Gylfadttir, Anton Ingason, Tricia Irwin, Daniel Ezra Johnson, Soohyun Kwon, Yong-cheol Lee, Marielle Lerner, Constantine Lignos, Laurel MacKenzie, Brittany McLaughlin, Hilary Prichard, Kobey Shwayder, Betsy Sneller, Joel Wallenberg, and Georgia Zellou for sharing their ideas and feedback with me on many occasions.

Not everyone who has been there for me in the past five years has been a linguist. I've counted myself very lucky to have the administrative support of Amy Forsyth, who never misses a beat but will always help me catch up when I do, and Sue Sheehan, who must be doing at least four people's jobs at the Linguistics Lab and yet still is the kind of sunshiny presence who brings in home-baked goodies for birthdays. I'm indebted to Janice Asher, Di O'Neill, and Robert Neroni for going above and beyond in their professional and personal investments in my health, happiness, and sanity. Also, as a someday cat lady, I feel no shame at thanking my cats for absorbing many tears of grad-school-induced angst.

Lastly I wish to thank my family. My grandparents, Jean and Louis Tamminga and Hendrika Hofland, have been my cheerleaders since I was little, greeting my every success with love and pride. My brother Aaron and my sister Mila deserve thanks for being the most easy-going, good-natured, warm and loving siblings a girl could ask for. And finally, the most heartfelt thank you goes to my parents, Ken and Nancy. I could not have been blessed with a more wonderful father and mother. They are a constant source of love, support, engagement, advice, and laughter in my life every day. I would not be who I am, or where I am, without them.

ABSTRACT

PERSISTENCE IN THE PRODUCTION OF LINGUISTIC VARIATION

Meredith Johanna Tamminga

David Embick and William Labov

This dissertation, which is situated in broad debates over the delineation of abstract grammatical knowledge from the use of language in context, argues for distinct but interacting contributions from grammatical, psychological, and social factors in the production of intraspeaker linguistic variability. The phenomenon under investigation is the tendency of speakers to repeat recently-used linguistic options in conversational speech, which I refer to as *persistence*. I take up three major themes: the use of persistence as evidence on the mental-representational unity of variable linguistic processes; the interaction of different loci of variation with different cognitively-rooted facilitatory effects; and the contextual sensitivity of persistence to both social and grammatical expectations.

The core results of this dissertation are based on data from 122 interviews drawn from the Philadelphia Neighborhood Corpus (Labov and Rosenfelder, 2011). I argue for distinct phonological and morphological processes in the production of the common morphophonological variables ING (working/workin), TD (old/ol), and DH (them/dem/em), with morphological variation showing generalized persistence while phonological variation is persistent only under conditions of lexical repetition. Specifically, I propose that verbal and nominal ING constitute distinct variables, as do past tense and monomorphemic TD, and that the alternation between stop and continuant consonants in DH is morphological in nature. The quantitative decay profiles of these variables, I suggest, tie their phonological versus morphological loci to their representation in episodic versus abstract memory systems.

Although the driving force behind persistence, in this view, is the operation of various general cognitive processes, I further argue that these processes reflect speakers' sociolinguistic awareness in a way that supports a holistic expectation-based view of persistence asymmetries. The quantitative results and new questions in this dissertation set the stage for continued progress toward an integrated model of how social, grammatical, and psychological forces contribute to the production of linguistic variation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Since its inception in the 1960s, quantitative variationist sociolinguistics has made substantial progress in understanding how both social and linguistic factors relate to language variation and, in turn, to language change. These advances have largely been made on the level of the speech community: a group of speakers who orient themselves towards a common set of evaluative norms and whose language use is, as a result, systematically structured. Hearkening back to the "grammars of the speech community" of Weinreich et al. 1968, Labov calls it "the central dogma of sociolinguistics that the community is conceptually and analytically prior to the individual" (2012, 266). This emphasis on the speech community has given rise to a rich literature detailing the driving role of external social factors such as class, gender, network, and cultural ideology in language variation and change. It has also substantially improved our understanding of the influence of internal linguistic factors on variation, with patterns of language use often revealing grammatical underpinnings. There is, however, another perspective on variation that remains poorly understood: the role of a speaker's cognitive processes, such as sentence planning or working memory, in their perception and especially production of language variation. The central goal of this dissertation is to advance our understanding of the interplay between psychological, grammatical, and social influences on language variation and change.

In the course of pursuing this goal, I advocate for a renewed focus on the human mind, as embodied in individual human beings, as the wellspring of language variation. Of course, the individual has never been absent from the study of the speech community. It is only through studies of the community that intricately stratified patterns of social class, gender, and style can be described; these patterns provide a backdrop against which the behavior of speakers with certain social characteristics can be understood. The Labovian view that the community supersedes the individual is set in opposition to a recurring focus on the *idiosyncrasy* of individuals. Weinreich et al. (1968) trace the latter view back to Hermann Paul and his view that the idiolect is the only true object of linguistic study. More recent calls to elevate the individual to greater prominence in sociolinguistics bear some resemblance to this 19th century viewpoint, putting at the forefront of the study of language variation "individuals creating unique ways to sound, to be, and to respond to specific rhetorical exigencies" (Johnstone, 2001, 123). The conception of language as "fundamentally particular, fundamentally the property of individuals" (Johnstone, 1999, 313) is not what I advocate here. The question I wish to address here is not what some specific individual knows or does, but rather what *all* individuals must know in order to speak in such a way that the patterns of the stratified speech community emerge. In other words, the focus is not on what differentiates individuals, but what unites them—a view entirely compatible in spirit with the primacy of the speech community, but also demanding of a fresh methodological approach. This is a perspective that privileges linguistics as a cognitive science rather than a social science or one of the humanities, and suggests that evidence from language variation makes a unique contribution to the puzzle of characterizing human linguistic competence.

The particular phenomenon under investigation in this dissertation is the tendency for a recently-used linguistic option to be used again, which I refer to as *persistence*. Such a tendency is relevant to the study of language variation and change because it impinges on the choice between the variants of a linguistic variable. The study of persistence in language variation has sometimes been advocated as a way to refine our description of the sources of the observed variation (Gries, 2005; Szmrecsanyi, 2006). Persistence effects are sufficiently robust that they often significantly improve statistical models of variation when included as a predictor. Although accounting for portions of observed variation is of course important, I suggest that the motivations for studying persistence are deeper.

First, persistence may provide compelling evidence on the central issue of 'sameness' in language. A core task of linguistic analysis is to give a principled adjudication of sameness on multiple levels. This is what Labov terms the "fundamental postulate of linguistics": that, despite extensive variability at every linguistic level, some utterances are the same (2004). The psycholinguistic priming literature recognizes that priming and related effects provide a dependent variable measuring the classification of things as the same or not in a way that does not require explicit judgments (Bock and Loebell, 1990; Branigan et al., 1995; Ferreira, 2003). I will suggest, following Estival (1985), that such subconscious reflections of structural unity emerge in naturalistic data as well. As persistence is a speaker's tendency to use the same linguistic option again, careful observations of persistence may shed some light on what counts as 'the same' and thus contribute to linguistic analysis.

Second, it is possible that persistence effects may be susceptible to both cognitive-universal and sociostylistic conditioning. Persistence is sometimes observed to be asymmetric, such that one variant in an alternation facilitates its own re-use more strongly than the other variant. The determinants of persistence magnitudes and asymmetries are not yet well understood. Here I suggest that the key is not what is the same, but what is different. The modulation of persistence asymmetries in response to shifts in context reflects speakers' dynamic estimation of what is unexpected. This estimation, I suggest, must be closely linked to a speaker's knowledge of stylistic markedness. Among the major sociolinguistic parameters such as gender, age, social class, and style, the latter is the only one that is manipulated in the course of speech production by any given individual.[1] The existence of persistence phenomena is unlikely to be attributable purely to style-shifting, though, reminding us that psycholinguistic factors are also at work in producing intraspeaker variability. The sensitivity of persistence asymmetries to context lies at the interface between universal processing effects and stylistic competence. Improving our understanding of persistence strength, then, may lead both to refinements in our ability to quantify style and to a clearer picture of the components of language use.

1.1 A note on terminology

The study of repetitiveness in variant choice has fallen under a wide variety of terms – persistence, priming, perseverance, parallelism, etc. – which seemingly change with every new publication. I use **persistence** here as a neutral term that has also been recently used by Szmrecsanyi in his dissertation and related publications (2005; 2006). Estival notes that the term *syntactic priming* is often used for this effect, at least when it operates on syntactic variables, because "it is assumed to be due to a tendency to use again a syntactic construction when it has already been produced or processed" (1985, 8). The term *priming* reflects a connection to the experimental psycholinguistic literature, where priming is the term used for shortened reaction times after previous exposure to an identical or related stimulus in experimental settings. The psycholinguistic literature has also turned to the study of production priming, showing that speakers who have heard particular syntactic construction are more likely to use that construction in a picture description task (see e.g. Bock (1986); somewhat ironically, Bock actually uses the term *persistence*). Gries (2005) in particular has been keen to make this connection explicit and claim that the mechanisms

[1] I take style here as an umbrella term encompassing a wide range of speaker behaviors relating to topic, interlocutor, stance, etc.

at play are one and the same.

I follow Szmrecsanyi in preferring a somewhat neutral term to describe repetitiveness as observed in corpus data. Avoiding the use of the term *priming* until it is empirically motivated allows for a more interesting discussion of whether, and to what extent, persistence effects can be attributed to the same mechanisms of priming in the psycholinguistic sense. Unfortunately the term *persistence* does not come with a transitive verb (analogous to "A primes B"), nor does it have a pair of nouns for the objects that are assumed to trigger and be affected by the phenomenon (analogous to "prime" and "target"). As a compromise, then, I will use *persistence* to refer to the phenomenon under investigation but freely borrow the terms *prime* and *target* to refer to the relevant tokens. The discussion of the relationship between the corpus-based observation of persistence and the experimentally observed priming effect is one that I hope this dissertation can help further inform.

1.2 An overview of the argument

The following chapters consider the grammatical, psychological, and sociostylistic perspectives on persistence, respectively. My primary focus throughout is on a set of morphophonological variables that are familiar in the sociolinguistics literature. To a first approximation the variables are ING (*working* ∼ *workin'*), TD (*old* ∼ *ol'*), and DH (*them* ∼ *dem* ∼ *'em*). In chapter 3 I begin by showing that persistence does not arise uniformly across the grammatical contexts that condition these variables. For example, there is persistence within but not across the nominal and verbal categories of ING. I take this as evidence that the surface variation normally subsumed under these variables actually originates from multiple points of variation, in effect deconstructing single variables into multiple variables. Within each of the newly-restricted variables, I focus particularly on whether the persistence effect is abstracted or lexically specific. The generalization I propose on the basis of this evidence is that variables with a morphological locus are persistent across their occurrence while phonological variables elicit persistence only under conditions of lexical repetition from prime to target.

In chapter 4, I extend the view on the persistence behavior of the morphophonological variables to include the decay of both the generalized and lexically-specific persistence effects as the prime and target get further apart. I show that the phonology–morphology distinction I proposed in chapter 3 has a temporal realization as well, with the lexically-specific persistence effect characteristic of phonological variation being apparently insensitive to the lag time from prime to target. This contrasts with the more generalized forms of persistence that I suggested arise from morphological variation, which decay over time, albeit relatively slowly compared to some of the effects of repetition priming of lexical items. My argument in this chapter is that the different persistence features of these types of variation reflect the way the variables are stored and recalled. In making this argument I delve into the temporal characteristics of experimental priming results, both repetition priming and structural priming.

In chapter 5 I ask what other forces besides decay over time can have an impact on the magnitude of persistence effects. By showing that verbal ING persistence is asymmetric in a way that is consistent across the speech community, I make the case that sociostylistic factors contribute to speakers' expectations about variation. This sociostylistic influence is set in opposition to linguistically-driven asymmetries in both morphological and phonological variation. Ultimately I conclude that a holistic view of unexpectedness, rather than one determined by frequencies in even carefully-delineated contexts, is the right way to go in understanding the forces that shape the size of persistence effects. This conclusion further strengthens the perspective of persistence as simultaneously implicating social, grammatical, and psychological factors.

Chapter 6 concludes by considering the conceptual disentanglement of the relationship between priming and style-shifting as two factors related to the use of variables in real time. In addition to surveying the new evidence from the dissertation that pertains to this question, I make several methodological suggestions for how future work might continue to pursue the discrimination of these factors. Finally, I outline a proposal that expands the suggestions made in chapters 3 and 4 regarding variable type and persistence behavior into a broader framework of predictions on the interaction between the grammatical architecture and the processing system. My contention is that quantitative data from intraspeaker variation can serve as a unique form of evidence on the nature of human linguistic systems.

1.3 Some background on sequences of variable observations

By way of background I begin with a review of the corpus persistence literature, an umbrella under which I unify the overlapping sociolinguistic and corpus linguistic literatures on persistence effects. This is the body of work that this dissertation is most clearly rooted in. Next, I give a partial introduction to the experimental structural priming literature, which may be thought of as a controlled experimental analog to the corpus persistence literature, especially those aspects of it that focus on syntactic, rather than lower-level, persistence. Finally, I devote some attention to a selection of recent papers that pick up more complex questions of repetition and temporal dynamics in variable production, which foreshadows the issues I discuss in chapter 4.

1.3.1 Corpus persistence

The body of research that most directly precedes the results in this dissertation is the literature on persistence tendencies in naturalistic data from sociolinguistic corpora. The observation that subsequent tokens of a variable may be in some way related to each other is not new. The earliest study to identify a persistence effect, to my knowledge, is Sankoff and Laberge (1978). Calling "the interactions of neighboring tokens...an interesting phenomenon in its own right" (Sankoff and Laberge, 1978, 119), the authors aim to characterize this effect on three variables in Montreal: the pronominal alternations between *on* and *tu/vous* for general indefinite human reference, *on* and *ils* for exclusive indefinite reference, and *nous* and *on* for first person plural. They divide potentially-linked tokens of each variable into three categories: embedding-constrained tokens, which share a referent within a clause; sequence-constrained tokens, which share a referent across adjacent clauses; and unconstrained tokens, which are more distant (the authors do not state whether the referent needs to be the same in this latter category). Although they hypothesize that these categories should form a cline of closeness along which the likelihood of a switch in variant choice should become more likely, the data on embedding-constrained and unconstrained tokens are too sparse for in-depth analysis. Their results on sequence-constrained tokens, however, provide strong evidence for what looks like a persistence effect, with speakers switching only about one-third as often as the null hypothesis of independence would predict. They do not report any apparent asymmetries, concluding in the case of the general indefinite pronoun variable that "an almost identical pattern holds true for switches in the opposite direction" (Sankoff and Laberge, 1978, 122).

Another influential early demonstration of such an effect comes from Poplack 1980 and Poplack 1984. In this study of the factors conditioning variable inflectional /s/- and /n/-deletion in Puerto Rican Spanish, Poplack puts forth the functionalist argument that deletion of /s/ in plural noun phrases and /n/ on plural verbs will be blocked when there is no disambiguating information in the pragmatic context, and permitted when the plurality can be inferred and thus doesn't need to be marked for communicative purposes. In the process of furnishing evidence for this argument, she uncovers the apparently counter-functional effect that preceding /s/-deletions within the noun phrase favor further deletion, while preceding retentions disfavor deletion. In her terms, "One marker leads to more, but zeros lead to zeros" (Poplack, 1980, 378). Her primary interest in this result is its apparent counterfunctionality (since too much deletion of agreement markers may presumably lead to miscommunication). Poplack and Tagliamonte, in their study of verbal /s/ in non-standard English, similarly demonstrate that "once a zero is used…another zero is most likely" (1989). They treat this as one of many constraints on variable verbal /s/ without interrogating its implications in depth.

The focus on /s/-deletion in Romance languages was continued with a series of studies of persistence in Brazilian Portuguese. Scherre and Naro discuss "a parallel processing effect, such that marking leads to further marking and lack of marking leads to further lack of marking" (1991, 23). In this study of subject-verb and subject-predicate adjective agreement, they divide the tokens of the dependent variable into three groups: verb preceded by a marked verb, verb preceded by an unmarked verb, and isolated or first in the series. The unit of analysis for dependence across tokens or isolation is sequences of clauses with the same subject, which could be broken by interlocuter interruption or a distance of greater than 10 clauses. This is the discourse level. They find the same effect for both verbs and predicate adjectives in such sequences: isolated or initial occurrences have a deletion rate near the overall deletion rate and a varbrul factor weight near 0.50; tokens following a marked token show an increased likelihood of being marked and have a factor weight disfavoring the probability of deletion; and tokens following an unmarked token show an increased likelihood of deletion and have a correspondingly high factor weight favoring deletion. They also look at the effect of subject marking on verb marking or adjective marking, as well as verb marking on adjective marking, within the clause, finding the same effect but concluding that the clausal level shows weaker parallelism in marking than the discourse level (in other words, that persistence of inflection from verb to verb is stronger than persistence from subject to verb). An apparent result which they do not discuss but which is relevant to my dissertation is that on the clausal level at least, the persistence effect of non-marking appears to be stronger than the persistence effect of marking. In all three cases (verb-by-subject, adjective-by-subject, adjective-by-verb) the frequency of marking is only slightly above the overall rate when the preceding token was marked, but dramatically lower than the overall rate when the preceding token was unmarked. It appears, then, that there is a persistence asymmetry in these results which holds in the same direction for both verbs, which have an overall marking rate of 78%, and adjectives, which have an overall marking rate of 51%.

The overall conclusions of Scherre and Naro (1991) are that marking tendencies are counter-functional, that independent tokens of a variable are not statistically independent, and that "formal parallelism...should be considered as a serious candidate for a universal of language use and processing" (1991, 30). In a follow-up study from the next year, the same authors investigate the relative strengths of persistence (which they begin calling the 'serial effect') and more traditional internal and external factors. They do this by splitting their dataset into two sections: the isolated or initial occurrences in one data subset, and the tokens that are (a non-initial) part of a sequence in a separate data subset. They then run separate analyses for these two subsets and find that the effect of how much education the speakers have is a significant conditioning factor in the subset of isolated and initial tokens but not in the serial tokens (Scherre and Naro, 1992). They link this to similar findings across a variety of other studies, where it is apparently the case that factor groups representing persistence are usually the first to be selected (and thus the strongest) by a step-up/step-down logistic regression in GoldVarb. The important message from this paper is that persistence effects are not marginal or hard to detect; they are generally large in magnitude and robust to different approaches to coding and analyzing the data. Scherre (2001) revisits the proposal from Scherre and Naro (1991) that persistence (or as she calls it again here, parallelism) should be considered a universal processing effect. The premise for this study is a careful replication of Poplack (1980), with the same coding scheme and approach to the data on /s/-deletion in different positions within noun phrases. She concludes that the persistence effect she observes "is clearly not mechanical" and that its nature "remains to be completely understood" (Scherre, 2001, 104).

There are other /s/-deletion studies too numerous to list here (but see Hernández-Campoy and Trudgill 2002 for an extensive list of references). The interest in Romance /s/-deletion includes an investigation of cross-variable persistence – the question of whether there is a correlation between variable null subject pronouns and variable verbal agreement marking (Cameron, 1993). The connection here is intended to be a functional one, where there is a reason to suppose the two processes might be linked grammatically. This makes it somewhat different from potential cross-variable persistence for decidedly unrelated (except, perhaps, stylistically) variables such as TD and ING.

The previous papers discussed thus far are primarily within the realm of morphophonology. Weiner and Labov 1983 show that the persistence effect is not limited to morphophonology, demonstrating that a verb is more likely to occur in the passive voice if there is a passive construction within the preceding five clauses. They describe this effect as "a mechanical tendency to preserve parallel structure," further noting that "there is undoubtedly a stylistic factor operating" (Weiner and Labov, 1983, 56). This study of the constraints governing selection of the active or passive voice, which the authors treat as functional alternants in most cases, found that while the external factors of style, sex, class, ethnicity, and age had small to insignificant effects on the selection of the active or the passive, the internal effects of information structure (given versus new) and parallel surface structure were both significant in a logistic regression at $p < 0.001$. They conclude that "the ordering of surface syntax across clauses is the predominant linguistic influence on this choice" (Weiner and Labov, 1983, 52). Since a subset of the preceding clauses with parallel structure also provide a coreferential subject, they additionally consider the independent effect of a preceding parallel structure that does not have a coreferential subject and conclude that the existence of a passive structure anywhere in the five clauses preceding a token strongly favors the choice of the passive in that token.

The same variable is revisited in more detail by Estival (1985). She considers a wider array of discourse factors to see whether their combined effect might turn out to be the true cause of the persistence effect detected by Labov and Weiner. She also asks whether the effect holds across different types of passives or whether it shows a within-type restriction. The two types of passives she considers are transformational passives ("John is believed to have left") and lexical passives ("John is interested in music" - examples hers). She finds that each type of passive makes it more likely that that type will be used again, but that the effect does not hold across the two different types of passives. She concludes that transformational passives promote transformational passives but not lexical passives, and vice versa, leading Estival to conclude that "the effect we have been studying [is] a syntactic priming effect" (1985, 21).

Estival also discovers an apparent asymmetry in the priming data. When investigating cases where the same verb is repeated within five clauses, she finds that "having recently used a verb in the passive strongly disfavors using it again in the active, while having used a verb in the active does not so strongly disfavor reusing it in the passive" (Estival, 1985, 15). When discourse-based repetitions (such as answering a question) are excluded and only purely lexical repetitions are considered, the effect is even stronger, such that there are no cases where a verb is first used in the passive voice and then repeated in the active voice.

Gries (2005) is a recent study of persistence unique for the explicitness with which he draws the connection between psycholinguistic priming in experimental work and corpus-based observations of persistence effects in use. He defines *syntactic priming* as "the tendency to reuse syntactic constructions" (Gries, 2005, 365), nodding to but ultimately rejecting an argument by Szmrecsanyi (2005) that *persistence* would be more appropriately used as a general term. His literature review covers experimental work in syntactic priming in great detail, but makes no reference to the extensive previous literature from sociolinguistics on persistence effects across multiple phonological, morphophonological, syntactic, and morphosyntactic variables. This emphasis in the literature review is presumably shaped by his stated intention to take up the questions posed by the experimental syntactic priming literature.

The syntactic phenomena Gries studies are the dative alternation and particle placement. Besides demonstrating that both of these alternations show persistence effects in the corpora he uses, he also undertakes a consideration of the role of verb identity in persistence. The core result, relative to the questions at hand in this dissertation, is that "both the dative alternation and particle placement exhibit a similar strength of the priming effect, and in both cases one construction primes more strongly than the other" (Gries, 2005, 383). In the case of the dative alternation, ditransitives have a stronger persistence effect than prepositional datives, while for particle placement it is the particle-final variant that more strongly facilitates subsequent use of the same variant. He also shows that individual verbs differ in their sensitivity to persistence based on their preference for different constructions. In both cases (but particularly in the first one) it might be an open question whether the alternations should be considered variables in the strict sense, but from a use point of view they might be treated as variables (Lavandera, 1978; Labov, 1978).

The most substantial piece of research produced to date on the topic of persistence is Szmrecsanyi (2006). Some results from this book are also published in Szmrecsanyi (2005). Szmrecsanyi lays out three goals of studying persistence in corpus data: "1. To show that corpus data...can match psycholinguistic data; 2. To suggest a methodology to integrate persistence into variationist research designs; 3. To demonstrate that consideration of the phenomenon can increase the linguist's ability to account for linguistic variation, and to predict speakers' linguistic choices more accurately" (Szmrecsanyi, 2005, 116). A notable strength of this work is the sophistication of the statistical modeling, as well as the degree to which Szmrecsanyi carefully operationalizes and accounts for the influence of conditioning factors other than persistence, such as information structure, dialect region, register, age, textual complexity, frequency, and lemma identity. Of the five syntactic and morphosyntactic variables that he studies in a range of English corpora, he reports asymmetries for three (future marking, particle placement, complementation) and no asymmetries for two (comparatives, genitives). In all three cases that show asymmetries, there are differences across the corpora in the direction and magnitude of these asymmetries. To take complementation strategy as an example, in four out of five corpora, gerundial complementation ("likes singing") is "stickier" (as Szmrecsanyi puts it) than infinitival complementation ("likes to sing"). In the fifth corpus, however, gerundial complementation is much less persistent – so much so, in fact, that it is less persistent than infinitival complementation in this corpus. Szmrecsanyi tentatively notes stylistic differences between the corpora in a way that suggests he suspects a link between persistence and style, but this possibility is not developed at any length.

1.3.2 Experimental structural priming

Next I turn to the another distinct body of literature that is of relevance for this dissertation: psycholinguistic research on experimental structural priming. Priming in psycholinguistics is more traditionally thought of as lexical or semantic, wherein prior presentation of 'dog' speeds reaction times to lexical decision on 'dog' (repetition priming), 'cat' (semantic priming), or 'dogs' (morphological priming). The structural, or syntactic, priming literature targets not the priming of stored lexical entries and their relationship with each other in the lexicon, but rather syntactic structure. Pickering and Branigan define structural priming as "the phenomenon whereby the act of processing an utterance with a particular form facilitates processing a subsequent utterance with the same or a related form" (1999, 136). Work in this domain often, but not always, focuses on tendencies in production rather than in latencies in perception. Early methodologies, beginning with Bock (1986) which was in turn inspired by the corpus results from Estival (1985) and Weiner and Labov (1983) described above, used a picture description task, but Pickering and Ferreira (2008) name sentence recall, written sentence completion, and spoken sentence completion as additional methodologies that have been added to the toolbox since then.

Bock's (1986) original methodology, which heavily influenced subsequent studies, is to first have participants familiarize themselves with a booklet of illustrations and listen to a set of sentences. After this familiarization phase, subjects are told they are participating in a picture and sentence recognition task, and will be asked to make forced-choice decisions about whether they have previously been exposed to sentences and pictures. In the trials of interest, the subjects hear a sentence that they have not previously heard, then are asked to decide if they recognize it. This sentence constitutes the prime. Then they turn the page in their booklet and see a picture, which they must first describe and then identify as familiar or not. The syntactic structure they choose for the description is the experimental target.

In her first experiment, Bock finds that "prepositional primes increased the incidence of prepositional utterances by 23% relative to their frequency following double-object primes, and double object primes increased the incidence of prepositional utterances by 22% relative to their frequency following prepositional primes" (Bock, 1986, 364). There is some ambiguity to this statement about relative change, which can be resolved through examination of the accompanying table. Prepositional datives occur at a rate of 48% following prepositional primes and 25% following double object primes. The 23% change Bock identifies is thus the difference between 48 and 25 in percentage points, rather than a true relative difference. The same of course holds for the double objects: they occur at a rate of 53% following double object primes and a rate of 31% following prepositional primes, hence the difference of 22 percentage points. Bock also includes a condition where the prime is an intransitive sentence, which she uses to assess the unprimed frequency of the two types of datives. She finds no particular preference for one or the other, with both occurring roughly 40% of the time.

Subsequent experimental work in the same vein as Bock (1986) has been extensive and relates directly to the themes in this dissertation, such as structural relationships, temporal decay, and asymmetries. One experimental study is of particular note from the sociolinguistic perspective because it is intended to elicit structural priming on nonstandard dialectal features. Squires (2013a) uses a paradigm in which experimental participants must guess which picture is referred to by the partially-obscured utterance they hear. The variables she investigates, *there's*+NP and NP+*don't*, are well-suited to investigation in this manner because the relevant points of variation are their combinations with either singular or plural NPs, therefore they can be related to pictures of singleton objects or groups of objects. She also enriches the structural priming paradigm by introducing accompanying images of speakers wearing socially-marked clothing and standing in front of houses that suggest different socioeconomic backgrounds. Her findings are reflective of a situation where linguistic elements of variation constrain social interpretations, but not vice versa.

1.3.3 Dynamic approaches

Labov touches on the mystery of how these patterns emerge from the speaker's production, noting: "This is remarkable when we consider the irregular fluctuations of the variables that seem to mark the individual sections of speech. For example, here are the occurrences of (th) in casual speech, in the order that they occurred: 1 2 2 1 2 2 1 1 1 1; and here are the occurrences in careful speech: 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1. There seems to be no pattern or system within this sequence – yet it fits into the larger pattern shown in the array of styles'" (Labov, 1966, 77). By drawing attention to the translation of consecutive strings of variants into a structured pattern over larger sections of speech, this quotation foreshadows the questions I ask in this dissertation.

In the realm of sociolinguistics, the issue of sequences of variation has begun attracting scattered interest in both perception and production. The perception of variation is the topic of Labov et al. (2011), which focuses on the notion of a "sociolinguistic monitor". The authors contend that such a monitor is not an abstract construct but rather "an observable property displayed by our subjects" (Labov et al., 2011, 432). The property in particular that the subjects display is their reaction to sequences of occurrences of ING in a spoken passage. As the authors point out, "the question remains open as to how sensitive listeners are to the fine-grained quantitative differences in production that are the normal output of variable rules" (2011, 432). This is the question that they seek to answer in their study, which is designed as an expansion of the matched guise paradigm. The usual matched guise method is to have experimental participants give their subject evaluation of words or phrases that differ minimally along some dimension of linguistic interest. Divergent evaluations can then be attributed to the manipulated dimension. What Labov et al. (2011) do is embed a series of 10 instances of ING in a news broadcast script that participants rate on a 'job suitability' scale. The scale asks them to rate how well-suited the speaker is to the job of news broadcaster, a job which is uniformly understood to require highly standard speech. All tokens are progressive participles (complements to the copula) and are cross-spliced to produce passages with different rates of /in/ and /ing/ tokens.

In the first experiment the primary point of evaluation is the overall proportion of /in/ versus /ing/ tokens, with the only order effect tapped being a comparison between /ing/-first and /in/-first block orders in the 50% condition, where the tokens of each variant are produced in sequences (that is, all the /in/ tokens before any /ing/ tokens and vice versa). Of particular interest, however, is their analysis of the real-time unfolding of these evaluations in a later experiment where, instead of giving a single judgment after listening to the full passage, participants are given a sliding scale and asked to update it continuously while listening to the test passage. The finding is that "the subject is monitoring the occurrences of (ING) with particular attention to the form that deviates from the norms for this situation" and that "debriefing questionnaires at the end of the experiment did not show a high level of conscious awareness of (ING)" (Labov et al., 2011, 456).

On the production side of sociolinguistic variation, a recent paper is Podesva 2007. Podesva aligns his work on the social meaning of style with a recent upsurge in sociolinguistic research focusing on the behavior of individuals within the speech community, arguing that "examining intraspeaker variation is desirable because… doing so facilitates a more direct observation of how speakers deploy linguistic features to achieve social ends" (2007, 2). Of particular interest is his discussion of the "temporal dynamism" exhibited by linguistic variables (Podesva, 2007, 5). He seeks "a means of representing how styles and their component linguistic features unfold over time" (Podesva, 2007, 6). The means he comes up with is what he terms a 'variation score', which represents different variables and their variants on vertical tiers with time represented horizontally.

(3) Figure 2. Variation score for Heath's barbecue setting

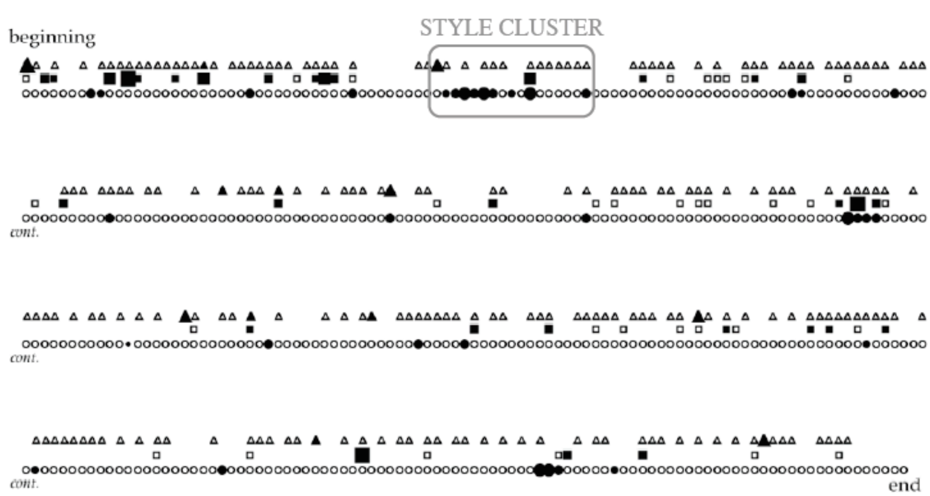

Figure 1.1: An example of a variation score (Podesva, 2007, 6)

In the variation score in 1.1, we see a new angle on the '1 2 2 1 2 2 1 1 1 1' and '2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1' sequences that Labov 1966 considered. In isolation they may appear to have no discernible pattern, but when layered on top of the sequences for other variables they begin to look less random. The 'style cluster' identified in figure 1.1 is identifiable not from a single tier alone but from the combination of the three tiers. The current weakness of this otherwise-promising approach, as Podesva acknowledges, is that without some measure of statistical significance it is difficult to distinguish meaningful style clusters from the vagaries of chance.

Another paper that takes an approach similar to Podesva's variation scores is Sharma and Rampton 2011. Sharma and Rampton consider a London speech community where multiple enregistered ethnolects coexist, making it somewhat different from the urban North American speech communities that have been the object of much sociolinguistic theorizing. Striking a balance between averaging over predetermined stylistic contexts and representing each individual token separately, they divide conversations into utterances of 5-10 words based on the boundaries of turn-constructional units, clauses, or footing shifts. They then calculate the percentage of variants belonging to each lect (Standard British English, Vernacular London English, and Indian English) across all of the variable tokens within the utterance, and graph the resulting Lectal Focusing Interaction metric against time. Figure 1.2 illustrates how quickly and dramatically a speaker is able to shift lects in the course of a single narrative.

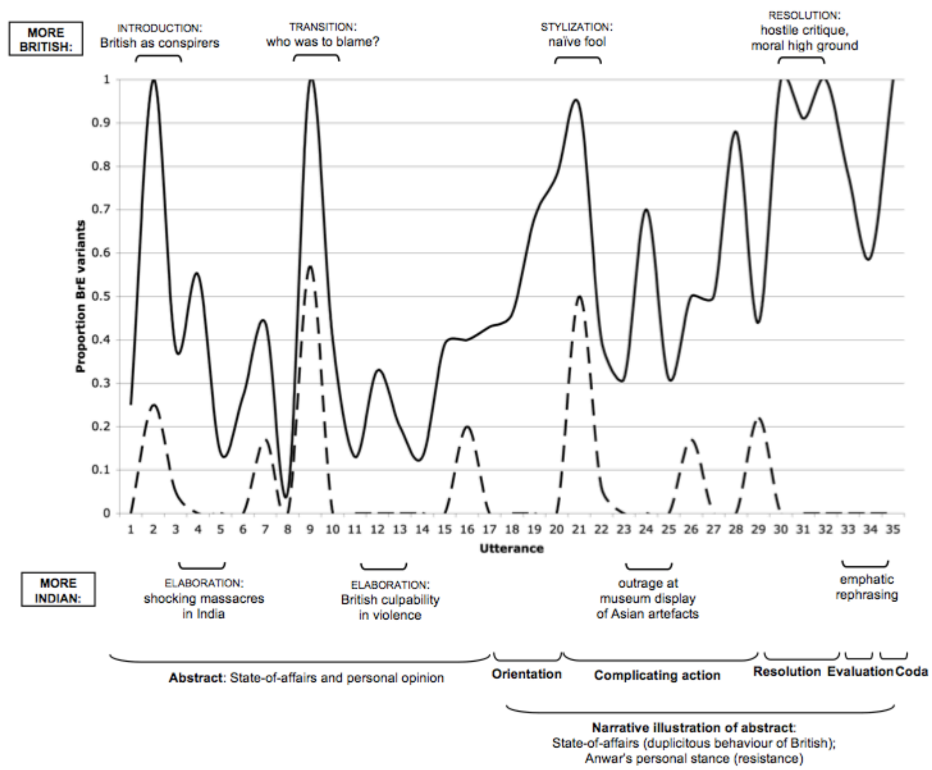

Figure 1.2: An example of Lectal Focusing in Interaction, (Sharma and Rampton, 2011, 13)

The experimental structural priming literature has also begun to investigate the effects of sequences of primes. Jaeger and Snider (2013) report on a series of studies that take multiple previous instances of the ditransitive alternation into consideration as a means of assessing an *expectation adaptation* account for syntactic priming. They distinguish between the expectations established by *prior* experience and the expectations established by *recent* experience. Prior experience in their terms means the overall expectedness of a construction in a way that is "neither individualized nor sensitive to the statistics of the current environment" (Jaeger and Snider, 2013, 60). Recent experience, on the other hand, denotes a more fine-grained picture of the stimuli presented in the experimental context within a narrow temporal window. Their first study is an analysis of naturalistic spoken data, and they include a predictor of the cumulative proportion of prepositional object construction choices in each dialogue up to the point of each observation. They do not find a main effect of this cumulative proportion, but neither do they find a particularly strong or significant priming effect at all. In their second study, they conduct a meta-analysis of three experimental studies using written presentation of primes. One of these studies is Kaschak (2007), which manipulates the proportions of the primes in blocks. They replicate the result that the proportion of prepositional objects in the prime phase has a significant main effect on the likelihood of producing a prepositional object as a target.

