UC San Diego
UC San Diego Electronic Theses and Dissertations

Title
Contextual and morphological effects in speech production

Permalink
https://escholarship.org/uc/item/5r9677hw

Author
Seyfarth, Scott James

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Contextual and morphological effects in speech production

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Linguistics

by

Scott James Seyfarth

Committee in Charge:

Professor Farrell Ackerman, Co-Chair
Professor Marc Garellek, Co-Chair
Professor Gabriela Caballero
Professor Sarah Creel
Professor Robert Malouf

2016

The Dissertation of Scott James Seyfarth is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Co-Chair

University of California, San Diego

2016

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

*** *

Chapter 2, in full, is a reprint of the material as it appears in Seyfarth (2014) [Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133 (1), 140–155. doi:10.1016/j.cognition.2014.06.013]. This work was reported, in part, at the 19th Architectures and Mechanisms for Language Processing conference.

Chapter 3, in full, is a reprint of the material as it appears in Seyfarth, Buz, and Jaeger (2016) [Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustical Society of America*, 139 (2), EL31–37. doi:10.1121/1.4942544]. This work was reported, in part, at the 28th Annual CUNY Conference on Human Sentence Processing. This chapter reports co-authored work, which is used here with permission from the co-authors. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, has been submitted for publication as Seyfarth, Garellek, Gillingham, Ackerman, and Malouf [Acoustic differences in morphologically-distinct homophones]. This work was reported, in part, at the 3rd American International Morphology Meeting. This chapter reports co-authored work, which is used here with permission from the co-authors. The dissertation author was the primary investigator and author of this paper.

2010	Bachelor of Arts, Linguistics
	Northwestern University
2010	Bachelor of Music, Musicology
	Northwestern University
2010–2011	Linguistics Research Center Foundation Fellow
	University of California, Santa Cruz
2016	Doctor of Philosophy, Linguistics
	University of California, San Diego

PUBLICATIONS

Garellek, M., & Seyfarth, S. (2016). Acoustic differences between English /t/ glottalization and phrasal creak. *Interspeech 2016*. San Francisco, CA.

Seyfarth, S., Buz, E., & Jaeger, T.F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustical Society of America*, 139 (2), EL31–37. doi:10.1121/1.4942544

Seyfarth, S. & Garellek, M. (2015). Coda glottalization in American English. In the Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow.

Malouf, R., Ackerman, F., & Seyfarth, S. (2015). Explaining the number hierarchy. In D. Noelle, R. Dale, A. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, & P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1500–1505). Austin, TX: Cognitive Science Society.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133 (1), 140–155. doi:10.1016/j.cognition.2014.06.013

Seyfarth, S. & Myslín, M. (2014). Discriminative learning predicts human recognition of English blend sources. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2064–2069). Austin, TX: Cognitive Science Society.

Seyfarth, S., Ackerman, F., & Malouf, R. (2014). Implicative organization facilitates morphological learning. In H. Leung, Z. O'Hagan, S. Bakst, A. Lutzross, J. Manker, N. Rolle, & K. Sardinha (Eds.), *Proceedings of the 40th Annual Meeting of the Berkeley Linguistics Society* (pp. 479–493). Berkeley, CA.

Rohde, H., Seyfarth, S., Clark, B., Jäger, G., & Kaufmann, S. (2012). Communicating with cost-based implicature: A game-theoretic approach to ambiguity. In S. Brown-Schmidt, J. Ginzburg, & S. Larsson (Eds.), *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 107–116). Paris, France.

ABSTRACT OF THE DISSERTATION

Contextual and morphological effects in speech production

by

Scott James Seyfarth

Doctor of Philosophy in Linguistics

University of California, San Diego, 2016

Farrell Ackerman, Co-Chair
Marc Garellek, Co-Chair

A native speaker knows how to produce an unlimited number of words and possible words in their language, which are related to each other by various semantic, phonological, and morphological properties. The production of an intended word is not independent of these relations: similar or associated words may influence speech processing and articulation in the short-term, and the cumulative effects may change the phonological forms of words in the long-term. Here, I present three studies investigating how such relations affect the acoustic durations of words, and how that interacts with phonological representation.

It is well-known that if a word can be easily predicted based on the surrounding words, it is likely to be produced in a shortened form. In the first study, I show that some words almost always occur in contexts where they can be easily

predicted, and are thus almost always produced in a shortened form. Through two corpus experiments, I demonstrate that these lexical patterns affect the phonological forms of words over the long-term: words that are typically predictable on the basis of nearby words become permanently shortened, and are produced with a shortened form even when they are not predictable.

In the second study, I examine how competition with phonologically-similar words affects hyperarticulation strategies. In a web-based experiment, I find that when speakers need to communicate a word clearly, they may actually reduce the duration of parts of that word when doing so would increase the acoustic contrast with a contextually-relevant minimal-pair competitor. The result provides evidence about how speakers use implicit knowledge of lexical contrasts in hyperarticulation.

In the final study, I investigate a mechanism through which paradigms of morphologically-related words might interact with each other in speech production. Phonetic paradigm uniformity proposes that the articulation of a word is influenced by the articulatory plans of morphologically-related words. In a laboratory experiment, I demonstrate that there is evidence for this influence based on the durations of morphologically-distinct homophones like FREES and FREEZE. Over the long-term, these patterns may help explain categorical changes in phonological structure.

Chapter 1

Introduction

1.1 Questions

A native speaker knows how to produce an unlimited number of words and possible words in their language. Existing words are related on various semantic, phonological, and morphological dimensions (among others), and new words may draw on these relationships. For example, the blend *Brexit* draws on phonological and semantic relationships with *British, Britain*, and *exit*, as well as the earlier blend *Grexit*.[1] The derived form *Brexiteer* takes advantage of a morphological pattern (*-eer*) inferred from the relation between words like *auctioneer, profiteer*, and *musketeer*. The meaning of *Brexiteer* can be inferred both from this pattern, and because of the word's likely co-occurrence with the name of a politician who has lobbied in favor of a Brexit. Further, a person hearing *Brexit* for the first time can articulate it immediately, by using knowledge of how sequences of articulatory gestures map to acoustics for the words in their language. These gestures are shared among words: the sequence /bɹ/ in *Brexit* is roughly the same as the one used in *bread* and *breath*, and it is easy for a speaker to adapt that sequence to the onset of a new word.

Words are thus not independent of each other, and knowing the patterns that they participate in is essential to knowing a language and producing meaningful utterances. How is the production of each intended word affected by access

---

[1]*Brexit* is the exit of the United Kingdom from the European Union, and *Grexit* is the exit of Greece from the eurozone.

1

to a lexicon of related words? In particular, how do the relations and similarities among words influence each word's ultimate articulation?

1.2 Outline of the dissertation

Lexical relations fall into two broad categories:

(1) Contextual or *syntagmatic* relations: the relations between words that occur in proximity to each other in an utterance, such as due to the fact they occur together in a fixed expression (*human, rights*) or because they often appear in the same discourse (*tire, wheel*) or the same phrase (*with, us*);

(2) *Paradigmatic* relations: the relations between words that share material and participate in some systematic contrast, such as a morphological contrast (*frees, freed*) or a phonological one (*dose, doze*)

This dissertation presents three studies on how syntagmatic and paradigmatic relations affect speech production. The specific focus of this work is on word and segment duration: how are the temporal properties of words influenced by these relations? Chapter 2 examines how the inter-predictability of words in the lexicon may permanently change phonological representations. Chapter 3 investigates how competition from minimal-pair words guides hyperarticulation in speech. Chapter 4 tests a way in which paradigms of morphologically-related words might interact with each other during articulation. Finally, the general discussion in Chapter 5 summarizes the main findings, and explores how they may contribute to a broader understanding of phonological representation and change in the long-term.

1.3 Place in the literature

1.3.1 Syntagmatic relations

A large subfield of psycholinguistic research investigates how related words are co-activated during the production of an intended target word (e.g., Dell, 1986; Rapp & Goldrick, 2000; Vitevitch, 2002; Goldrick & Blumstein, 2006; Bell, Brenier, Gregory, Girand & Jurafsky, 2009; Goldrick, Folk & Rapp, 2010). With regard to contextual relations, a major focus involves how the probability and availability of upcoming linguistic material affects the phonetic realization of that material. As one instance of this phenomenon, it is well-known that if a word is likely to occur given a particular context, it will be produced more quickly and with less prominence (*probabilistic reduction*; Lieberman, 1963; Jurafsky, Bell, Gregory & Raymond, 2001; Bell et al., 2009, and see §2.1.1). For example, the word *nine* is very likely to occur as the last word of the utterance in (3a), below. However, it is less likely to occur in the context of (3b)—in this context (absent other information), *nine* is no more probable than any other number word.

(3) a. A stitch in time saves *nine*.

b. The number that you will hear is *nine*.

In the predictable context of (3a), the word *nine* is pronounced in a reduced form that may include a shorter duration and reduced peak amplitude relative to the same word in (3b) (Lieberman, 1963). There are several compatible explanations for this phenomenon. Under one such proposal, pre-activation or priming of an intended word occurs in the context of related words, which facilitates access and speeds articulatory production (Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon

& Newlands, 2000; Bell et al., 2009, among others; see Jaeger & Buz, 2016 for a critical review of mechanisms).

These psycholinguistic relationships have consequences for linguistic structure. Usage-based models of linguistics investigate how syntagmatic lexical patterns, such as collocational or co-occurrence frequency, contribute to re-analysis or grammaticalization (Bybee, 2002, 2003, 2006; Dabrowska, 2008; Torres Cacoullos & Walker, 2009; Boye & Harder, 2012; Brown & Rivas, 2012, see also Gahl & Garnsey, 2004, 2006). For example, the frequent co-occurrence of the words *I don't know* causes it to be sometimes reduced to *I dunno*. This form has been re-analyzed as a pragmatic hedge expression [aɪɾənoʊ], with a flap that is acceptable in this form but is otherwise unusual in word-initial position (Bybee & Scheibman, 1999).