In their third study, Jaeger and Snider (2013) construct and conduct an original experiment that manipulates sequence of primes to be either balanced (regular alternation between one construction and the other) or blocked (consistent exposure to one construction in the first half of the experiment, then consistent exposure to the other construction in the second half. They find a significant effect of prime surprisal given the recent exposure — in other words, primes are more effective due to being more unexpected when they come right after a switch from a block of one variant to a block of another, than when both variants are being presented interspersed. I take this as an indication that ordered sequences do matter and will be of interest to study in later work. Their results also support an implicit learning account of structural priming such as that proposed by Chang et al. (2000).

Chapter 2

Data and methods

The following chapters present case studies making use of a number of corpora. The primary corpus is the Philadelphia Neighborhood Corpus (PNC). I also present an additional case study using data from the Frank Porter Graham Corpus, which is a corpus of child African American Vernacular English. My motivation for working with corpora of different linguistic varieties is to show that the effects I demonstrate are neither accidental nor limited to a particular speech community.

2.1 Data

2.1.1 Philadelphia Neighborhood Corpus

The Philadelphia Neighborhood Corpus of LING560 Studies (PNC) (Labov and Rosenfelder, 2011) is a collection of sociolinguistic interviews conducted between 1973 and 2010. This corpus furnishes the data for the core studies of how persistence in the morphophonological variables ING, TD, and DH interacts with their linguistic structure and sociostylistic conditioning. Interviews from the PNC are generally conducted in the speaker's home or another familiar community location, such as a bar, community center, or retirement home. They typically last around an hour, although some are as short as 20 minutes and some approach two hours. Following the theory and methods described in Labov (1984), the interviews focus on eliciting narratives of personal experience to direct the speaker's attention away from the interview task.

Not all of the 1,087 recordings currently available as part of this collection have been transcribed, which puts a practical constraint on the amount of data that can be readily extracted. I initially selected a core sample of 42 interviews, all with white speakers from the working-class Irish and Italian neighborhoods of South Philadelphia, that were at least 45 minutes long and had been transcribed in their entirety. This smaller sample is roughly balanced by sex and decade of birth, as shown in table **??**. Although the sample contains three generations, I do not expect to find apparent time effects because the variables of interest—TD, ING, and DH—are understood to be stable in Philadelphia (Labov, 2001a). Although a sample of this size is typical for a study of community-level patterns of variation, it proved to be insufficient for the investigation of grammatical effects in ING and TD. When considering the cross-tabulation of several grammatical categories for each of these variables, the data quickly becomes sparse in certain areas. I therefore augmented the sample with an additional 76 white speakers whose transcribed interviews I coded only for ING and TD, shown in table 2.2. The choice of these additional interviews for inclusion was based purely on the length of the transcript available, with the resulting sample reflecting both asymmetries in the demographic groups that make up the corpus and current transcription priorities in its development. There are more women than men included, and the early years of the corpus are somewhat overrepresented. The extra interviews were not coded for DH because the volume of data obtained from a single interview is so much larger, and takes correspondingly longer to code.

The transcribed interviews in PNC have been forced-aligned using the FAVE suite (Rosenfelder et al., 2011), which in turn allows efficient hand-coding of phonological vari-

Birth year	Male	Female	**Total**
Before 1930	5	5	**10**
1930–1959	11	10	**21**
After 1959	5	6	**11**
Total	**21**	**21**	**42**

Table 2.1: 42-speaker PNC sample for DH coding

Birth year	Male	Female	**Total**
Before 1930	18	28	**46**
1930–1959	18	29	**47**
After 1959	16	9	**25**
Total	**52**	**66**	**118**

Table 2.2: 118-speaker PNC sample for ING and TD coding

ables facilitated by a Praat script.[1] The aligner uses the Carnegie Mellon University Pronouncing Dictionary to assign phonemic pronunciations to the orthographically transcribed words, then assigns each phoneme to a segment of the recorded speech signal. A quirk of the CMU Dictionary is that it often includes multiple pronunciations for a lexical item. In some cases these include the casual variants that I wish to target, such as an apical-final pronunciation in *working*. This interferes with searching for tokens of the variables, so I cleaned up the dictionary using a Python script so that only the standard pronunciation is available during alignment. After this step is taken and the alignment is complete, the morphophonological variables can be defined over strings of phones that map to CMU Dictionary pronunciations, allowing the handCoder Praat script to query the phone tier and move automatically to each occurrence of the relevant environment in turn. The analyst is presented with the observation and the three words on either side of it, then makes an auditory judgement of the value of the dependent variable. The normal practice when coding phonological variables in this way is to set the search parameters so that tokens occurring

[1] Thanks to Joe Fruehwald for making his useful handCoder.Praat script available.

in contexts that will later be excluded are not picked up as part of the coding process. However, to allow for the revised approach to the variable context that I discuss in section 2.2.1, I included neutralization contexts in the search criteria (TD before a coronal stop, for example) and simply did not assign them a code for the dependent variable. This ensures later access to the timing and sequence of even possibly-irrelevant tokens.

The handCoder also collects certain datapoints about each observation automatically upon coding. Chief among these (from the perspective of persistence) are the timestamp of the observation and the identity of the lexical item. Grammatical coding can't be done automatically because the PNC transcripts PNC are neither morphologically annotated nor syntactically parsed. However, the handCoder does allow for the entry of additional fields, so I added codes for the grammatical classes (which will be described in chapter 3) while doing the auditory coding. In cases where the grammatical category could not be reliably determined from the words in the window presented by the handCoder, I checked the full transcript afterwards.

With these categories included, the output of the handCoder was read into R as a data frame. I then used R to code each token for several relevant values of the previous tokens. By way of illustration, consider the phrase "I hate workin' past five but I was runnin' late in the morning." The handCoder would find three ING tokens based on their CMU Dictionary pronunciations: *working*, *running*, and *morning*.[2] I would enter 0, 0, and 1 respectively as the dependent variables, since I use 0 for /in/ and 1 for /ing/ when coding ING. I would also enter 'gerund', 'progressive', and 'monomorpheme' for the grammatical categories of the three tokens respectively. Meanwhile, the handCoder would automatically note the word identity and the timestamp. The data at that point would consist of three observations of four variables: the dependent variable, the grammatical category, the word, and the

[2]The CMU Dictionary contains apical-final pronunciations for many ING words; I used a Python script to remove these from the dictionary prior to FAVE-alignment so that the entire variable context can be recovered from the phone tier.

Word	Variant	Gram.	Time	Prec.Variant	Prec.Word	Prec.Gram.	Lag
working	0 (/in/)	ger.	12	0 (/in/)	N/A	N/A	N/A
running	0 (/in/)	prog.	17	0 (/in/)	working	ger.	5
morning	1 (/ing/)	mono.	20	1 (/ing/)	running	prog.	3

Table 2.3: Toy demo of how persistence factors are coded

timestamp. Then in R, I create new variables for 'Preceding variant', 'Preceding word', 'Preceding grammatical category', and 'Lag' (the difference in seconds between the target observation and the preceding once, with distances measured from word ends). The result would be a data frame as in table 2.3. As the first line of table 2.3 makes clear, the first observation will not be a useful token because it has no prime. However, it can still serve as the prime for the second observation. The method that I have just described is the one that I use throughout, across different corpora and variables.

2.1.2 Frank Porter Graham Corpus

The Frank Porter Graham Corpus is a corpus of African American Vernacular English (AAVE) child speech. The corpus was collected as part of a study conducted by the Frank Porter Graham Child Development Institute in North Carolina. A cohort of children belonging to families that were given relocation vouchers to leave dangerous or impoverished neighborhoods have been recruited for annual evaluation from ages 4 to 17, beginning in 1990. The child participants in the corpus are all African American and native AAVE speakers. The interview data included in the analyses reported in chapter 6 come from interviews conducted with child/mother (for the younger children) or child/peer (for the older children) dyads. The data in this corpus is in a less refined format than the other corpora that I deal with. As a result, I use it only in a single case study based on the data collected and coded by Tamminga and McLaughlin (2013).

2.2 Measuring persistence

In order to delve into the fine-grained behavior of persistence effects, I find it necessary to give careful consideration to several methodological issues that are not required for standard variationist sociolinguistics. One issue is the complications posed for persistence by the definition of the envelope of variation. Making decisions about which tokens should be counted in an analysis and which should not is rarely a trivial task, but it increases in complexity when token candidates are dependent on each other in sequence. Another issue is that the notion of persistence strength has not been concretely defined. This makes it difficult to test hypotheses relating to the sensitivity of persistence to social and grammatical contexts, particularly those that propose an inverse relationship between variant rates and persistence strength. A third issue is that of statistical evaluation: while it is relatively straightforward to ask whether we can say with confidence that persistence is present, we are on shakier ground when it comes to the question of whether two persistence effects are of comparable or different magnitudes. This chapter considers these three points and draws conclusions regarding the best practices for measuring persistence, thereby setting the stage for the results in the remainder of the dissertation.

2.2.1 Redefining the envelope of variation

Almost all quantitative results in the study of language variation hinge on an unglamorous but critical first step: deciding what to count and what not to count. It is standard practice in the variationist paradigm to exclude a range of contexts for any variable that is studied quantitatively (Labov, 1972; Tagliamonte, 2006). Contexts where the linguistic feature in question is not variable are generally excluded from quantitative analysis of variation, while those that do vary are kept in the analysis. The set of the retained, variable, environments is often referred to as the *envelope of variation* or the *variable context*. The task of defining the envelope of variation sounds simpler than it is in practice; often an initial round of descriptive quantitative analysis must be completed and combined with linguistic intuition or elicited grammaticality judgments to determine exactly where a feature is or is not variable.

Once the variable context has been defined, the excluded contexts generally disappear permanently from any further consideration. For instance, previous research has shown that coronal stops are almost never subject to deletion after /r/ in Philadelphia English (Cofer, 1972), so words ending in /rt/ or /rd/ clusters are not counted as undeleted tokens of TD. A separate set of contexts are ones where the analyst (and by extension, the hearer) is not able to determine by ear (or other available measurement technique) which variant was used. These are generally termed *neutralization contexts* (note that 'neutralization' here does not have its technical phonological meaning). In the case of TD, for example, the strings "las' time" and "last time" are, for practical purposes in connected speech, indistinguishable. Tokens of TD occurring before coronal stops, affricates, and fricatives are therefore excluded as being neutralized. This type of exclusion is a methodological necessity and is not, to my knowledge, understood to carry any theoretical significance. The willingness of researchers to discard these contexts is based, though, on an implicit assumption that neutralization contexts will occur at consistent rates across the various grammatical or social subsets of the data. Whether this assumption is correct or not is an empirical question that I will not take up here.

While this approach to defining the variable context has been widely used in quantitative variationist research for decades, and in fact has been used in studies that include persistence as a predictor of interest, it is not perfectly suited to investigations of this topic. Of the two standard types of exclusions, invariant contexts and neutralization contexts, I suggest that the latter poses the more serious problem for persistence coding. If we interpret invariant contexts as not being part of the variable at all, they can be treated straightforwardly with exclusion in the coding of persistence, just as they are in normal sociolinguistic coding. In other words, we assume that the rule is representationally formulated such that, for example, coronal stop deletion is specified not to apply following /r/, and that this is the reason for the non-occurrence in that context. The invariant context, then, is irrelevant to the definition and operation of the variable and, by hypothesized extension, to its mental representation. The same sort of reasoning could be applied to the case of final stops that are not in clusters, which have been treated as a separate variable since Wolfram (1975), and indeed are very rarely deleted in the speech of white Philadelphians. It is possible, of course, that even invariant contexts could be relevant for persistence, possibly through a phonological priming mechanism. I will investigate this question further through the relationship between ING and non-variable engma in section 3.2.4.

The issue of neutralization contexts, however, is potentially more problematic when it is necessary to code tokens dependent on the tokens surrounding them in real time connected speech. We exclude neutralization contexts because of imperceptibility on the hearer's end. This does not rule out the possibility that the speaker intended (on some level) to produce either a deletion or retention. Say there are two unproblematic TD tokens — call them A and C — separated by a TD token in a neutralization context — call it B.

A ...... B (neut) C

If the task is to code each token for the value of the immediately preceding token, coding C will be problematic. Is B the immediately preceding token? Then we do not know the correct value to assign to the previous token because the previous token was neutralized. Or does B not even count as a token? If we exclude the neutralization contexts in the initial round of coding, as we would normally do, then it will look like the previous token is A. But we have no way of knowing whether that is or is not the correct preceding token, and if B was an instance of the variable in the speaker's head then the wrong preceding token will have been coded if we use A.

The most careful approach to this area of uncertainty is to exclude not only the tokens in neutralization environments, but also the tokens that depend on them. However, it is not necessary to exclude the following token entirely — that is, it can still form part of the analysis as a preceding token for some other token. The clearest way to think about this is to think of coding persistence as the formation of pairs, and *pairs* that include neutralized tokens are excluded (rather than simply the neutralized tokens themselves). As mentioned in section 1.1, I adopt the terminology of 'prime' and 'target' for the first and second items, respectively, in each pair. This terminology, by design, implies an influence from the first item, the prime, on the second item, the target. Most tokens are members of two pairs, being the prime for one pair and the target for another. In some cases a token will be excluded with respect to its membership in one pair but included with respect to its membership in another pair. From a practical standpoint, the way this conceptualization of primes and targets is put into place is to code all tokens in the variable context including neutralization environments – neutralized tokens can be coded X instead of 1 or 0. Then, after each token is subsequently coded for the value of the preceding token, all tokens with an X *either* for their own value *or* for the value of their prime are excluded.

Another issue that is unique in the coding of persistence is the role of the interlocutor in persistence. It seems likely that the use of the same variable by the speaker's interlocutor (in the PNC, for example, the interlocutor might be the interviewer) might affect subsequent values of that variable for the speaker. Indeed, Gries (2005) finds that persistence is particularly strong across speakers, although I would suggest that in the case of construction choice (syntactic variability) the discourse-motivated use of straightforward word-for-word repetition of the interlocutor's utterance is likely to contribute to this result and should, I submit, be considered a distinct, though possibly related, phenomenon. The question of persistence across speakers broaches the issue of accommodation, by which I mean a general tendency for people to behave more like each other as they interact, aligning their verbal behavior on a number of levels. Although this is a fascinating phenomenon that raises many critical issues, I do not address it in this dissertation. Just as I draw a distinction between the observation of variant clustering in speech and the facilitation effect of a recently-produced form, there is probably a conceptual distinction to be drawn between general alignment between speakers and the effect of having perceived and processed the linguistic objects used by the interlocutor. Investigating the relationship between accommodation and cross-speaker persistence (if, indeed, these do not turn out to be one and the same thing) will need to be a subject for further research. To avoid this issue, I do not include pairs of tokens that are interrupted by speech from other conversational participants, whether the interviewer, family members, friends, or neighbors. The use of a variant by the interlocutor might well affect the subsequent use of that variable by the speaker, but I make the simplifying assumption that any effect is not further transmitted to the next following use (that is, the second token of the variable from the speaker after its use by the interlocutor).

In practice, the exclusion of interruptions is accomplished by including time-stamped transcript lines from interlocutors in the data frame being coded for persistence in R. A variable token that occurs immediately after an interruption from a different speaker, then, functions just as the initial token in an interview does: it serves as a prime for the next target, but is not itself considered a target with a valid prime. In the case of ING I only include interlocutor transcript lines containing an ING token as interruptions, whereas for TD and DH it is more efficacious to simply treat all substantial speech from interlocutors as interrupting the formation of prime–target pairs. I do not, however, treat channel cues or feedback, such as laugher or pause fillers, as interruptions.

2.2.2 Testing the significance of persistence

Finally, I turn to the question of whether any observed persistence effect is statistically significant — that is to say, is it a sufficiently large effect, relative to the variability observed generally in the data, that it is unlikely to have occurred by chance? There are a number of angles and levels from which I approach this question. In the simple univariate case, I occasionally wish to ask whether there is a difference in variant rates between two (or more) subsets based on what the prime was. In this case I turn to the binomial confidence interval. The binomial confidence interval is appropriate for populations that follow a binomial distribution, which means that each observation is the outcome of a binomial trial. The confidence interval can then be understood similarly to one constructed from the normal distribution: it indicates that on 95 of 100 occasions, given the number of observations and their variability, the confidence interval would contain the population mean. Of the variety of methods possible for constructing binomial confidence intervals, I choose to use the Clopper-Pearson exact method. This method is based on the calculation of the exact binomial distribution, rather than an estimate of the distribution, hence its name. This is a conservative method for constructing a binomial confidence interval, meaning that it may be wider than is strictly necessary. I use the `binom` package in R (CITE) for this purpose.

In most of the cases I will discuss, however, it will be necessary to move quickly beyond the binomial confidence interval and turn instead to multiple logistic regression. A crucial issue that is not addressed by the univariate approach is that there is a relationship between prime and target created not by any true sequential dependence across tokens but by the fact that both are drawn from the same distribution. It is thus more likely that the members of the pair share a variant than it would be if all of the observations in the sample were drawn from the same distribution. To make this problem more concrete, consider a simple case with two speakers, one who uses /ing/ 75% of the time and one who uses /ing/ 25% of the time. Assume we have the same amount of data from each, such that the resulting data set has a mean /ing/ rate of 50%. If all the data were drawn from a binomial distribution with p=0.5, the expectation would be to find equal rates of /ing/ after /ing/ primes and /in/ primes. However, the /ing/-dominant speaker will have produced many more instances of /ing/ overall and subsequently many more /ing/–/ing/ pairs, while the /in/-dominant speaker will have produced many more instances of /in/ overall and subsequently many more /in/–/in/ pairs. There will appear to be priming, but it will be attributable to the mixture of the two distributions rather than to a true effect of any given prime on its subsequent target.

A reasonable approach to dealing with this issue is to use random intercepts by speaker in a mixed-effects logistic regression. As a quick sanity-check to ensure that the reasonableness of this approach is in fact in line with its effectiveness in practice, I produce a simulated dataset containing 100 observations each from a set of 37 different binomial distributions with probabilities ranging from 0.05 to 0.95 by 0.025. These numbers are chosen to be roughly comparable with the nature of the datasets I will use in the following chapters. There is no priming involved in the generation of this data, so when it is fit to a logistic regression we should not find a significant effect of the "prime" on the value of the "target". Of course, when we fit a model without a random effect of speaker, we find that there is a large and highly significant effect of the prime on the target. When we add a random intercept by speaker, this effect disappears. I thus proceed with using random speaker intercepts to account for the covariation between prime and target that is attributable to their being drawn from the same distribution. I

Chapter 3

Persistence in grammatical perspective

This chapter addresses the mediation of persistence effects by the grammatical structures underlying linguistic variables. Framing persistence as the tendency to repeat the same linguistic object or option as was recently used opens up the question of what counts as "the same". This is the sort of question that, upon first consideration, might seem to pose a nearly intractable methodological problem: must we define which instances of some surface-similar set of phenomena count as the same or not prior to attempting to measure persistence? On the contrary, one goal of this chapter is to show that there is already a solid basis for the expectation that the repetitiveness effects of persistence operate on a relatively abstract, rather than superficial, level. This allows for the use of persistence as a sort of diagnostic tool. If we can show that in cases where we have independent reasons to be fairly confident where the lines of structural demarcation are drawn, persistence operates within but not across those lines, it becomes reasonable to take observed persistence patterns as reflective of shared grammatical identity (or lack thereof) in cases where the structural analysis is not known. This will be particularly useful in the case of variables whose exact definition is a point of contention, as well as in the case of constructions observed in long-complete language changes for which we no longer have access to speaker intuitions.

I begin by outlining previous results from both experimental and corpus-based research suggesting that we should expect to see a role for grammatical identity in the quantitative behavior of persistence. I then present analyses of the role of persistence in three variables which are generally understood as implicating some combination of morphological and phonological variation. The three variables are ING (*working* $\sim$ *workin'*), TD (*old* $\sim$ *ol'*) and DH (*them* $\sim$ *dem*). Immediately at stake is the description of each variable as involving probabilistic phonological processes versus competition between stored allomorphs — my approach to cashing out the notion of "phonological variation" in contrast to "morphological variation". I give the greatest attention to persistence in ING, arguing that the observed sensitivity of this effect to the relationship between the morphological structure of the prime and the target points to morphological variability. TD and DH are both compared and contrasted with ING in turn; an attempt to fully reconcile the results from all three variables is presented in chapter 4 when I consider the role of different types of memory in persistence.

By demonstrating that persistence effects are constrained by the structural relationships underlying observed variation, I aim to reinforce a certain perspective on grammar and language use. In this perspective, the use of language in context is a distinct phenomenon from an underlying generative grammatical system, but reflects the properties of the objects supplied by the grammar. Evidence from language use is therefore not at all irrelevant to the study of grammar, as it can provide a unique window on grammatical structure. The *do*-support case in particular reinforces the utility of studying the quantitative properties of both synchronic and diachronic language use as evidence to improve our understanding of grammatical structures, particularly when other sorts of evidence are inaccessible or equivocal. When quantitative results such as those from the study of persistence provide an independent line of evidence for non-obvious grammatical analyses, the conclusions we can draw about grammatical structure are strengthened.

3.1 Prospects for a role of grammar in persistence

3.1.1 The nature of the argument

The experimental structural priming literature discussed in chapter 1 has made progress in exploiting priming relationships to address the question of structural identity. The relationship between persistence and the underlying mental representations of grammatical objects has attracted more interest in the structural priming literature than in corpus studies. Under a definition of structural priming as "the phenomenon whereby the act of processing an utterance with a particular form facilitates processing a subsequent utterance with the same or a related form" (Pickering and Branigan, 1999, 136), the question of what counts as "the same" or as "related" is both relevant and potentially addressable. The general mode of reasoning is that experimental observations of priming relationships between different syntactic structures are taken as evidence for the nature of the mental representation of the underlying linguistic objects. Branigan et al. lay out the argument explicitly:

> If the processing of a stimulus affects the processing of another stimulus, then the two stimuli must be related along a dimension that is relevant to the cognitive system. Under certain circumstances, we can conclude that they are represented in a related manner. If the relationship between two stimuli is syntactic, then we can use this relationship as a way of understanding what syntactic information is represented, and how that information can interact with other information. (Branigan et al., 1995, 490)

Pickering and Branigan also argue this point when they "claim that syntactic priming taps into knowledge of language itself, and as such can inform linguistic theories that are concerned with accounting for knowledge of language" (1999, 140). They specifically advocate the use of priming studies to determine relationships between sentences. Note that this is a question that is not susceptible to introspection-based methods — there is no "relatedness judgment" comparable to an acceptability judgement. This is not, of course, to say that syntactic analysis is unable to address questions of structural relationships, but rather to point out that such relationships cannot be probed straightforwardly. There is thus a need for as many independent sources of evidence as possible for theoretical accounts of the relationships between surface-similar sentences. When different sources of evidence for a theoretical analysis converge, it considerably strengthens our confidence in that analysis.

In this chapter and the next, I bring this line of reasoning into the realm of morphophonological variables. Just as syntax faces questions of structural identity that are unanswerable with the usual methods of syntactic analysis, the study of morphophonological variation faces questions of variable definition that have not found satisfactory answers in traditional quantitative methods. The variables at hand have received extensive quantitative and qualitative scrutiny in the sociolinguistics literature, yet their basic structural definitions remain subject to debate. I suggest here that evidence from persistence can be brought to bear on the relationship between variable and invariant contexts, the operation of apparently-similar processes in distinct environments, and the relative contributions of phonology and morphology to these variables. The study of persistence can serve as a tool for understanding the relationships between grammatical strings.

Pickering and Ferreira discuss a possible extension of the use of persistence to diagnose relationships: that in the face of overwhelming general evidence for structural priming, a failure to observe priming by the same methods may indicate that the prime and target are not representationally related (2008, 429). There are a couple of caveats to bear in mind when pursuing this type of argument. First, there is the ever-present statistical issue that a non-significant result does not constitute positive evidence for the null hypothesis. We thus lack the statistical tools to demonstrate that a null result with respect to the appearance of persistence is meaningful rather than the result of chance (although see Gallistel (2009) for a Bayesian attempt to resolve this long-standing problem). Second, there is always the possibility that an apparent lack of an effect is the outcome of two countervailing effects. Allen and Badecker (2002) offer a cautionary tale along these lines, showing that the oft-claimed lack of morphological priming in irregular verb paradigms could actually be attributed to the combination of morphological priming and inhibition from orthographic similarity. The documented but poorly-understood existence of anti-persistence — what Szmrecsanyi (2006) calls "horror aequi" effects — makes this possibility particularly plausible. I will nonetheless proceed with this line of reasoning in subsequent sections, keeping these two caveats open as possible interpretations whenever a lack of persistence is a crucial result. The working hypothesis throughout this chapter, then, is that when we believe two surface-similar strings to be outputs of the same underlying grammatical structures, speakers should show a tendency to repeat the structure.

3.1.2 Previous observations of grammar in persistence

There are a number of studies, particularly experimental ones, that have taken up the same type of question as I raise here: the use of priming as a diagnostic for relatedness across surface-similar strings where the relatedness is in question. The original thrust of the structural priming literature, as discussed in chapter 1, was to show that the priming effect is, in fact, a syntactic one, rather than a more superficial effect of word repetition. For example, does "The cat was chased by the dog" prime "The man was arrested by the cop"? Although this is a fundamental first step along this line of reasoning because it shows that the persistence effect does arise in cases where we know the sentences to be structurally identical, I focus here on the handful of results showing that surface similarity does not produce priming in the absence of a shared grammatical representation.

An early result of this type is Estival's (1985) exploration of different kinds of passives. Building on the result from Weiner and Labov (1983) that passive sentences facilitate the use of more passives in conversational interview speech, Estival asks whether the effect holds across different types of passives or whether it shows a within-type restriction. The two types of passives she considers are transformational passives ("John is believed to have left") and lexical passives ("John is interested in music" — examples hers). She finds that each type of passive makes it more likely that that type will be used again, but that the effect does not hold across the two different types of passives. This study took place at a time when the nature of these different passives was in question, but Estival's result is in line with the modern consensus that these two types of passives are structurally distinct (see Embick (2004) for discussion).

An early experimental demonstration that priming reflects structural identity beyond simple surface repetition comes from Bock and Loebell (1990). They show through the use of a picture description task that phrases headed by *to* facilitate the choice of the prepositional object word order in describing double-object pictures. This observation holds regardless of whether the goal of the prime PP is a beneficiary ("The wealthy widow gave her Mercedes to the church") or a location ("The wealthy widow drove her Mercedes to the church"). The same priming is not observed, on the other hand, when the prime is an infinitival *to* phrase ("Susan brought a book to study").

In a more recent paper, Ferreira (2003) investigates the sensitivity of optional sentence complementizer *that* to priming by other functions of *that*. Using a sentence recall task, he shows that when the prime *that* is a sentence complementizer like the variable target is ("The company insured that the farm was covered for two million dollars"), it boosts the rate at which the subject choose to use *that*. But when the prime is a determiner ("The company insured that farm for two million dollars") or the head of a noun complement ("The theory that penguins built the igloos was completely false"), it does not produce this priming effect.

Finally, in Tamminga and Ecay (2014) we demonstrate the relevance of persistence effects to the study of a failed change in Early Modern English: the use of periphrastic *do* in affirmative declarative clauses. The historical syntactic change within which the affirmative declarative *do* case is situated is the emergence in Early Modern English of *do*support, the use of *do* as an auxiliary rather than a main verb. In Middle English, all verbs raised from their original position inside the verb phrase to a higher tense position, T. The modern grammar of English does not have verb-raising for finite main verbs, instead allowing two possibilities: tense-lowering onto the verb when the verb position and the tense position are in a sufficiently local relation, or the insertion of the dummy auxiliary *do* when tense-lowering is blocked by negation or when the tense head moves to above the subject as for a yes–no question. In the early 16th century, near the beginning of the change towards a language containing *do*-support instead of verb raising, auxiliary *do* also began to be used with affirmative declarative sentences. Although this use increased during the 16th century, it reversed direction after about 1575 and did not come to be part of the modern grammar of English (note that this use is not the same as emphatic or contrastive *do*). Ecay (2012) suggests that affirmative declarative *do* can be attributed to an intermediate grammatical option where *do* is merged in little *v* as an external argument marker. Under his analysis, some portion of the modern *do*-support environments, at least early in the change, are also attributable to this grammar.

In Tamminga and Ecay (2014), we suggest that Ecay's account carries with it several predictions for the persistence behavior of *do*. If it is true that at the beginning of the change, *do* represents the same functional head in both affirmative declaratives and the modern *do* contexts, then there should be persistence of *do* between these types of clauses. The persistence should not hold up, however, when the proposed reanalysis severs the underlying structural unity of these surface forms. Using data from the Penn Corpora of Historical English, we show that in the first time period, as predicted, there is a robust persistence effect where the use of a modern-type *do* boosts the probability of the subsequent use of *do* in an affirmative declarative. This effect drops off precipitously, so that in the later two time periods there is a much smaller effect of persistence. The fact that the persistence effect does not disappear entirely reflects the impossibility of the intermediate grammar being lost overnight; indeed, the continued presence of *do* in affirmative declaratives, albeit at very low rates, is evidence that the intermediate grammar is still in existence throughout the entire period covered and should continue to occasionally be the source of *do* in modern-type environments as well. In most cases where surface *do* is observed in the later periods, however, it is attributable to the modern grammar and hence does not affect the probability of using the intermediate grammar. This is another case, then, where the persistence effect appears to target quite abstract morphosyntactic units in a way that reflects their underlying structural unity.

These results highlight the sensitivity of priming, at least in a laboratory context, to the grammatical representation underlying surface strings. Priming effects arise only when the grammatical representation is shared, and fail to arise given surface similarity arising from distinct grammatical representations. In conjunction with the result from Estival showing that persistence in the use of passive structures is similarly sensitive to grammatical structure in natural conversational speech, I interpret these results as a basis upon which it is reasonable to move to the level of morphophonological variation and expect grammatical structure to play a role as well. I thus turn next to the investigation of persistence in two classic morphophonological variables, ING and TD. The data for each variable come from the 122 interviews from the Philadelphia Neighborhood Corpus, as discussed in chapter 2.

3.2 Persistence in ING

3.2.1 The definition of the variable

Although ING is one of the most familiar sociolinguistic variables, defining it precisely and circumscribing its context of occurrence are not easy tasks. In neutral terms, ING is the alternation between unstressed /ing/ and /in/, which might be exemplified using a progressive participle like *working*: the standard engma-final pronunciation *working* alternates with the /n/-final pronunciation commonly spelled *workin'*. The grammatical conditioning of the variable has been emphasized in literature on the topic, with the basic generalization being that the /ing/ variant is more frequent with noun-like things and the /in/ variant is more frequent with verb-like things. However, Labov noted 40 years ago that "there are a number of technical questions in the definition of this variable which have not always been given the attention they deserve" (1972, 238); I contend that the same could still be said even though important progress has been made in understanding the variable's origins (Houston, 1985; Labov, 1989) and social perception (Campbell-Kibler, 2011; Labov et al., 2011). The core unsettled question is whether the variation in ING is phonological or morphological.

For ING, as for other variables, it is necessary to properly define the variable in order to study its quantitative behavior, but also necessary to study the variable's quantitative behavior in order to define it properly. This problem is brought into particularly sharp focus by the demands of persistence coding, where the potential sequential dependence of tokens means that decisions about tokens in certain contexts cannot be put off as outside the realm of the study. Finding the answer to this ultimately-empirical question, though, is indispensable for making decisions about the inclusion or exclusion of different sets of lexical items within the definition of the variable.

Early work treated the ING variable as phonological in nature, admonishing for instance against "the naive approach...neglecting unstressed *-ing* in *something* and *nothing*" (Labov, 1972, 238) and giving the definition of the variable using variable-rule notation over distinctive features. It was not long, though, before the variable was recast as a choice between two allomorphs, /ing/ and /in/. This morphological account has perhaps been favored in recent work, at least in part because it is consistent with the variable's origin in the conflation and competition of two distinct Old English suffixes, participial *-inde/ende* and the verbal noun ending *-inge/ynge* (Houston, 1985; Labov, 1989). A consequence of the morphological account, though, is that the alternation between /ing/ and /in/ in presumably-simple nouns, like *morning* and *ceiling*, as well as in the quantifiers *something* and *nothing*, must be attributed to a different source because these words do not contain a distinct morpheme to serve as the locus of variable allomorphy. This is somewhat dissatisfying in that not only does the variation across these cases appear in identical form on the surface, but also these forms seem intuitively to serve the same stylistic purpose regardless of whether they are plausibly analyzed as a distinct morpheme or not. Whether or not the latter issue is a problem depends on what approach to style is assumed.

The tension between these two types of accounts continues to appear throughout recent decades. Labov (2001a) defines the variable in terms of the unstressed syllable, suggesting a phonological view, while Labov et al. (2011) explicitly take the position that the variable is morphological. Labov (2001c) encapsulates the field's uncertainty. He first states that "the overall envelope of variation includes words with the suffix *-ing*, monomorphemic nouns like *morning* and *ceiling*, and compounds with *-thing*: *nothing, something*. It therefore appears that it must be defined as a phonological alternation that affects all unstressed /ing/ syllables" (Labov, 2001c, 79). Later in the same chapter, however, he asserts that "*something* and *nothing* are excluded from the nominal sub-group because they exhibit much higher levels of apical realization, and appear to form a different sociolinguistic variable" (Labov, 2001c, 79). Since it has long been recognized that variation can occur at different rates in grammatical contexts without needing to treat the various contexts as separate variables, I would suggest that evidence independent of variant rates is of importance for making these kinds of definitional decisions. However, some of the more recent ING studies, perhaps out of frustration with this seeming unsolvable problem, sidestep the difficult question of what is "in" and "out" of the variable context.

I offer these observations to highlight the lack of consensus that remains after a half-century of intensive quantitative inquiry into the nature of this variable. In section 3.2.3 I will argue that the persistence behavior of different types of ING tokens can guide us towards a more empirically-grounded definition of the variable, although questions will remain. First, though, I will revisit the definition of the grammatical categories themselves. While the main point of grammatical interest has been differences in ING rates across the categories, and the main point of contention has been whether all of these categories need to be distinguished or even belong inside the envelope of variation, a preliminary step is the definition of those categories. It is not always clear that different studies have taken the same approach to classifying tokens from the messy world of real speech into any particular set of ING categories. In the following section I propose to reorganize the basic ING grammatical categories and pair them with usable classification diagnostics as well as current morphological analyses.

3.2.2 A new approach to ING grammatical coding

In this section I introduce a reworking of the grammatical structure of the ING variable coding to capture a modern morphosyntactic analysis of the -ing suffix at a usable level of granularity. One goal of the coding scheme I adopt is to avoid the hyper-fractionation of accounts such as Houston's (1985), which break the variable down into so many categories (Houston, for example, lists 32 different categories, although they can be grouped into 8 larger categories) that it is difficult to make a reasonable generalization about any

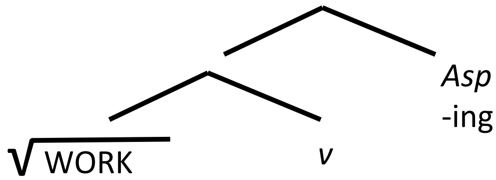

Figure 3.1: Progressive ING structure

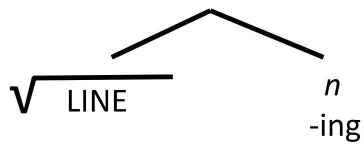

Figure 3.2: Root-attached ING structure

one category. Another motivation for revising the grammatical coding task is that previous accounts apply an ever-shifting range of terminology and typically give no useful diagnostics for which sentences might belong in which category. From the point of view of an analyst seeking to make use of such a coding scheme, the example sentences inevitably make the categories seem clear in a way that is lost as soon as real-world examples are compared. Adding yet another set of terms into the mix may seem counterproductive, but I contend that the new scheme is justified by its inclusion of simple (though not foolproof) diagnostics. Although the coding of grammatical categories in ING remains at times more art than science, the approach that I outline here should allow the categorization of most real sentence instances in a replicable way.

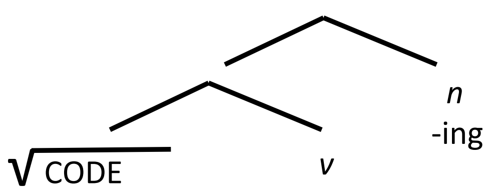

Figure 3.3: Gerund ING structure

$$\sqrt{\text{CEILING}}$$

Figure 3.4: Monomorpheme ING structure

The structural analyses that I give for the categories of the new coding scheme are based loosely on the morphosyntactic framework laid out in Embick (2008). I distinguish five grammatical categories of ING: monomorpheme, root-attached, gerund, progressive, quantifier. The first one maps more or less directly onto the category that has typically been called "nouns" or "monomorphemes". What I mean by monomorpheme is really a root, in the Distributed Morphology (Embick, 2008) sense. This category thus carries with it the implicit claim that the items in the category are stored in the lexicon as whole objects and are inserted as roots into the syntactic structure, as in 3.3. I could have called this category the roots, but I wished to avoid confusion between the terms "root" and "root-attached". Examples of words that I classify as roots are *morning, pudding, awning, ceiling* and *during*.

There are some grey areas between monomorphemes and the next category, the root-attached items. The structure for root-attached items is given in 3.2. In this structure there is an -ing suffix that is either the nominalizing or adjectivalizing head. Nouns and adjectives that can be transparently segmented but are still essentially compositional in their meaning, such as "interesting", are placed in this category. The grey zone is in the analysis of words like "morning", where technically one might see a root "morn" but it seems implausible that speakers actually analyze "morning" as "morn" + "ing". Ultimately this kind of decision is psycholinguistic and would need to be assessed experimentally on a word-by-word basis; it is entirely possible that the decomposition may even vary from speaker to speaker. In the absence of this type of evidence, I treat words as monomorphemes if they do not contain a stem that is itself a word in common usage. Examples of words that are most often classified as root-attached are the *building* (as in an apartment building), *housing, clothing*, and *lining*.

The next category I refer to as gerunds, although again this is a cover term that may not map directly onto other uses of the term ‘gerund’. Its structure is illustrated in 3.3. In this structure the suffix -ing is, by hypothesis, in a nominalizing head attached to a verbalizing head above the root, potentially with other projections intervening between the nominalizing and verbalizing heads. What this category has in common with the progressives is the presence of a verbal head within the internal composition of the word. This is illustrated in figure 3.4. In addition to the canonical progressives, where -ing is the suffix of a verb that is the complement to “to be”, this category includes other instances that share the structure of -ing being in an Aspect head attached to a verbalizing head above the root. There are several diagnostics for distinguishing between the gerund and progressive categories. First, I should clarify that in this coding scheme gerunds do not need to have an overt agent, as in the canonical syntax textbook examples like “John’s destroying the city made me angry”. A simple diagnostic for gerunds is that they can have a noun substituted for them. For example, “She does computer coding” is comparable to “She does computer work” and therefore is classified as a gerund.

However, some things that look like they would otherwise be gerunds are classified as root-attached if they have a special, non-combinatorial meaning. For example, I classify “drinking” as root-attached in the case where it refers implicitly to the consumption of alcohol rather than simply to the consumption of liquid. Transitive gerunds take objects in the accusative case, which is a useful diagnostic for transitive verbs (substitute a pronoun to check the case) but obviously less so for intransitive verbs. Progressive items can only be non-stative, so that if a verb is stative, such as “knowing”, it must be a gerund unless it is being coerced into a particularly unusual reading. Finally, progressive verbs can be understood as being parallel to a descriptive secondary predicate in that they can have a *pro* subject, making adjectival substitution a useful diagnostic for the progressive category. For example, "They stayed home studying" is comparable to "They stayed home sick," and therefore classified as a progressive token.

I also take a different approach to the quantifiers than some previous work. *Something* and *nothing* can be realized with the typical ING variants, /ing/ and /in/, but they can also be realized in more reduced forms with glottal stops and syllabic nasals. These forms are often treated as further reductions of tokens containing /in/, but it is at least equally plausible that they are distinct vocabulary items being inserted prior to any phonological manipulation. In that case, they should bleed any variation between nasal segments or suffixes. The occurrence of the reduced forms seems to be preferred by certain individuals rather than distributed throughout the sample, but is not sufficiently frequent to support a full-fledged analysis. In the interest of being conservative, I therefore treat the reduced forms as exclusion contexts rather than instances of /in/. Note that the quantifiers *everything* and *anything* are standardly excluded from the envelope of ING variation; Labov states that they "always show velar articulation, presumably reflecting secondary stress or tertiary stress on the final syllable" (2001c, 79). I have not come across any examples to contradict the factual statement here. It is not clear that the suggestion about stress is correct, however. In a quick check of a set of verbs that share the stress pattern of *everything* and *anything*, I found that a range of lexically-specific rates indicating that the apical realization is quite possible under this stress pattern (happening: 82% /ing/; hollering: 33% /ing/; listening: 57% /ing/). The difference between robustly-variable *something, nothing* and non-variable *everything, anything*, then, is quite difficult to account for.

Figure 3.1 presents the basic token counts and rates of the /ing/ variant across the categories I have just defined. The rates are consistent with the generalization that more noun-like ING categories have higher rates of /ing/; the token counts make clear that these categories are also less well represented in the data. The question that I turn to now is

Category	Mean /ing/ rate	N
Quantifier	0.33	901
Progressive	0.37	4407
Gerund	0.53	1343
Monomorpheme	0.69	238
Root-attached	0.83	625

Table 3.1: Basic rates and number of observations for the grammatical categories of ING

the degree to which membership of ING tokens in these categories affects their persistence behavior.

3.2.3 Persistence across grammatical categories

This chapter is not the first investigation of persistence in ING. Abramowicz (2007) finds that the previous ING token is a significant predictor of ING, with /ing/ facilitating subsequent use of /ing/ and /in/ likewise facilitating /in/. He hypothesizes that if a speaker uses /ing/, then "the following token — whatever is its category — is more likely to have velar pronunciation, too" (Abramowicz, 2007, 31). He treats his own data according to this hypothesis, not dividing it by grammatical category, but he does raise the possibility that the categories may play a role in determining the persistence effect. Here I take up this question in more detail, and with a greater volume of data at my disposal. The questions to be addressed are:

1. Does ING show persistence?
2. If so, is the persistence of ING sensitive to the grammatical category of the trigger and target?
3. If so, is the grammatically sensitive persistence effect independent of word repetition?

To answer these questions, I conduct both univariate and multivariate analyses as discussed in chapter 2. In the multivariate models I include standard linguistic predictors as controls, which I will not discuss at length here. The control predictors I include for ING, when modeling the dependent variable of target variant (as opposed to variant repetition from prime to target), are the preceding phonological context (coronal obstruent, coronal nasal, velar nasal, other), the following phonological context (pause, not pause), log of speech rate (vowels per second within the 7-word window presented by the handCoder Praat script), speaker age at the time of interview, and speaker sex. Log speech rate and age are both centered. I use the base-2 log because it is very easy to convert back into seconds, making it more intuitive to understand the scale of the lag term (in the univariate perspective at least). The results of these control predictors are generally in line with the expectations established by decades of research.

Figure 3.5 shows the basic effect of persistence in ING, collapsed over all of the grammatical categories and all of the speakers — essentially a replication of Abramowicz (2007). The x-axis represents the number of seconds between prime and target in log space, while the y-axis represents the dependent variable of target variant realization. Red points, and the smoothing line over them, represent tokens where the prime was an /in/, while blue points and the blue smoothing line represent tokens where the prime was an /ing/. Towards the left hand side of the graph, where the prime and target are in close proximity (less than a second apart, in some cases), we see that there is a very large difference between the mean rate of /ing/ after /ing/ primes, between 75 and 80 %, and the mean rate of /ing/ after /in/ primes, which veers close to 0%. The fact that the lines appear perfectly straight is not stipulated (that is, I did not fit a linear smooth; the loess smooth is free to wiggle); rather it indicates that the decay function appears to be exponential (since it is linear in log space).

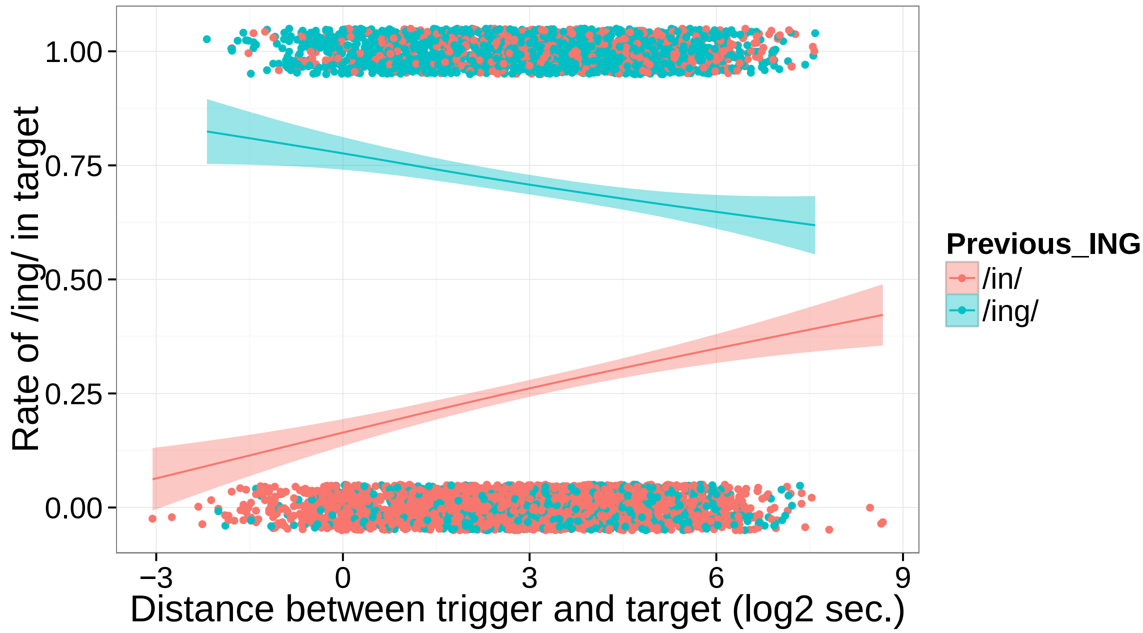

Figure 3.5: Univariate effect of persistence in ING over all speakers and grammatical categories, with Loess smoothing splines

Having seen that, as expected, ING is robustly persistent, I turn to the more interesting question of how the persistence effect interacts with grammatical category. Does it matter whether the prime and target are of the same grammatical category or not? And if persistence is affected by the prime and target being of different grammatical categories, are there differences depending on *which* categories the prime and target belong to? Understanding the persistence behavior of ING across the full 5x5 contingency table of prime–target combination possibilities would not be a simple task. Furthermore, the data in some of these cells, once the interruptions of interlocutors and other category instances are dealt with, is sparse. I therefore conduct a series of more restricted analyses that are more quantitatively tractable.

First, I look at the case where the prime and target are of matched grammatical category. Because the amount of data available is smaller here than when I considered the full dataset,

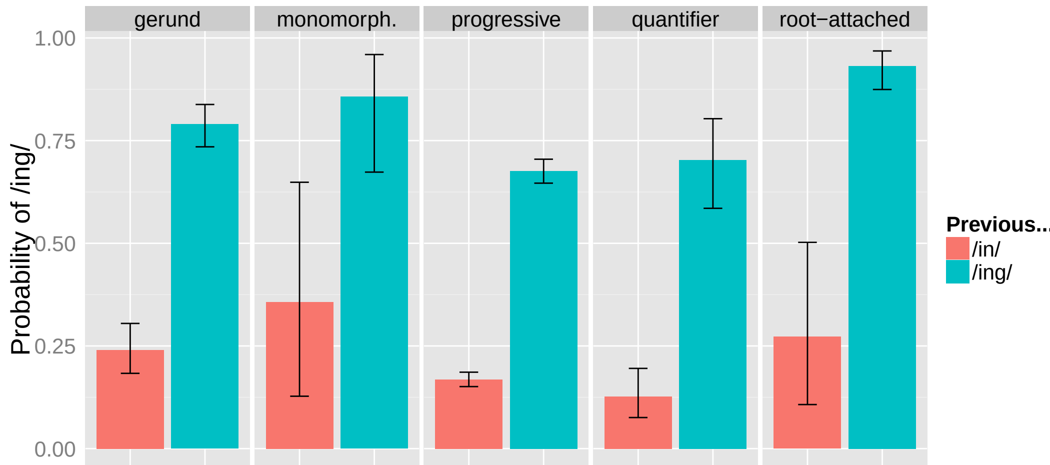

Figure 3.6: Persistence in ING with prime–target pairs matched for grammatical category, error bars are 95% Clopper-Pearson CIs

I abstract away from the decay effect by simply pooling all the data in each subset regardless of distance between prime and target. Figure 3.6 shows the relationship between /ing/-primed and /in/-primed /ing/ rates across the five matched categories. In each case, the only pairs that are included are ones that have no intervening instance of any other kind of ING; in other words, these are cases where a speaker produced, say, a progressive token of ING and then shortly afterward produced another progressive token of ING without having produced any other forms of ING in the interim. What we see is that the same robust persistence effect as we saw across all tokens is once again present when the prime and target are of the same grammatical category. Note that the large error bars for the monomorpheme and root-attached categories reflect the relative paucity of data in these categories.

Next, I consider the role of progressives, by far the most numerous category, in producing repetition in further progressive instances as well as other types of ING. We can see the descriptive univariate results of prime variant by grammatical category of the target in figure 3.7. While progressive, gerund, and quantifier ING targets appear to be strongly affected by the variant used in the progressive prime, root and root-attached targets are only weakly affected, if at all. These results are in line with the results when I fit a generalized linear model in R (using lme4) with a random intercept by speaker, a random slope for the prime variant, target variant (/ing/ = 1) as dependent variable, and the control predictors described above. The structure with the random slopes for prime variant by speaker is as recommended in Barr et al. (2013); it absorbs a certain amount of random variability not only across speakers, but also in the relationship between /ing/-persistence and /in/-persistence across speakers. All of the ING models that I will present here include the control predictors of preceding segment (coronal obstruent, velar nasal, other), following segment (pause or other), speaker age, speaker sex, and the log of speech rate. Because the control predictors have many levels and are not relevant to the discussion here, I elide them from the presentation of model results throughout, but full model results can be found in the appendix. Additionally, I fit the full set of interactions with the lag, or time between prime and target, included, but will not present those parameters until chapter 4, at which point it will be of interest to present them. Again, these interaction terms can be seen in the appendix as well. The estimated parameters for the predictors of interest are given in table 3.2

The reference level for the previous token in the model fit presented in table 3.2 is /in/, and the reference level for the grammatical category is progressive. The control of the prime type, rather than the target type, allows us to see the familiar effect of grammatical class on the dependent variable. Each of the other grammatical categories except the quantifiers has significantly more /ing/ than the progressives, with the order from least to most /ing/ being quantifiers, progressives, gerunds, monomorphemes, and root-attached. The significant positive effect of a previous /ing/ indicates that in the default case, when both the trigger and target are progressive, there is a strong persistence effect in that speakers tend to repeat whichever variant they just used.

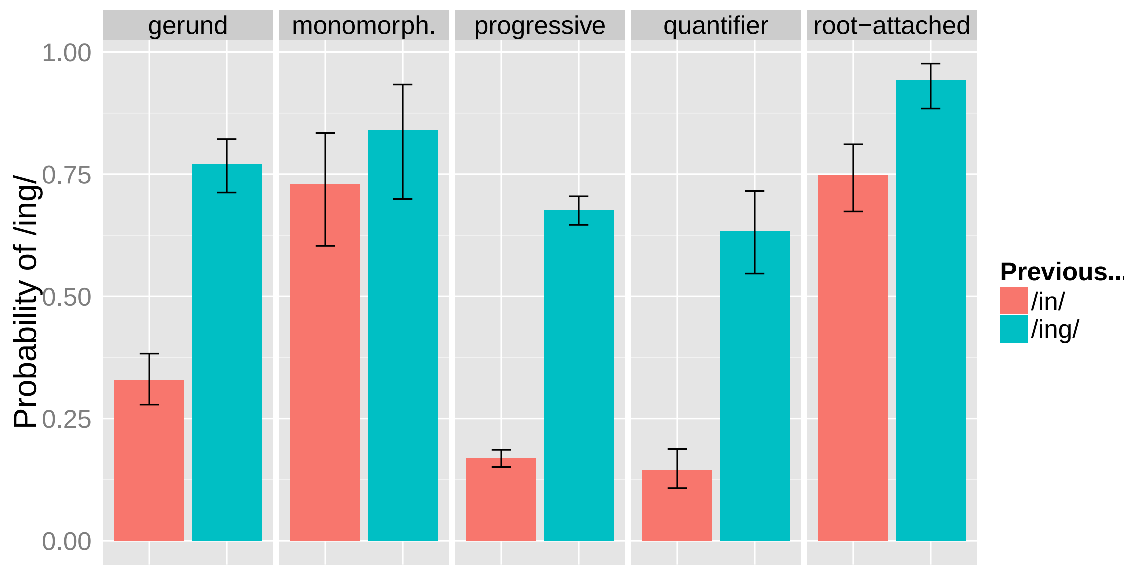

Figure 3.7: Persistence in ING across target grammatical categories with progressive primes, error bars are 95% Clopper-Pearson CIs

	Estimate	Std. Error	z value	p-value
/ing/ prime	1.11	0.13	8.44	< 0.001
Gerund target	0.89	0.15	5.78	< 0.001
Monomorph. target	2.37	0.35	6.67	< 0.001
Root-attached target	3.07	0.23	13.11	< 0.001
Quantifier target	-0.72	0.20	-3.58	< 0.001
/ing/ prime x gerund target	-0.38	0.25	-1.49	0.137
/ing/ prime x monomorph. target	-2.11	0.60	-3.50	< 0.001
/ing/ prime x root-attached target	-1.12	0.51	-2.21	0.027
/ing/ prime x quantifier target	0.06	0.31	0.18	0.856

Table 3.2: Estimated coefficients for ING parameters relevant to persistence, with progressive primes across target grammatical categories (N = 4238)

The model fit crucially includes an interaction term for the previous variant by the target's grammatical class. This is the term which allows us to ask whether the effect of a progressive trigger differs by the grammatical category of the target. The interaction terms of an /ing/ prime with a gerund target and a quantifier target do not differ significantly from the individual influences of the prime/target predictors, indicating that the significant effect of a progressive /ing/ prime is as robust on gerund and quantifier targets as it is on progressive targets. The coefficients for the monomorpheme interaction term and the root-attached interaction term are both negative, so their contribution is to reduce the overall persistence effect in those categories. Although the significance of the interaction term indicates only that there is a significant reduction in the amount of persistence, not that it disappears, we can see from the values of the coefficients that they are large enough to nearly wipe out the apparent persistence effect.[1] In other words, it seems that there is an asymmetry whereby the persistence effect from progressive triggers to gerund and quantifier targets is equally as strong as to progressive targets, but the persistence effect is not in evidence between progressive triggers and monomorpheme and root-attached targets.

In the case where the prime is of the next most commonly-occurring grammatical category, gerunds, the total number of observations is 1301. Although the number of observations is much smaller, Figure 3.8 shows that the same pattern as figure 3.7 is present: like progressive primes, gerund primes produce persistence in gerund, progressive, and quantifier targets but not in monomorpheme or root-attached targets. Future work based on datasets an order of magnitude larger than this one might pursue a truly complex analysis of persistence across every possible combination of categories.

The evidence presented here on ING demonstrates that ING is strongly persistent, with the use of /ing/ promoting more /ing/ and the use of /in/ promoting more /in/. This basic ef-

[1]Although perhaps not entirely in the root-attached case; I suspect this is a result of the difficulty of correctly and consistently diagnosing root-attached tokens

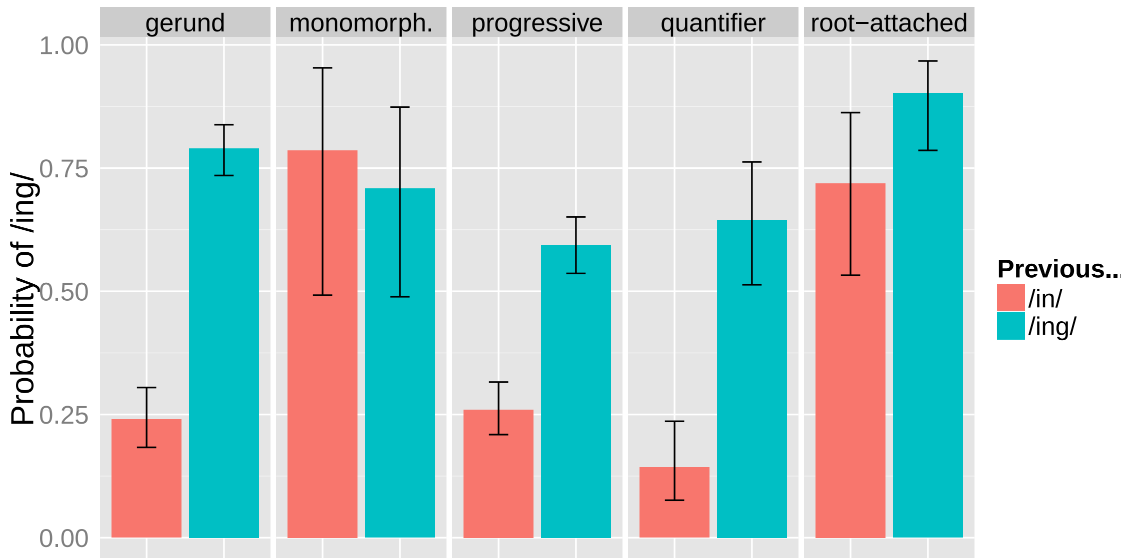

Figure 3.8: Persistence in ING across target grammatical categories with gerund primes, error bars are 95% Clopper-Pearson CIs

fect is not, however, insensitive to the grammatical category of the trigger and target tokens. It appears that, from the perspective of persistence, gerunds, progressives, and quantifiers share something that the more noun-like categories, the root-attached and monomorpheme categories of ING, do not. Although the structures that I posited in section 3.2.2 are not strictly identical for progressives and gerunds, the commonality is that in both cases the ING is a suffix attaches to a verbalizing head *v*.[2]. There are thus structural grounds for treating both of these types of ING as the same in that each is a suffix in a local relationship with *v*. I will henceforth refer to these as the "verbal" categories, and the monomorphemes and root-attached categories as "nominal". I contend that the persistence effects here are therefore consistent with the emerging consensus that verbal suffix ING is distinct from nominal ING. In other words, I argue that this evidence comes down on the side of a mor-

[2]Although gerunds have been observed to pattern closely with monomorphemic nouns in previous work (Labov, 2001c), many instances of what would be called gerunds in older classification schemes would be in the root-attached category in my scheme

phological analysis of the majority of the variation attested in ING; such an account requires variability between /ing/ and /in/ in non-verbal contexts to be treated under the aegis of a distinct variable. The phonological approach to ING, though, would not be able to account for the sensitivity of persistence to the grammatical categories, as it should be blind to the morphosyntactic context of the previous token. A remaining puzzle is the behavior of the quantifiers. It appears that they are facilitated by, and can facilitate, the use of the verbal ING suffix. This is a question that would be well-suited to experimental investigation, but which will remain an unanswered mystery in this dissertation.

Although gerunds and progressives prime each other robustly, close inspection of figures 3.7 and 3.8 the effect does appear to be slightly smaller when gerunds prime progressives and vice versa than when there is within-category priming. I also have not yet shown that the persistence effect is independent of lexical repetition. I combine the verbal categories to produce a large enough dataset upon which to base an analysis of grammatical and lexical repetition. Most of the verbal ING words cannot appear as nominal INGs, and vice versa, but there is substantial overlap in surface lexical form across the two verbal categories. Descriptively, figure 3.9 shows that the effect is indeed robust whether the lexical item is repeated from target to prime or not. As with grammatical repetition, though, it does appear to be somewhat weaker when the prime and target do not share a lexical item. The most familiar example of this type of similarity boost, the lexical boost, is the repetition of the verb in verb alternations in experimental structural priming paradigms. As a general example, "the boy gave the girl an apple" would promote re-use of the double object construction to a greater degree if a subject was asked to describe a picture involving *giving* than if they were asked to describe a picture involving *baking*, although both are expected to show priming. I suggest that the effect of verb identity in repetition strengthening the effect of ING persistence is directly analogous to this phenomenon. In ING, we see that there is a general repetition effect regardless of whether the verb stem involved is repeated

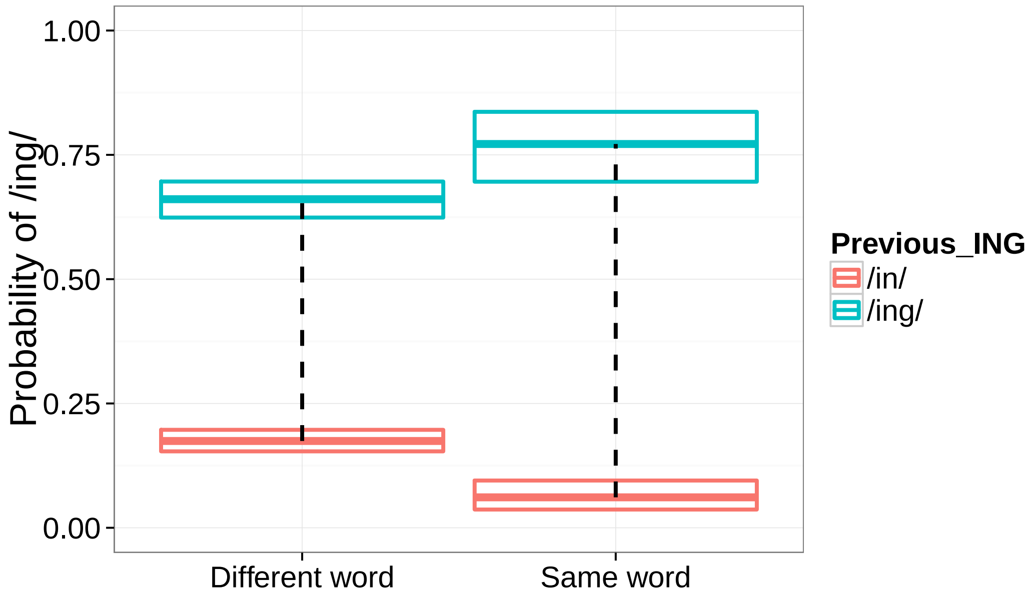

Figure 3.9: Persistence in ING progressives with prime–target pairs matched or not matched for lexical item, error bars are 95% Clopper-Pearson CIs

or not, but the effect is significantly strengthened by the repetition of the same verb.

To assess the apparent strengthening of persistence when prime and target share a word or grammatical category, I turn the perspective of the models around slightly. Rather than taking the variant used in the target as the dependent variable, I take as the dependent variable whether or not the variant used in the prime is then repeated in the target. This approach allows for the investigation of which factors favor repetition generally without requiring a complex set of difficult-to-interpret interaction terms. The parameters of a model including both grammatical and lexical repetition predictors are given in table 3.3. Bear in mind that the underlying dataset is the combination of progressive and gerund tokens only. As before, the results come from fitting a generalized linear model with a random intercept for speaker and a random slope for prime variant; the control predictors for speaker demographic and phonological factors are not expected to systematically affect a tendency towards repetitiveness and are not included.

	Estimate	Std. Error	z value	p-value
/in/ prime	-1.59	0.19	-8.326	< 0.001
Same word	0.72	0.33	2.154	0.031

Table 3.3: Estimated coefficients for ING parameters relevant to persistence, with progressive primes across target grammatical categories (N = 4238)

First, we see that the effect of the previous token is still robustly significant even when the lexical items of the trigger and target are different, indicating that what is subject to persistence is a smaller or more abstract unit than the word. There is also a considerable boost for the strength of the persistence effect under the condition of lexical repetition, which dovetails with the well-documented 'lexical boost' effect from the experimental structural priming literature (Hartsuiker et al., 2008; Snider, 2008; Jaeger and Snider, 2013). Similarly, there is a significant effect of grammatical category repetition: although there is persistence from gerunds to progressives and vice versa, the persistence is stronger when it is from gerunds to gerunds or progressives to progressives. The lexical repetition effect is almost twice as large as the grammatical repetition effect, though. Finally, this analysis indicates that the decay over time we saw in figure 3.5 is real and significant, with the negative coefficient for lag indicating a weakening of the persistence effect as the prime and target get further apart.

3.2.4 Is the persistence attributable to phonological priming?

Thus far I have been treating the persistence in ING as morphological in nature, and linking it as a result to a morphological account of the variation in ING in general. However, a possibility that has not yet been ruled out is that the persistence should be characterized as the repetition of phonological objects — whether gestures, phonemes, or syllables — rather than the repetition of abstract allomorphs. Now, I would argue that the morphologically-sensitive persistence results in the preceding subsection themselves constitute evidence against this possibility. In this section, though, I marshal an independent line of evidence against a phonological explanation for the persistence effects just demonstrated.

In many cases of morpho-phonological variability, it would be difficult to pull out a phonological influence because the relevant phonological components — for example, coronal stops in TD, as discussed below — occur so frequently and in so many different contexts and lexical items that they surely saturate whatever expectation or excitation can be assigned to them. The case of ING, on the other hand, has the advantage of implicating a relatively rare phoneme: engma. This allows for the investigation of whether there is an effect of a recent engma on the choice between /ing/ and /in/. More explicitly, the goal here is to distinguish between two possibilities. I have just shown that having recently produced a verbal instance of ING with /ing/ boosts the probability of a speaker choosing /ing/ over /in/ for the next ING token. One possibility, the one I have been implying thus far, is that the speaker is biased towards /ing/ in the target because of the recent choice of /ing/ over /in/, but the other possibility is that the speaker is biased towards /ing/ because they have recently articulated an engma.

To address this possibility, I extracted from the dataset at hand all instances of engma that occurred within the 15 seconds preceding a token of the ING variable. I excluded the cases where this procedure captured instances of the ING variable as the preceding engma. Of the remaining instances, 509 were instances of engma after a high lax front vowel and 392 were engma after a different vowel. The quality of the vowel preceding the engma might be expected to matter because phonological priming has been suggested to target rimes, either in addition to or instead of distinct phonemes. The cases where engma occurs after the high lax front vowel share the same segmental rime content as the tokens of the ING variable. I then simply calculated the rates of /ing/ in ING tokens after these two engma-preceded contexts and compared them to the rate of /ing/ in ING tokens that

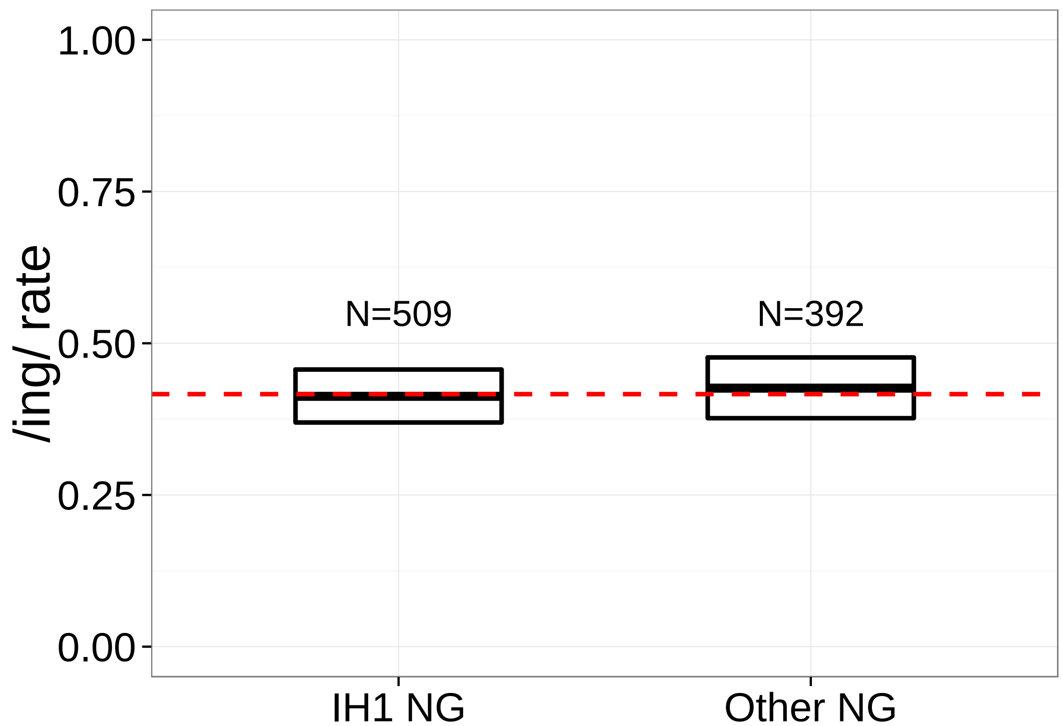

Figure 3.10: Rates of /ing/ in ING variable tokens after non-ING engmas, error bars are 95% Clopper-Pearson binomial CIs, dashed red line is ING rate without any preceding engma

had no engma of any sort within the preceding 15 seconds. These rates are illustrated in figure 3.10, with the rates after instances of engma in two contexts being presented in black with 95% confidence intervals and the red dashed horizontal line representing the rate of /ing/ when there is no preceding engma. It is quite evident from this graph that there is no case to be made for phonological priming as the underlying mechanism for the ING-persisence effects discussed in the previous section; although it is not possible with such a simplistic analysis to prove that there is *no* phonological priming, especially since phonological priming if it does exist may be a relatively small and short-lived effect, it seems safe to conclude that it does not appear in any robust way in this kind of naturalistic data and thus would be insufficient to account for the ING persistence results.