The studies in Chapter 2 explore the consequences of syntagmatic probability-driven reduction on phonological forms, beyond the role of specific collocations. The proposal is that frequent probabilistic reduction accumulates in memory, which leads to permanent reduction of words that typically occur in high-probability lexical contexts. This finding may contribute to an explanation of the phenomenon in which the most predictable words have the fewest speech segments (Piantadosi, Tily & Gibson, 2011). The research fits within a growing body of work that highlights the role of inter-word usage context in understanding the nature of lexically-idiosyncratic sound change (§5.2.1).

1.3.2 Paradigmatic relations

Previous psycholinguistic work on paradigmatic effects on speech production has examined two kinds of patterns: phonemic contrasts (e.g., a word's *phonological neighborhood*, which typically refers to the set of words that contrast with

a given word on exactly one segment), and morphological paradigms. For example, many studies argue that production latencies (how quickly a speaker can begin producing a word) and articulatory durations are affected by the number of phonological neighbors that a word has (Gahl, Yao & Johnson, 2012; Gahl & Strand, 2016; Fricke, 2013; Sadat, Martin, Costa & Alario, 2014; Fricke, Baese-Berk & Goldrick, 2016; Buz & Jaeger, 2015; Vitevitch, 2002; Vitevitch & Sommers, 2003; Fox, Reilly & Blumstein, 2015). Other work investigates how the co-presence of a single neighbor influences an intended word's phonetic realization (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012; Schertz, 2013; Buz, Tanenhaus & Jaeger, 2016, and see §3.1, §3.4).

However, the mechanism for the durational effects is not yet clear (Jaeger & Buz, 2016). Among various possibilities, competition-centric accounts propose that they are caused by co-activation of phonologically-related words (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012; Goldrick, Vaughn & Murphy, 2013; Fricke, 2013; Fricke et al., 2016), which feeds back into or competes with the activation of the intended word (Baese-Berk & Goldrick, 2009; Gahl et al., 2012). The activation level of a word is thought to be related to its articulatory duration (Baese-Berk & Goldrick, 2009; Gahl et al., 2012; Fricke, 2013, though see Jaeger & Buz, 2016). Alternatively, communication-centric accounts propose that the durational effects result from a need to enhance a particular phonological contrast (Schertz, 2013; Buz et al., 2016, cf. Lindblom, 1990). For example, a word might be lengthened if it has a large number of neighbors, or if it must be communicated in the context of a single, easily-confusable neighbor, in order to increase the perceptual distance between that word and its neighbors (though see Gahl & Strand, 2016) and thereby improve the likelihood of communicative success.

The experiment presented in Chapter 3 informs the communication-centric

account: when a coda-voicing minimal-pair word is presented in the same context, speakers adjust the durations of *dose* and *doze* in two distinct ways that each enhance the relevant contrast (shortening the vowel, and prolonging voicing in the coda, respectively). This suggests that speakers take advantage of their implicit knowledge about the phonological relationship between the two words to guide their articulations (see also Buz et al., 2016, and other research reviewed in Chapter 3).

In terms of morphological paradigms, it has often been proposed that an intended word is influenced by the phonetic targets of morphologically-related words. In particular, it is believed that a morphologically-complex word may be pronounced with some subtle phonetic echo of its relatives. For example, the German word *Rad* 'wheel' is nominally pronounced [ʁaːt], but the final [t] is actually pronounced with some phonetic cues associated with voiced [d]. This may be due to the influence of its relative *Räder* [ʁæːdɐ] 'wheels', which has a voiced [d] in the corresponding position (see §4.1.1). However, this pattern can also be described as *incomplete devoicing*. Under this proposal, there is no phonetic interaction between morphological relatives. Instead, *Rad* has underlying voicing, and surfaces with partial voicing because of incomplete application of a phonological rule (e.g., d → t / __ #). To help disentangle these competing proposals, the data in Chapter 4 provide novel evidence for morphology-phonetics interactions that cannot clearly be accounted by reference to an underlying form.

Understanding the conditions under which morphology-phonetics interactions occur is an important area of inquiry, both within psycholinguistics and within the study of language change. For example, one strand of empirical work suggests that the relative surface frequency of paradigm members may be crucial (Hay, 2001, 2003, 2007; Schuppler, van Dommelen, Koreman & Ernestus, 2012;

Cohen, 2014, 2015, and see Hay & Baayen, 2005). In the tradition of phonology and historical linguistics, one name for morphology-conditioned phonetic patterns is *subphonemic analogy* (Bloomfield, 1933, cited in Garrett, 2015), such as when a phonological pattern is extended to morphological relatives that would not otherwise participate in that pattern. As one instance, in Scottish English, /ai/ is lengthened to [aːe] in open syllables as in *tie*. Through analogy, the nucleus in *tied* is also lengthened, even though it lacks the appropriate conditioning environment (Scobbie, Turk & Hewlett, 1999; Ladd, 2016). Chapter 4 investigates a set of synchronic predictions involving such analogical (or *paradigm uniformity*) effects on acoustic durations between morphological relatives. More generally, the discussion in §5.2.2 explores how these effects might play a role in phonological change.

1.4   Bibliography

Baese-Berk, M. & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language*, *42*(1), 1–22.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Bloomfield, L. (1933). *Language*. New York: Henry Holt.

Boye, K. & Harder, P. (2012). A usage-based theory of grammatical status and grammaticalization. *Language*, *88*(1), 1–44.

Brown, E. L. & Rivas, J. (2012). Grammatical relation probability: How usage patterns shape analogy. *Language Variation and Change*, *24*(03), 317–341.

Buz, E. & Jaeger, T. F. (2015). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, 1–21.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent productions. *Journal of Memory and Language*.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Bybee, J. (2003). Mechanisms of Change in Grammaticization: the Role of Frequency. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 602–623). Oxford: Blackwell.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711–733.

Bybee, J. & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, *37*(4), 575–596.

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement: Contextual and paradigmatic effects on morpheme pronunciation. *Morphology*.

Cohen, C. (2015). Context and paradigms: Two patterns of probabilistic pronunciation variation in Russian agreement suffixes. *The Mental Lexicon*, *10*(3), 313–338.

Dabrowska, E. (2008). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language*, *58*(4), 931–951.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283.

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, *83*, 97–117.

Fricke, M. D. (2013). *Phonological encoding and phonetic duration.* PhD thesis, University of California, Berkeley.

Fricke, M. D., Baese-Berk, M. M., & Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience.*

Gahl, S. & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, *80*, 748–775.

Gahl, S. & Garnsey, S. M. (2006). Knowledge of Grammar Includes Knowledge of Syntactic Probabilities. *Language*, *82*(2), 405–410.

Gahl, S. & Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language.*

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806.

Garrett, A. (2015). Sound change. In C. Bowern & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 227–248). Routledge.

Goldrick, M. & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683.

Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, *62*(2), 113–134.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*, *134*(2), EL172–EL177.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, *39*(6; ISSU 376), 1041–1070.

Hay, J. (2003). *Causes and consequences of word structure*. Routledge.

Hay, J. (2007). The phonetics of 'un'. In J. Munat (Ed.), *Lexical creativity, texts and contexts* (pp. 39–57).

Hay, J. & Baayen, R. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, *9*(7), 342–348.

Jaeger, T. F. & Buz, E. (2016). Signal reduction and linguistic encoding. In *Handbook of Psycholinguistics*. Wiley-Blackwell.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the emergence of linguistic structure*.

Kirov, C. & Wilson, C. (2012). The specificity of online variation in speech production. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, Sapporo, Japan.

Ladd, D. R. (2016). Lexical allophones. In *Perspectives on Marginal Contrasts*, Ithaca, NY.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*(3), 172–187.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&h theory. In *Speech production and speech modelling* (pp. 403–439). Springer.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526–3529.

Rapp, B. & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499.

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3-4), 249–263.

Schuppler, B., van Dommelen, W. A., Koreman, J., & Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, *40*(4), 595–607.

Scobbie, J. M., Turk, A., & Hewlett, N. (1999). Morphemes, Phonetics and Lexical Items: The Case of the Scottish Vowel Length Rule. In *Proceedings of the XIVth International Congress of Phonetic Sciences.*, volume 2, (pp. 1617–1620). International Congress of Phonetic Sciences.

Torres Cacoullos, R. & Walker, J. (2009). The present of the English future: Grammatical variation and collocations in discourse. *Language*, *85*(2), 321–354.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735–747.

Vitevitch, M. S. & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4), 491–504.

Chapter 2

Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation

ABSTRACT

Language-users reduce words in predictable contexts. Previous research indicates that reduction may be stored in lexical representation if a word is often reduced. Because representation influences production regardless of context, production should be biased by how often each word has been reduced in the speaker's prior experience. This study investigates whether speakers have a context-independent bias to reduce low-informativity words, which are usually predictable and therefore usually reduced. Content word durations were extracted from the Buckeye and Switchboard speech corpora, and analyzed for probabilistic reduction effects using a language model based on spontaneous speech in the Fisher corpus. The analysis supported the hypothesis: low-informativity words have shorter durations, even when the effects of local contextual predictability, frequency, speech rate, and several other variables are controlled for. Additional models that compared word types against only other words of the same segmental length further supported this conclusion. Words that usually appear in predictable contexts are reduced in all contexts, even those in which they are unpredictable. The result supports representational models in which reduction is stored, and where sufficiently frequent reduction biases later production. The finding provides new evidence that probabilistic reduction interacts with lexical representation.

2.1 Introduction

2.1.1 Probabilistic reduction

In speech production, language-users reduce words when they are predictable in the local context, as well as when they are frequent overall (Lieberman, 1963; Whalen, 1991; Gahl, 2008). This reduction manifests as a broad array of articulatory and acoustic effects, including differences in word and syllable duration, vowel dispersion and quality, plosive voice onset time, syllable deletion, and language-specific segmental deletion, among others (Bell, Jurafsky, Fosler-Lussier, Girand, Gregory & Gildea, 2003; Aylett & Turk, 2006; Baker & Bradlow, 2009; Hooper, 1976; Bybee, 2002; Jurafsky, Bell, Gregory & Raymond, 2001; Everett, Miller, Nelson, Soare & Vinson, 2011; Clopper & Pierrehumbert, 2008; Yao, 2009; Gahl & Garnsey, 2004; Bybee, 2006; Tily, Gahl, Arnon, Snider, Kothari & Bresnan, 2009; Kuperman & Bresnan, 2012; Demberg, Sayeed, Gorinski & Engonopoulos, 2012; Moore-Cantwell, 2013). These phenomena have been known for over a century (see Bell, Brenier, Gregory, Girand & Jurafsky, 2009, for a review), and are usually described together as the *probabilistic reduction hypothesis*—words with higher probability are articulatorily reduced, for a variety of local and global probabilistic measures.

The cause of probabilistic reduction is not fully understood, although it can be accounted for in several different (and compatible) models of speech production. For example, such reduction may indicate that speakers actively manage their productions to balance audience-design considerations with articulatory efficiency (Lindblom, 1990). Under this theory, speakers hyper-articulate unpredictable words in order to improve listeners' chances of parsing words that they

have low expectations for. They hypo-articulate words that listeners can easily pre-
dict based on the context, in order to save on articulatory effort. Smooth-signal or
uniform-information-density versions of this theory frame this behavior as speak-
ers' preference for keeping a constant rate of information transfer (Aylett & Turk,
2004; Pluymaekers, Ernestus & Baayen, 2005; Levy & Jaeger, 2007). Speakers
spend more time on unpredictable words, which are informative, and relatively
little time on predictable words, which provide less new information.

An alternative account for probabilistic reduction is based in speaker-internal
processing factors (Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon & Newlands,
2000; Munson, 2007; Bell et al., 2009). Under this theory, words are activated more
strongly by their phonological, semantic, and syntactic associates. This facilitates
retrieval and speeds production (Gahl, Yao & Johnson 2012, cf. Baese-Berk &
Goldrick 2009). For example, Kahn & Arnold (2012) show that a linguistic prime
causes speakers to reduce a word target even when audience-design factors are
controlled for, while a non-linguistic prime does not trigger reduction.

2.1.2 Is reduction stored in lexical representation?

An important question is whether probabilistic reduction is exclusively an
online effect, or whether it is also represented offline in the lexicon. It is generally
argued that unreduced citation forms have a privileged representational status
(Ernestus, Baayen & Schreuder, 2002; Kemps, Ernestus, Schreuder & Baayen,
2004; Ranbom & Connine, 2007). However, there is evidence that reduced forms
are also represented. Lavoie (2002) and Johnson (2007) show that words with
homophonous citation forms can have very dissimilar distributions of reduction
variants in conversational speech, and each word may in fact have special reduced
variants that are unattested for its homophone. For example, [fɹ] and [fə] are

attested variants of *for* but not of *four*. This suggests that reduced forms are to some extent word-specific, and therefore associated with lexical representation, rather than created exclusively online during production.

Furthermore, language-users have a processing advantage for common reduced forms of a word. This advantage is relative to how often the word is reduced (Connine & Pinnow, 2006; Connine, Ranbom & Patterson, 2008). For example, French *genou* [ʒənu] is often realized in a reduced form [ʒnu], which lacks an audible schwa. On the other hand, *querelle* [kəʁɛl] is more often realized with a full schwa in the first syllable. In isolated word production, speakers are faster to produce forms like [ʒnu] than [kʁɛl], all else held equal, where [ʒnu] but not [kʁɛl] is a common word-specific reduction (Racine & Grosjean, 2005; Bürki, Ernestus & Frauenfelder, 2010). In lexical decision experiments, Ranbom & Connine (2007) and Pitt, Dilley & Tat (2011) show that listeners are faster to classify reduced forms like English *gentle* [dʒɛɾ̃l̩], with a nasal flap, than [dʒɛnʔl̩], where the flap but not the glottal stop is a usual reduction of [t] in words like *gentle*. These findings indicate that reduced variants, when they are typical realizations of a word, are likely stored in representation (Pitt, 2009; Ernestus, 2014).

There are at least three ways this storage might be implemented. First, storage of reduction might involve multiple phonologically-abstract, categorical variants, which include both unreduced and reduced forms of a word (as described above). Second, individual productions of reduced words might be stored as exemplars with fine-grained phonetic detail, including acoustic reduction (Pierrehumbert, 2002; Johnson, 2007). Third, reduction might be represented indirectly via changes to articulatory timing relations that are lexically specified (Browman & Goldstein, 1990; Byrd, 1996; Lavoie, 2002).

Is probabilistic reduction stored in lexical representation? Reduction as-

sociated with high contextual probability is standardly treated as an online phenomenon, such as a kind of priming or else active management of information density, as in 2.1.1. The evidence discussed here suggests that reduction is stored when it occurs often enough. Therefore, if a word is very often reduced because it typically occurs in high-probability contexts, language-users may store this reduction in lexical representation as well.

2.1.3 Informativity

In usage, some words almost always occur in predictable contexts, whereas others are unlikely in each of the contexts that they occur in, even though they might be relatively frequent overall. For example, the word *current* usually occurs in the context of *current events* or *the current situation*, and is therefore usually predictable in context. On the other hand, the word *nowadays* has roughly the same log-frequency overall as *current*, but *nowadays* occurs in a wide variety of contexts (see figure 2.1). Thus, on average, *nowadays* is more unpredictable in each of its contexts.

The average predictability of a word in context is its *informativity* (Cohen Priva, 2008; Piantadosi, Tily & Gibson, 2011). Word informativity is formally defined as:

$$-\sum_c P(C = c \mid W = w) \log P(W = w \mid C = c) \tag{2.1}$$

In equation 2.1, c is a context and w is a word type. Context is usually operationalized simply as the n preceding or following words in an utterance. The informativity of a word type is the averaged probability with which a word will occur given each of the contexts that it can occur in. This average is weighted
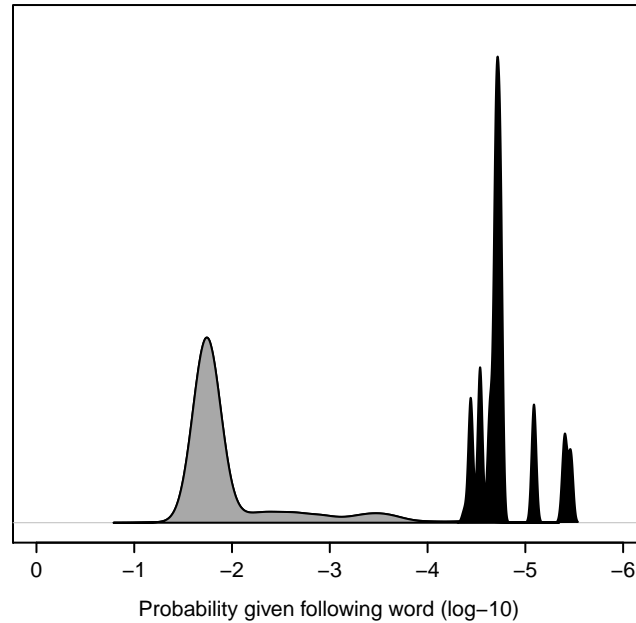
Figure 2.1: Density plot showing the (log-10) probability of tokens of *current* (gray) and *nowadays* (black) given the following word as context; predictability is higher to the left. Tokens of *current* usually occur in predictable contexts, and so *current* has more density on the left (low informativity); tokens of *nowadays* are usually unpredictable, and so *nowadays* has more density on the right (high informativity). Tokens taken from Fisher; probabilities from COCA.

by the frequency with which the word occurs in each context. Usually-predictable words (like *current*) have low informativity, because they tend to provide less new information in actual communicative use. Usually-unpredictable words (*nowadays*) have high informativity, because in actual use they tend to be surprising and informative.

Because low-informativity words are usually predictable, they are also usually reduced. On the other hand, high-informativity words are rarely reduced. The experiments described in 2.1.2 demonstrate that reduced forms of a word are more accessible if a reduced form is a typical realization of that word. If probabilistic reduction is stored, reduction of low-informativity words should be more accessible than reduction of high-informativity words. In a model that assumes

abstract variants, reduced variants of low-informativity words should be retrieved and produced in spontaneous speech relatively more often than could otherwise be explained. In general terms, the prediction is that speakers should have a stronger bias for reducing a word if that word is usually reduced and usually predictable elsewhere. The current study evaluates this hypothesis through an analysis of conversational speech corpora.

2.1.4 Evidence for linguistic informativity effects

2.1.4.1 Sub-lexical informativity

There is evidence from language-specific phonetics that average contextual predictability does affect sub-lexical representation and processing. Using speech production data from English and Dutch, Aylett & Turk (2004, 2006) and van Son & Pols (2003) show that greater within-word predictability is associated with shorter segment and syllable token duration, which is consistent with the larger picture of probabilistic reduction. Building on this research, Cohen Priva (2008, 2012) demonstrates that certain features of English segment types are best modeled by the average contextual predictability of each type within words across the lexicon. Consonants that are on average less predictable in context have longer durations and are deleted less often than consonants that are more predictable, even when they do occur in highly predictable contexts. This paper extends these findings from segment realization to word realization.

There is related within-word evidence from perception for the claim that language-users are biased by average contextual probabilities, in addition to local ones. Lee & Goldrick (2008) show that the errors that English and Korean speakers make in recalling nonce words are dependent not only on immediate segment

transition probabilities, but also on the typical reliability of different intra-syllable boundaries in each of the two languages. This suggests that the representation of sub-syllabic units (the body and rime) is shaped by the average predictability within each unit.