3.3 Persistence in TD

I now turn to the study of persistence in TD. TD is the variation between retention and deletion of final apical stops /t, d/ in consonant clusters (Labov et al. 1968; Guy 1980, *inter alia*). I will refer to the process as /t,d/-deletion or the variable as a whole as TD. In some words the deletion targets a phoneme within a morpheme, as in *old* or *west*, while in others it targets a phoneme that also constitutes a morpheme, as in *grazed* or *kicked*. The TD literature has accordingly distinguished between deletion in monomorphemes and regular past tense verbs, with the former showing reliably higher rates of deletion than the latter. A third class of words, the 'semiweak' or 'derivational' past tense verbs, are those that have both a final inflectional /t/ or /d/ and a vowel change in the stem, as in *kept* or *lost*. The semiweak verbs undergo deletion at an intermediate rate. Unlike in the ING case, and despite ongoing research into the grammatical conditioning of TD, the possible role of allomorphy in this case has been considered marginal rather than central. Patrick (1991) argues for a null past tense allomorph in Jamaican Creole, and Fruehwald (2012) suggests that a similar account might be used to make sense of the intermediate status of semiweak verbs in Philadelphia English. Otherwise, however, the variation is most often attributed to a variable phonological deletion process that is in some way responsive to morphological context in its rate of application. I show in this section that TD also differs from ING in that it shows persistence only when the lexical item is repeated, and suggest that a possible reason for this difference is the locus of deletion being phonological, rather than morphological.

For TD coding, glottalization and palatalization are counted as indicating /t,d/ presence. Tokens with a following coronal stop, affricate, or fricative are labeled as neutralization contexts where the evidence for /t,d/ presence is not audible. Tokens with a preceding /n/ and a following /s/ are also treated as neutralization contexts because they categorically give

Category	Mean retention rate	N
Monomorpheme	0.41	4163
Regular past	0.70	2025

Table 3.4: Basic rates and token counts for the grammatical categories of TD under discussion

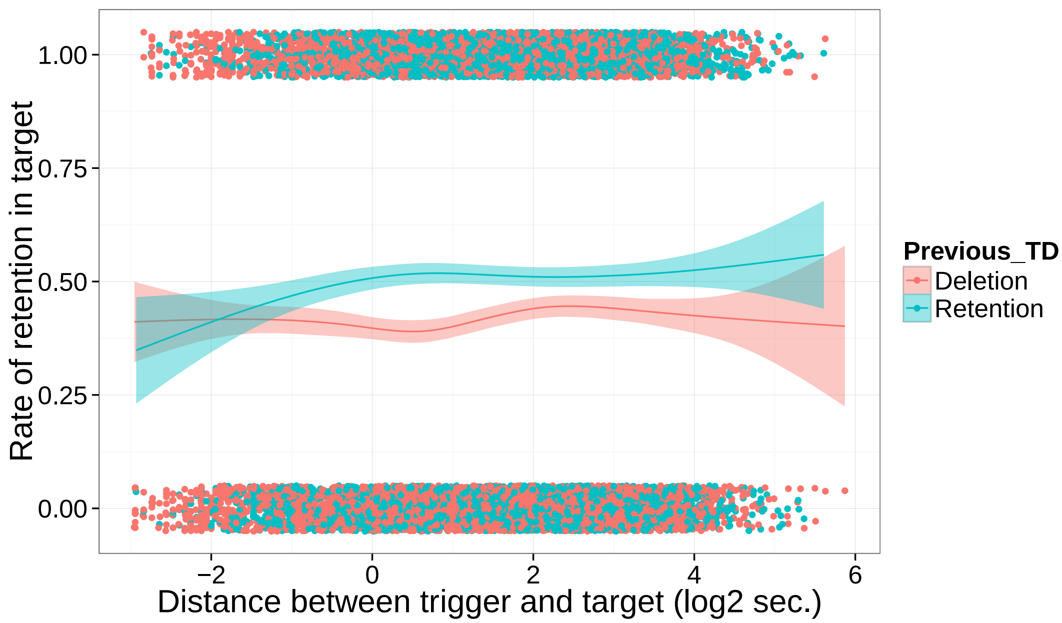

Figure 3.11: Basic effect of persistence in TD over all speakers and grammatical categories, with Loess smoothing splines (N = 6110)

rise to an excrescent /t/. The treatment of these neutralization contexts is as described in section 2.2.1, with instances of neutralization contexts interrupting potential prime–target pairs. The TD variable does not include clusters where the segment preceding the /t/ or /d/ is /r/, as these do not undergo deletion in Philadelphia. The lexical item *and* is also excluded based on previous research suggesting that it is best treated as exceptional in its underlying representation (Neu, 1980; Guy, 2007). The clusters including /r/ and the word *and* are both excluded entirely from the consideration of the variable, again for the reasons described in section 2.2.1.

Figure 3.11 shows the basic persistence effect for TD, with the data points for all speakers from all grammatical categories combined. The graph can be read in the same way as described for figure 3.5. Although there does appear to be evidence for some degree of persistence in that the blue and red splines do diverge in the expected direction, the effect is not nearly as striking as with ING. One possibility to investigate is that there may be a grammatical effect, as there was in ING. While the overall ING dataset is dominated by progressive and gerund tokens that prime themselves and each other, allowing the persistence effect to appear in the overall sample, the TD dataset is split more evenly between two categories that are both robustly represented (shown in figure 3.4) but which might not prime each other, if the verbal-versus-nominal pattern seen in ING were to reappear here. This could dilute the amount of persistence evident in the dataset.

Unfortunately the occurrence of semiweak verbs is too rare to model statistically, so the dataset that I deal with here includes only the set of prime–target pairs where the primes are monomorphemes or regular past tense and the targets are also monomorphemes or regular past tense (giving a 2x2 set of possibilities for the grammatical categories of the prime and target). As a first check on this possibility, figure 3.12 divides the data based on whether the prime and target are or are not matched for grammatical category. The graph is in line with the possibility that lack of persistence in the case of non-matched prime and target grammatical category could be diluting the overall persistence effect, and also seems congruent with the ING grammatical category result.

However, when the analysis is extended one step further to assess the role of lexical repetition in persistence, the result is different than what was seen for ING. Based on figure 3.13, it appears that the persistence in the matching cases of TD, unlike in ING, is being driven by variant repetition under the condition of lexical repetition, not by a more generalized effect. Rather than there being an additional boost to the persistence effect when the prime and target are the same word, there appears to be little if any persistence effect in

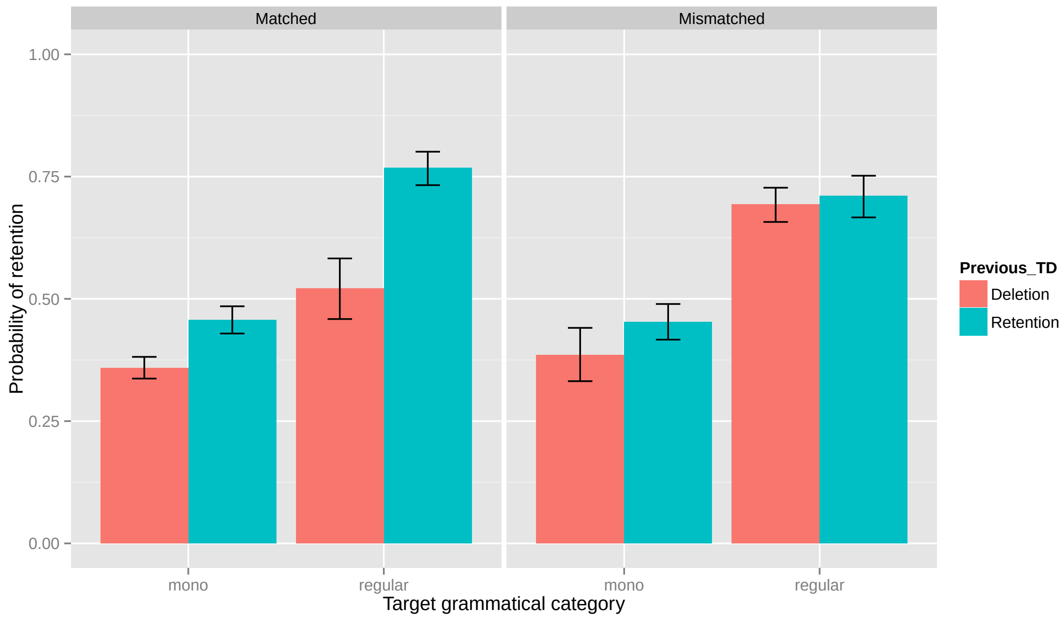

Figure 3.12: Persistence in TD across prime–target pairs matched or mismatched for grammatical category, error bars are 95% Clopper-Pearson CIs

action at all if the prime and target are *not* the same word.

To assess whether the descriptive results I have just described hold up when the other factors affecting TD are taken into consideration, I turn once again to fitting generalized linear mixed effects models in R with random intercepts by speaker and random slopes for prime variant by speaker. Treating the full set of logically possible and theoretically interesting interactions simultaneously is unwieldy and results in model non-convergence, so I fit separate models for the grammatically matched and grammatically mismatched conditions. The control predictors that I include but do not discuss further are the stop identity itself (voiced versus voiceless alveolar stop), preceding segment (stop, affricate, sibilant, non-sibilant fricative, nasal, liquid), following segment (fricative, liquid, glide, other consonant, vowel, pause), voicing match (whether or not the alveolar stop has the same voicing specification as the previous consonant in the cluster), and log speech rate (calculated the same as for ING). The social demographic factors of speaker sex and birth

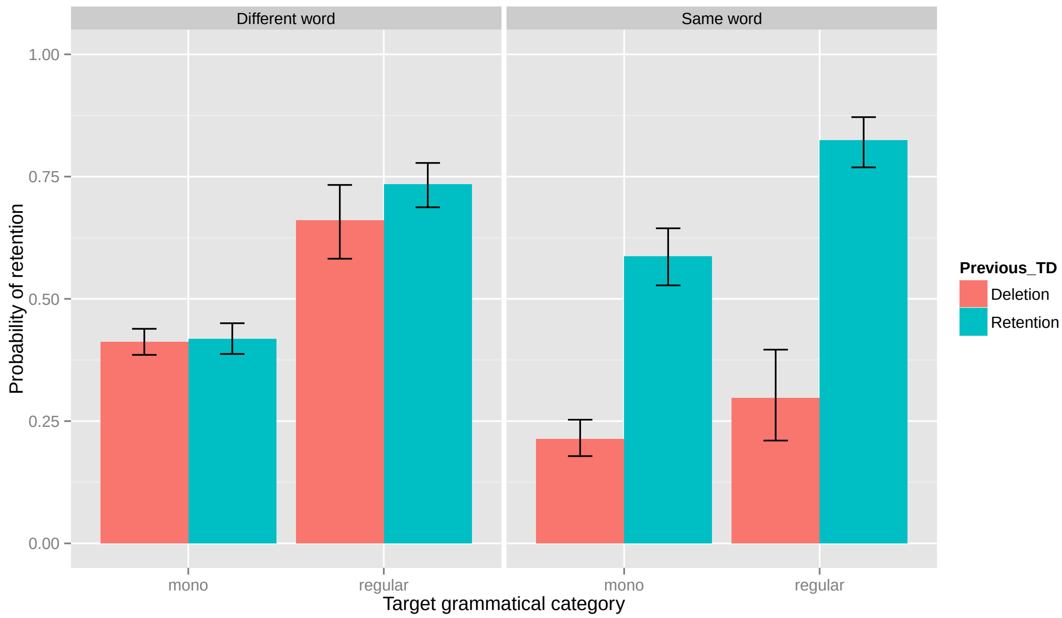

Figure 3.13: Persistence in TD across prime–target pairs matched or mismatched for lexical item, error bars are 95% Clopper-Pearson CIs

year also cause the models to fail to converge so I do not include them; in any case, social factors are expected to be of somewhat less importance for TD than ING.

Table 3.5 presents the estimated parameters of a logistic regression fit to the data from the grammatically mismatched pairs. The significant main effect of the regular past target simply indicates the well-established higher rate of stop retention in past tense compared to monomorpheme tokens. There is no evidence here for any effect of the prime variant on the target variant, either in the default monomorpheme case or in past tense targets as seen in the non-significant interaction. The multivariate results here are thus in line with the observation from above that there is not persistence from past tense primes to monomorpheme targets or vice versa.

To investigate the matched conditions, in which we do expect to find persistence at least under circumstances of lexical repetition, I split the data from the matched prime–target pairs by the grammatical category of the prime and target. Table 3.6 shows the main effects

	Estimate	Std. Error	z value	p-value
Regular past target	0.62	0.15	4.05	< 0.001
Deletion prime	-0.20	0.15	-1.36	0.174
Regular target x deletion prime	0.15	0.21	0.69	0.489

Table 3.5: Estimated coefficients for TD parameters relevant to persistence, with grammatically mismatched primes and targets

	Estimate	Std. Error	z value	p-value
Deletion prime	-0.07	0.13	-0.50	0.614
Same word	0.56	0.21	2.62	0.009
Deletion prime x same word	-1.31	0.29	-4.49	< 0.001

Table 3.6: Estimated coefficients for TD parameters relevant to persistence, with monomorphemic primes and targets

	Estimate	Std. Error	z value	p-value
Deletion prime	-0.80	0.42	-1.89	0.059
Same word	0.16	0.39	0.40	0.688
Deletion prime x same word	-1.11	0.63	-1.77	0.077

Table 3.7: Estimated coefficients for TD parameters relevant to persistence, with regular past primes and targets

and interaction of the prime variant and lexical repetition for monomorphemic pairs. The interaction of prime and lexical repetition, in combination with the lack of significant main effect for prime, means that there is persistence when two monomorphemes in a prime–target pair are the same word, but not when they are different words. This is in line with the lack of generalized persistence absent lexical repetition that was suggested by figure 3.13.

Table 3.7, though, tells a different story. When it comes to regular past tense pairs, it seems that there may indeed by an effect of the prime variant on the target even when there is not lexical repetition, with a lexical boost just as with ING. Although the p-values fall just above the standard 0.05 level for significance, I do not consider this good evidence to disregard the effect of these predictors, especially given the relatively smaller amount of data available for past tense pairs. The tendency towards this persistence effect is not entirely undetectable in the graphs, where it is partially masked by the effect of the distance between prime and target, which is the topic of the next chapter. I suggest that the quantitative evidence here points to very different behavior of past tense and monomorphemic TD. In later sections I will explore the possibility that past tense TD is a morphological variable while monomorphemic TD is subject only to phonological variation (which presumably applies in the past tense case as well).

3.4 Persistence in DH

The third and fourth morphophonological variables that I turn to are both found in the function words beginning with voiced alveolar fricatives. I will refer to them as DH-stop and DH-null, and I will treat them as two separate variables although it would be possible to unify them as a single variable. DH-stop is the use of a non-continuant segment (a stop or flap) in the place of initial eth, giving rise to the well-known but stigmatized forms written ‘dis’ and ‘dat’ for *this* and *that*. Previous research indicates that, unlike in other nearby speech communities such as New York, only word-initial and voiced alveolar fricatives are variable in Philadelphia (so that *think* does not become 'tink' and *father* does not become 'fadder' (Labov, 2001c). DH-stop has often been compared to ING in terms of its stylistic sensitivity in Philadelphia; see for example Labov (2001a). One advantage of DH-stop over ING and TD, as a variable to study, is that it occurs many times in any stretch of speech, since all of the contexts for its use are high frequency function words. Of course, this also means that coding DH-stop is much more time-consuming (from a time-per-interview perspective) than coding ING or TD; accordingly, I coded only 42 interviews for DH-stop rather than the full set of 122 used for ING and TD.

As a methodological note, I count any degree of perceptible continuancy in a DH-stop token to be indicative of the relatively standard instance of the variable, following Labov (2001c), who notes that affricate-like tokens do not arouse the same sort of stigmatization as a fully non-continuant (stop) variant does. DH-null occurs in the same set of function words as DH-stop does, but instead of the variant being a non-continuant segment it is null, so that for example *them* is pronounced ''em'. When examining the influence of lexical repetition in both DH variables, I ignore contraction so that, for example, 'they'll' is treated as an instance of 'they'. I also exclude the lexical item *the* because it is of such high frequency that it swamps the rest of the data set, and is also quite difficult to code because many instances of it are so fleeting that they are reduced to near-imperceptibility. I do, however, treat the occurrence of *the* as an exclusion context that interrupts the formation of prime–target pairs, so that the potential impact of DH variability in *the* on other DH words is mitigated. Finally, I implicitly assume the ordering where DH-null comes before DH-stop, so that the proportions for DH-null represent the rate of null tokens out of all DH tokens in the relevant categories whereas the proportions for DH-stop represent the rate of null tokens out of only the tokens that remain after DH-null tokens are excluded. This relationship is represented schematically in figure 3.14.

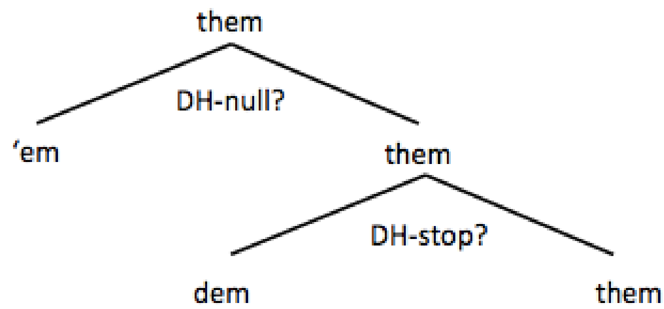

Figure 3.14: Assumed relationship between DH-null and DH-stop

The previous two variables have shown interesting patterns in the relationship between persistence and lexical repetition. Comparing this interaction in DH-null and DH-stop reveals that there is a difference between the two DH variables. Figures 3.15 and 3.16 show the raw proportions of how persistence is affected by lexical repetition. From these figures, it appears that while there is a persistence effect in both cases, for DH-stop it is a generalized persistence effect reminiscent of the one seen with verbal ING, whereas for DH-null it is an effect that appears only when the prime and target are the same word, just as we saw for monomorphemic TD.

Tables 3.8 and 3.9 present the estimated parameters from the multivariate modeling in the same vein as the previous variables. For both DH-stop and DH-null, the control predictors are speaker sex, speaker age (centered), preceding segment (vowel, consonant, pause), and speech rate (same measure as for ING and TD). These models also include the lag term and its interaction, which will be discussed in the next chapter, as well as random intercepts by speaker and random slopes by prime variant. Again, the full model results are available in the appendix. In table 3.8, the main effect of a stop shows that there is a significant persistence effect for DH-stop even when the prime and target are different words. The positive main effect of lexical repetition and the negative interaction of the stop prime and same word indicate that there is a lexical boost, for the effects of both stop

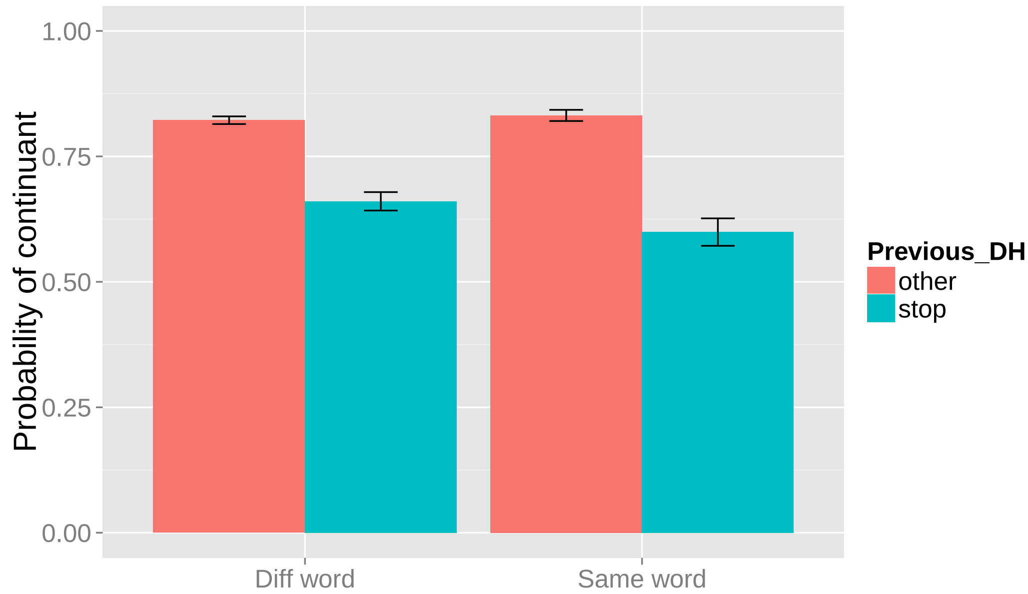

Figure 3.15: Persistence in stop/continuant variability in DH-stop, error bars are 95% Clopper-Pearson CIs

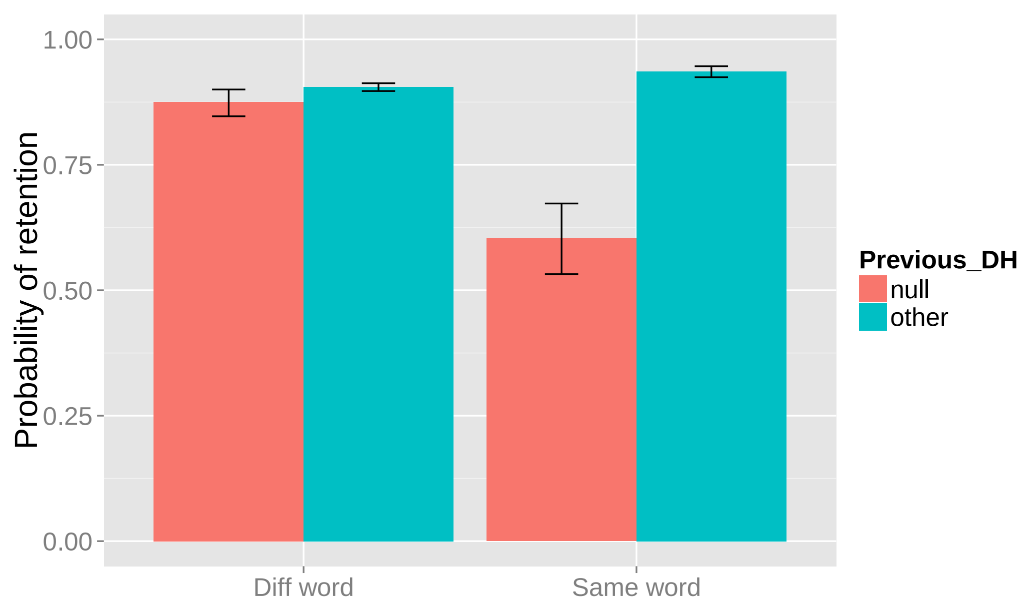

Figure 3.16: Persistence in /dh/ presence/absence in DH-null, error bars are 95% Clopper-Pearson CIs

	Estimate	Std. Error	z value	p-value
Stop prime	-0.45	0.12	-3.82	< 0.001
Same word	0.25	0.11	2.37	0.018
Stop prime x same word	-0.68	0.19	-3.61	< 0.001

Table 3.8: Estimated coefficients for DH-stop parameters relevant to persistence (N = 7462)

	Estimate	Std. Error	z value	p-value
Null prime	-0.09	0.17	-0.53	0.598
Same word	0.44	0.15	2.98	0.003
Null prime x same word	-2.08	0.29	-7.18	< 0.001

Table 3.9: Estimated coefficients for DH-null parameters relevant to persistence (N = 8409)

and continuant primes. For DH-null, in contrast, table 3.9 shows that the main effect of a null prime variant is not significant, but its interaction with lexical repetition is: just as in monomorphemic TD, the persistence effect in DH-null is restricted to cases where the prime and target are the same word.

3.5 Persistence in phonology and morphology

Sections 3.2 and 3.3 have shown that the persistence characteristics of TD are quite different from the persistence characteristics of ING. For ING, persistence is independent of lexical repetition, while for TD, persistence arises only in cases where the trigger and target are the same word. What might account for this difference? Two fairly obvious possibilities based on previous analyses of these variables are their relative degree of socio-stylistic evaluation and their loci within the grammatical system. ING is often held up as a socially meaningful, stylistically dynamic variable *par excellence*, correlating well with social class, gender, social networks, social evaluation, stylistic context, and more. TD, on the other hand, is most often discussed with respect to its linguistic conditioning, except when it has been argued to be relatively insensitive to social and stylistic context (Labov, 2001b; Hazen, 2011). It could be, then, that we find what appears to be more robust persistence with more stylistically-dynamic variables, potentially as a result of stylistic clustering rather than true sequential dependence across tokens. Making such an account fit with the patterns of lexical repetition seen in this chapter, however, would not be straightforward.

A second possibility, opened up by my argument in favor of a morphological locus of variation for verbal ING, is that the difference in persistence behavior stems from the distinction between phonological versus morphological variation. My suggestion here is that morphological variation induces generalized persistence while phonological variation gives rise to only lexically-specific persistence. In chapter 4 I will expand on this suggestion with respect to the temporal decay of the different variables, and consider how such a distinction might be explained with respect to the operation of different memory systems interacting with the structure of the lexicon and the grammatical architecture. Here, however, I will note only that the investigation of the non-verbal categories of ING will be of particular importance for this suggestion. There is both the empirical question of whether nominal and root-attached ING behave like TD in showing only lexically-specific persistence, and the more theoretical question of whether making a distinction between nominal and root-attached as categories is justified at all. Unfortunately, answering these questions more deeply will require a substantially larger dataset or perhaps experimental work; once the prime and target are constrained to be of the same nominal category and then further sliced up by lexical repetition and prime variant, the number of tokens is too low to draw even reliable descriptive statistics from.

When DH-stop and DH-null, as discussed in section 3.4, are considered in tandem with ING and TD, we can see that they line up in somewhat unexpected ways. Although on the face of it one might expect to see alignment between DH-null, being by hypothesis morphological, with ING and DH-stop, being by hypothesis phonological, with TD, in fact the opposite pattern obtains. DH-stop shows generalized persistence with a lexical boost, while DH-null is lexically restricted. The behavior of DH-null might be understood as a quite different variable with a word-level locus, which would then look more like a case where each lexical item constitutes its own variable for the purposes of persistence. In other words, if the choice a speaker makes is between 'them' and ''em' rather than between overt and null versions of the /dh/, then the use of ''em' would not have any reason to impact a subsequent choice between 'that' and ''at' because they do not share a structural relationship. This would then be consistent with the persistence pattern we see, and despite the surface similarity the lexically-specific TD persistence would have its origins in a different framework. Indeed, I suspect that nominal ING may show a similar pattern, with some lexical items having two forms, one velar-final and one apical-final, and others having only the velar-final version. This suspicion is partly driven by the observation that there are five nominal ING lexical items (*wedding, beginning, clothing, evening*, and *housing*) that occur at least five times in the dataset without occurring in the apical form. Although standard variationist practice might exclude these, I object to that practice on the grounds that it does not distinguish between lexical items that *cannot* alternate (for example, the place names *Reading, PA* and *Flushing, NY* are not supposed to be variable (Labov, 2001c)) and ones that simply happen not to because they occur rarely and are expected to have low rates of /in/ anyhow. This is a subject for future work, making the prediction that nominal ING persistence should be lexically restricted as with DH-null. The behavior of DH-stop, in the framework I have been developing here, may seem somewhat more puzzling. Its resemblance to the verbal ING and past tense TD patterns suggests an analysis in which the variation in DH-stop is actually morphological. I will revisit this possibility in the next chapter.

All of the variables that I consider here sit in some way at the interface of phonology and morphology. Note that one of the most prominent variables in the corpus persistence literature, /s/-deletion in Spanish and Brazilian Portuguese, could also be characterized, in at least some contexts, as morphological rather than phonological variation (that is, the variation could be described as a suffix alternation). It would be instructive to revisit this variable with an eye to teasing apart its persistence behavior in monomorphemes and morphologically complex words. However, while this interface position makes such variables of great interest for the complexity of their behavior and the possibility that the surface variation we see may be decomposable into multiple processes, it does tend to muddy the waters when considering the interpretation of the results.

If the distinction I am beginning to outline in this chapter is to stand, it will be of critical importance to anchor these results to comparable behavior from less ambiguously-located variables. A crucial testing ground for the hypothesis that phonological variation gives rise to only lexically-specific persistence is low-level phonological processes that are not understood to have any plausible morphological locus of variation, such as TH-fronting in British English. Conversely, it will be important to ask whether variables that target higher levels of the grammar or more abstract forms show the generalized persistence that I hypothesize characterizes morphological variation, particularly given my situation of the morphological structure of ING within a framework that treats morphology and syntax as a single system (Embick, 2008). The behavior of morphological variation under such a framework should bear a striking resemblance to at least some kinds of syntactic variation.

3.6 Chapter summary

The primary goal of this chapter has been to demonstrate a connection between the phenomenon of persistence and questions of grammatical structure, both categorical and variable. There is evidence for the mediation of observed persistence effects by grammatical structure: clear evidence for the circumscription of persistence by grammatical contexts for surface-similar instances of variation, and evidence that is at least consistent with the possibility that variation at different levels of the grammar gives rise to different types of persistence behavior. The sensitivity of persistence to linguistic structure, insofar as my proposal can be shown to hold across new cases, makes it a useful source of quantitative evidence for distinguishing morphological from phonological variation and for identifying the underlying unity or disunity of surface-similar strings. Specifically, I have suggested that persistence appears in a generalized form across instances of a variable only when the variation is morphological. In contrast, it does not appear in variation that is the outcome of a probabilistic phonological process, except when the prime and target are the same word. Persistence can thus be used as a source of evidence on two points. One is the distinction between variation at different levels of the grammar, particularly morphology and phonology; the other is the diagnosis of multiple underlying sources for variation that appears unified on the surface. Moreover, the very observation that persistence reflects underlying, rather than surface, structures gives us reason to see a connection between persistence and the psychological factors at play in the production of language. The next chapter extends the analysis of ING, TD, and DH to make explicit and strengthen the connection between the persistence patterns observed here and their cognitive bases.

Chapter 4

Persistence in cognitive perspective

In chapter 3 I argued that persistence reflects the underlying grammatical origins of observed variation and demonstrated that different variables have different quantitative persistence patterns, particularly with respect to the interaction of persistence and lexical repetition. Although those analyses did include a lag term to capture the distance between prime and target, plus the interaction of the lag term with the prime variant and lexical repetition effects, I did not present or discuss the lag-related model coefficients in that chapter. Here I take up that analysis with the same set of variables (that is, verbal ING, monomorphemic and past tense TD, DH-stop, and DH-null), aiming to characterize the basic decay patterns of the persistence effects that were demonstrated in chapter 3. This analysis is motivated by the interest in the experimental literature in using decay profiles to disentangle priming effects with different underlying mechanisms.

4.1 Decay in the priming literature

The degree to which priming effects last over time has been of substantial interest recently in the structural priming literature. There is an apparent split between results where the effect is no longer detectable with even a single intervening trial (Wheeldon and Smith, 2003; Branigan et al., 1999), and results where the effect is detectable at some distance (Bock and Griffin, 2000; Branigan et al., 2000), possibly even up to a week (Kaschak et al., 2011). To some extent these differences may be attributable to the methods used. For example, the week-long ditransitive priming effect that Kaschak et al. (2011) observe in their first experiment is after an exposure phase where the participants are exposed to 20 instances of one type of prime and 0 instances of the other (i.e., 20 double object sentences and 0 prepositional object sentences, or vice versa). This dramatic bias may induce a stronger effect than in the case where the design is to present a single prime and have the participant respond to a related target some moments later in the experiment; although the time delay is much shorter, the emphasis on the prime is also substantially weaker.

Another direction that has been pursued in an attempt to reconcile the temporal differences across studies is the development of multi-factorial analyses of structural priming. The lexical boost, where a syntactic structure makes a more effective prime if its verb is repeated in the target, has played a key role in the development of such analyses. Hartsuiker et al. (2008), for example, suggest that the seeming disparity between Branigan et al. (1999) and Branigan et al. (2000) is due to the latter study using primes and targets with lexical overlap while the former uses lexically distinct primes and targets. What Hartsuiker et al. show is that the basic effect of structural priming is relatively long-lasting (persisting at a stable magnitude through up to 6 trials), while the enhancement of the effect provided by lexical repetition is present only when the prime and target trials are adjacent.

There have been many similar efforts to explain differential decays across different features of repetition priming ("cat" primes "cat") and morphological priming ("cats" primes "cat") of words by localizing different effects to different systems. The most common proposal is that word recognition implicates a short-lived activation effect and a long-term episodic effect (see Tenpenny (1995) and Versace and Nevers (2003) for more comprehen-

	Estimate	Std. Error	z value	p-value
/in/ prime	-1.59	0.19	-8.326	< 0.001
Same word	0.72	0.33	2.154	0.031
Lag (log2 sec.)	-0.08	0.03	-2.623	0.009
/in/ prime x same word	-1.28	0.47	-2.731	0.006
/in/ prime x lag	0.19	0.04	4.952	< 0.001
Same word x lag	-0.06	0.09	-0.681	0.496
/in/ prime x same word x lag	0.08	0.14	0.575	0.565

Table 4.1: Estimated coefficients for ING parameters relevant to persistence, with verbal primes and targets (N = 4238)

sive overviews, as well as more recent work from Kouider and Dupoux (2009)). I will discuss the exact temporal findings of such results in section 4.5, in addition to suggesting that there are not two but three sources for priming effects to unify the set of results across the repetition priming and structural priming literatures and make sense of the complex set of persistence decay profiles of the morphophonological variables at stake in this dissertation.

4.2 The decay of ING persistence

Table 4.1 presents, from the same model as in table 3.3 in chapter 3, the parameters relevant to the question of how the persistence effect within verbal prime–target ING pairs unfolds over time. The default level for the prime variant is /ing/ and the default level for lexical repetition is non-repetition. Recall from chapter 3 that ING shows generalized persistence with a lexical boost: the significant main effect of an /in/ prime indicates persistence even when the prime and target are different words and the significant main effect of lexical repetition indicates that when the prime contains /ing/ and the target contains the same word as the prime, there is an extra boost to how likely the target is to contain /ing/. Newly presented here are two results regarding the lag predictor. First, the main effect of lag

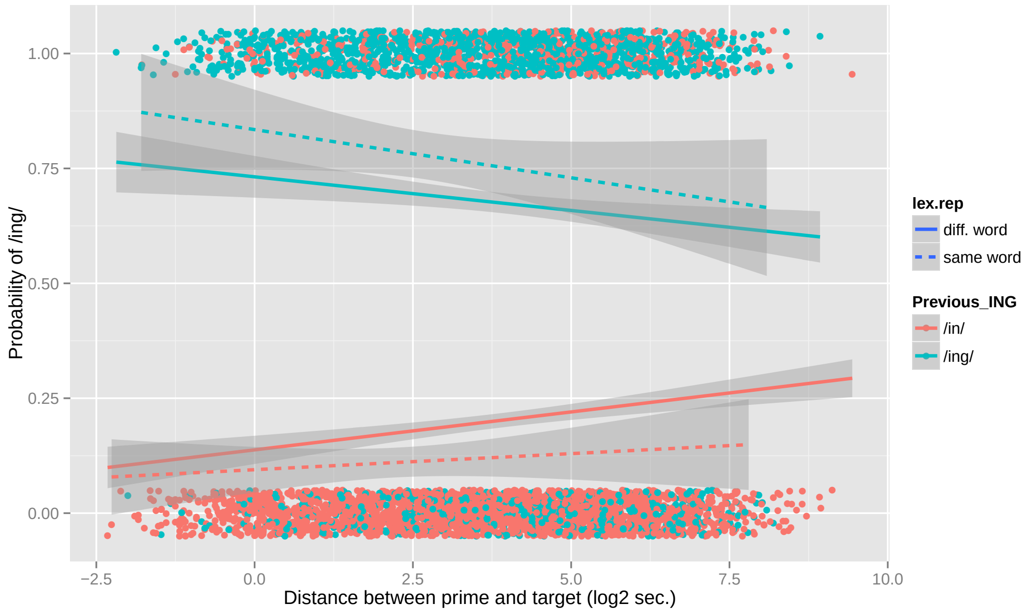

Figure 4.1: Persistence over lag time in verbal ING with lexical repetition

indicates that the likelihood of an /ing/ target after an /ing/ prime lessens over time. Second, the positive interaction between an /in/ prime and lag indicates that the persistence effect from /in/ also lessens over time. However, neither interaction containing both lag and lexical repetition is significant, indicating that the effect of lexical repetition is orthogonal to the decay of persistence over time. In other words, there is greater persistence in the lexical repetition condition than when there is not lexical repetition, but as the prime and target get further apart the strength of persistence declines in both the repetition and non-repetition conditions simultaneously. This effect can be seen, to a first approximation in its univariate form, in figure 4.1. To understand the distances under discussion, figure 4.2 shows the density of the lag times in the verbal pair data, which are characterized by a mean of 34.9 seconds, a maximum of 694.6 seconds (about 12 minutes), and a standard deviation of 51.7 seconds. Pairs with lexical repetition tend to be closer together than pairs with different prime and target words, but because they are taken into account simultaneously in the model, it is not likely that this distinction is the direct cause of the lexical boost.