Together, these results provide some evidence that language-users represent average contextual trends—not just local relationships—and that this representation influences both production and perception below the word level. However, similar processing effects have yet to be shown at the word level.

2.1.4.2 Word lengths

Piantadosi et al. (2011) and Mahowald, Fedorenko, Piantadosi & Gibson (2013) show that word informativity is correlated with word lengths. Words that are usually predictable in context have fewer letters or segments than words that are usually unpredictable. This correlation is independent of—and greater than—the known correlation between word lengths and frequency (Zipf, 1935). Frequent and predictable words are thus both shorter in length and also reduced in production. Several authors have proposed that these two phenomena are connected, and that probability-conditioned reduction leads to permanent representational change (Bybee, 2003; Lindblom, Guion, Hura, Moon & Willerman, 1995; Mowrey & Pagliuca, 1995; Pierrehumbert, 2001).

2.1.5 The current study

This paper evaluates whether speakers have a bias favoring reduced productions of words that are typically encountered in reduced forms. If context-driven probabilistic reduction is represented at the word level, productions of usually-predictable and therefore usually-reduced words should be more reduced across

contexts. This effect should be proportional to each word's informativity, since informativity measures how often that word is reduced online in predictable contexts. On the other hand, if contextual predictability only causes reduction on-line, usually-predictable words should only be reduced when they are actually predictable.

One way to look for the effect is the following: if a word that is usually predictable occurs in an unpredictable context, it should appear more reduced than would be otherwise expected for that context. Equivalently, if a word that is usually unpredictable occurs in a predictable context, it should appear less reduced than would be expected for that context, since its representation is biased towards a clear form. This can be operationalized as the effect of word informativity on acoustic duration: high-informativity words—words that are usually unpredictable—should have longer durations than those words which are usually predictable, when all other factors are held equal.

This hypothesis is evaluated on word durations extracted from the Buckeye and Switchboard corpora using a series of linear mixed-effects regression models. Word informativity is included as a variable in the model, and the analysis should show that this variable is significantly associated with greater word duration if the hypothesis is true. The control variables include local probabilistic reduction—it is mathematically true that if words are shorter when they appear in predictable contexts, then the average token of a usually-predictable word will be shorter than the average token of a rarely-predictable word. However, the effects of local reduction are included as a separate parameter in the model. Therefore, if low-informativity words are only shorter on average because most or all of their tokens occur in predictable contexts, there will be no independent effect of the type-level informativity variable: token contextual probability will capture all of the possible

variation arising from context.

Thus, once local probability is factored out of every word token's duration, a non-significant informativity effect would suggest that words have durations that can ultimately be derived from length, segmental content, syllable count, raw frequency, etc. On the other hand, if low-informativity words are shorter across all contexts, regardless of local token probability, then informativity should capture a significant amount of the remaining variance.

2.2   Materials and methods

2.2.1   Word duration data

Word durations were extracted from two sources of natural English speech, which were analyzed separately: the Buckeye Corpus of Conversational Speech (Pitt, Dilley, Johnson, Kiesling, Raymond, Hume & Fosler-Lussier, 2007) and the NXT Switchboard Annotations (Calhoun, Carletta, Jurafsky, Nissim, Ostendorf & Zaenen, 2009) based on Switchboard-1 Release 2 (Godfrey & Holliman, 1997).

The Buckeye Corpus is a collection of interviews conducted around 1999–2000 in Columbus, Ohio. There are forty speakers in the corpus, each of whom was recorded for about one hour under the initial pretense that they were participating in a focus group on local issues. The speakers were balanced for age and sex, but all were white natives of Central Ohio belonging roughly to the middle and working class. The corpus itself contains the original audio recordings as well as several types of transcriptions and annotations. Word durations for this study were taken from the timestamps provided for the word-level annotations. Additionally, each word token is annotated with two different segmental transcriptions. First,

each token includes a dictionary-based transcription, which is the canonical or citation form for the word type, generated from automatic alignment software. Second, each individual token includes a close phonetic transcription, created by an annotator who hand-corrected the software-generated segments and timestamps for each token.

The Switchboard Corpus is a collection of telephone conversations conducted as a corporate research project in 1990–1991. The NXT-formatted subset used here included 642 annotated conversations between 358 speakers (Calhoun, Carletta, Brenier, Mayo, Jurafsky, Steedman & Beaver, 2010). Speakers from all areas of the United States were recruited through internal corporate and government listservs, through public electronic bulletin boards, and by peer-to-peer recruitment. No effort was made to balance speakers by age, sex, region, or socioeconomic class. Speakers were assigned to conversations with individuals they had not previously spoken with, and pairs were provided with one of 70 general-interest conversation topics selected by an automated operator. Word durations for this study were taken from the corpus annotation timestamps, which were created by hand-correction of automated word alignment. Below the word level, only dictionary-based segmental transcriptions were available for the words in Switchboard.

Only content words were included in the analysis, as it has been shown that function and content words respond differently to predictability effects (Bell et al., 2009), and function words are generally considered to be processed differently than content words (e.g., Levelt, Roelofs & Meyer, 1999). Some content word tokens were excluded from the analysis for prosodic or other reasons, following standard practice (e.g., Bell et al., 2003, 2009; Gahl et al., 2012; Jurafsky et al., 2001; Jurafsky, Bell & Girand, 2002). For the purposes of the following exclusions and

for calculating speech rate, an utterance was defined as a stretch of speech by a single speaker that is delimited by pauses, disfluencies, or other interruptions greater than or equal to 500 milliseconds. Tokens were excluded if they were adjacent to a disfluency, a pause, or a filled pause; if they were utterance-initial or utterance-final; if the word was cliticized (e.g., *cousin's*); if the word type or bigram context was not found in the language model; if the utterance speech rate was more than 2.5 standard deviations away from the speaker's mean; if the word token duration was more than 2.5 standard deviations away from the word type's mean; or if there were no vowels or syllabic consonants in the word token's close phonetic transcription (in Buckeye). In the Buckeye Corpus, all data from speaker 35 were excluded due to a large number of transcription and alignment errors (Gahl et al., 2012).

2.2.2 Probabilistic language model

To test a hypothesis about word predictability, it is necessary to estimate inter-word probabilities from a source that belongs to the same genre of language to be analyzed. Research has shown that word probabilities estimated from corpora of the same language register as the one to be modeled are much better associated with different word processing variables than probabilities estimated from larger but dissimilar corpora (Brysbaert, Keuleers & New, 2011; Brysbaert & New, 2009; Francom and Ussishkin, submitted). The Fisher English Training Part 2 Transcripts have a history of use by previous researchers looking to estimate probabilities in natural conversational speech (e.g., Arnon & Snider, 2010). Furthermore, they are a good genre-of-speech match with Buckeye and Switchboard, since all three corpora involve recorded informal conversations between individuals who are meeting for the first time.

The Fisher Part 2 corpus is a collection of English telephone conversations created at the Linguistic Data Consortium to aid speech-to-text dialogue systems (Cieri, Graff, Kimball, Miller & Walker, 2005). There are 5,849 conversations of up to ten minutes each, totaling over 12 million words. Speakers were randomly assigned a conversation partner and one of 100 topics, with an effort to balance speakers by sex, age (although speakers older than 50 are under-represented), and geographical region (roughly one-fifth from the US North, Midland, South, and West dialect regions; with one-fifth from Canada or speaking non-US or non-native English varieties). Each speaker participated in usually 1–3 conversations in order to maximize inter-speaker variation within the corpus, and topics were selected for a range of vocabulary. The contents of Buckeye, Switchboard, and Fisher do not overlap.

Two bigram language models were calculated based on the Fisher transcripts. These models list the probabilities that each word will occur, given either the word before it (in one model) or the word after it (in the second model). Bigram models are standardly used in studies of predictability-based phonetic reduction, and focused research on predictability measures has shown negligible improvement in predicting reduction from trigram or more complicated models (Jurafsky et al., 2001). The probabilities were smoothed with the modified Kneser-Ney method described by Chen & Goodman (1998), using the SRILM Toolkit (Stolcke, 2002; Stolcke, Zheng, Wang & Abrash, 2011) with smoothing parameters optimized by the toolkit. The final estimates were used as measures of local contextual predictability. All other probabilistic measures were also estimated from the Fisher transcripts.

2.2.3 Variables

Each word token in the two corpora was annotated based on both the Fisher probability data as well as the type- and token-specific variables described here. First, word informativity for each word type w was calculated using equation 2.1, with probabilities taken from the smoothed Fisher language models. Two estimates of word informativity were calculated, with context c taken as either the preceding word or following word, respectively. Word informativity here is therefore the average of a word's bigram probability across the contexts that it occurs in, weighted by how frequently it occurs in each of those contexts. Informativity was calculated in bans, which uses log base 10.

Previous research has demonstrated that predictability given the following n-gram context is associated with greater and more reliable reduction effects than predictability given the preceding n-gram context. In fact, predictability given preceding context has often been reported as failing to reach significance in predicting English duration reduction (Jurafsky et al., 2001; Bell et al., 2009; Gahl, 2008), except for high-frequency function words. However, informativity in written corpora is usually calculated based on the preceding context (Piantadosi et al., 2011), and so preceding informativity is also included here.

Tokens were also annotated for a variety of control variables taken from previous literature on models of word duration. The collected data were analyzed with a series of linear mixed-effects models containing these variables as parameters, in order to evaluate the direction and significance of the association between informativity and word duration. The exact statistical procedure used to analyze the data is described in section 2.2.4.

Word durations, and all continuous control variables, were log-transformed

(base 10) and centered around their respective means within each model. The distributions of these variables were found to be more normal in log space, and this practice follows previous research (e.g. Bell et al., 2009; see Kuperman & Bresnan, 2012 for discussion). Informativity results are qualitatively the same with and without log-transformation. In addition to the variables described in this section, each model listed below also includes per-word random intercepts, per-speaker random intercepts, and correlated per-speaker informativity slopes as controls for individual word-type and speaker idiosyncrasies.

Baseline duration: In order to calculate a baseline expected duration for each word token, the Modular Architecture for Research on Speech Synthesis (MARY) text-to-speech system (Schröder & Trouvain, 2003) was used to analyze each utterance. This baseline was selected to control for the segmental length, content, and context of each word form. This method follows Demberg et al. (2012), who also use durations from the MARYTTS system as a baseline control when they evaluate the effects of syntactic predictability. Previous work on predictability effects has also used orthographic length, syllable count, simple segmental length, expected word durations estimated by summing average segment durations, and/or per-word random intercepts as statistical controls for word form length and content. Alternative analyses using different baselines measures are discussed in section 2.3.3.4.

The cmu-slt-hsmm voice package was used to calculate acoustic parameters for each segment and word token. This package has been trained on part of the CMU ARCTIC database (Komineck & Black, 2003) to estimate segment and word durations based on the phonological features of the current and adjacent segments, syllable structure and position, and word stress. To some extent, the package also models phrase accents and prosody, based on part-of-speech and utterance

boundaries indicated with punctuation.

Utterances were sent to the MARYTTS system for analysis, and word durations were extracted from the realized acoustic parameters for each utterance. Demberg et al. (2012) show that this method of analyzing full utterances, which allows sentential context to be included, generates word duration estimates that are superior to single-word analysis. The final baseline estimates were log-transformed and centered.

Syllable count: Number of syllabic segments in the word type's transcription; log-transformed and centered.

Speech rate: Number of syllabic segments per second in each utterance; log-transformed and centered.

Bigram probability: Two variables for the conditional probability of a word given the previous or following word, as estimated from the smoothed language models; log-transformed and centered.

Word frequency: Raw token count of a word in the Fisher transcripts; log-transformed and centered.

Orthographic length: Number of letters in the word's orthography; log-transformed and centered. Previous research has indicated that orthographic length may have an independent effect on word duration (Warner, Jongman, Sereno & Kemps, 2004; Gahl, 2008), and it is important to include this variable to control for the previously-observed association between orthographic length and word informativity (Piantadosi et al., 2011).

Part of speech: Coded as noun, verb, adjective, or adverb based on the annotations provided in the Buckeye and Switchboard corpora. Other parts of speech and proper nouns were excluded. In the models, this variable was treatment-coded with noun as the base level.

2.2.4   Model procedure

Linear mixed-effects models were fit to the Buckeye and Switchboard word duration data, using the full set of variables as predictors. Analysis was conducted using the lme4 package in R (Bates, Maechler & Bolker, 2013; R Core Team, 2013).

To help guard against model overfitting, backward model selection was done to remove predictors that did not significantly improve a model. Following this procedure, after each full model was fit, it was compared against a set of models that each had one fewer predictor. Each model in this set had a different predictor removed. If the full model was not significantly better than each of these models ($\alpha = 0.15$ based on log-likelihood fit), the predictor that contributed the least improvement to fit was removed from the full model. These steps were repeated until the final model was significantly better than all possible alternatives with one fewer predictor. The final model was then compared against the original to confirm that it fit the data as well as the original. Final p-values for each effect were calculated by log-likelihood ratio tests that compared the fit of the final model with and without each variable.

For each corpus, a model was fit to all valid content word tokens from that corpus. However, word informativity is highly correlated with segment count (Piantadosi et al., 2011). Since duration is also predicted by segment count, this raises the concern that any effect of informativity on duration might be simply because informative words tend to have more segments. As a precaution against this confound, a baseline expected duration is included as a control variable, as described in 2.2.3. However, as a further precaution, additional models were fit over content words matched for a single length in each corpus. For example, one model was fit only to words with two segments, another model was fit only to words with

three segments, etc. In this way, words could be compared exclusively with other words of the same segmental length. If informativity has an independent effect on word duration beyond simply the association with segment count, it should show up in every one of these models as well.

2.3   Results

2.3.1   Study 1: Buckeye

The data from Buckeye included 41,167 word tokens meeting the inclusion criteria, distributed among 3,429 types. Two predictors were found not to significantly improve fit, and were removed in order: (1) informativity given the previous word ($p > 0.5$), and (2) word frequency ($p > 0.5$). Random per-speaker informativity slopes provided a significant improvement in fit ($p < 0.0001$), and were retained the model. The final fixed and random effects estimates appear in Tables 2.1 and 2.2, respectively, with durations in base-10 log seconds. Correlations between each pair of continuous variables appear in appendix table 2.7, and density plots showing the distribution of probabilistic variables appear in Figure 2.5.

Crucially, informativity given the following word was found to be significantly associated with word duration. Six additional models were fit over subsets of words grouped according to their dictionary transcription length, so that informativity effects on duration could be evaluated independently of segment count. A summary of the results appears in Table 2.3. As before, informativity given the following word was found to be reliably associated with duration for words of up to seven segments long. Informativity given the previous word reached significance in

the predicted direction for 2–segment words, but was non-significant for all other lengths. Predictors that did not improve fit and were removed from each additional model are listed in appendix Table 2.8. For every model with a significant informativity effect in table 2.3, informativity was a reliable predictor (at least $p < 0.05$) both before and after the non-significant predictors were removed.

Table 2.1: Fixed effects summary for model of Buckeye word durations.

	β	SE	t	$p(\chi^2)$
INTERCEPT	0.0257	0.0057	4.48	—
BASELINE DURATION	0.5879	0.0150	39.32	< 0.0001
SYLLABLE COUNT	0.0592	0.0104	5.71	< 0.0001
SPEECH RATE	-0.3406	0.0077	-43.97	< 0.0001
BIGRAM PROB. GIVEN PREVIOUS	-0.0102	0.0007	-15.00	< 0.0001
BIGRAM PROB. GIVEN FOLLOWING	-0.0205	0.0007	-30.55	< 0.0001
ORTHOGRAPHIC LENGTH	0.0437	0.0167	2.62	0.0089
PART OF SPEECH = ADJECTIVE	0.0033	0.0032	1.04	(< 0.0001)
PART OF SPEECH = ADVERB	-0.0172	0.0042	-4.09	—
PART OF SPEECH = VERB	-0.0275	0.0022	-12.54	—
INFORMATIVITY GIVEN FOLLOWING	0.0244	0.0023	10.77	< 0.0001

Table 2.2: Random effects summary for model of Buckeye word durations.

RANDOM EFFECT	SD	Cor.
Word (intercept)	0.043	—
Speaker (intercept)	0.033	—
Speaker (inf., following)	0.007	0.061
Residual	0.098	—

2.3.2   Study 2: Switchboard

The data from Switchboard included 107,981 word tokens meeting the inclusion criteria, distributed among 4,997 types. Of the ten predictors, only word frequency did not significantly improve the model at $\alpha = 0.15$ and was removed

Table 2.3: Informativity results for Buckeye models that compared words matched for segment count. # Seg = segment count; Types = content word types in Buckeye with that number of segments; Tokens = content word tokens with that number of segments.

# Seg	Types	Tokens	INF. GIVEN PREVIOUS				INF. GIVEN FOLLOWING			
			β	SE	t	$p(\chi^2)$	β	SE	t	$p(\chi^2)$
2	51	2,734	0.0569	0.0217	2.62	0.0116	0.0805	0.0210	3.83	0.0002
3	536	13,282	—	—	—	—	0.0271	0.0050	5.47	0.0001
4	737	12,552	—	—	—	—	0.0198	0.0044	4.53	0.0001
5	660	4,660	—	—	—	—	0.0347	0.0046	7.50	0.0001
6	494	3,324	0.0143	0.0075	1.89	0.0616	0.0335	0.0080	4.22	0.0001
7	379	2,260	—	—	—	—	0.0252	0.0060	4.19	0.0001

($p > 0.7$). The fixed effects summary is given in Table 2.4, and the random effects summary appears in Table 2.5 (all durations in base-10 log seconds). Random per-speaker informativity slopes significantly improved the fit of the model ($p < 0.0001$). Correlations between each pair of continuous variables, and density plots for probabilistic variables, are given in the appendix. As in Buckeye, informativity given the following word captured a significant amount of duration variance ($p < 0.0001$). In this model, informativity given the previous word also reached statistical significance ($p < 0.05$).

A summary of the results of duration models over words matched for segment count appears in Table 2.6. The results from Switchboard replicated those from Buckeye: informativity given the following word was significantly associated with word duration in the predicted duration for words of 2–7 segments. Informativity given the previous word reached significance for words of seven segments, but was eliminated or marginal for all other lengths. A complete list of non-significant predictors that were pruned from each model is given in appendix Table 2.10. For every model in which informativity was statistically significant after non-significant predictors were removed, it was also significant when these predictors were retained in the model, except for the model of 2–segment words

($p < 0.11$ before non-significant predictors were removed).

Table 2.4: Fixed effects summary for model of Switchboard word durations.

	β	SE	t	$p(\chi^2)$
INTERCEPT	0.0287	0.0023	12.62	—
BASELINE DURATION	0.5363	0.0102	52.34	< 0.0001
SYLLABLE COUNT	0.0492	0.0070	7.01	< 0.0001
SPEECH RATE	−0.3260	0.0044	−74.79	< 0.0001
BIGRAM PROB. GIVEN PREVIOUS	−0.0082	0.0005	−18.17	< 0.0001
BIGRAM PROB. GIVEN FOLLOWING	−0.0227	0.0004	−53.91	< 0.0001
ORTHOGRAPHIC LENGTH	0.1343	0.0115	11.69	< 0.0001
PART OF SPEECH = ADJECTIVE	−0.0051	0.0021	−2.36	(< 0.0001)
PART OF SPEECH = ADVERB	−0.0186	0.0026	−7.18	—
PART OF SPEECH = VERB	−0.0410	0.0017	−23.87	—
INFORMATIVITY GIVEN PREVIOUS	0.0040	0.0016	2.48	0.0131
INFORMATIVITY GIVEN FOLLOWING	0.0142	0.0016	8.72	< 0.0001

Table 2.5: Random effects summary for model of Switchboard word durations.