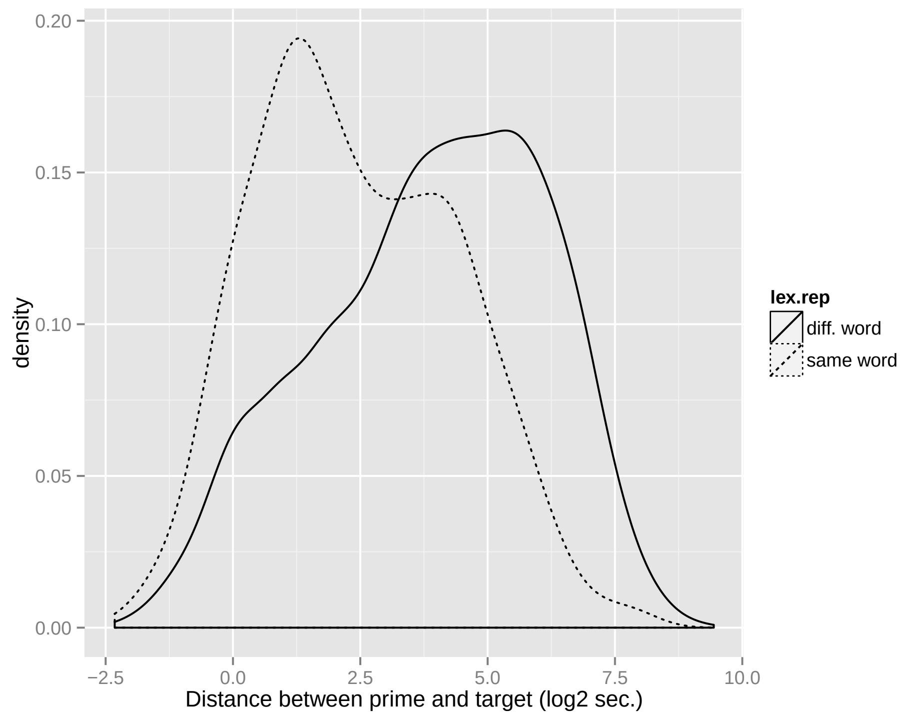

Figure 4.2: Distribution of lag times by lexical repetition in verbal ING pairs

4.3 The decay of TD persistence

In table 4.2 I apply the same structured model to monomorphemic TD (again, this section presents additional predictors from the same model as in figure 3.6 in chapter 3). Just as the basic effect of persistence in monomorphemic TD is distinct from verbal ING, with persistence appearing only under circumstances of lexical repetition, so too are its temporal properties. In fact, there is no evidence for decay over time in the lexically-specific monomorphemic TD persistence effect seen in chapter 3, despite the appearance of such an effect in the univariate perspective of figure 4.4. None of the model terms including the lag predictor are significant, and indeed the predicted coefficients and effect sizes are very small. It appears that no matter how far apart the prime and target for a monomorphemic TD pair are, at least within the bounds of the dataset available here, the prime has a constant influence on the target.

	Estimate	Std. Error	z value	p-value
Deletion prime	-0.07	0.13	-0.50	0.614
Same word	0.56	0.21	2.62	0.009
Lag (log2 sec.)	-0.00	0.05	0.00	0.997
Deletion prime x same word	-1.31	0.29	-4.49	< 0.001
Deletion prime x lag	0.05	0.06	0.86	0.389
Same word x lag	-0.04	0.12	-0.33	0.742
Deletion prime x same word x lag	0.12	0.16	0.78	0.434

Table 4.2: Estimated coefficients for TD parameters relevant to persistence, with monomorphemic primes and targets

To give a quick characterization of the distribution of lag distances imposed by conversational turn-taking in this dataset, the mean lag time is 4.7 seconds, the maximum is 45.5 seconds, and the standard deviation is 5.1 seconds. Note that this is a substantially shorter set of lag times than in the ING data, reflecting the greater frequency of TD tokens. A more nuanced understanding of the distribution of lag times in the monomorphemic TD pairs can be gleaned from figure 4.3, where it can be seen that there is a greater proportion of pairs with a lag between 1 and 2 seconds in the lexical repetition condition compared to the more broadly distributed non-repetition context. In some respects this mirrors the observation from ING, although the time spans are of course much shorter here.

Table 4.3 and figure 4.5 show that the weak generalized persistence effect on past tense TD, as discussed in chapter 3, also seems to decay over time, although only the effect of a deletion prime is susceptible to this decay. As with verbal ING, there is not evidence for any interaction of the lexical boost with the decay, so that the persistence effect decays over time in both the same word and different word conditions.

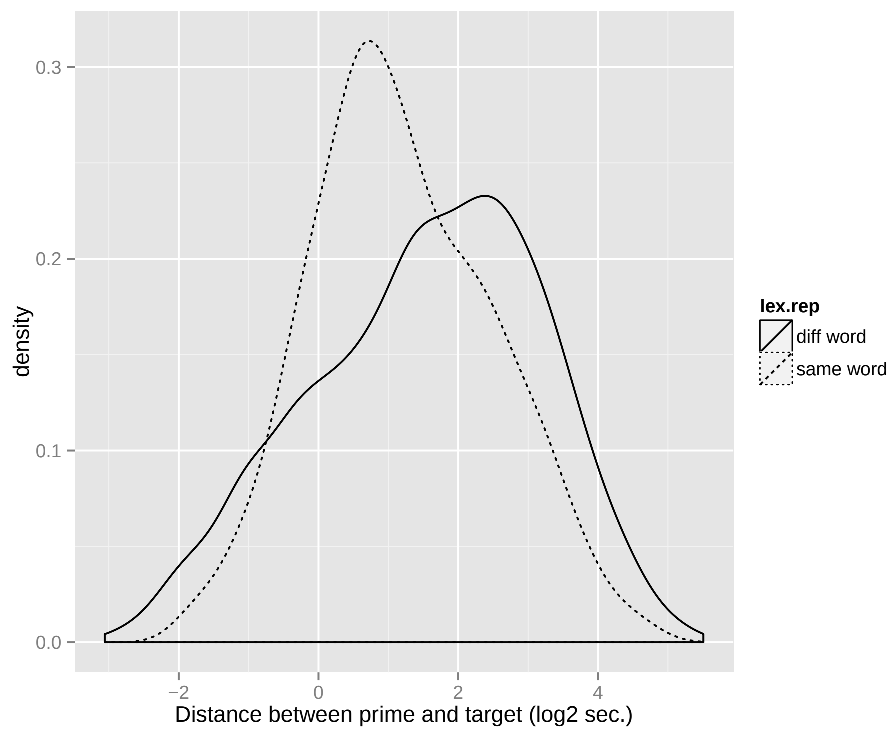

Figure 4.3: Distribution of lag times by lexical repetition in monomorphemic TD pairs

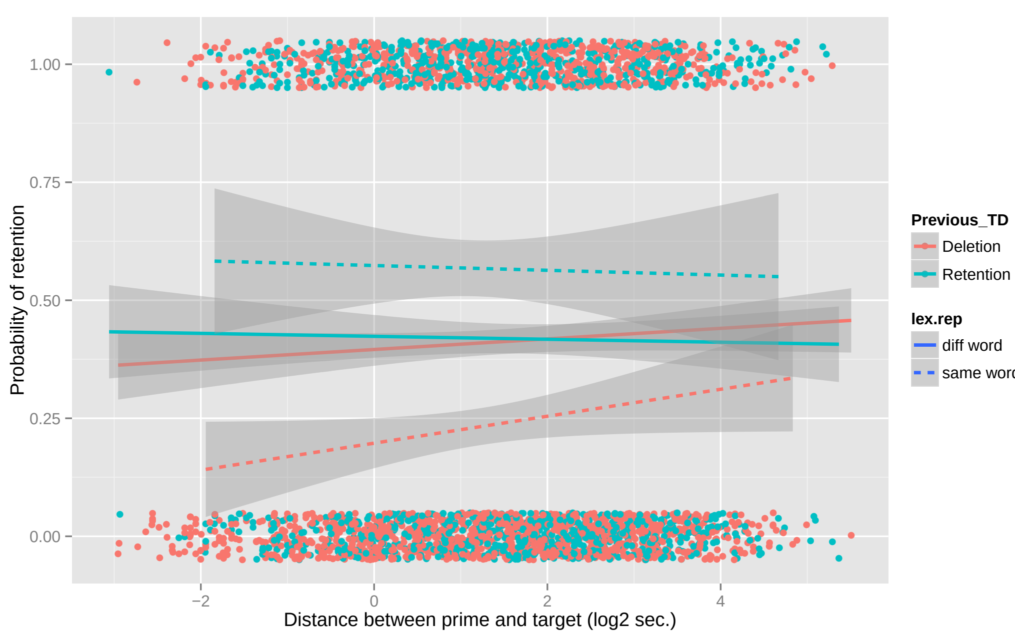

Figure 4.4: Persistence over lag time in monomorphemic TD with lexical repetition

	Estimate	Std. Error	z value	p-value
Deletion prime	-0.80	0.42	-1.89	0.059
Same word	0.16	0.39	0.40	0.688
Lag (log2 sec.)	-0.11	0.11	-1.02	0.306
Deletion prime x same word	-1.11	0.63	-1.77	0.077
Deletion prime x lag	0.40	0.20	1.96	0.050
Same word x lag	0.18	0.20	0.91	0.361
Deletion prime x same word x lag	-0.19	0.33	-0.57	0.567

Table 4.3: Estimated coefficients for TD parameters relevant to persistence, with regular past primes and targets

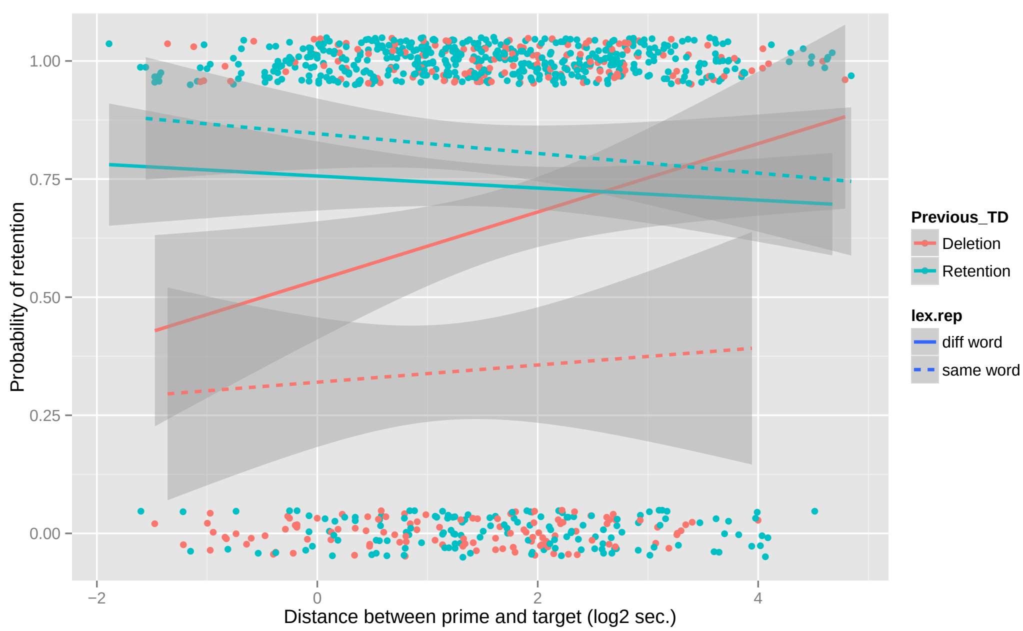

Figure 4.5: Persistence over lag time in regular past tense TD with lexical repetition

4.4 The decay of DH persistence

Finally, I turn once more to the two DH variables, DH-stop and DH-null. Recall from chapter 3 that DH-stop shows generalized persistence, whereas DH-null is persistent only with lexical repetition. Table 4.4 shows that the generalized DH-stop persistence, as with verbal ING in section 4.2 and past tense TD in section 4.3, decays over time and does not interact with lexical repetition, indicating once more that the decay takes place in both the repetition and non-repetition contexts at rates that are not significantly different. This further alignment of the patterning of DH-stop with verbal ING and past tense TD suggests that the basis of this variable may be morphological. This is somewhat unexpected, as the DH-stop variable has always, to my knowledge, been treated as phonological. It is not, however, completely without basis. Understanding DH-stop variability as morphological, in light of the generalized persistence effect it elicits, requires there to be a morpheme corresponding with the variability but not subsuming the non-shared parts of the DH words — in other words, the alternation needs to be localized to a morpheme consisting of the /dh/. That there exists such a morpheme is the position taken on syntactic and cross-linguistic grounds in, for example, Leu (2008). Furthermore, a moment of reflection on the correspondences between deictic and interrogative pronouns (*there* $\sim$ *where*, *then* $\sim$ *when*, etc.) shows that this notion is not without a relatively transparent basis in English. Although this issue is not, to my knowledge, resolved and uncontroversial in the theoretical literature, I suggest that with further refinement and a range of evidence from additional variables, quantitative data from phenomena such as the persistence patterning might make a contribution not just to the circumscription of variables but to the description of morphosyntactic structures.

As for DH-null, in chapter 3 I suggested that it might be preferable to attribute the DH-null behavior to a lexically-specific alternation as opposed to the low-level phonological

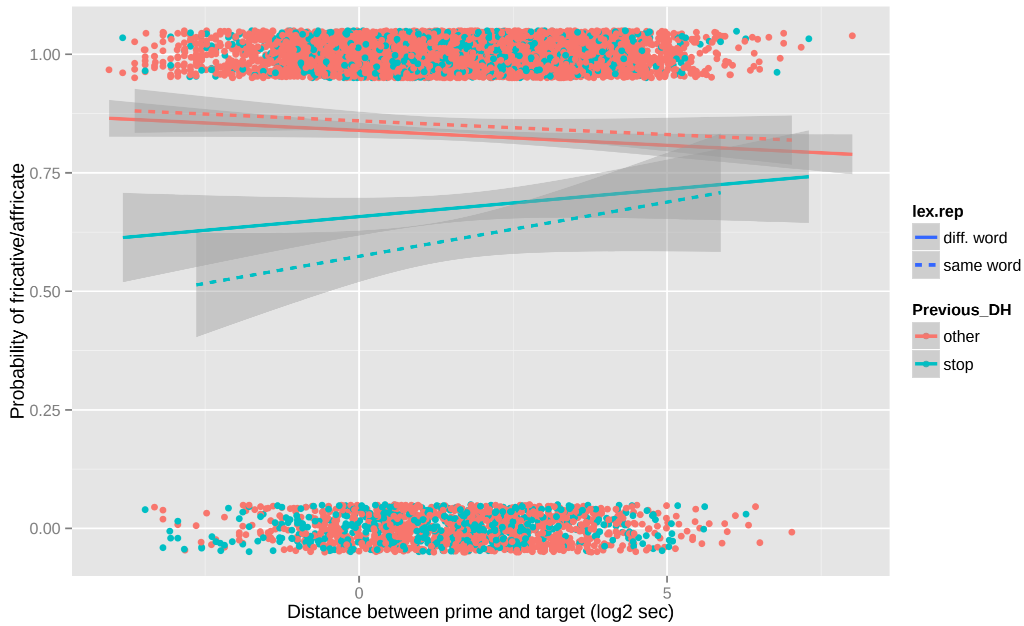

Figure 4.6: Persistence over lag time in DH-stop with lexical repetition

deletion I ascribe to monomorphemic TD. Table 4.5 provides more concrete evidence supporting this suggestion: unlike the lexically-specific persistence in monomorphemic TD that is robust and non-declining throughout the full range of lag times present in the data, the lexically-specific persistence effect in DH-null decays in the same way as the persistence in verbal ING, past tense TD, and DH-stop. Reconciling the locus of this variable with DH-stop is a project for future work.

The distribution of lag times in both of the DH variables, seen in figures 4.8 and 4.9, bear a closer resemblance to TD than ING even though the decay profiles of both DH-stop and DH-null more closely resemble the decay profile of verbal ING. For DH-stop, the mean lag is 6.2 seconds with a maximum of 257.3 and a standard deviation of 6.2, while for DH-null the mean lag is 5.7 seconds with a maximum of 257.3 and a standard deviation of 9.9. This suggests that the distribution of lag times is not the sole factor allowing us to detect, or preventing us from detecting, the decay behavior of the persistence effect in monomorphemic TD.

	Estimate	Std. Error	z value	p-value
Stop prime	-0.45	0.12	-3.82	< 0.001
Same word	0.25	0.11	2.37	0.018
Lag (log2 sec.)	-0.04	0.03	-1.42	0.155
Stop prime x same word	-0.68	0.19	-3.61	< 0.001
Stop prime x lag	0.09	0.05	1.97	0.049
Same word x lag	-0.01	0.05	-0.25	0.804
Stop prime x same word x lag	0.04	0.08	0.49	0.623

Table 4.4: Estimated coefficients for DH-stop parameters relevant to persistence (N = 7462)

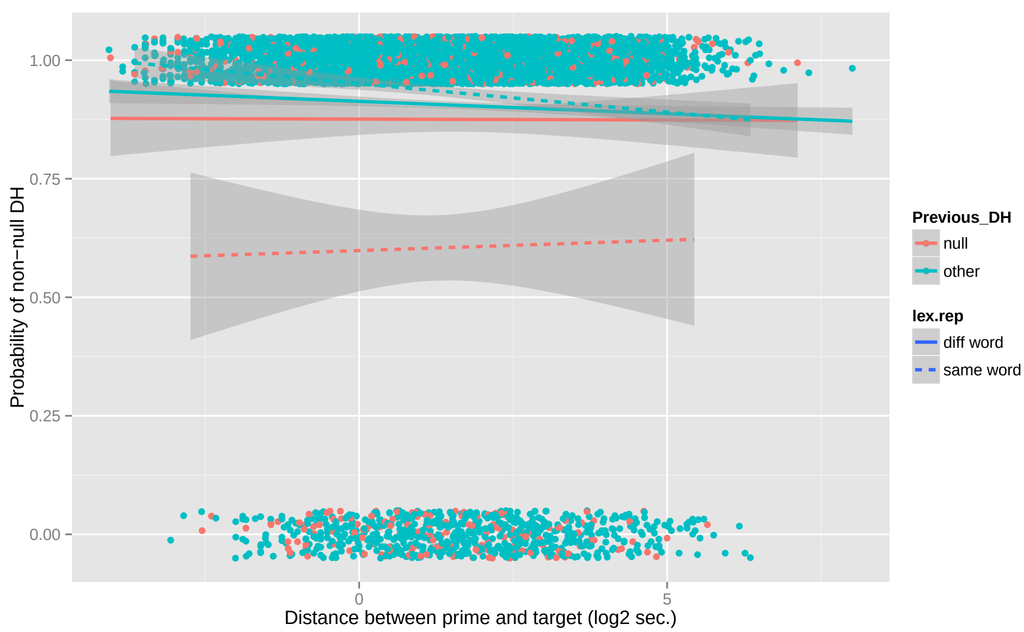

Figure 4.7: Persistence over lag time in DH-null with lexical repetition

	Estimate	Std. Error	z value	p-value
Null prime	-0.09	0.17	-0.53	0.598
Same word	0.44	0.15	2.98	0.003
Lag (log2 sec.)	-0.09	0.03	-3.59	$<$ 0.001
Null prime x same word	-2.08	0.29	-7.18	$<$ 0.001
Null prime x lag	0.04	0.07	0.65	0.514
Same word x lag	-0.10	0.06	-1.65	0.100
Null prime x same word x lag	0.14	0.13	1.14	0.255

Table 4.5: Estimated coefficients for DH-null parameters relevant to persistence (N = 8409)

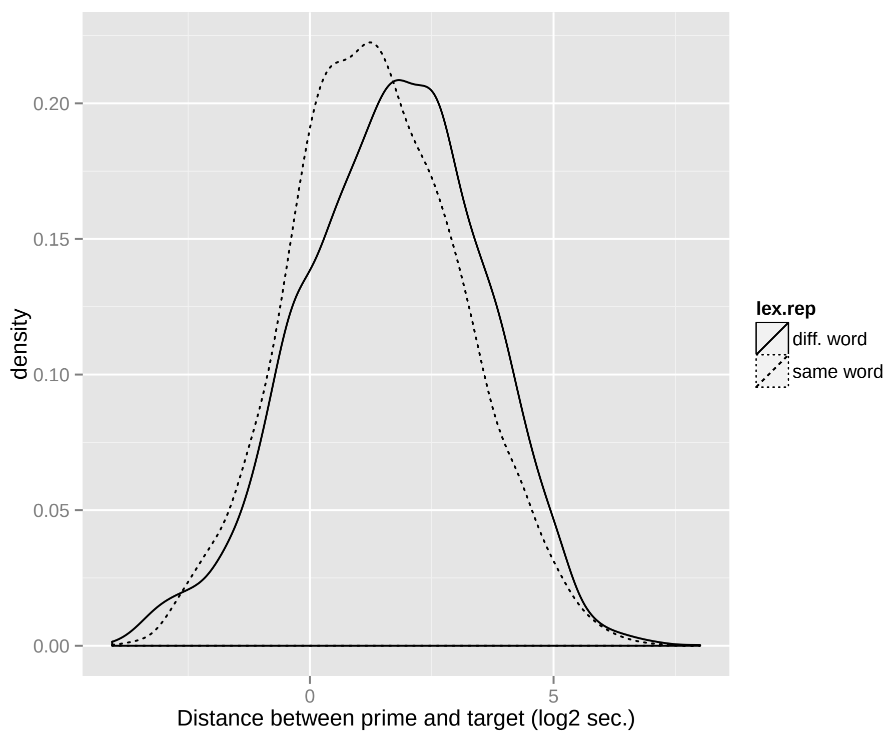

Figure 4.8: Persistence over lag time in DH-stop with lexical repetition

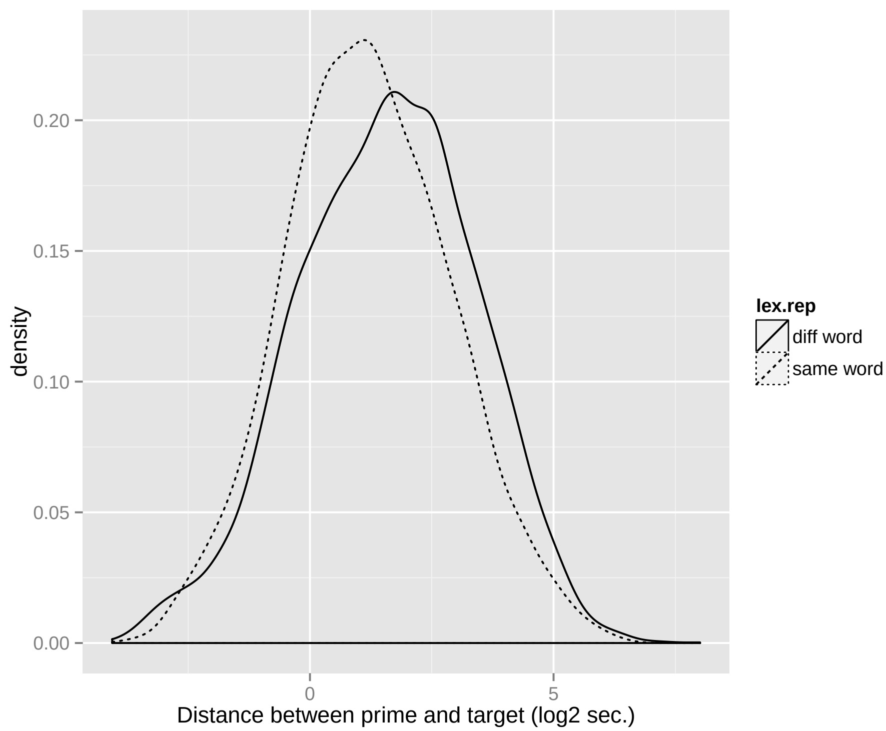

Figure 4.9: Persistence over lag time in DH-stop with lexical repetition

Variable name	Persistence?	Lexical boost?	Decay?
Nominal ING	Yes	???	???
Verbal ING	Generalized	Yes	Yes
Monomorphemic TD	Lexically-restricted	N/A	No
Past tense TD	Generalized	Yes	Yes
DH-stop	Generalized	Yes	Yes
DH-null	Lexically-restricted	N/A	Yes

Table 4.6: Variables and their persistence patterns

4.5 Memory and processing in persistence

Thus far in this chapter and the previous one, I have considered a number of different variables that show a range of persistence effects. By way of summary, table 4.6 presents an overview of the variable definitions, types, and persistence characteristics that I have argued for. By "lexically restricted" I mean that the persistence effect appears only when prime and target are the same word.

In chapter 3, I suggested that the different patterns can be attributed to different variable loci, with morphological variation being persistent in general and phonological variation being persistent only with lexical repetition. The suggestion as framed in section 3.5 is essentially descriptive; it ascribes quantitative properties to different grammatical levels without accounting for the differences. I suggest in this section that the descriptive account of a difference between morphological and phonological behavior can be expanded to have explanatory power. In order to develop and support this suggestion, I first turn to a consideration of the temporal decay properties demonstrated in the experimental priming literature. There are three different types of priming relationships that I examine: repetition priming (*cat* primes *cat*), morphological priming (*cats* primes *cat*), and structural priming (as discussed at length in previous chapters).

On the basis of a series of priming experiments, Forster and Davis (1984) posit the involvement of a short-term repetition effect, which is "an automatic consequence of repeated access of the same lexical entry',' and a long-term effect, which is "totally mediated by episodic factors" (Forster and Davis, 1984, 694). Their primary evidence for this claim is the difference between masked and unmasked priming. In a masked priming experiment, the prime item is presented so quickly (generally on the order of 60 milliseconds) that it is not consciously perceived by the participant — they are unable to report what word they saw, or whether it was the same or different than the target word that follows for a fully perceptible length of time afterwards. Unmasked priming, in contrast, allows the participant to perceive the prime and therefore, presumably, have access to the episodic memory of having perceived that prime. Forster and Davis show that the long-term effect is observable only in unmasked priming but not in masked priming. Under masked conditions, they find a 35-52 millisecond priming advantage at a lag of 0 items (target immediately after prime), a 17 millisecond advantage at a lag of 1 item, and no significant advantage at a lag of 17 items, concluding that "most of the decay takes place within the first second" (Forster and Davis, 1984, 693). Early work on repetition priming using unmasked tasks show long-lasting effects; Scarborough et al. (1979), for example, finds a 32 millisecond advantage after an average of 170 items. Forster and Davis' argument, then, is that the long-lasting effects of unmasked repetition priming must be attributable to episodic factors whereas the brief effect of masked repetition priming is attributable to abstract lexical access facilitation.

The discussion around abstract and episodic effects is complicated, though, by more recent work. Kouider and Dupoux (2009) show that in the experimental morphological priming paradigm, the effects of morphological priming (facilitation of *cat* after *cats*, for example, which presumably targets the lexical entry for a feline) are long-lasting while the strength of repetition priming (facilitation of *cat* after *cat*) decays over time. They attribute this difference to the activation of morphemes in abstract memory, which is long-lasting, versus the storage of surface pronunciation features only in episodic memory, which is short-lasting — the surface episodic similarity in repetition priming gives an additional, but temporary, boost to the abstract lexical effect, the activation of the lexical entry for a feline. This opposite the hypothesized long-term nature of episodic priming and short-term nature of abstract priming discussed in Forster and Davis (1984).

When turning to the temporal properties of structural priming, it is relevant to consider the properties of both the syntactic repetition effect and the boost that it receives from lexical repetition across prime and target. The relative tenacity of this lexical boost has recently been taken under investigation in the literature (Konopka and Bock, 2005; Hartsuiker et al., 2008). Using a picture-description task to elicit instances of the ditransitive alternation in Dutch, which is analogous to the English distransitive alternation, Hartsuiker et al. (2008) show that the generalized effect of structural priming is long-lasting but the lexical boost is not. At a lag of 0 they find a priming advantage of around 50 percentage points more of the primed construction with lexical repetition compared to around 25 percentage points more of the primed construction without lexical repetition. By a lag of two intervening trials, the lexical boost has disappeared, with priming giving around a 15 percentage point advantage to the primed construction regardless of whether or not there is lexical repetition. The lexical boost, then, appears robustly only when the prime and target are adjacent and decays rapidly with intervening trials. Hartsuiker et al. interpret these results as supporting a multi-factorial account of the origins of structural priming.

It is tempting to liken the lexical boost on structural priming to the lexical boost observed in the morphophonological variables under investigation here. The choice between a double object construction and a prepositional object construction seems intuitively analogous to the choice between one pronunciation of a word and another. My proposal, however, is that both the generalized and lexically-restricted effects observed in ING, TD, and DH should be ascribed to the same mechanisms at work in the lexical, rather than structural, forms of priming. The observation that DH-stop patterns with verbal ING and past tense TD not only in its generality but also in its decay continues to suggest a unified analysis for all three of these variables. If these three variables have a morphological locus of variation it is because they implicate a selection between different allomorphs of distinct morphemes. I suggest, then, that the choice of a stored lexical item, be it a root like *cat* or an affix like *-ing*, can be facilitated through the same mechanisms at play in lexical priming.[1]

The argument that accessing a lexical entry for a root like *cat* implicates the same processes as accessing an affix like *-ing* rests on an assumption about the lexicon where affixes are stored as lexical entries in a way that is not distinct from the storage of roots. But is there actually evidence for affix priming in the same vein as the extensive literature I have cited on structural priming and repetition priming? The literature on this topic is perhaps less extensive than those literatures, but it does exist and points strongly in the direction of priming for morphological components in general. Marslen-Wilson et al. (1996) demonstrate effects of priming across words that are unrelated except for a shared derivational morpheme: *happiness* primes *darkness*. They also argue convincingly that this is not merely an effect of orthographic or phonological overlap by showing that *happiness* does not prime words like *harness*. Deutsch et al. (1998) find similar priming effects in a non-concatenative morphological system, showing that the verbal word patterns that convey morphosyntactic information in Hebrew can be primed. The dissociation of the priming of non-root morphemes from the priming of linear strings that share phonological and semantic content helps strengthen the argument that what is being primed in these cases is truly a morphological unit.

However, VanWagenen (2005) points out that there is still semantic content in affixes despite it being perhaps somewhat vaguer than the semantic content of roots tends to be,

[1] The question of what exactly those mechanisms *are*, I will leave for the psycholinguists to debate.

raising the possibility that the apparent priming effect across words sharing an affix is actually a semantic priming effect. She capitalizes on the fact that phonological and semantic priming decay to invisibility at long distances to design an experiment investigating the priming relationship between derivational suffixes in English. The results of her experiment support the original interpretation of Marslen-Wilson et al. (1996): morphologically-related prime–target pairs (*humanism* and *heroism*) show a priming advantage of 20 milliseconds over the baseline of unprimed words, while semantically-related (*valor* and *heroism*) and phonologically-related (*heresy* and *heroism*) pairs receive no significant advantage. She further suggests that "morphological priming is identity priming" (VanWagenen, 2005, 29), a conclusion which accords well with my analysis of affix-allomorphy persistence in morphological variables.

Turning now to the case of monomorphemic TD, there is not a competition between two stored forms. Rather, a fairly late phonological or perhaps phonetic deletion process applies superficially. There is no grammatical object corresponding to the deleted form and stored in the lexicon to host the facilitation, and therefore no effect of persistence — at least not in the generalized case. Why, though, should there be an apparent persistence effect in monomorphemic TD under conditions of lexical repetition if the argument is that monomorphemic TD does not involve a stored representation? Here I point to the relationship between repetition priming and morphological priming demonstrated by Kouider and Dupoux (2009). The episodic properties that they suggest make *cat* a better prime for *cat* than *cats* is are what I suggest may be at work here

4.6 Chapter summary

The goal of this chapter, in a general sense, has been to pursue the integration of sociolinguistic and psycholinguistic phenomena. I have argued that verbal ING, past tense TD, DH-stop involve competition between stored allomorphs, which gives rise to an abstract persistence effect, while monomorphemic TD involves a variable phonological process that triggers an episode-based persistence effect. To put it more generally, the persistence evidence relates the grammatical structure underpinning these variables to the way that they are stored and recalled, suggesting that morphological alternants are represented in abstract memory as lexical entries while phonological variants are represented only in episodic memory. Quantitative evidence from persistence may thus play a role in understanding the nature of the grammatical architecture by linking psychological processes evident in natural speech production to different types of grammatical objects and structures and serving as a window on the interface between the grammatical and psycholinguistic components of human language.

Just as importantly, this chapter also makes a contribution to the question of how sociolinguistic variability is produced by individual speakers. Although this is quite a new question from the sociolinguistic perspective and therefore one that remains wide open, the observation that domain-neutral cognitive processes such as priming (whatever the underlying mechanisms) play a role in the outcome of individual observations of variables demands a change in our approach to the quantitative study of variation. When rates of variant use are attributed to different socio-demographic groups, there is no room for psychological mechanisms because the speech community and its component social strata, networks, or communities of practice are not embodied units endowed with lexicon-storing brains. Stating sociolinguistic generalizations at the community level thus poses a dual risk: the risk of overstating the importance of social factors in shaping patterns of language variation, and the risk of missing generalizations about the psychological contributions to variation that are not just interesting but may hold true explanatory power.

None of this is to advocate for the removal of social or community-level considerations from the study of intraspeaker variation. Indeed, to do so would be to fly in the face of decades of work documenting the relationship between our human social lives and imaginations and the variation that characterizes the speech of individuals. In the next chapter I put forth the suggestion that the socio-stylistic evaluation of different variables can directly impact the operation of the psycholinguistic priming effects I have discussed here.

Chapter 5

Persistence in sociostylistic perspective

So far, in chapters 3 and 4, I have addressed the issues of a) whether persistence is present or absent given different combinations of primes and targets, and b) whether the persistence effect weakens over time. This chapter turns to the question of what other factors, besides distance between prime and target, can drive the strength of persistence. If a persistence effect is present, what factors control whether it is strong or weak? In addition to asking what determines the overall strength of persistence, it is of interest to broach the possibility of persistence being asymmetrical. What I mean by asymmetrical is that one variant serves as a very effective prime, facilitating its own subsequent re-use strongly, while another is less effective in that regard. For example, Estival (1985) and Bock (1986) note that there is no evidence for actives priming actives in the same way that passives prime passives. Why should some primes be effective primes and some primes have little to no effect even when there is evidence for a structural relationship between them?