RANDOM EFFECT	SD	Cor.
Word (intercept)	0.039	—
Speaker (intercept)	0.028	—
Speaker (inf., previous)	0.006	−0.158
Speaker (inf., following)	0.007	0.089
Residual	0.100	—

Table 2.6: Informativity results for Switchboard models that compared words matched for segment count.

			INF. GIVEN PREVIOUS				INF. GIVEN FOLLOWING			
# Seg	Types	Tokens	β	SE	t	$p(\chi^2)$	β	SE	t	$p(\chi^2)$
2	81	6,136	—	—	—	—	0.0270	0.0119	2.27	0.0238
3	665	35,956	—	—	—	—	0.0158	0.0036	4.39	0.0001
4	983	29,892	0.0063	0.0034	1.84	0.0670	0.0143	0.0035	4.11	0.0001
5	940	14,586	0.0061	0.0035	1.75	0.0806	0.0173	0.0034	5.07	0.0001
6	758	8,372	—	—	—	—	0.0220	0.0037	5.98	0.0001
7	607	6,539	0.0097	0.0047	2.06	0.0400	0.0145	0.0041	3.53	0.0005

2.3.3   Additional description and exploratory analysis

Figure 2.2 shows a sample of word types that appeared in the analysis. Log-frequency appears on the vertical axis and informativity given the following word is on the horizontal axis, with both measures calculated from the Fisher language model. *Nowadays* appears on the right side of the chart, with an informativity of 3.93 bans. *Current* has an informativity of 1.13 bans, which would put it to the left of the plot area.
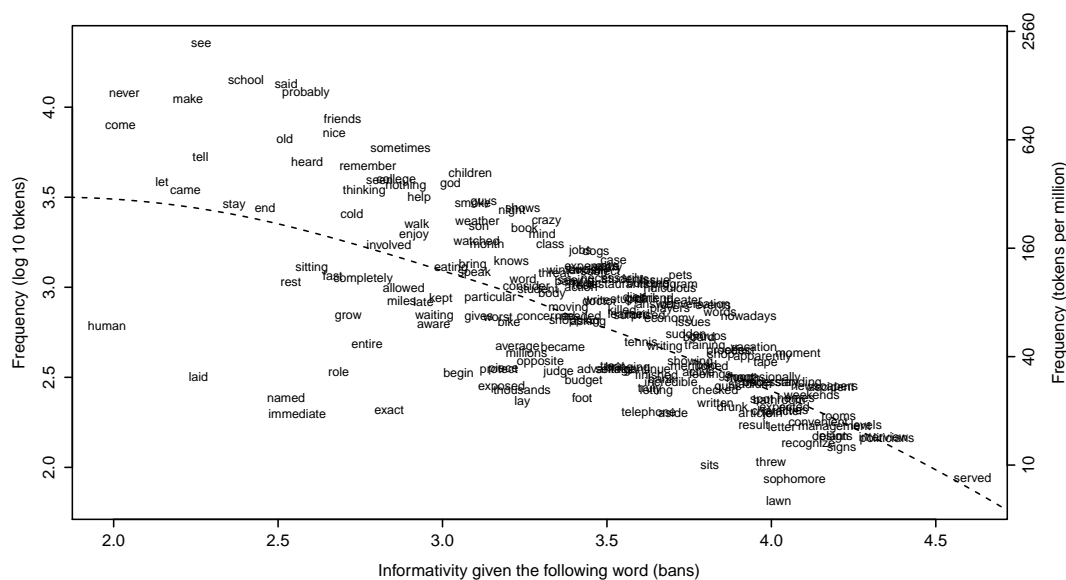
Figure 2.2: Sample of 200 word types that were observed in the combined data at least 10 times. Word types on the left are on average more predictable in context, and word types towards the top are more frequent. Dashed line shows local trend for all word types with at least 10 observations.

2.3.3.1   Extension of previous work on correlations with informativity

For all content word types across the two corpora, the correlation between orthographic length and informativity given the previous word (Spearman's $\rho = 0.27$) is numerically similar to what was reported by Piantadosi et al. (2011) for

all English word types (note that coefficients in 2.6 are calculated over tokens). This is marginally higher than the correlation between orthographic length and frequency ($\rho = 0.26$). Since function words were excluded from the data here, this finding helps address a possible concern that the correlation was influenced by the distribution of function words versus content words (see Mahowald et al., 2013). Furthermore, the correlation between orthographic length and informativity given the following word is slightly higher ($\rho = 0.29$).

2.3.3.2 Word frequency

Word frequency did not consistently improve the fit of the duration models, contra earlier research. However, this was found to be due to the inclusion of per-word random intercepts in the current models, which captured most of the effect of frequency and some of the other per-type variables. Removing the per-word intercept parameter revealed an apparently reliable effect of word frequency in the *positive* direction (Buckeye: $\beta = 0.0092$, $t = 4.45$; Switchboard: $\beta = 0.0088$, $t = 7.04$). Furthermore, removing per-word intercepts made both informativity effects appear very reliable in the predicted direction (Buckeye: given previous: $\beta = 0.0087$, $t = 3.85$; given following: $\beta = 0.0402$, $t = 16.27$; Switchboard: given previous: $\beta = 0.0139$, $t = 11.65$; given following: $\beta = 0.0289$, $t = 24.75$).

The positive word frequency estimate does not suggest that more frequent words have longer durations. Instead, word frequency is acting as a suppressor variable for informativity (Friedman & Wall, 2005; Wurm & Fisicaro, 2014). Informativity is strongly correlated with word durations (Pearson's $r_{Y1} = 0.52$ for Buckeye) and frequency is less well correlated with duration ($r_{Y2} = -0.28$). However, informativity and frequency are well correlated with each other ($r_{12} = -0.58$). In the case where the inequality $r_{12} > {r_{Y2}}/{r_{Y1}}$ holds (here: $0.58 > {0.28}/{0.52}$) or is

nearly satisfied[1], the variable that is less well correlated with the outcome (frequency) may change sign.

Because frequency is much better correlated with informativity than with duration, in the model frequency is used to explain some of the error associated with informativity, rather than explaining a unique portion of duration variance. Wurm & Fisicaro (2014) show that, on simulated data with two correlated variables that fall into this region, the impact on the variable with the larger effect (here, informativity) is that this variable may have a slightly inflated estimate β, but that there is also a large loss of statistical power for detecting the variable's effect. Adelman, Brown & Quesada (2006) also report that frequency acts as a suppressor when it is entered into a model with a measure of contextual diversity (to predict reading times; see also McDonald & Shillcock, 2001), which is a variable that is conceptually similar to informativity.[2]

However, in the more conservative analysis—which includes the per-word intercept parameter—frequency simply fails to reach statistical significance at $\alpha = 0.15$ and is removed from the model. With this parameter, frequency was also found to be non-significant in each of the models by segment count for Switchboard, as reported in appendix table 2.10, and non-significant in four of the seven models by segment count for Buckeye, as reported in 2.8.

To evaluate whether a different corpus might produce different results, alternative bigram language models (forward and backward) were constructed from the

---

[1]In a model with more than two predictors, the exact boundary also depends on the other relations among the predictors and with the outcome variable. Also, note that calculations here are done for frequency and duration prior to log-transformation, but the same condition exists for correlations after log-transformation.

[2]It is possible to remove this correlation and change the direction of the frequency effect by residualizing informativity on frequency. However, residualization does not change the estimates of the residualized variable (informativity), has a number of other undesirable effects on the model, and makes the interpretation of both residualized and residualizer variables problematic (Wurm & Fisicaro, 2014).

450-million-word Corpus of Contemporary American English (Davies, 2008). This corpus is larger, but it is mixed-genre and includes mostly written sources. This analysis included the original random effects structure and all of the parameters in the original model, except that the probabilistic measures (local bigram probability, frequency, and informativity) were replaced by estimates from the COCA language model. The COCA language model was smoothed using the Witten-Bell method (Witten & Bell, 1991), since counts for very low-frequency bigrams in COCA were not available.[3] In this analysis, frequency was found to be a reliable predictor in the expected direction (Buckeye: $\beta = -0.009$, $t = -2.38$, Switchboard: $\beta = -0.0111$, $t = -4.06$), while informativity given the following word retained its significance (Buckeye: $\beta = 0.0170$, $t = 4.65$, Switchboard: $\beta = 0.0081$, $t = 3.40$).

2.3.3.3   Predictability versus lexicalized two-word expressions

Given that the informativity measure is derived from bigram predictability, one question is whether predictability-driven reduction might be caused by lexicalized bigrams, rather than predictability-in-context per se. For example, one hypothesis might be that *human* is reduced in contexts like *human being*, *human rights*, *human nature*, *human life*, etc. not because *human* is predictable in those contexts, but because a bigram like *human rights* is stored as a single lexical entry. If such bigrams are processed as a single unit, it would be reasonable to expect them to be produced more quickly than two more independent words.

This question primarily has to do with the interpretation of bigram predictability, not informativity. If a word like *human* or *current* occurs in a common

---

[3]The informativity effects were robust to the smoothing technique used on the original Fisher language model. For both datasets, model estimates for both bigram probability and informativity were similar when unsmoothed probabilities were used instead.

expression, it will be predictable in that context. Therefore, its reduction can be explained by its local bigram probability. If a word like *human* mainly occurs in such expressions, it will also have a low informativity value, since it is predictable on average. However, this could not be taken as evidence for an informativity effect. Wherever *human* occurs in a predictable expression, its reduction can be explained by local bigram probability. In order for there to be statistical evidence for an informativity effect, *human* must also be shortened in unpredictable contexts (or lengthened in predictable ones) beyond what the model would otherwise predict given the local relations between words.

Some previous work has addressed whether word reduction associated with local predictability should be interpreted as an effect of lexicalized bigrams. Bell et al. (2009) calculate pointwise mutual information for each bigram type in their data. This measure quantifies how dependent two words are on each other. They exclude tokens that occur in bigrams with a mutual information value in the top 15% of their data, but find that their bigram-predictability estimates are not qualitatively different with these data excluded. This suggests that predictability-driven reduction cannot be exclusively attributed to lexicalized bigrams. Jurafsky et al. (2001) study reduction of function word durations and vowel quality. They exclude tokens that have a bigram probability greater than the median bigram probability in their data. They find that local bigram predictability given the *following* word is robust to this exclusion, whereas local predictability given the *previous* word is still a significant predictor of duration reduction, but not of vowel reduction. Note that, for content word durations, Bell et al. (2009) did not originally find a significant effect of predictability given the previous word.

2.3.3.4 Alternative baseline durations

Recent work on probabilistic reduction has used corpus-derived estimates of word duration as a baseline for each word type. For example, Bell et al. (2009) and Gahl et al. (2012) calculate expected durations first by determining the mean duration of each segment type across the corpus. Then, for the segments in each word's transcription, these mean segment durations are summed together to generate a baseline expectation for the word's total duration. For example, in the word *fish* [fɪʃ], the baseline duration would be the mean corpus duration of [f], plus the mean corpus duration of [ɪ], plus the mean corpus duration of [ʃ]. An alternative analysis was conducted using this measure as a baseline, rather than the MARYTTS duration model. For Buckeye, close phonetic transcriptions were available, and these were used in place of dictionary transcriptions. This means that expected durations in Buckeye were calculated based on each word token's transcription, rather than the word type's dictionary citation form. However, using dictionary transcriptions for Buckeye did not qualitatively change the results. With this baseline measure, informativity given the following word was significant (at least $p < 0.05$) in both of the full corpus models and all of the models stratified by segment count, for both corpora. Informativity given the previous word was also significant in some models using this baseline, but overall the effect was inconsistent.

A similar alternative baseline might also be adapted to take into account some of the effects of segmental context. Buz & Jaeger (p.c.) calculate baselines first by extracting mean durations of each segment type conditioned on the previous segment type. For example, the segment [t] has a different mean duration following [s] than following [n] or [ə], which reflects general articulatory constraints as well as

language-specific phonetics and phonology. An alternative analysis was conducted using this biphone-sensitive measure as a baseline. For each word type, the mean durations of each segment type given the previous segment in the transcription were summed together. For example, in the phrase *a fish* [ə fɪʃ], the baseline duration of the token *fish* was taken to be the mean duration of [f] when it follows [ə] in the corpus, plus the mean duration of [ɪ] when it follows [f], plus the mean duration of [ʃ] when it follows [ɪ].

Using these baselines, in the Switchboard analysis, informativity given the following word was significant in the full corpus model and all of the models stratified by segment count. In Buckeye, for which close phonetic transcriptions were available, informativity given the following word was significant (at least $p < 0.05$) in the full model and all of the models stratified by segment count, except for the model of 6–segment words ($p < 0.10$). When dictionary transcriptions were used to generate the Buckeye baselines, the informativity effect in the model of 6–segment words was also significant. As before, informativity given the previous word was inconsistent across analyses and word lengths.

2.3.3.5 Segment deletion and duration

The effect of informativity on word durations might involve segment compression, and it might also involve categorical segment deletion. Figure 2.3 shows the mean percentage of segments that were transcribed as deleted from word tokens in the Buckeye data at each segment count. These numbers are very similar to those reported for content words by Johnson (2004). For that study, close transcriptions were only available for about ⅓ of the full corpus. However, the data here also involve additional exclusions (such as exclusion of pause-adjacent tokens; see section 2.2.1). As observed by Johnson (2004) and replicated here, short con-

tent words with four or fewer segments in their dictionary transcriptions are rarely transcribed as having segments deleted ($< 5\%$ of segments deleted; although see also the segment deviation analysis in that study). Since there was an effect of informativity on word durations for these short words, it is more likely that the effect for these words involves compression rather than deletion.
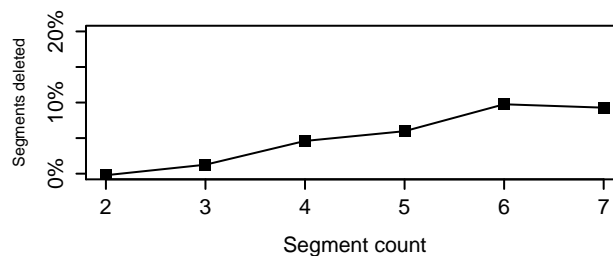
Figure 2.3: Mean percentage of segments deleted for word tokens of each segment count, based on comparisons of the dictionary and close phonetic transcriptions in the Study 1 Buckeye data.

Figure 2.4 shows the Spearman correlation coefficients between word informativity and the percentage of segments that were deleted in Buckeye tokens, divided by dictionary segment count. The figure also shows the coefficients between informativity and duration at each segment count. While there is a reliable correlation across different word lengths between informativity given the following word and duration, informativity is not well associated with deletion (as transcribed in the corpus) for shorter words of less than five segments. For longer words, full segment deletion probably also contributes to the duration effect.

To explore these two possible components of the duration effect, an alternative analysis was carried out using only those Buckeye word tokens which had a close phonetic transcription that was completely identical to the word type's dictionary transcription (20,296 of the original 41,167 tokens meeting the inclusion criteria). For these data, informativity given the following word had a statistically significant effect (at least $p < 0.05$) in the full model and for words of each segment
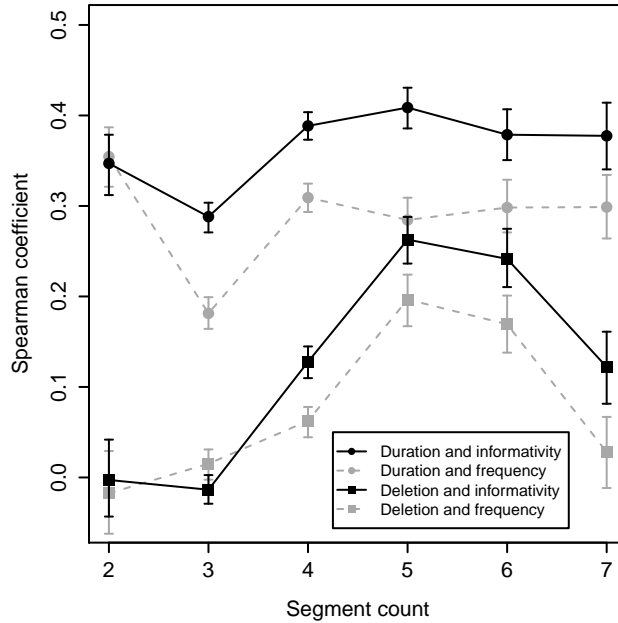
Figure 2.4: Spearman correlations between (i) informativity given the following word and (ii) duration (black circles) or percentage of segments deleted (black squares). Calculated over Study 1 Buckeye word tokens at each dictionary transcription length; error bars show bootstrapped 95% confidence intervals. Additional gray lines show correlations between frequency and duration or percentage of segments deleted. Lower informativity corresponds with shorter durations and higher deletion rates; deletion coefficients here are inverted for comparison with duration coefficients.

count, except for 7–segment words, where the effect was non-significant ($p > 0.4$; with 408 tokens and 122 different types). The effect sizes for informativity were roughly the same as in the primary analysis; for 5 and 7–segment words the estimate was much smaller. Because informativity was found to influence duration for these words which had no segments deleted, compression is likely a key component of the informativity effect on word durations. However, deletion is not ruled out as a part of the effect, especially for longer word types for which full segment deletion is transcribed more often in general.

However, the analysis is potentially much more complex when considering segment deletion in particular, instead of whole-word durations. For example,

language-specific categorical deletion processes are typically restricted to particular contexts or classes of words, such as [t] reduction in English (Pitt et al., 2011) or schwa deletion in English or French (Connine et al., 2008; Racine & Grosjean, 2005). Many deletion processes are also sensitive to additional sub-lexical factors such as morphological structure (e.g., Labov, Cohen, Robins & Lewis, 1968; Guy, 1991; although see also Sugahara & Turk, 2009). A targeted study is likely better suited to analyze the interaction of word informativity with deletion processes (see Cohen Priva, 2012, especially Ch. 3).

2.4   General discussion

The main result of this paper is that the average contextual predictability of a word—word informativity—has a significant effect on that word's acoustic duration. A word that is usually unpredictable has a longer duration than a word that is usually predictable, independent of local contextual predictability, frequency, or segment count. The effect size is comparable to reduction associated with local predictability. The effect was reliable for informativity given the following word, but not given the previous word. This difference is consistent with previous research on local probabilistic reduction in content words.

Since informativity captured a significant amount of the variance beyond local bigram probability, it is not the case that predictable words simply have shorter tokens on average because most of their tokens are reduced. Instead, even after the local probability of every token is taken into account, there is an additional effect of each word's average predictability in the hypothesized direction. If most tokens of a word are reduced in duration because they appear in high-probability contexts, the duration of that word was also found to be shorter overall,

independent of token context.

2.4.1 Accounting for the informativity effect

Previous research has suggested that if a word is reduced often enough, speakers will encode that reduction in their representation of the word. This stored reduction will bias future productions of that word in all contexts. The current results support this prediction for probabilistic reduction. If a word is reduced very often because it often occurs in high-probability contexts, productions of that word are biased towards a more reduced form, even when it is not produced in a high-probability context. The informativity finding might be captured by several possible speech processing models, described below.