One suggestion that has been made, and recently tested more carefully, is that the persistence effect may show inverse frequency relationships. Under the inverse frequency hypothesis, triggers that are less frequent have a stronger priming effect than ones that are more frequent. Deriving specific predictions from this hypothesis, though, requires us to determine what the correct denominator is for the calculation of frequency. Jaeger and Snider (2007) put it into plain language when they point out that, to test the idea that less probable primes are more effective, the relevant question to ask is: "probable given what?".

For example, although actives are used more frequently than passives not only overall but presumably also in most contexts, for many other alternations or variables there are some contexts where one alternant is quantitatively preferred and other contexts where the other alternant would be more common. The types of context that have primarily been investigated from this perspective are linguistic ones. Jaeger and Snider (2007) and Jaeger and Snider (2013) look at verb-specific preferences for different constructions such as ditransitives and passives. Some individual ditransitive verbs are more often found with a double object, and others are more often found with a prepositional object. The empirical result from Jaeger and Snider is that the unexpectedness of a verb occurring with a specific construction favors the re-use of that construction. If a verb is usually used with a double object and is instead, on some occasion and for whatever reason, used with a prepositional object, then this will further boost the persistence effect that would already favor the next ditransitive verb occuring with a prepositional object, independently of the target verb's preferences.

In this chapter I put forward the suggestion that sociostylistic considerations may be part of the cluster of factors that contribute to the evaluation of what is "expected" and therefore impact the strength of the persistence effect. This suggests that expectation, rather than frequency per se, is the right way to think about these types of asymmetries, although often frequencies calculated over different denominators are relevant for expectation-setting. The types of syntactic alternations that have previously been used to document inverse frequency effects in persistence are generally thought to be less stylistically sensitive than some of the morphophonological variables that have been discussed so far in this dissertation. This makes the morphophonological variables better suited to the investigation of these types of effects in social contexts. So for example, if an individual speaker prefers the /in/ variant of ING even while that is the minority variant in the community as a whole, might that speaker have a stronger persistence effect for their dispreferred variant /ing/ relative to /in/? Or if a speaker is in a relatively formal situation that favors /ing/, does /in/ have a stronger persistence effect than if the same speaker were in a casual situation that favors /in/?

This chapter begins with a community-level look at ING, focusing on the behavior of individuals rather than the variable's grammatical underpinnings. I ask whether the stylistic and social contextual expectedness of ING can affect the persistence effect in the same way that verbal construction expectations affect the persistence effect in the double object construction (Jaeger and Snider, 2013). Then I compare the sociostylistic contextual effects on persistence to the effects of linguistic expectation, asking whether the impact of socially-conditioned expectations is different from linguistically-conditioned expectation. This section considers data from a new variable, third person singular verb agreement in African American Vernacular English (*he walk* versus *he walks*), and from TD deletion.

One disadvantage of using naturalistic production data is that the rare contexts, which are the interesting ones from the point of view of looking for an inverse frequency effect, occur, by definition, rarely. They therefore rarely constitute the triggers for trigger-target pairs and may not provide enough data to analyze statistically if they are considered in variables or contexts that do not occur frequently enough. As a result, I will primarily limit my investigations to the most frequently occurring types of linguistic contexts, both as triggers and as targets, and focus the question of expectedness on the sociostylistic dimension as this is novel compared to other approaches and has the potential to be complementary to them. What I will suggest is that linguistic predictors trigger an inverse frequency effect, so that linguistic contexts affect persistence asymmetries in a way that is strictly predictable from rates of occurrence, whereas the impact of sociostylistic meaning is not strictly predictable from frequencies, even frequency specific to sociostylistic context, but rather is sensitive to the social evaluation of the forms at the community level.

5.1 Socially-mediated asymmetries

In this section I consider whether the distribution and/or meaning of variability within the speech community affects the magnitude and asymmetry of persistence effects. There are several ways that this suggestion might be observed, at least in principle. One possibility is that for any given individual, their habitual tendency towards one variant or the other — the somewhat mysterious 'baseline' or, in older terminology, 'input probability' — sets up an expectation within that individual's own speech. If so, then each speaker should show an effect where their own dispreferred variant makes a stronger prime when it is used. Another possibility is that for any given individual *in any given context*, their dominant variant in that context is a weaker prime while they are in that context. For some individuals, then, they might always still have an asymmetry in the same direction if their majority and minority variants never switch, while for other individuals there might be switching of the direction of the asymmetry due to switching of the majority and minority variants in different sociostylistic contexts. From the point of view of the data available here, though, it would not be possible to differentiate between these two possibilities because we have evidence from each individual only in a single setting.

A third possibility is that the probability of a variant that drives the inverse frequency effect is the general probability of that variant in the speech community as a whole, in which case everyone should have roughly the same persistence asymmetry. Finally, a fourth possibility is that the social evaluation of variants is a factor driving persistence asymmetries, such that a variable that is either socially stigmatized or stylistically mismatched to context reflects the discrepancy in producing an inflated persistence effect, even if that does not

	Estimate	Std. Error	z value	p-value
(Intercept)	1.44	0.21	6.75	< 0.001
/ing/ prime	-1.12	0.38	-2.92	0.004
Lag (log2 sec.)	-0.11	0.03	-4.30	< 0.001
Same word	0.62	0.32	1.97	0.049
Prime stem surprisal	-0.11	0.08	-1.36	0.175
/ing/ prime x lag	0.04	0.04	0.94	0.349
/ing/ prime x same word	0.30	0.46	0.65	0.519
Lag x same word	0.00	0.10	0.01	0.990
/ing/ prime x lag x same word	-0.10	0.14	-0.71	0.479

Table 5.1: Estimated coefficients for verbal ING with repetition from prime to target as dependent variable (N = 4869)

strictly reflect variant use frequencies. In the following sections I will present evidence from ING that accords best with this fourth possibility.

5.1.1 Asymmetries in ING

The first question that arises is how to measure the strength of persistence in a way that is not susceptible to the data distributional biases that arise when one variant or the other is dominant. To understand what the problem is here, consider a quite careful, well-educated speaker who uses a lot of the /ing/ variant. We might hypothesize that this speaker will have a strong persistence effect when they do use /in/. However, of the two subsets that the data comprise, /in/-primed and /ing/-primed, it is clear that this speaker's /in/-primed subset would be very small, and the estimate of its dependent variable rate quite unstable, because there would be few /in/ tokens to serve as triggers. Their /ing/-primed subset, on the other hand, would be relatively large. Conversely, for a speaker who is generally a more casual speaker (whether for class-based, stylistic, or idiosyncratic reasons) and therefore mostly uses /in/, there will be few /ing/ tokens to serve as primes and therefore it will be difficult to estimate the rate of /ing/ use after /ing/ primes for this speaker.

The first step that I take to mitigate this difficulty is to exclude all speakers for whom the total number of observations in the /ing/-primed *or* /in/-primed subset is fewer than 10. I then rearrange the models of the type I was fitting in chapters 3 and 4 so that the dependent variable is not the variant used in the target, but the outcome of whether the target variant is or is not the same variant as the prime variant. This approach allows the inclusion of the prime variant as a predictor, which encapsulates the hypothesis that one variant in the prime may be more effective at promoting its own repetition than the other. The confirmation of this hypothesis is the key result from table 5.1. The large effect of the intercept indicates the general persistence effect in the ING variable: that the prime variant is more likely to be repeated than not repeated. A similarly large and similarly significant effect, though, is that there is a difference between the effect of /ing/ and the effect of /in/. The direction of the effect is negative for /ing/ as a prime compared to a baseline of /in/, which means that /in/ is "stickier", or a more effective prime, than /ing/. In other words, *relative to /in/*, /ing/ disfavors repetition of the prime variant in the target. The non-significance of the interaction terms indicate that the asymmetry does not interact with the lexical repetition or decay effects discussed in chapter 3.

What is of particular interest about the case of ING is that, setting aside the speakers at the furthest ends of the distribution who are difficult to investigate because of the imbalance inherent in their data across the /ing/-primed and /in/-primed subsets, there is a shared asymmetry at the community level. Speakers who use /in/ more than /ing/ in their interviews demonstrate the same asymmetry as speakers who use /ing/ more than /in/ — for both types of speakers it is the /in/ variant rather than the /ing/ variant that seems to trigger a stronger persistence effect. Indeed, it is not even clear whether there is a robust persistence effect triggered by /ing/ at all, except perhaps in the speech of the most dominant /in/ users. This is seen in figure 5.1. In this figure the size of each point is weighted by the number of observations, as is the confidence interval on the smoothing line. Each speaker in the

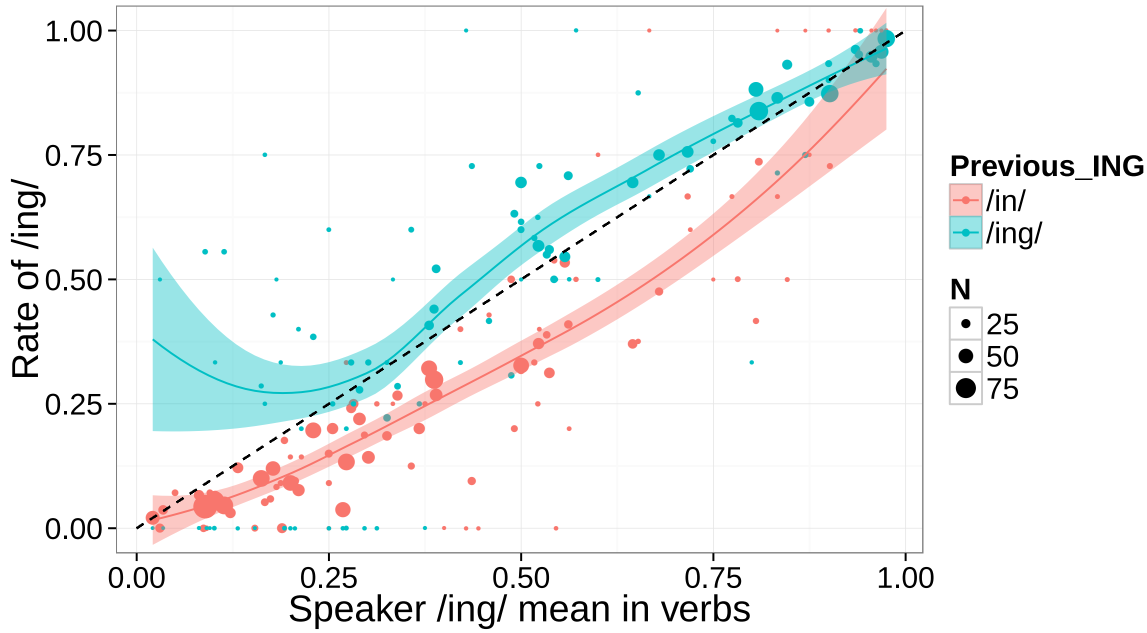

Figure 5.1: By-speaker values for ING, after /ing/ and /in/ triggers

community data is represented by two points, a blue point representing their /ing/-primed /ing/ rate and a red point representing their /in/-primed /ing/ rate. The dashed black line at x=y represents the null hypothesis that there is no effect of the prime variant on the target. If there were no persistence the red and blue dots would cluster along that line. Instead, there is an upward divergence of the blue line, representing a tendency for speakers to have more /ing/ after /ing/ than they should given their baseline, and a more dramatic downward divergence of the red line, representing the larger tendency for speakers to have less /ing/ after /in/ than their baseline would predict. This view thus reflects the regression result that /in/ priming is stronger than /ing/ priming. While the effect appears perhaps slightly smaller in the low-/ing/ range of the speakers, bear in mind that this is also where each speaker's /ing/ mean rate is dominated by their /in/-primed tokens, which will pull their mean towards the /in/-primed rate and make it more difficult for the /in/-primed rate to appear different from the putative baseline. In other words, even where we should by the basic asymmetrical features of the data see the opposite tendency for low-/ing/ speakers, the /in/-persistence effect trumps it and surfaces.

Additionally, notice that there is not evidence for an effect of stem surprisal on ING. Stem surprisal in this case is calculated as the negative log of the reciprocal of the probability of a variant: -log(1/p). The probability *p* is the probability of getting /ing/ or /in/ with that specific verb stem. The surprisal values are calculated across the entire dataset, as there are not enough instances of individual verbs within the data from individual speakers to calculate surprisal by speaker. This should at least capture major differences across lexical items that generally tend towards affiliation with one variant or the other for, maybe, stylistic reasons (for example, see Fischer's (1958) observation of a difference between words that children use with teachers versus while playing in the schoolyard). This is different from the findings of Jaeger and Snider (2013) on the effect of verb surprisal on the choice of the ditransitive alternant, where different verbs seem to have preferences for different frames and the use of a frame that is dispreferred for some verb makes a more effective prime of that frame than if it were used with a verb that prefers that frame. It might have been the case that when a verb occurs more often with /ing/ a token of that verb with /in/ is unexpected and therefore elicits a strong priming effect; this is not the observed result.

I suggest that the asymmetry seen here is rooted in the social evaluation of the variants involved in ING variation — that it is another reflection of the "consensus that the /in/ variant is most appropriate for casual speech, and least appropriate for formal styles" described almost half a century ago by Labov (1966, 398). This type of shared judgment has been considered one of the hallmarks of the speech community and is generally observed in the form of social groups with a range of baseline probabilities shifting in tandem towards more prestigious norms as they move towards more careful speech styles. Here my proposal is that the persistence effect is modulated by the community members' shared judgment that the /in/ variant is out of place in an interview context, or perhaps their shared judgment of nonstandardness of the /in/ variant in general. The data from sociolinguistic interviews, particularly relatively stylistically coherent ones, are insufficient to tease apart these two possibilities.

Either way, the important point is that the basis for the asymmetry is not the speakers' own preferences but rather their awareness of the community norms. If this proposal is correct, it suggests that sociolinguistic awareness is deeply rooted and cognitively pervasive, being able to influence the operation of the persistence effect that I have suggested is attributable to low-level processes like lexical access and implicit learning. Although this is perhaps a quite different view on the role of sociolinguistic awareness than is dominant in the field today, I would argue that it gives no less profound a role to the social salience of language variation. This proposal is also in line with recent neuroimaging evidence on ING from Loudermilk (2013).

5.2 Cross-variable clustering

The observation that there is a role for sociostylistic evaluation in persistence may raise the issue of whether persistence effects observed in corpus data might be attributed solely to stylistic covariation, or perhaps to priming with a social locus (facilitation of the sociostylistic meaning of the variants). I will discuss this issue at some length in chapter 5. Here I investigate the covariation of the morphophonological variables treated in previous chapters with each other. Both verbal ING and DH-stop have been shown to be stylistically dynamic in Philadelphia English (Labov, 2001a). Moreover, Labov (2001a) shows that both ING and DH-stop rates correlate with the same set of descriptors for the environments defining careful and casual speech. The sections of speech characterized by high rates of the casual variant of ING are the same sections of speech characterized by high rates of the casual variant of DH, in line with an attention-to-speech approach to style shifting (Labov, 1966). Given this stylistic affiliation, if the persistence effects within variables are due to stylistic clustering, then it should be possible to detect what looks like persistence across stylistically-related variables. If we treat instances of DH as primes and instances of ING as targets, or vice versa, we might expect to see an effect whereby the use of a stop variant for DH appears to "prime" the use of /in/ for ING, as both of these variants are associated with more casual speech and therefore should co-occur in some non-trivial way.

To test this prediction I combined the datasets for ING and DH, sorted the combined dataset by time of token occurrence within each interview, and then coded them for persistence as if they were all occurrences of the same variable. I then selected out subsets of this tri-variable dataset where the "prime" was ING and the "target" was DH, or vice versa. I then further subsetted the data using the same methods as the previous chapters, looking at whether, say, the /in/-primed DH subset has a different rate of stops than the /ing/-primed subset does. There is a significant Pearson correlation of 0.46 ($p = 0.002$) across speakers' baseline rates of each of these variables; particularly because this is not a sample that is carefully controlled for social factors such as socioeconomic class, this is to be expected given the social stratification of both variables. To account for this source of covariation I conducted the descriptive univariate analyses, shown in figures 5.2 and 5.3, for both types of prime–target pair on a by-speaker level, rather than in the aggregate. These figures are constructed in the same way as figure 5.1 and the point of interest in both is that, unlike in figure 5.1, the red and blue lines cling together around the imaginary null hypothesis line of $x=y$. This suggests a lack of any persistence effect across variables that would be comparable to the one seen within a variable. Although the lack of effect is a little unclear in this perspective, with the blue and red lines being somewhat wobbly, the null result is more fully supported by the multivariate models shown in tables 5.2 and 5.3. Like the within-variable models in chapters 2 and 3, these have a random intercept by speaker and

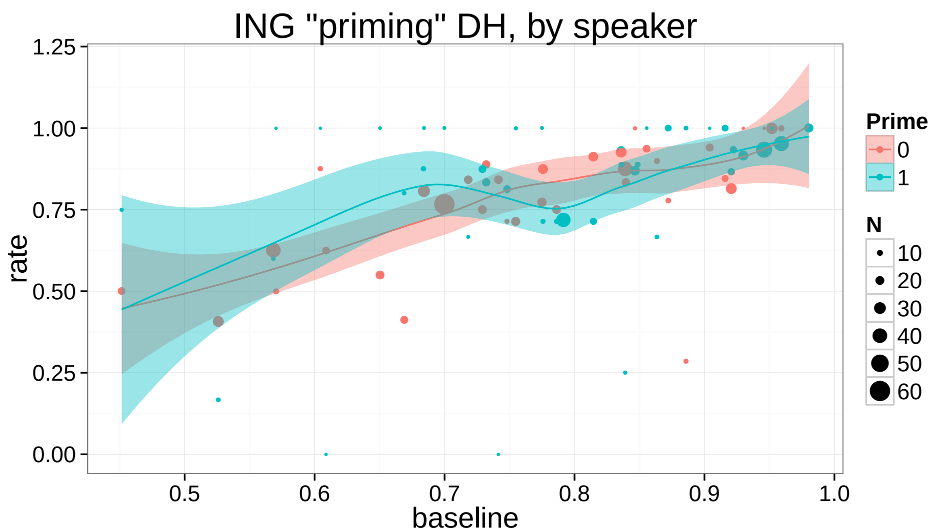

Figure 5.2: Effect of ING on DH-stop, by speaker

	Estimate	Std. Error	z value	p-value
(Intercept)	-1.10	0.42	-2.61	0.009
/ing/ prime	0.40	0.33	1.24	0.216
Lag (log2 sec)	-0.02	0.05	-0.37	0.714
/ing/ prime x lag	0.02	0.06	0.20	0.767

Table 5.2: Estimated coefficients for the effect of verbal ING on DH-stop (N = 1134)

random slope for prime variant by speaker. The effect of the prime variant is not significant in either the ING-on-DH or DH-on-ING direction, nor are the interaction terms with the lag predictor.

This null result is of interest because it gives weight to the idea that simple stylistic co-variation is inadequate to account for persistence patterns observed in corpus data. It also suggests that what is being primed is not a stylistic evaluation — that is, it is not likely that the social meaning is itself the locus of the priming effect. Although different variables may have distinct or shifting stylistic meanings, there is sufficient evidence for their meaningful co-occurrence in speech, presumably derived from overlapping social meaning, that both style-based clustering and priming of social meanings should appear across variables if

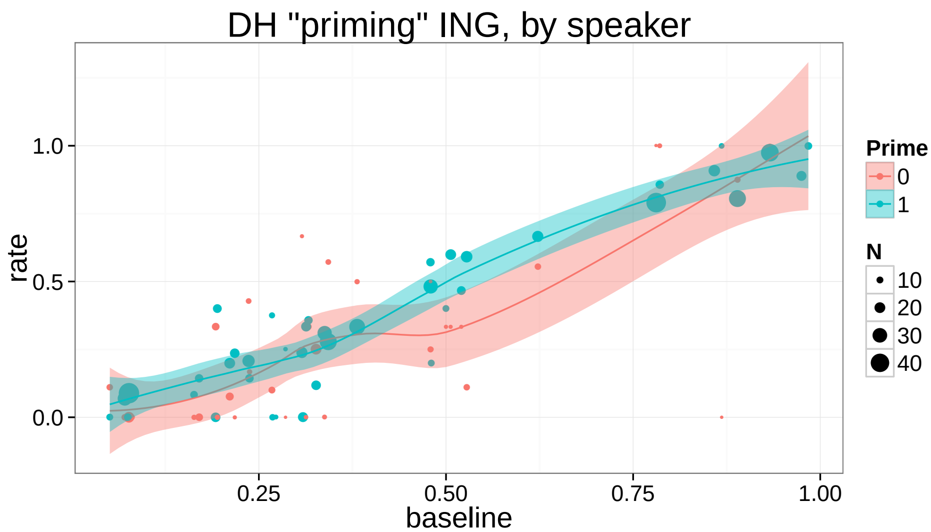

Figure 5.3: Effect of DH on ING, by speaker

	Estimate	Std. Error	z value	p-value
(Intercept)	1.55	0.18	8.49	< 0.001
Continuant prime	0.27	0.25	1.05	0.294
Lag	-0.02	0.02	-1.23	0.220
Cont. prime x lag	0.02	0.04	0.52	0.602

Table 5.3: Estimated coefficients for the effect of DH-stop on verbal ING (N = 1032)

they are the fundamental mechanism underlying persistence effects. This is not to rule out the possibility that these alternative sources of clustering or persistence might be in play at points, but rather to argue that they are insufficient to account for within-variable persistence.

5.3 Linguistically-mediated asymmetries

So far in this chapter I have argued that social expectations play a role in persistence asymmetries. I did not find an effect of verb-specific surprisal on ING comparable to the one demonstrated by Jaeger and Snider (2013) for the ditransitive alternation. But verb-specific surprisal is not the only way that linguistic factors might condition a speaker's expectations about linguistic variant use. This section investigates whether linguistically-determined contextual expectations at the level of morphophonological variable conditioning can be demonstrated.

5.3.1 Verbal -s in AAVE

I first present a slightly-reanalyzed version of the results reported in Tamminga & McLaughlin (2013). The topic of that paper is persistence in variable inflection of third-person singular verbs in African American Vernacular English (e.g. 'He walk' $\sim$ 'He walks'). Just as in the other cases considered here, this is a frequently-occurring variable with a long history of sociolinguistic study (CITE). Rates of 3sg -s presence are fairly low in casual AAVE speech, between 50 and 25%. One of the major factors that has recently been shown to condition the rate at which -s is present is the animacy of the subject. Previous work attributed the basic difference between the high likelihood of absence with a phrase like "He walk" and the lower likelihood of absence with a phrase like "The clock ticks" as being due to the difference between pronoun and full noun phrase subjects.

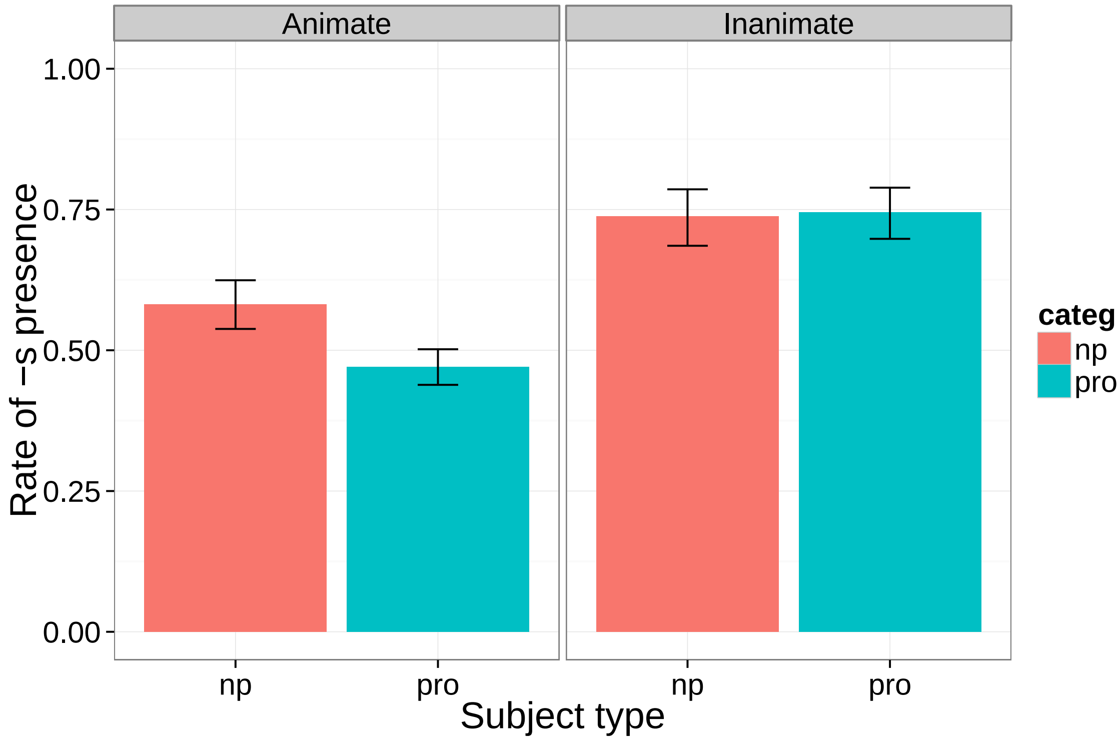

Figure 5.4: Basic animacy vs subject type effect (N = 2345)

However, McLaughlin has shown that the difference is actually due to a factor that was confounded with structural subject type: whether the subject was human or nonhuman. Unfortunately the role of animate nonhuman subjects in this framework is still unclear as there is not enough data to form a clear quantitative picture on their behavior, but for the time being I will follow McLaughlin in using the term 'animacy' to describe the difference between animate human subjects and inanimate non-human subjects. The breakdown by subject type (pronoun vs. NP) and animacy (animate vs inanimate) as reported by McLaughlin is shown in figure 5.4. The data in all of the below figures in this subsection come from the Frank Porter Graham Corpus as described in chapter 1. Unlike in the other case studies in this work, though, the data here includes prime–target pairs across individuals as well as within the speech of single individuals.

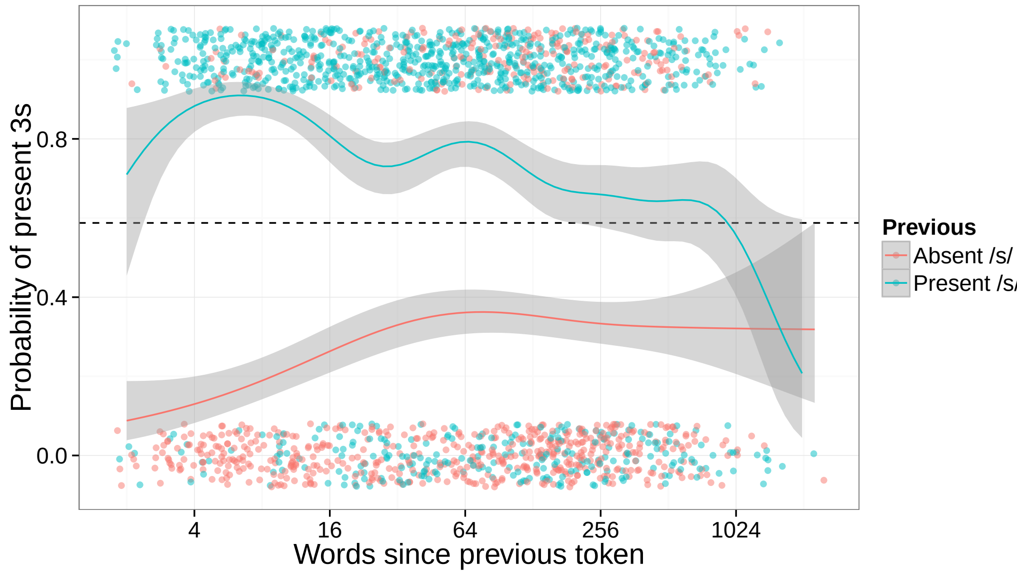

Figure 5.5: Persistence of 3sg -s by lag distance

Figure 5.5 shows that the basic persistence effect is robust in 3sg -s, and that it decays over time (with time here being counted in orthographic words, rather than seconds). In this graph we can already begin to identify some degree of persistence asymmetry, with a perhaps-larger effect for -s absence than presence as a strong prime.

In the context of structural mediation of persistence asymmetries though, the interesting question is whether the expectations created by grammatical conditioning of the variation can modulate the strength of the persistence effect. We see in figure 5.6 that, collapsing over the pronoun/NP distinction, the rate of -s presence when the verbal subject is animate is right around 50% whereas the rate of -s presence when the verbal subject is inanimate is approximately 75%. The hypothesis given other results here, then, is that when a prime is inanimate the use of overt -s should be unsurprising and therefore produce a relatively smaller persistence effect than the absence of -s. In the case of an animate prime, both options are about equally likely so the effect should be balanced (whether large or small may depend on whether this effect is driven more by surprisal or entropy).

Figures 5.7 and 5.8 break down the effect of a preceding overt or null -s on a target verb

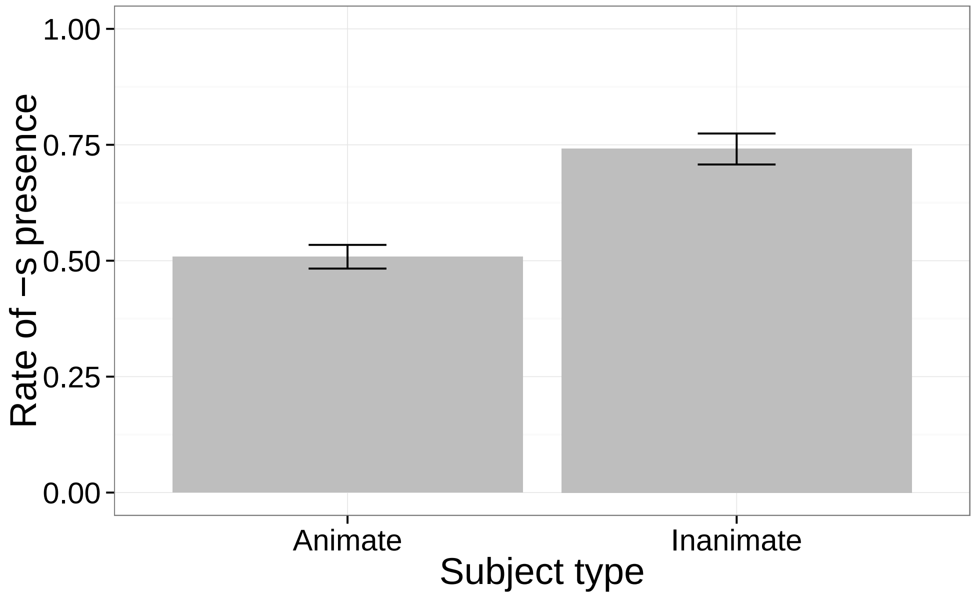

Figure 5.6: Basic animacy -s rates(N = 2345)

by the animacy of the prime subject. The results align with the hypothesis as described. The effect of an animate prime is in figure 5.7. What we see is a robust persistence effect for both absence and presence of 3sg -s in the prime. Note, of course, that this holds across both target types, animate and inanimate, which in light of the discussion in the previous chapter is in line with our assumption that this is a single variable with a single linguistic locus that is simply conditioned by other grammatical factors in the local context. The result of equally sized persistence effects for the two prime variants contrasts with the result in figure 5.8. There we see that when a prime is inanimate, and hence -s presence is the more expected variant (at least from a frequency-based point of view), there is only a weak effect of persistence triggered by -s presence and a stronger effect of persistence triggered by the rarer -s absence variant.

Although this is a relatively straightforward result it is important in light of the problems associated with token count biases introduced by variant frequencies. In cases such as the study of what effect of an individual's preference for /in/ or /ing/ has on the impact of

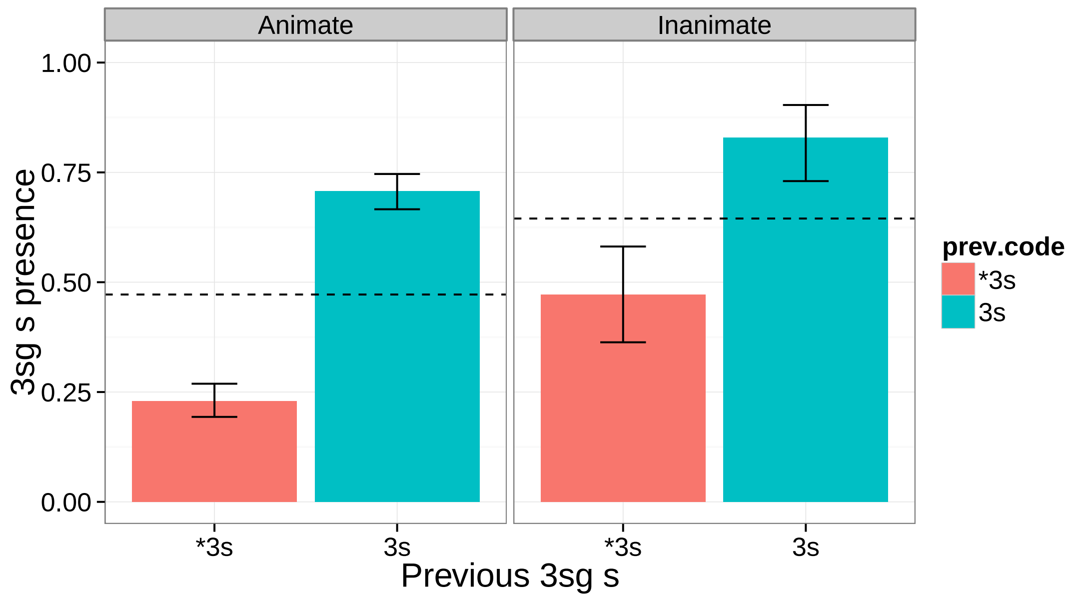

Figure 5.7: Effect of an animate prime on 3sg -s

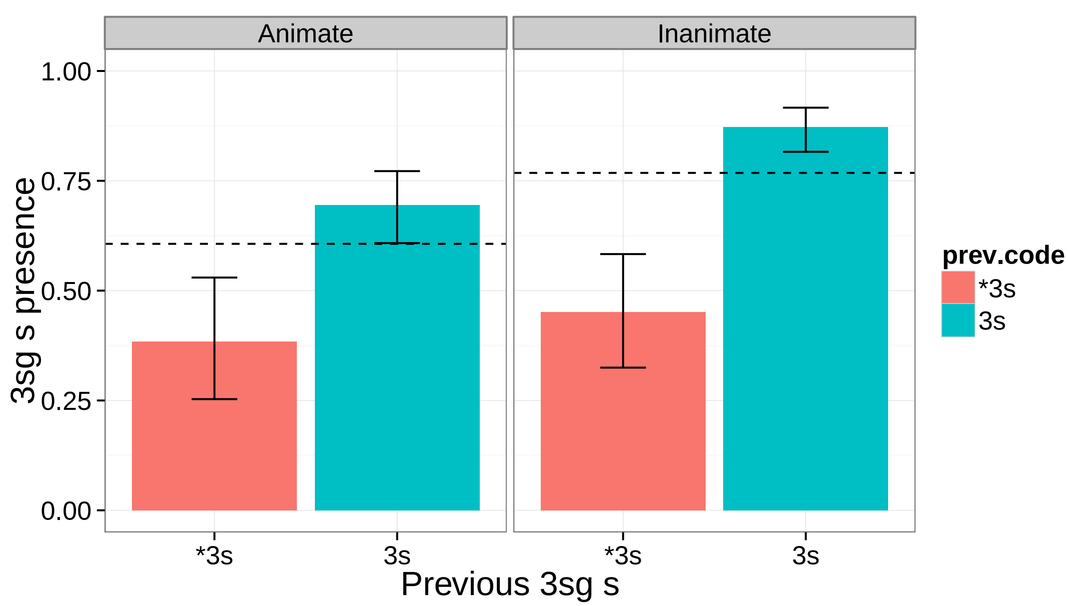

Figure 5.8: Effect of an inanimate prime on 3sg -s

their dispreferred variant as a prime, there is a confound. We want to know whether the minority variant induces a stronger persistence effect, which we intuitively would want to measure as the observation that there is a larger distance between the baseline and the minority-primed subset rate. But then there is the problem that the minority-primed subset is substantially smaller than the majority-primed subset so the majority-primed subset will dominate in the calculation of the mean and thus make it look like the majority-primed subset is always closer to the "baseline". In this case the majority/minority distinction cuts across the contexts and therefore at least partially sidesteps this confound. There is still an uneven distribution, but it is not the dominating factor because we are comparing across contexts rather than across priming conditions.

This is also an interesting case study because it uses children instead of adults. The behavior of these children, pooled over ages 4 to 17, shows robust and nuanced persistence effects. Future work might explore the developmental trajectory of these effects but certainly this seems favorable towards an account where relatively automatic persistence accounts arise relatively early in development.

5.3.2 The phonological context of TD

In chapter 4 I showed that the deletion of coronal stops is a variable process that does not give rise to persistence effects beyond the context of lexical repetition. I suggested that this type of persistence behavior may be correlated with phonological, as opposed to morphological, variation. If so, we might also ask whether TD shows a different type of behavior from AAVE verbal -s in terms of linguistically-mediated persistence asymmetries, even when there is persistence to be observed because the lexical item is repeated from prime to target. One of the most robust conditioning factors on TD deletion is the phonological segment following the token. Following consonants favor deletion, while following vowels promote retention of the stop. This is quite a large effect: in the data set at hand, there is 12.5% retention before consonants and 51.3% retention before vowels. We might thus expect to see a noticeable persistence asymmetry comparable to the one observed in verbal -s immediately above. The retention of a coronal stop prior to a consonant in the prime should be quite surprising indeed and thus produce a substantial favoring of subsequent retention, especially compared to the effect of deletion before a consonant. The near-equal balance of retention and deletion before a vowel in the prime, on the other hand, should give rise to a relatively balanced persistence effect on the target.

In figure 5.9 we see that this is not the case. The data in this figure are prime-target pairs where both prime and target are instances of the same monomorpheme. This is the best-represented context where we do see some effect of persistence. At a glance, the following segment of the prime doesn't seem to have any particularly large effect on the persistence — visually, the relationship between the red and blue bars is about the same across the two following-segment conditions. If anything, there is an effect that runs counter to the surprisal-based predictions: the confidence interval on the bar of the more *surprising* variant relative to each following-segment condition contains the overall average, suggesting a non-significant effect. Given the relatively lowish token counts, however, the safest conclusion is probably that we do not find any evidence for a surprisal-based asymmetry based on the linguistic context of the prime in the case of TD. This is thus another case in which the behavior of TD is somewhat different from the behavior of a morphological variable, in this case verbal -s.

5.4 Chapter summary

The results in this chapter suggest that the basic currency of persistence asymmetry is expectation. Crucially, the determining factor is not strictly frequency. Usually frequent variants are the expected ones for any given context, but in the case of ING we saw that

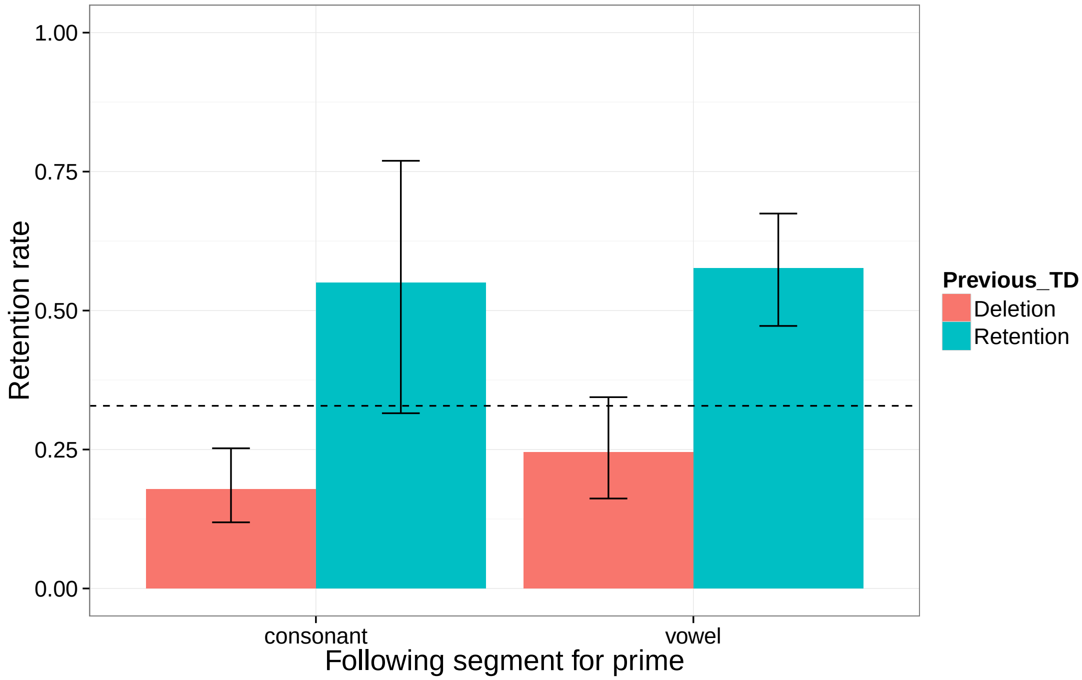

Figure 5.9: Effect of a previous TD token on a target by the segment following the prime

it is the /in/ variant that drives the persistence effect regardless of whether an individual speaker shows a preference for /ing/ or /in/ in the interview context. I suggest that this is a reflection of the shared community-level evaluation of the /in/ variant as being the odd one out somehow — whether as the non-standard variant in general, or the variant that is seen as incongruous with the interview situation, these data can't say. The fact that speech communities share evaluations of variables has long been known from data on uniform style shifting behavior and subjective reaction tests; indeed, it is considered one of the defining features of a speech community. This, then, looks like a cognitive encoding of that shared evaluation at the individual level.

A more straightforward inverse frequency effect is observed when we turn to the linguistic conditioning of morphological variation in AAVE. Here, the asymmetry observed is one where the use of the -s variant that is unexpected (from a frequency-based perspective) given the animacy of the prime subject produces stronger priming than the use of the expected variant. It is apparent that this effect is not an overall effect of one variant or the other being the stronger prime, as seen for ING at the community level, from the context where both variants are equally expected and produce equal sized persistence effects. Linguistically-driven asymmetry of this type does not appear in the case of TD and its sensitivity to phonological context. I showed in chapters 2 and 3 that monomorphemic TD does not show persistence effects comparable to those in DH-stop and verbal ING, and argued that this is due to TD's phonological locus. The failure of TD to be sensitive to the speaker's expectations given the prime, then, is consistent with the suggestion that the persistence in TD is more episodic and superficial.

The overall picture is one where production priming of linguistically-stored objects is sensitive to the speaker's expectations given a combination of the frequency of variants in different linguistic contexts and the appropriateness of variants in different sociostylistic contexts. Both sociostylistic context and linguistic factors have an effect on the priming of the variables at play. This suggests that these have independently-identifiable effects but that they can impact each other, meaning that there is a complex relationship between social, grammatical, and psycholinguistic planning/processing factors in the production of the patterns of variability that we observe in conversational speech. Future work might fruitfully attend to the integration of these types of information in real time.

Chapter 6

Persistence extended

6.1 Major findings

In my investigation of the grammatical, psychological, and sociostylistic facets of persistence in conversational speech, I have advanced several major arguments. First, I have shown that persistence is sensitive to grammatical status in ways that may reveal the unity of process underlying familiar morphophonological variables, or conversely call their coherence into question. This argument takes from the experimental structural priming literature the observation that repetition necessarily presupposes sameness, with the question of interest being on what level this sameness is relevant to speech planning and processing. The persistence evidence on what has been called ING in the literature reveals a division between the verbal and nominal contexts of ING that mirrors the observed rate differences in those contexts: there is no persistence from verbal to nominal ING tokens even though there is persistence within different verbal categories. This lack of cross-contextual persistence, I argue, supports the morphological account of ING, where verbal ING variation is allomorphic in nature (i.e., the variation is between the suffixes /ing/ and /in/) and nominal ING must be treated as a separate variable. The reasoning here is that a purely phonological account of ING variation would have no means of capturing the failure of uniform persistence — if the use of the apical in, say, *ceiling* is the same as the use of the apical in *working*, it should be able to affect the choice of the variant in *working* just as well as *working* does in *thinking*. By appealing to the already-existing morphological analysis of ING, it is possible to explain some of the persistence facts in a natural way. The allomorph selection facilitates future choices of the same allomorph but does not affect a phonological place-of-articulation decision in an unrelated word, despite the surface resemblances. This analysis has been unable to account for the persistence behavior of the quantifiers, which appear to be linked to verbal ING. I suggest that further examination of the quantifiers and verbs, likely in an experimental setting, will be necessary to fully reconcile all of the ING categories.

Turning to TD, another sociolinguistic showcase variable, I again found a distinction between the cases where the material targeted for variable deletion is and is not coterminous with a past tense suffix. This time, however, the distinction is not only that there is no cross-category persistence but also that monomorphemic TD is persistent only when the prime and target are the same lexical item, while past tense TD shows weak but generalized persistence. I suggested that the verbal (that is, regular past tense) portion of TD should be put in the same category as verbal ING, whereas the monomorphemic type of TD should be put in a different category, one characterized by more superficial variation that does not give rise to abstracted forms of persistence. It appears that DH-stop also elicits the same pattern as verbal ING and past tense TD, leading me to suggest that this variable is not phonological but actually morphological. DH-null, in contrast, shows lexical specificity like TD.

These results form the basis of the demonstration that persistence is a useful diagnostic for grammatical status. This is not just a methodological convenience; rather, it is important to observe that this type of quantitative evidence for linguistic structure is made possible by the constraints that such structure places on variation. Or, to put it more strongly, language variation operates within linguistic systems, through the manipulation or competition of objects generated by the grammar. Whatever the theoretical analysis given to variation in general, it is clear that it relates directly to grammatical structure and its elucidation cannot be meaningfully separated from questions of how to characterize the structure of human linguistic knowledge.

The second argument that I have brought forward in this dissertation is that the temporal properties of persistence, notably its decay over time, may reveal elements of the psychological origins of persistence. These temporal properties correspond well to the hypothesized distinction between phonological and morphological variables' interactions with persistence. I showed that the same monomorphemic TD variation that was notable in chapter 3 for its failure to extend beyond the condition of lexical repetition is also notable for its apparent failure to decay as target and prime get further apart. This contrasts with the decay with lag time that is evident in all the other variable case studies, including the lexically-restricted case of DH-null. I linked the different temporal decay and lexically specific profiles of these case studies to the different memory and lexical access systems that have been implicated in the literature on repetition priming and structural priming.

Specifically I suggested that the morphophonological variable persistence evidence can be incorporated into a tripartite model of the cognitive bases of priming. An initial fleeting effect of word repetition, perhaps reflecting heightened activation from lexical access, provides facilitation that is detectable in masked repetition priming paradigms as a brief episode-less priming effect but only appears in the corpus variation case as an early upward shifting of the lexical boost provided primarily over time by episodic memory. This long-term episodic memory, I suggest, may be, as has been suggested previously, the basis of the long-term repetition priming effect in unmasked paradigms but it can also serve as the basis of the non-decaying lexically-specific effect of persistence in TD deletion. The robust observations of structural priming in other corpora and experiments and morphological persistence in my data are neither lexically-specific nor fleeting; I place them in a third category which is abstract and relatively long-lasting, which might be attributable to the mechanisms of implicit learning as suggested by some in the experimental structural priming literature.

The general point of this chapter was to ascribe different persistence decay patterns to a distinction between grammatical objects such as allomorphs, stored in abstract memory (presumably, or by definition, in the lexicon), and phonologically-altered surface forms (stored in episodic memory in some state of greater or lesser abstraction). By doing so, I suggest that it will be possible to give some portion of the variation observed in conversational speech an explanation, rather than simply a description. This type of explanatory account is particularly valuable because it makes strong and falsifiable predictions, largely because the general cognitive mechanisms to which I am ascribing persistence effects should be more or less universal, operating consistently across variables and speech communities. For instance, it predicts that a phonological variable that has no plausible origin whatsoever in morphological structure — perhaps /th/-fronting (*think* ∼ *fink*) in dialects that have it — should not be able to produce persistence effects that generalize beyond cases of lexical repetition. There is some evidence on just this variable that suggests that my theory may already be on its way to falsification (Clark, 2014). However, this work has not yet ruled out lexical repetition contexts as the potentially sole source of the demonstrated persistence effect. The prediction from the analysis given here, then, is that the persistence effect on /th/-fronting should turn out to closely resemble the effect on monomorphemic TD: it should be restricted to cases where the prime and target are the same lexical item and should show little to no evidence for decay over time.

Finally, the third major argument that I have pursued, in chapter 5, is that expectation, rather than frequency, is the correct conceptualization for the basis of persistence asymmetries. The reason I argue this is that I show a role for sociostylistic evaluation, in the case of verbal ING, in driving persistence asymmetries, and that this role is not directly attributable to the frequencies with which variants are used. Regardless of whether individual speakers use more /ing/ or more /in/ in their interviews, they show a tendency to have stronger persistence resulting from the use of /in/ than from the use of /ing/. I ascribe this asymmetry to the shared community-level evaluation of /in/ as less appropriate for certain types of speech contexts, likely including the context of being interviewed and recorded by a researcher from a prestigious local university. An alternative but related possibility is that /in/ is simply marked as a nonstandard variant and thus produces such an asymmetry; because I have no evidence from a context in which /in/ would be the widely-accepted "appropriate" variant (watching sports at a bar perhaps?), I cannot distinguish between these possibilities. However, I consider the former to be the more limited and therefore less bold hypothesis given the currently available data.

I combine the evidence from non-frequency-driven stylistic expectations with evidence from linguistic contextual expectations. In line with previous work showing that verb-specific surprisal of structural alternants drives the strength of priming in syntactic variation, I find a contextually-refined role for frequency in setting up linguistically-based asymmetries in the kind of abstract priming that I argued characterized both allomorphic variable persistence and structural priming. In the case of AAVE 3rd singular -s, a grammatical context that favors -s realization shows stronger persistence from null-s primes, whereas a grammatical context that favors null-s shows stronger persistence from realized-s primes.

We see here that frequency given a grammatical context is an appropriate level of analysis for inverse frequency effects in linguistic conditioning of persistence asymmetry. This contrasts, however, with the finding that ING does not have verb-specific surprisal effects, perhaps suggesting that syntactic alternations but not morphological ones are tied to verb identity. It also contrasts with the finding that, yet again, monomorphemic TD is the odd one out: despite a strong frequency difference across phonological contexts, there is no apparent asymmetry stemming from the effect of ‘surprising’ deletion prior to a vowel or ‘surprising’ retention prior to a consonant. My suggestion is that both linguistic-structural and sociostylistic considerations interact in determining what kind of variant a speaker expects, even if they are themselves speaking, but that phonological variation is not deep enough to trigger this kind of expectation violation persistence boost. I have argued that the psychological bases of persistence should be universal, but they are also susceptible to being impacted in the details of their implementation by a speaker’s fine-grained contextual knowledge about the way language works in the real world.

This set of arguments integrates the consideration of grammatical objects and architecture, domain-general cognitive processes as applied to grammatical objects within the architecture, and the interaction of contextual knowledge on the operation of these grammatically sensitive psychological processes. In the widest view, my goal is to show that the study of variation is integral to the study of language structure and processing, but that this does not mean any of these factors are reducible to the others. The ultimate goal of the dissertation is to tie these findings to a more complete understanding of the relationships between grammatical representation and derivation, sociolinguistic variation, and universal psycholinguistic factors. Researchers working from a perspective privileging one of these elements have sometimes tended to either subsume all linguistic phenomena under their perspective or reject the other elements as wholly unrelated. I take the position that neither of these extremes will prevail in a fully-articulated theory of the human linguistic capacity. Complex multi-factorial accounts that allow for different modules and systems with different cognitive and temporal properties will be needed to deal with the complex set of empirical results about the way language is represented, computed, and understood. To develop such accounts, we will need not only theoretical argumentation from acceptability ratings or grammaticality judgments and cross-linguistic typologies, but also quantitative naturalistic data that is drawn from real-world contexts and carefully-controlled experimental data to clarify the behavior of rarely-occurring but probative contexts.

6.2 Open questions

Like most dissertations, this one leaves more questions unanswered than answered: some of them are longstanding issues in the literature, others are new questions that are raised by the results and framework of this study.

A key question that I have made reference to but not yet treated at length is what the relationship is between style and priming. The reason that style is relevant is that, like priming, it is another source of clustering behavior in conversational speech. This question references a debate over whether persistence observed in corpus data can be attributed to priming *qua* priming, or whether it is impossible to make such an attribution because the messy and uneven nature of corpus data will always leave open the possibility of some other explanation. As mentioned in the introduction, the phenomenon I call persistence has gone and continues to go by many other names, one of the most common of which is "priming". Throughout the preceding chapters I have attempted to maintain the use of the term "persistence" to describe the observation of statistical repetitiveness in corpus data. The use of the term "priming", though, is quite common in the sociolinguistics literature, which tends to be quite clear about attributing the observation of variant clustering in conversational speech to psycholinguistic mechanisms. By invoking the concept of priming this literature equates experimental psycholinguistic phenomena with quantitative patterns of conversational speech. From the experimental side, such an equation has not always been welcomed. The use of corpus data makes impossible the implementation of strict controls as in experimental design, leaving open alternative explanations. Most commonly, claims that repetitiveness in corpus data reflects priming are countered by the possibility that clustering may actually be due to stylistic coherence in speech or text.

Skepticism about the nature of corpus results on persistence has come from both corpus linguists and experimental psycholinguists. Sankoff and Laberge, in their early identification of the non-independence of successive tokens of standard and non-standard French pronouns, consider the possibility that the effect they have isolated is actually stylistic, asking, "Is sequence constraint a form of text coherence, albeit quantitative, operating on coreferential variables in adjacent syntactic units, or does it simply reflect stylistic homogeneity within relatively short passages of time?" (1978, 122). In concluding in favor of the former interpretation, they point out that the simple switch rate in unconstrained tokens is very close to the predicted switch rate under the null hypothesis, which leads them to believe that the effect is not one of stylistic coherence. Estival acknowledges that "in order to show that [syntactic priming] exists, we must first make sure that we have isolated it from the effect of other factors" (1985, 8); she takes a mixed approach to this effort by pulling out subsets of the data and including discourse-level predictors. Szmrecsanyi expands on the variety of factors besides psycholinguistic priming that might be at play, pointing out that "in naturalistic data, speakers' output may exhibit persistence effects for reasons of rhetoric, politeness, or thematic coherence, to aid the process of gap filling in creating and processing elliptical utterances, to open up question-answer pairs, because speakers feel like intentionally repeating items from previous discourse – but it is not easily possible to disentangle the above motivations through corpus study in a waterproof fashion" (2006, 144). This is essentially the same criticism as leveled by Branigan et al. and Pickering and Branigan (1999), the former of whom suggest that "Weiner and Labov's (1983) results might just reflect shifts in the register used during the interviews which they studied" (1995, 492).

The suggestion that stylistic covariation is sufficient to account for the existence of persistence effects in corpus data has itself been questioned in turn, particularly from the direction of sociolinguists and corpus linguists. Scherre 2001, Cameron and Flores-Ferrán 2004, and Gries 2005 all advocate for a psycholinguistic interpretation of persistence effects observed in language use. Cameron and Flores-Ferrán, for example, point to the pyscholinguistic theory of spreading activation (Dell, 1986) to account for persistence effects in Spanish null subjects (2004). Estival's (1985) early demonstration that the persistence effect on the passive alternation, identified by Weiner and Labov 1983, is significant even after a variety of discourse-level factors are controlled makes some progress towards addressing the criticisms of those who doubt that such effects could be observed in corpus data. Gries' protestation that "explaining the frequent cases of syntactic priming by hundreds of sudden register/formality changes does not seem very plausible" (2005, 387) has rather less basis, particularly given a modern understanding of style as being highly sensitive to fine-grained shifts in topic and social stance. Also, an assumption of "hundreds of sudden register/formality changes" is not necessarily a requirement for apparent priming to fall out of data that has no sequential dependence across tokens but that does have shifts in the baseline rate. A stretch of speech that has high rates of one variant in the first half and then high rates of the other variant in the second half could still produce an effect where, say, /ing/ is more likely after /ing/ and /in/ is more likely after /in/.

Meanwhile, Jaeger and Snider note that the tasks typically used in experimental priming studies may also themselves be problematic because they might induce explicit learning and thus reflect cognitive processes distinct from priming (2007, 28) (where *priming* here refers to the presumed neural basis; essentially they suggest that priming in the behavioral sense may not be a unified phenomenon). Furthermore, I note that if priming in the psycholinguistic sense is to be understood as having ecological validity, it *should* appear in conversational speech because it should be an active force in speech production.

But the question of the relationship between style and persistence goes beyond the question of whether persistence is epiphenomenal of stylistic clustering. Style is the primary sociolinguistic factor that actually conditions variation within individuals even though quantitative sociolinguistics as a whole is generally careful to note that the type of variation it concerns itself with is intraspeaker variation. For any given speaker, the set of sociodemographic characteristics that, by hypothesis, characterizes their baseline rate of use sets them more or less statically, with some leeway for, say, age-grading or social mobility. I have implicitly been assuming a conceptualization of the relationship between style and priming that attributes shifts in a speaker's "baseline" or "target" as being the domain of style-shifting and repetitiveness from instance to instance of any given variable to priming. Under this type of conceptualization, style is a property of larger chunks of speech that gives you some sort of large-grain clustering of tokens in conversation, while priming is a property of sequences of tokens that gives you finer-grained repetitiveness within those larger clusters from style. Such a view of style is essentially founded on early work from Labov (1966), which uses different interview tasks (answering questions conversationally, recounting brushes with death, reading written passages and word lists, pronouncing explicit minimal pairs) to manipulate the amount of attention that a speaker is paying to speech. As the speaker's attention is drawn to their speech they increase their self-monitoring and move towards a more prestigious and careful standard. This attention to speech applies equivalently across different variables.

One important open question in moving towards a careful articulation of the relationship between style from priming is what kinds of methodological techniques might be useful for discriminating between these factors. Style is typically explored using conversational data drawn from a well-understood social context, so that knowledge about the social context can be applied to the interpretation of variation that seems to covary with it. Since this attention-to-speech model was developed in the 1960s, sociolinguists working on speech style have demonstrated many ways in which style is more multi-dimensional than simple attention to speech. However, almost all of these complexities have been illustrated in qualitative case studies of small numbers of individuals — quite often single individuals (Podesva, 2007; Johnstone, 2009). The problem is that this leaves us without a model of style that can be incorporated into more generalizable models of human language performance more generally. I suggest we will need to begin by reverting to a more basic approach to style here, one that understands style as upwards or downwards shifts in variant probabilities in response to the demands of greater or less casualness in speech exerted by the context, topic, or interlocutor. This is not to deny the value of multifaceted qualitative approaches to style, but rather to make an initial operationalization of the phenomenon of "style-shifting" possible in order to enable the testing of claims about what kinds of patterns it can produce.

Although the potential relationship between style-shifting and observed persistence has been appealed to many times in discussions of persistence and priming, it has not been formalized in a way that allows anyone to make any definitive statements in either direction. Instead we have, on the one hand researchers who call corpus results into question by suggesting that apparent priming "might" be attributable to style-based clustering, and on the other hand researchers who wish to appeal to psycholinguistic explanations for corpus data complaining that such a suggestion doesn't "seem" plausible. I contend that we can move this discussion forward by investigating the quantitative profiles of both style-shifting and priming separately. The literature on experimental priming, which takes place using consistent tasks in the socially-stable context of a university lab, can form the basis of our generalizations about the quantitative characteristics of priming, presumably isolated from style-shifting. The converse isolation, that of style-shifting from potential priming effects, would be more challenging using natural or even experimental data.

I therefore suggest that we will need to turn instead to the use of simulation. What might a style-shifting simulation look like? Making the methodological decisions about simulating style shifting is a way of playing with the quantitative properties that a model of style might have. I consider that the use of simulation might provide data derived without sequential dependence across tokens as a testing ground within which to look for evidence of apparent persistence. The argument would be that simulating simple style-shifting can give us some idea of how much clustering can derive from stylistic co-variation of neighboring tokens and how much might or must instead be attributed to priming. Preliminary investigations and attempts along these lines have run into a number of questions about what assumptions would make sense to stipulate in such a simulation. How large should the style boost above or below a speaker's baseline be? How frequently and regularly should simulated speakers shift between styles? Should asymmetries in the style shift boost be built in? In other words, these are questions about the currency of stylistic style shifting and its temporal grain.

In addition to simulation, I agree with previous attempts to make clear the parallels between results such as mine or other corpus persistence demonstrations and the known characteristics of experimentally-elicited priming, both lexical and structural. As the resemblance between the properties of priming and the properties of persistence are shown to be more detailed, the belief that they have the same source is strengthened. It should be borne in mind, though, that a distance effect would presumably characterize any style-clustering account of corpus persistence as well. If prime and target being more likely to contain matching variants is due to their shared covariation with the stylistic context, then all else being equal, primes and targets that are closer together should be more likely to be in the same stylistic context than primes that are further apart. The general observation that persistence decays is entirely compatible with a priming-based account but is not likely to be probative of the distinction between priming and stylistic clustering.

As the picture is brought into sharper focus to show differential decay functions across different structural and lexical combinations of primes and targets, however, as I did in chapter 4, it begins to take on a much more natural explanation in the locus of these kinds of repetitiveness effects in different storage and memory systems, rather than style-based clustering. Refining a model of style to make these kinds of differences emerge would require a curious set of interactions between style and lexical identity. For example, how might the observation that monomorphemic TD words are persistent based on lexical identity regardless of the distance between them, at least within some range of reasonable distances within a conversation, be cashed out in a model that has no sequential dependence across the instances? A possible path towards an alternative would be to imagine that each individual monomorpheme that undergoes TD deletion has a unique value for its probability of deletion, perhaps related to that word's stylistic valence. Even instances at a distance of the same TD word might then appear to be related to each other via their both being drawn from the same distribution. But while it seems sensible to see such variation in play with sets of words, some which are high-deletion and some which are low-deletion, the range of the probability space is likely not differentiable enough to prevent this kind of covariation from emerging more generally across TD words that happen to have overlapping distributions in their likelihoods. In other words, such a stylistic mechanism might be able to elicit apparent persistence within lexical items but it would be difficult to restrict it sufficiently to explain why there is not more generalized persistence in the case of TD. Localizing the effect to facilitation at a level where word identity matters, such as episodic memory for whole words, solves this problem.

Each facet of an attempt to disentangle style and priming carries substantial caveat. Taken together, though, I argue that the evidence points towards the substantiation of a mostly priming-based analysis of persistence effects. Because there is substantial evidence for the independent existence of both style-shifting and priming, both are likely to be in play to some degree during speech. In the bigger picture, then, the task is to elucidate the difference between these two repetition-favoring mechanisms, to discriminate quantitatively between their effects, and to elaborate a model of the production of variation which makes concrete the contribution and interaction of the two.

Once we do assume a crucial role for priming in corpus observations of persistence, a longstanding question that this dissertation could never have hoped to answer becomes quite relevant: what exactly *is* priming? In chapter 3 I put forward a typological suggestion based on the descriptive, lexically-specific, and temporal properties of what I suggest are different facilitatory effects. I sketched out a possible cognitive mechanism for each of these effects — activation arising from lexical access, implicit learning on grammatical elements stored in abstract memory, fine-grained episodic representations — but the basis of this sketch was merely the suggestion of these mechanisms in previous work on different kinds of priming. There is nothing in this data to indicate that those must be the correct mechanisms at play.

Furthermore, I have made reference at various points to the domain-generality of priming. Priming is far from solely being a psycholinguistic phenomenon, existing in other cognitive domains such as visual object recognition (Kristjánsson and Campana, 2010) and social affect (Bower and Forgas, 2000). A full account of what cognitive mechanisms underlie priming in psycholinguistics must eventually reconcile itself with facilitatory effects in these other types of domains as well. One important direction for this type of work in the future will likely be neuroimaging, but there is yet plenty of room for innovative behavioral work as well. The study of linguistic priming stands to both benefit from and contribute to a domain-general story of priming, and I would suggest that a theory of linguistic priming would benefit from the inclusion of more sociolinguistic variables into the set of phenomena that are studied under the aegis of priming. Almost all of the experimental work that is done on structural priming has been done on very standard and familiar syntactic alternations such as the passive alternation and the ditransitive alternation. A greater diversity of phenomena, and therefore a potentially more discriminatory or revealing or complex set of results, might be achieved by looking to the sociolinguistics literature for variables at different grammatical levels (Sumner and Samuel, 2005) or from nonstandard dialects (Squires, 2013b).

6.3 Directions in the temporal dynamics of variation

I take the position that in order to ultimately make progress both disentangling and reconciling our understanding of different sources of repetitiveness and clustering in speakers' variable behavior, we will need to adopt an approach that goes beyond the critical integration of existing methodologies and paradigms and instead brings new quantitative sophistication to the study of the temporal dynamics of variation within individuals' speech. In advocating for this line of inquiry, I draw an analogy to mid-century views on sociolinguistic variation and the emergence of what is sometimes known as the variationist paradigm.

Prior to the advent of quantitative sociolinguistics in the 1960s, variability in the speech of individuals was considered to be unprincipled, generally relegated to concepts like "optional" or "free variation". Consider, for example, the words of Allan Hubbell describing variable r-fulness in New York City English: "The pronunciation of a very large number of New Yorkers exhibits a pattern in these words that might most accurately be described as the complete absence of any pattern... Such speakers sometimes pronounce /r/ before a consonant or a pause and sometimes omit it, in a thoroughly haphazard fashion... The speaker hears both types of pronunciation about him all the time, both seem almost equally natural to him, and it is a matter of pure chance which one comes to his lips" (1950, 48). The picture here is one of indiscriminate and meaningless coin-flipping. Attitudes such as Hubbell's form the backdrop for Labov's early quantitative sociolinguistic studies in Martha's Vineyard (1972), the Lower East Side(1966), and Harlem (1968), which showed that in the apparent chaos of the variation there was deep, highly-regular systematicity that corresponded with both social and linguistic structure even while simultaneously maintain-ing its probabilistic nature no matter how finely the structural categories were differentiated. This observation of orderly heterogeneity opened up what has now become a half century of productive work refining and expanding our understanding of how social and linguistic structure relates to probabilistic variation.

The suggestion I am making, then, is that the study of sequences of variation in individuals' production is in need of a quantitative revolution comparable to the Labovian revolution at the level of the speech community. Labov himself, this time, both foreshadows and provides the foil for this budding paradigm shift. He compares two stretches of speech from a single Lower East Side speaker, stretches which he characterizes as representing the two dominant styles in the sociolinguistic interview: careful and casual speech. After showing how the rates of different variables track the difference between the two styles and fit into an array based on the attention-to-speech model of style, he looks more closely at the strings of observations in each style (casual: 1 2 2 1 2 2 1 1 1 1 versus careful: 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1) and concludes that "there seems to be no pattern or system within this sequence — yet it fits into the larger pattern shown in the array of styles" (1966, 77). The results of this dissertation make clear that there is plenty to say about patterns and systems within sequences such as these. Even in the small case he presents, for example, we can detect that the 2s and 1s are not distributed evenly and seem instead to show the hallmarks of persistence, with a long stretch of 1s in the middle of the careful sequence, for example.

When we turn to these sequences as meaningful objects of study in their own right, we can identify new types of questions to ask. I suggest two such questions here and illustrate them with simple moving averages (with a window the size of 1/20th of each interview). The data are on TD and DH variation from anonymous individuals in the PNC samples that formed the backbone of this study. First, we open up the possibility that two speakers who end up with a mean rate that is the same might appear to have very different

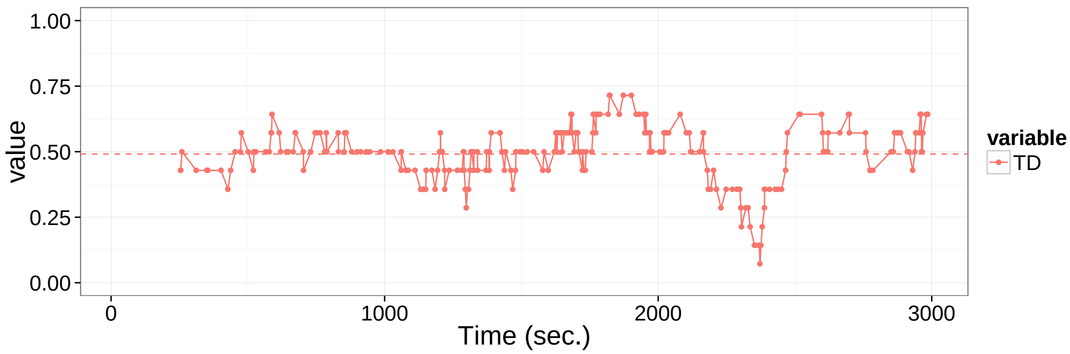

Figure 6.1: Simple moving averages of TD for an individual speaker, low dynamism

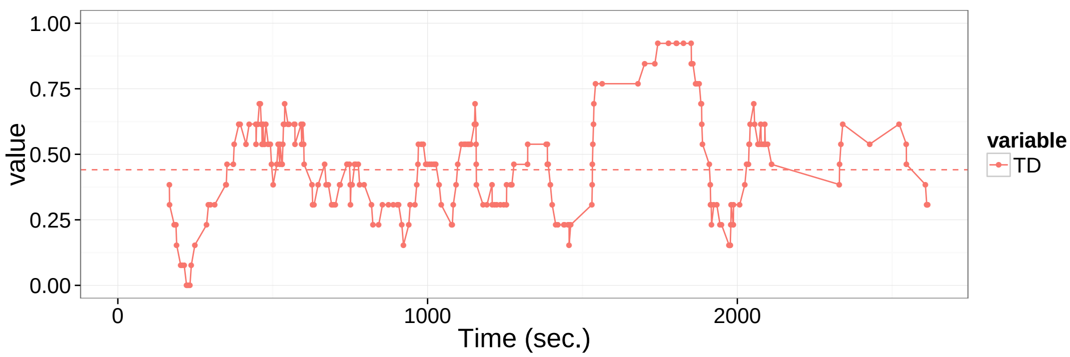

Figure 6.2: Simple moving averages of TD for an individual speaker, high dynamism

degrees of dynamism when their data is viewed in sequence rather than in aggregate. To see what this might look like, compare figures 6.1 and 6.2. Both of these speakers have aggregate means of right around 50% retention for TD. However, the mean in the case of figure 6.1 results from a moving average that hovers close to 50% throughout, indicating that the speaker is distributing his deletion instances somewhat evenly, whereas the same mean in the case of figure 6.2 results from averaging over prolonged periods of high and low deletion rates. A simple first step in quantifying this axis of interspeaker variation (that is, differences between speakers or interviews in how divergent their moving average is from their baseline) would be to subtract the overall mean from the moving average at each observation and then divide by the number of observations, resulting in a mean amount of divergence. The details of this approach will need to be worked out.

The second axis of variation on the level of sequences that I suggest here is the question of what I have previously called microcovariation: the fine-grained correlations across sequences of different variables. As we look at moving averages of different variables, can we see them tracking each other or possibly even repelling each other? Indeed, figure 6.3 illustrates an example of the former possibility: as her DH continuancy rates go up so do her TD retention rates, and as her DH rates go down so do her TD rates. Meanwhile, figure 6.4 presents what appears to be the opposite scenario: increases in DH stopping are paired with decreases in TD deletion, and vice versa. It might be fruitful to think about such patterns using derivatives of the slopes of these moving average curves, as has been done with vowel trajectories recently in, for example, Fruehwald (2013). Or, we might find statistical tools from econometrics, such as autocorrelation, useful. The variation score from Podesva (2007) and the lectal focusing calculations in Sharma and Rampton (2011) are both precursors of this approach that deserve to be expanded upon, refined, and incorporated into a full-fledged model of the temporal dynamics of variation.

I should note here that seems quite likely that the stylistic demands of context, topic, and interlocutor play a major role in shaping the characteristics of individual interviews from the perspectives of both dynamism and microcovariation. But understanding how those stylistic demands shape the unfolding of variability in real time and interact with the more universal influences of memory and processing is in fact the major goal of this line of inquiry. In other words, I do not suggest that previous approaches to style shifting and persistence are on the wrong track, but that this type of questioning can help integrate them and give them an explanatory basis.

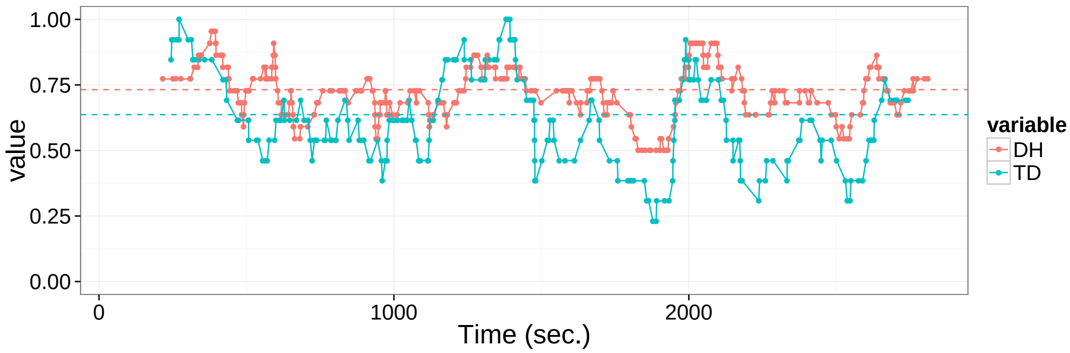

Figure 6.3: Simple moving averages of TD and DH for an individual speaker, convergent sequences

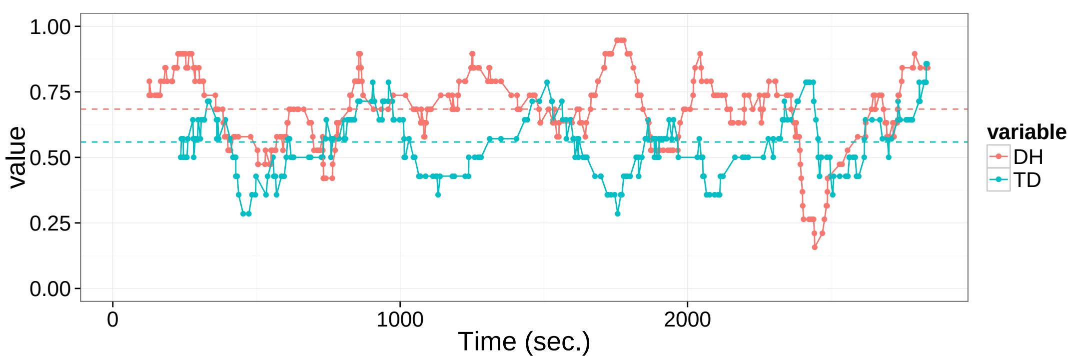

Figure 6.4: Simple moving averages of TD and DH for an individual speaker, divergent sequences

6.4 Variation and the grammatical architecture

Turning my attention to an even larger picture than the study of temporal dynamics, I hope that this dissertation can form a small piece of the puzzle of how we might conceptually and empirically discern the different types of knowledge that speakers bring to the task of producing and perceiving linguistic variation. Atlhough most of the attention thus far in what is beginning to be known as the study of *sociolinguistic cognition* (Campbell-Kibler, 2010; Loudermilk, 2013) has been on the perception side (see (Clopper and Pisoni, 2005; Campbell-Kibler, 2006; Sumner and Samuel, 2009; Babel, 2012; Squires, 2013b)), *inter alia*), there has recently been a small swell of interest in this project on the production side as well.

Preston, for example, sketches a model where linguistic variants can be generated either grammar-internally or from multiple grammars and the social and psycholinguistic factors impacting the probability of a variant's selection exist outside of the grammar(s) proper (2004). His sketch of this model is reproduced in figure 6.5. Preston's model admittedly leaves certain crucial details to future inquiry. It is consistent, though, with MacKenzie's suggestion that we should distinguish language use from grammar (2012) and with Labov et al.'s inference of the existence of a "sociolinguistic monitor which tracks, stores, and processes information on linguistic variation" (2011, 435). Chapter 3 of this dissertation might be thought of as dealing with the manner in which the 'Processing' component of Preston's model acts on the 'Grammar' component, while chapter 4 touched on an interaction between the 'Processing' and 'Sociocultural selection device' components. The case studies I have discussed here in terms of persistence clearly implicate some of these same questions and modules, shoring up the motivation for working out such models but also beginning to fill in some of the details.

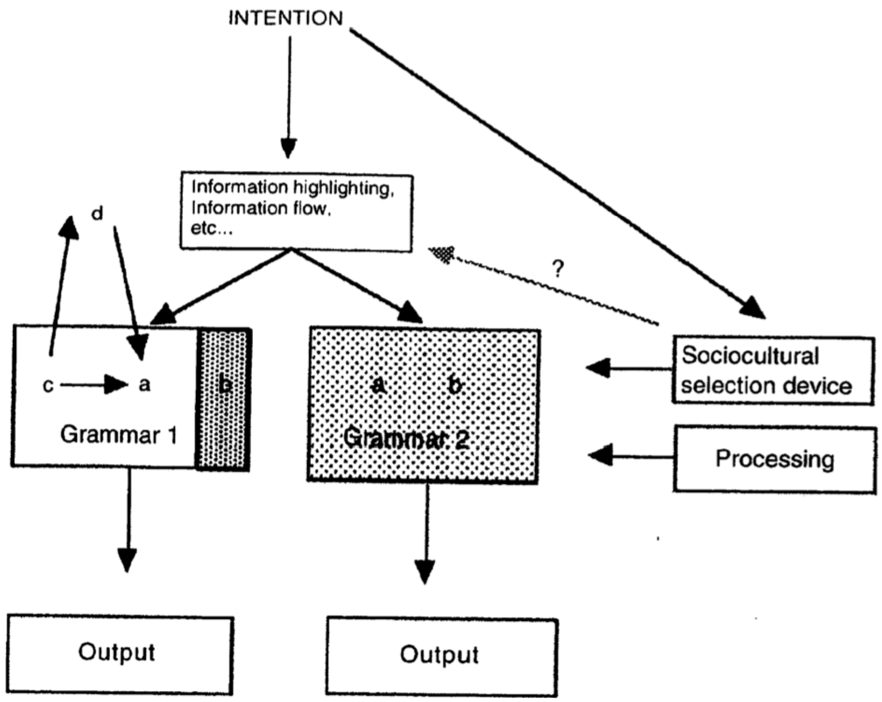

Figure 6.5: Preston's "Elaborated Level III Psycholinguistic Model", reproduced from figure 8.6, (2004, 156).

The evidence that I have presented on the interaction of persistence with variables located at different levels of the grammar suggests that one shortcoming of a model like the one Preston sketches is that it treats all variables in the same way. The linguistic options labeled 'a', 'b', and 'c' in figure 6.5 might just as well be syntactic alternants, allomorphs, variable phonological rules, synonyms, or whatever else — they are put in the same boxes with external forces such as processing and sociocultural selection impinging on them in ways that are presumably consistent. My discussion of the role of memory systems in persistence suggests that grammatical objects stored in abstract memory (in the lexicon) are treated differently by the processing component than more superficial variables are. This observation maps on to a more general distinction between early- and late-locus variables that we propose in Tamminga and MacKenzie (2014).

The suggestion from that paper is that there are two major categories of variables.

Early-locus variables involve the selection of stored grammatical objects, which might mean functional heads as in the case of some syntactic alternations, or allomorphs as in the case of some morphological variables. Late-locus variables, in contrast, involve changes made to the form of the objects that were already selected for derivation in the morphosyntax. The basic division maps relatively clearly onto the already intuitively understood division between morphosyntactic variation on the one hand and phonetic/phonological variation on the other, but I argue that this is a deep division that groups phonetic and phonological variation together even though there may be differences in, for example, their gradience versus categoricalness, and in contrast groups syntactic and morphological variation together. That is, it is a suggestion that the highest-level division in a typology of variables should be the dividing line between morphology and phonology, and that the content of this division is whether variable involves the selection or manipulation of grammatical objects. This is a proposal that presumes a certain theoretical approach to the behavior of alternation and competition in grammatical architectures; see Embick (2008) for a more elaborated description of the type of assumptions that make early/late distinction proposed here coherent.

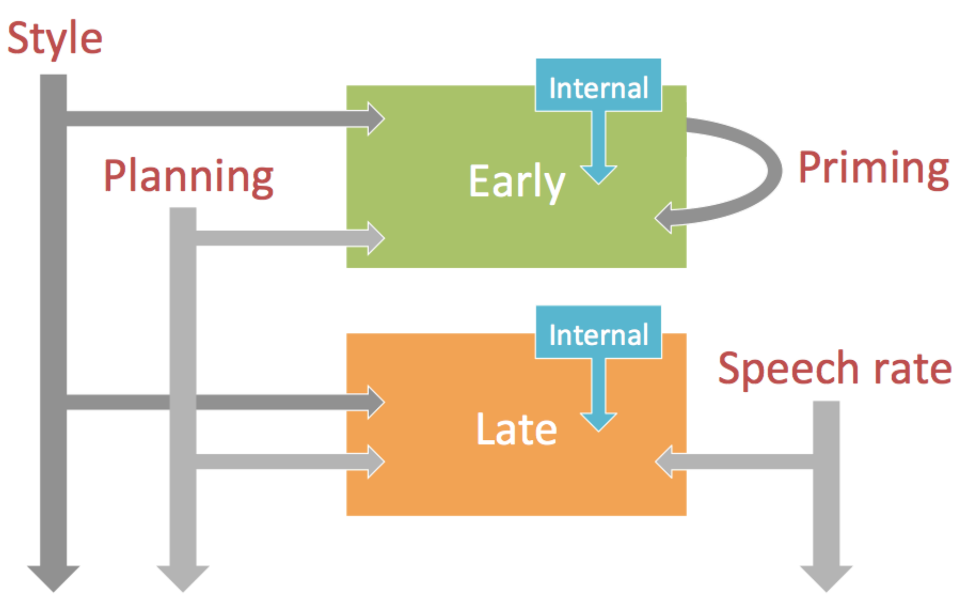

Figure 6.6: Proposed interaction of early/late-locus variables and extra grammatical forces from Tamminga and MacKenzie (2014).

This suggestion is motivated by the desire to capture a prediction that these two types of variables can diverge in their interaction with what has been called the extragrammatical conditioning of variation (MacKenzie and Tamminga, 2013). While arguments for the inclusion of variability into the mechanisms of the productive grammar itself have focused on the ways in which the conditioning of variation resembles probabilistic versions of categorical linguistic rules (Guy and Boberg, 1997; Guy, 1997), there are also ways in which variation is conditioned by factors that are almost certainly not internal to the grammar. Priming is an excellent example in light of its long-distance effects and domain-generality. The other examples of extragrammatical factors conditioning variation that we discuss in Tamminga and MacKenzie (2014) are style, speech rate, and production planning. Our suggestion in that paper, as seen in figure 6.6, was that priming can affect only early-locus variation because it requires a grammatical object to be activated. Although here I have suggested that this view can be refined to reflect different facilitatory mechanisms under the umbrella term "persistence", the scenario that I suggested in chapter 4 still maps well onto the early/late-locus distinction.

In this discussion of extragrammatical effects on variable linguistic behavior, I would emphasize that, in Kiparsky's charming formulation, "I am *not* asking whether [variable rule application] should be shoved under some rug where linguists put things that do not interest them" (1972, 223). Rather, my view is that the susceptibility of variation to non-grammatical influences, at the same time as it is a phenomenon that is entirely indebted to the products of the grammar, frees variable phenomena to reflect on the interaction between grammar and language use in ways that are impossible by grammar-internal properties alone.

6.5 Conclusion

The issues I have discussed in this chapter represent different directions for the extension of the relevance of priming to various bigger-picture goals. In one view, persistence represents the naturalistic reflection of priming, a critical piece of evidence for how the human mind processes language. Continued work on persistence *qua* priming thus stands to improve our general understanding of language processing by offering new variables, variables that straddle interfaces, and socially-sensitive nonstandard variables to the set of the usual suspects in the structural priming literature. It also is key for demonstrating that the controlled experimental results on priming match up with speakers' behavior in the real world, lending ecological validity to the use of experimental techniques to study language as it would be naturally used in context.

Another view is that persistence is part of the general temporal dynamics of language variation, a newly-broadened area of research that offers the promise of beginning to clarify the cognitive underpinnings of language variation specifically, asking how speakers are able to produce linguistic variation that gives rise to stratified speech communities. This view of persistence can then be paired with continued improvement of our understanding of how variation interfaces with a range of other social and cognitive factors, such as style-shifting, self-monitoring, speech rate, articulatory demands, and production planning.

A final point to be made about each of these views and approaches, as well as about the full set of results from this dissertation, is that surface frequencies of variant use are not themselves an object of study. Wherever we turn, no matter which variable we look at, the variability that we find attested in language use turns out to be rooted in multiple systems and origins. A fully elaborated model of language variation will ultimately require continuous attention to the multi-layered nature of apparently unified surface variation, as well as the influence of more general cognitive forces in shaping variation.

Appendix A

	Estimate	Std. Error	z value	p-value
(Intercept)	-1.43	0.59	-2.42	0.016
Voiceless	0.18	0.25	0.71	0.479
Following fricative	0.77	0.27	2.83	0.005
Following glide	2.00	0.20	10.18	< 0.001
Following liquid	1.08	0.28	3.87	< 0.001
Following pause	2.00	0.18	11.29	< 0.001
Following vowel	2.45	0.17	14.40	< 0.001
Preceding fricative	0.65	0.46	1.41	0.158
Preceding liquid	-0.03	0.45	-0.06	0.952
Preceding nasal	-0.86	0.44	-1.96	0.050
Preceding sibilant	-0.70	0.45	-1.55	0.120
Preceding stop	0.04	0.46	0.09	0.925
Homovoiced	-0.01	0.30	-0.02	0.982
Rate (log2 V/sec., centered)	-0.33	0.08	-4.09	< 0.001
Regular past target	0.62	0.15	4.05	< 0.001
Deletion prime	-0.20	0.15	-1.36	0.174
LogLag	0.04	0.03	1.44	0.150
Regular target x deletion prime	0.15	0.21	0.69	0.489

Table A.1: Estimated coefficients for TD parameters, with grammatically-mismatched primes and targets

	Estimate	Std. Error	z value	p-value
(Intercept)	-0.18	0.55	-0.33	0.743
Following fricative	0.41	0.27	1.53	0.127
Following glide	1.41	0.16	8.63	< 0.001
Following liquid	0.27	0.28	0.97	0.333
Following pause	2.14	0.15	14.26	< 0.001
Following vowel	2.35	0.15	15.66	< 0.001
Preceding liquid	-0.63	0.54	-1.19	0.236
Preceding nasal	-1.31	0.53	-2.50	0.012
Preceding sibilant	-0.99	0.52	-1.88	0.060
Preceding stop	-0.07	0.55	-0.12	0.901
Homovoiced	-0.94	0.12	-7.76	< 0.001
Rate (log2 V/sec., centered)	-0.26	0.06	-4.09	< 0.001
Deletion prime	-0.07	0.13	-0.50	0.614
Same word	0.56	0.21	2.62	0.009
Lag (log2 sec.)	-0.00	0.05	-0.00	0.997
Deletion prime x same word	-1.31	0.29	-4.49	< 0.001
Deletion prime x lag	0.05	0.06	0.86	0.389
Same word x lag	-0.04	0.12	-0.33	0.742
Deletion prime x same word x lag	0.12	0.16	0.78	0.434

Table A.2: Estimated coefficients for TD parameters, with monomorphemic primes and targets

	Estimate	Std. Error	z value	p-value
(Intercept)	-3.85	0.82	-4.68	< 0.011
Voiceless	0.61	0.33	1.81	0.070
Following fricative	1.54	0.50	3.09	0.002
Following glide	1.93	0.37	5.27	< 0.011
Following liquid	1.00	0.56	1.80	0.072
Following pause	2.81	0.38	7.49	< 0.011
Following vowel	2.76	0.33	8.47	< 0.011
Preceding fricative	1.50	0.52	2.88	0.004
Preceding liquid	1.19	0.56	2.12	0.034
Preceding nasal	0.84	0.54	1.55	0.120
Preceding sibilant	0.11	0.52	0.22	0.828
Preceding stop	1.21	0.56	2.14	0.032
Homovoiced	1.92	0.48	3.99	< 0.011
Rate (log2 V/sec., centered)	-0.27	0.15	-1.78	0.074
Deletion prime	-0.80	0.42	-1.89	0.059
Same word	0.16	0.39	0.40	0.688
Lag (log2 sec.)	-0.11	0.11	-1.02	0.306
Deletion prime x same word	-1.11	0.63	-1.77	0.077
Deletion prime x lag	0.40	0.20	1.96	0.050
Same word x lag	0.18	0.20	0.91	0.361
Deletion prime x same word x lag	-0.19	0.33	-0.57	0.567

Table A.3: Estimated coefficients for TD parameters, with regular past primes and targets

	Estimate	Std. Error	z value	p-value
(Intercept)	0.23	0.31	0.76	0.450
Preceding coronal obstruent	-0.50	0.22	-2.22	0.026
Other preceding segment	-1.41	0.21	-6.59	< 0.001
Preceding velar nasal	-1.07	0.49	-2.19	0.028
Following pause	0.45	0.11	4.13	< 0.001
Age (centered)	-0.01	0.01	-0.85	0.396
Male speaker	-1.18	0.33	-3.57	< 0.001
Log of speech rate (centered)	-0.30	0.06	-4.70	< 0.001
/ing/ prime	1.11	0.13	8.44	< 0.001
Gerund target	0.89	0.15	5.78	< 0.001
Monomorph. target	2.37	0.35	6.67	< 0.001
Root-attached target	3.07	0.23	13.11	< 0.001
Quantifier target	-0.72	0.20	-3.58	< 0.001
/ing/ prime x gerund target	-0.38	0.25	-1.49	0.137
/ing/ prime x monomorph. target	-2.11	0.60	-3.50	< 0.001
/ing/ prime x root-attached target	-1.12	0.51	-2.21	0.027
/ing/ prime x quantifier target	0.06	0.31	0.18	0.856

Table A.4: Estimated coefficients for ING parameters, with progressive primes across target grammatical categories (N=4238)

	Estimate	Std. Error	z value	p-value
(Intercept)	2.00	0.34	5.870	< 0.001
Preceding coronal obstruent	-0.63	0.21	-2.995	0.003
Other preceding segment	-1.27	0.20	-6.301	< 0.001
Preceding velar nasal	-1.62	0.43	-3.786	< 0.001
Following pause	0.53	0.11	4.994	< 0.001
Following velar consonant	0.16	0.17	0.936	0.349
Age (centered)	0.00	0.01	0.199	0.842
Male speaker	-0.91	0.36	-2.551	0.011
Progressive target	-0.59	0.09	-6.221	< 0.001
Rate (log2 V/sec., centered)	-0.35	0.06	-5.516	< 0.001
/in/ prime	-1.59	0.19	-8.326	< 0.001
Same word	0.72	0.33	2.154	0.031
Lag (log2 sec.)	-0.08	0.03	-2.623	0.009
/in/ prime x same word	-1.28	0.47	-2.731	0.006
/ing/ prime x lag	0.19	0.04	4.952	< 0.001
Same word x lag	-0.06	0.09	-0.681	0.496
/in/ prime x same word x lag	0.08	0.14	0.575	0.565

Table A.5: Estimated coefficients for ING parameters, with progressive primes across target grammatical categories (N=4238)

	Estimate	Std. Error	z value	p-value
(Intercept)	2.28	0.18	12.81	< 0.001
Male speaker	-0.56	0.23	-2.39	0.017
Age (centered)	-0.01	0.01	-2.25	0.025
Preceding pause	-0.76	0.07	-10.48	< 0.001
Preceding vowel	0.03	0.09	0.28	0.777
Rate (log2 V/sec., centered)	0.00	0.04	0.03	0.978
Stop prime	-0.45	0.12	-3.82	< 0.001
Same word	0.25	0.11	2.37	0.018
Lag (log2 sec.)	-0.04	0.03	-1.42	0.155
Stop prime x same word	-0.68	0.19	-3.61	< 0.001
Stop prime x lag	0.09	0.05	1.97	0.049
Same word x lag	-0.01	0.05	-0.25	0.804
Stop prime x same word x lag	0.04	0.08	0.49	0.623

Table A.6: Estimated coefficients for DH-stop parameters (N=7462)

	Estimate	Std. Error	z value	p-value
(Intercept)	2.55	0.16	16.01	< 0.001
Male speaker	-0.85	0.20	-4.19	< 0.001
Age (centered)	-0.01	0.01	-1.21	0.228
Preceding pause	0.93	0.10	9.66	< 0.001
Preceding vowel	1.40	0.15	9.62	< 0.001
Rate (log2 V/sec., centered)	0.11	0.05	2.21	0.027
Null prime	-0.09	0.17	-0.53	0.598
Same word	0.44	0.15	2.98	0.003
Lag (log2 sec.)	-0.09	0.03	-3.59	< 0.001
Null prime x same word	-2.08	0.29	-7.18	< 0.001
Null prime x lag	0.04	0.07	0.65	0.514
Same word x lag	-0.10	0.06	-1.65	0.100
Null prime x same word x lag	0.14	0.13	1.14	0.255

Table A.7: Estimated coefficients for DH-null parameters (N=8409)

Bibliography

Abramowicz, L. (2007). Sociolinguistics meets Exemplar Theory: Frequency and recency effects in (ing). *University of Pennsylvania Working Papers in Linguistics*, 13:27–37.

Allen, M. and Badecker, W. (2002). Inflection regularity: probing the nature of lexical representation in a cross-model priming task. *Journal of Memory and Language*, 46:705–722.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40:177–189.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68:255–278.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18:355–387.

Bock, K. and Griffin, Z. M. (2000). The persistence of structural priming: transient activation or implicit learning? *Journal of Experimental Psychology*, 129(2):177–192.

Bock, K. and Loebell, H. (1990). Framing sentences. *Cognition*, 35:1–39.

Bower, G. H. and Forgas, J. P. (2000). Affect, memory, and social cognition. In Eich, E., Kihlstrom, J. F., Bower, G. H., Forgas, J. P., and Niedenthal, P., editors, *Cognition and emotion*, pages 87–168. Oxford University Press.

Branigan, H., Pickering, M., and Cleland, A. (1999). Syntactic priming in written production: evidence for rapid decay. *Psychonomic Bulletin and Review*, 6(4):635–640.

Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., and Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24(6):489–506.

Branigan, H. P., Pickering, M. J., Stewart, A. J., and McLean, J. F. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory and Cognition*, 28(8):1297–1302.

Cameron, R. (1993). Ambiguous agreement, functional compensation, and nonspecific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change*, 5:305–334.

Cameron, R. and Flores-Ferrán, N. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context*, 1(1):41–65.

Campbell-Kibler, K. (2006). *Listener perceptions of sociolinguistic variables: The case of (ING)*. PhD thesis, Stanford University.

Campbell-Kibler, K. (2010). New directions in sociolinguistic cognition. *University of Pennsylvania Working Papers in Linguistics*, 15(2):31–39.

Campbell-Kibler, K. (2011). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change*, 22:423–441.

Chang, F., Dell, G. S., Bock, K., and Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29(2):217–229.

Clark, L. (2014). Phonological repetition effects in natural conversation: evidence from TH-fronting in Fife. In Lawson, R., editor, *Sociolinguistics in Scotland*, chapter 8. Palgrave Macmillan.

Clopper, C. G. and Pisoni, D. B. (2005). Perception of dialect variation. In Pisoni, D. B. and Remez, R., editors, *The Handbook of Speech Perception*, chapter 13, pages 313–337. Blackwell.

Cofer, T. (1972). *Linguistic variability in a Philadelphia speech community*. PhD thesis, University of Pennsylvania.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93:283–321.

Deutsch, A., Frost, R., and Forster, K. I. (1998). Verbs and nouns are organized and accessed differently in the mental lexicon: evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24(5):1238–1255.

Ecay, A. (2012). Three-way competition and the emergence of do-support in English. Paper presented at Diachronic and Generative Syntax 14, University of Lisbon, July 4.

Embick, D. (2004). On the structure of resultative participles in English. *Linguistic Inquiry*, 35(3):355–392.

Embick, D. (2008). Variation and morphosyntactic theory: Competition fractionated. *Language and Linguistics Compass*, 2/1:59–78.

Estival, D. (1985). Syntactic priming of the passive in English. *Text*, 5(1/2):7–21.

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language*, 48:379–398.

Fischer, J. (1958). Social influence of a linguistic variant. *Word*, 14:47–56.

Forster, K. I. and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10(4):680–698.

Fruehwald, J. (2012). Redevelopment of a morphological class. *University of Pennsylvania Working Papers in Linguistics*, 18(1):77–86.

Fruehwald, J. (2013). *The phonological influence on phonetic change*. PhD thesis, University of Pennsylvania.

Gallistel, C. (2009). The importance of proving the null. *Psychological Review*, 116(2):439–453.

Gries, S. T. (2005). Syntactic priming: a corpus-based approach. *Journal of Psycholinguistic Research*, 34(4):365–399.

Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In Labov, W., editor, *Locating language in time and space*, chapter 1, pages 1–36. Academic Press, New York.

Guy, G. R. (1997). Competence, performance, and the generative grammar of variation. In Hinskens, F., van Hout, R., and Wetzels, W. L., editors, *Variation, change, and phonological theory*, pages 125–144. John Benjamins Publishing Co.

Guy, G. R. (2007). Lexical exceptions in variable phonology. *University of Pennsylvania Working Papers in Linguistics*, 13(2):109–119.

Guy, G. R. and Boberg, C. (1997). Inherent variability and the obligatory contour principle. *Language Variation and Change*, 9:149–164.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroek, S., and Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, 58:214–238.

Hazen, K. (2011). Flying high above the social radar: Coronal stop deletion in modern Appalachia. *Language Variation and Change*, 23:105–137.

Hernández-Campoy, J. M. and Trudgill, P. (2002). Functional compensation and southern peninsular Spanish /s/ loss. *Folia Linguistica Historica*, 13:31–57.

Houston, A. (1985). *Continuity and change in English morphology: the variable (ING)*. PhD thesis, University of Pennsylvania.

Hubbell, A. (1950). *The Pronunciation of English in New York City: Consonants and Vowels*. King's Crown Press.

Jaeger, T. F. and Snider, N. (2007). Implicit learning and syntactic persistence: surprisal and cumulativity. *University of Rochester Working Papers in the Language Sciences*, 3(1):26–44.

Jaeger, T. F. and Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127:57–83.

Johnstone, B. (1999). Lingual biography and linguistic variation. *Language Sciences*, 21:313–321.

Johnstone, B. (2001). The individual. In Duranti, A., editor, *Key terms in language and culture*, pages 122–125. Blackwell.

Johnstone, B. (2009). Stance, style, and the linguistic individual. In Jaffe, A., editor, *Stance: sociolinguistic perspectives*, chapter 2, pages 29–52. Oxford University Press.

Kaschak, M. P. (2007). Long-term structural priming affects subsequent patterns of language production. *Memory and Cognition*, 35(5):925–937.

Kaschak, M. P., Kutta, T. J., and Jones, J. L. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin and Review*, 18:1133–1139.

Kiparsky, P. (1972). Explanation in phonology. In Peters, S., editor, *Goals of Linguistic Theory*, pages 189–227. Prentice-Hall, Cinnaminson, NJ.

Konopka, A. E. and Bock, K. (2005). Helping syntax out: How much do words do? Paper presented at the eighteenth annual CUNY conference on human sentence processing. Tucson, AZ.

Kouider, S. and Dupoux, E. (2009). Episodic accessibility and morphological processing: Evidence from long-term auditory priming. *Acta Psychologica*, 130:38–47.

Kristjánsson, A. and Campana, G. (2010). Where perception meets memory: a review of repetition priming in visual search tasks. *Atten Percept Psychophys*, 72(1):5–18.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (1978). Where does the linguistic variable stop? a response to beatriz lavandera. *Working Papers in Sociolinguistics*, 44:6–23.

Labov, W. (1984). Field methods of the project on Linguistic Change and Variation. In Baugh, J. and Sherzer, J., editors, *Language in use: readings in sociolinguistics*, pages 28–53. Prentice-Hall.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1:85–97.

Labov, W. (2001a). The anatomy of style-shifting. In Eckert, P. and Rickford, J. R., editors, *Style and Sociolinguistic Variation*, chapter 5, pages 85–108. Cambridge University Press.

Labov, W. (2001b). Applying our knowledge of African-American English to the problem of raising reading levels in inner-city schools. In Lanehart, S., editor, *Sociocultural and historical contexts of African American English*, pages 299–330. John Benjamins.

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

Labov, W. (2004). Quantitative analysis of linguistic variation. In Ammon, U., Dittmar, N., and Mattheier, K. J., editors, *Sociolinguistics/soziolinguistik: An international handbook of the science of language and society*, volume 1, pages 6–21. de Gruyter, Berlin.

Labov, W. (2006 [1966]). *The social stratification of English in New York City*. Cambridge University Press.

Labov, W. (2012). What is to be learned: the community as the focus of social cognition. *Review of Cognitive Linguistics*, 10(2):265–293.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., and Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4):431–463.

Labov, W., Cohen, P., Robins, C., and Lewis, J. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Final report, Cooperative Research Project 3288, Vols. I and II.

Labov, W. and Rosenfelder, I. (2011). The Philadelphia Neighborhood Corpus of LING 560 studies, 1972-2010. With support of NSF contract 921643.

Lavandera, B. R. (1978). Where does the sociolinguistic variable stop? *Language in Society*, 7(2):171–182.

Leu, T. (2008). *The internal syntax of determiners*. PhD thesis, New York University.

Loudermilk, B. C. (2013). Psycholinguistic approaches. In Bayley, R., Cameron, R., and Lucas, C., editors, *The Oxford Handbook of Sociolinguistics*. Oxford University Press.

MacKenzie, L. (2012). *Locating variation above the phonology*. PhD thesis, University of Pennsylvania.

MacKenzie, L. and Tamminga, M. (2013). Two case studies on the non-local conditioning of variation. Paper presented at the Linguistics Society of America annual meeting, Jan. 3.

Marslen-Wilson, W., Ford, M., Older, L., and Zhou, X. (1996). The combinatorial lexicon: priming derivational affixes. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society: July 12-15, 1996, University of California, San Diego*, volume 18, pages 223–227. Psychology Press.

Neu, H. (1980). Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In Labov, W., editor, *Locating language in time and space*, chapter 2, pages 37–54. Academic Press, New York.

Patrick, P. L. (1991). Creoles at the intersection of variable processes: (td)-deletion and past-marking in the Jamaican mesolect. *Language Variation and Change*, 3(2):171–189.

Pickering and Branigan (1999). Syntactic priming in language production. *Trends Cogn Sci*, 3(4):136–141.

Pickering, M. J. and Ferreira, V. S. (2008). Structural priming: A critical review. *Psycholinguistic Bulletin*, 134(3):427–459.

Podesva, R. J. (2007). Three sources of stylistic meaning. In *Proceedings of SALSA 2007*.

Poplack, S. (1980). Deletion and disambiguation in Puerto Rican Spanish. *Language*, 56(2):371–385.

Poplack, S. (1984). Variable concord and sentential plural marking in Puerto Rican Spanish. *Hispanic Review*, 52(2):205–222.

Poplack, S. and Tagliamonte, S. (1989). There's no tense like the present: Verbal -s inflection in early Black English. *Language Variation and Change*, 1:47–84.

Preston, D. (2004). Three kinds of sociolinguistics: A psycholinguistic perspective. In Fought, C., editor, *Sociolinguistic variation: critical reflections*, chapter 8. Oxford University Press.

Rosenfelder, I., Fruehwald, J., Evanini, K., and Yuan, J. (2011). *FAVE Program Suite [Forced alignment and vowel extraction]*.

Sankoff, D. and Laberge, S. (1978). Statistical dependence among successive occurrences of a variable in discourse. In Sankoff, D., editor, *Linguistic variation: models and methods*, chapter 8, pages 119–126. Academic Press.

Scarborough, D. L., Gerard, L., and Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory and Cognition*, 7(1):3–12.

Scherre, M. (2001). Phrase-level parallelism effect on noun phrase number agreement. *Language Variation and Change*, 13(01):91–107.

Scherre, M. and Naro, A. (1991). Marking in discourse: "Birds of a feather". *Language Variation and Change*, 3:23–32.

Scherre, M. and Naro, A. (1992). The serial effect on internal and external variables. *Language Variation and Change*, 4(01):1–13.

Sharma, D. and Rampton, B. (2011). Lectal focusing in interaction: A new methodology for the study of superdiverse speech. *Queen Mary's Occasional Papers Advancing Linguistics*, 2:1–22.

Snider, N. (2008). *An exemplar model of syntactic priming*. PhD thesis, Stanford University.

Squires, L. (2013a). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics*, 17(2):200–237.

Squires, L. (2013b). Talker specificity and the perception of grammatical variation. *Language, cognition, and neuroscience*.

Sumner, M. and Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52:322–338.

Sumner, M. and Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60:487–501.

Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-linguistic analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1(1):113–149.

Szmrecsanyi, B. (2006). *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Mouton de Gruyter.

Tagliamonte, S. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.

Tamminga, M. and Ecay, A. (2014). Priming effects in language change as diagnostics of grammatical structure. Paper presented at the Penn Linguistics Conference, March 29, University of Pennsylvania.

Tamminga, M. and MacKenzie, L. (2014). Elaborating extragrammatical effects on variation. Paper presented at the Linguistics Society of America annual meeting, Minneapolis, Jan. 3.

Tamminga, M. and McLaughlin, B. (2013). The sensitivity of persistence to subject animacy in AAVE third singular /s/. Paper presented at UK Language Variation and Change 9, University of Sheffield, Sept. 3.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review*, 2(3):339–363.

VanWagenen, S. (2005). The morphologically organized mental lexicon: further experimental evidence. Master's thesis, UCLA.

Versace, R. and Nevers, B. (2003). Word frequency effect on repetition priming as a function of prime duration and delay between the prime and target. *British Journal of Psychology*, 94:389–408.

Weiner, E. J. and Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics*, 19(1):29–58.

Weinreich, U., Labov, W., and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In Lehmann, W. and Malkiel, Y., editors, *Directions for historical linguistics: A symposium*, pages 97–195. University of Texas Press, Austin and London.

Wheeldon, L. and Smith, M. (2003). Phrase structure priming: a short-lived effect. *Language and cognitive processes*, 18(4):431–442.

Wolfram, W. (1975). Variable constraints and rule relations. In Fasold, R. W. and Shuy, R. W., editors, *Analyzing variation in language: papers from the second colloquium on New Ways of Analyzing Variation*, pages 70–88. Georgetown University Press.