2.4.1.1 Exemplar-based and combined exemplar–abstract models

In an exemplar model, all phonetic detail of each incoming word token is stored as an exemplar of that word (Goldinger, 1996, 1998; Johnson, 1997, 2006, 2007). Representation is constructed from the distribution of previously-encountered exemplars. The forms that are most often heard and stored therefore have a greater influence on this distribution. Since low-informativity words are very commonly reduced, the distribution of exemplars of a low-informativity word will be biased towards reduced forms. To generate a production target, one or more exemplars are sampled from this distribution (Goldinger, 2000; Pierrehumbert, 2001). In some intermediate models, word exemplars that belong to a single category are compressed into a phonologically-abstract secondary representation that also influences the generation of a production target (Goldinger, 2007; Ernestus & Baayen, 2011; German, Carlson & Pierrehumbert, 2013; Ernestus, 2014). The production target is then passed to a system for phonetic implementation,

during which motor-planning, articulatory, or other effects influence the final realization (Ernestus, 2014).

When speakers produce a low-informativity word, they will be more likely to sample reduced exemplars of that word when generating a target, even if the context would not otherwise trigger reduction (Goldinger, 2000; Pierrehumbert, 2002). During phonetic implementation, online probabilistic reduction may occur (Ernestus, 2014), causing the word to be further reduced. Thus, low-informativity words should be reduced due to offline informativity effects—a higher likelihood of sampling reduced exemplars—as well as online contextual ones, which apply during phonetic implementation.

2.4.1.2   Abstract models with multiple variants

The results can also be described by a model with abstract phonological representations. In such a model, each word representation includes an unreduced citation form, and may also include several reduced variants that are sufficiently common in a language-user's experience. The importance that each variant form has for perception and production depends on their relative frequencies (Connine et al., 2008; Racine & Grosjean, 2005; Bürki et al., 2010; Pitt et al., 2011; Ranbom & Connine, 2007). It may also depend on factors such as orthography (Ranbom & Connine, 2007, 2011), a communicative pressure that favors unreduced forms (Pitt et al., 2011), and probabilistic knowledge about articulatory contexts (Mitterer & McQueen, 2009).

If a reduced variant of a word type has a higher relative frequency, it will be more accessible in production (Bürki et al., 2010), and more often accessed in spontaneous speech. In this case, the informativity effect would represent the proportion with which the unreduced variant is selected in favor of the reduced

variant. For example, consider a scenario in which the word *current* has two abstract variant forms stored in lexical representation. One form is the unreduced citation variant [kɹ.ɪnt], and the other is a reduced variant [kɹ̃ʔ]. The analysis presented here models word duration, not the selection of variant forms. Duration is not represented in these abstract forms, but it is empirically true that the unreduced variant has an average duration of 350ms, while the reduced variant has an average duration of 200ms. If speakers usually select the unreduced variant, on average the word type will have a longer duration that is closer to 350ms. If speakers usually select the reduced variant, on average the word type will have a shorter duration that is closer to 200ms.

Given that unreduced variants can reasonably be assumed to always be longer than reduced variants (or at least not shorter), the informativity parameter then describes the tendency for speakers to select unreduced variants in general. If informativity has a positive coefficient, that means that the model predicts that high-informativity word types will have longer durations on average, because speakers prefer to select the unreduced variant more often than the reduced variant.

2.4.1.3 Rational speech production

There are other representational models that might result in an informativity effect. For example, it may be the case that each word type has a default phonetic target for production. If a speaker chooses to deviate from this default, such as to reduce or hyper-articulate a word, it is costly to do so in motor planning (even though reduction might save on articulatory effort) and the speaker may not always reach their deviant target. Therefore, in order to minimize both planning and articulatory costs and maximize effectiveness, a rational speaker will select a default target for each type that is most similar to the tokens that usually

need to be produced. If reduction is usually called for, then the rational speaker will choose a reduced default form, so that it is only rarely necessary to deviate from that form and incur costs (thanks to Roger Levy for this suggestion). In this model, informativity is a goal-oriented effect chosen directly by the speaker, rather than an indirect consequence of representation.

2.4.1.4 Efficient articulation

A related model might also refer to word-specific articulatory timing specifications. In this model, words are specified for tighter or looser alignments of each necessary articulatory gesture. This results in a range of possible reductions during fast speech that is unique to each word (Lavoie, 2002). Low-informativity words usually occur in predictable contexts. In such contexts, a speaker is more likely to be understood, and consequently is more likely to accept an imprecise production of a target word as an adequate acoustic realization of that word. A speaker will gradually learn through experience that some or all gestures in such a word can have looser timings, yet still produce an acceptable acoustic form (following Jaeger & Ferreira, 2013; Jaeger, 2013). Over time, the word will acquire less strict gestural alignment specifications. Because of this, gestural overlap (and, potentially, a greater degree of acoustic reduction) will become more likely for this word in all contexts.

For example, Jaeger & Ferreira (2013) suggest that it is primarily reductions of low-confusability words or word forms that become acceptable variants in common usage. By definition, low-informativity words tend to be more predictable and are unlikely to be very confusable in context.

2.4.1.5   Direct knowledge of average word probabilities

It may instead be the case that an informativity metric is represented directly at the word level. In other words, language-users would not store reduction or reduced variant forms, but instead would directly track how predictable a word type is on average. This would be stored in addition to (implicit or explicit) knowledge about specific inter-word relationships. In production, speakers would then use their knowledge of both factors, plus frequency, to determine online how much articulatory effort to give to a word. The results of Lee & Goldrick (2008), who argued that speakers perceptually track both average predictability as well as local predictability of segments within syllables, suggest that language-users may in some way use informativity-like knowledge in perception.

Extending this account to production would require that informativity-driven reduction be a consequence of audience-design considerations. An online processing model of probabilistic reduction is more difficult to reconcile with an informativity effect. In these models, locally-predictable words gain a boost in activation from nearby words or related constructions that are usually associated with them, which speeds production. The current results show that low-informativity words are shorter even in unpredictable contexts where no nearby syntactic or semantic associates would cause pre-activation. An online processing model of informativity would thus require an additional mechanism to explain how pre-activation occurs even in contexts where there are no associated words to activate the target.

An online audience-design account of informativity is possible. In this account, speakers are aware of the fact that their own model of word probabilities may not be exactly the same as their listeners' models. Since informativity is an

averaged probability, it would be prudent for the speaker to adjust an extreme estimate of local probability by also weighing how probable the target word usually is, on average. Note that the primary informativity effect involved predictability given the following word, not the previous word. If speakers reduce words when they are predictable for the listener, this must be accommodated by research showing that listeners interpret words based on the following context as well as the preceding context (Szostak & Pitt, 2013).

In this account, speakers would store probabilistic knowledge of how predictable each word tends to be, rather than storing the effects of reduction. This would still likely entail lexical representation of informativity, since this knowledge is word-specific. This account would also require that speakers then balance at least three sources of probability (context-specific, average context, and frequency) in choosing how much articulation a word requires for efficient communication. By contrast, an offline representational account of informativity gives rise to the observed effect and requires neither active goal-oriented behavior on the part of the speaker nor fine-grained negotiation of multiple word-specific probabilities during each articulation.

2.4.2   Summary

Previous research has shown that words are reduced when they occur in predictable contexts. The results of this paper show that words which are typically predictable in context are reduced even when they occur in unpredictable contexts. This phenomenon is predicted by models in which probabilistic reduction is stored in representation, and this stored reduction has a stronger effect on processing when it is relatively more frequent.

The effect was found to be robust across English word types of different

lengths and segmental content, for two large corpora and various implementations of a probabilistic language model. None of a large set of control variables could fully account for the relationship between informativity and duration. Future work might evaluate different representational accounts with laboratory production studies, in not only English but also other languages. In particular, such work might help explain the correlation between word lengths and informativity across languages.

2.5   Acknowledgments

* * *

Chapter 2, in full, is a reprint of the material as it appears in Seyfarth (2014) [Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133 (1), 140–155. doi:10.1016/j.cognition.2014.06.013]. This work was reported, in part, at the 19th Architectures and Mechanisms for Language Processing conference.

2.6 Predictor summaries
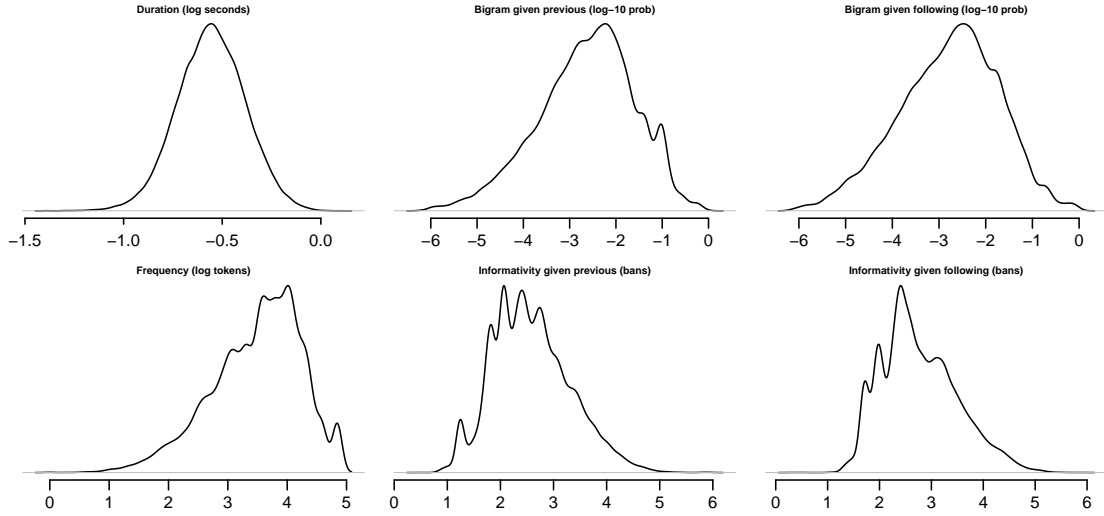
Figure 2.5: Density plots showing the distribution of the probabilistic variables in the Buckeye Corpus. Variables were centered before being entered into the model.
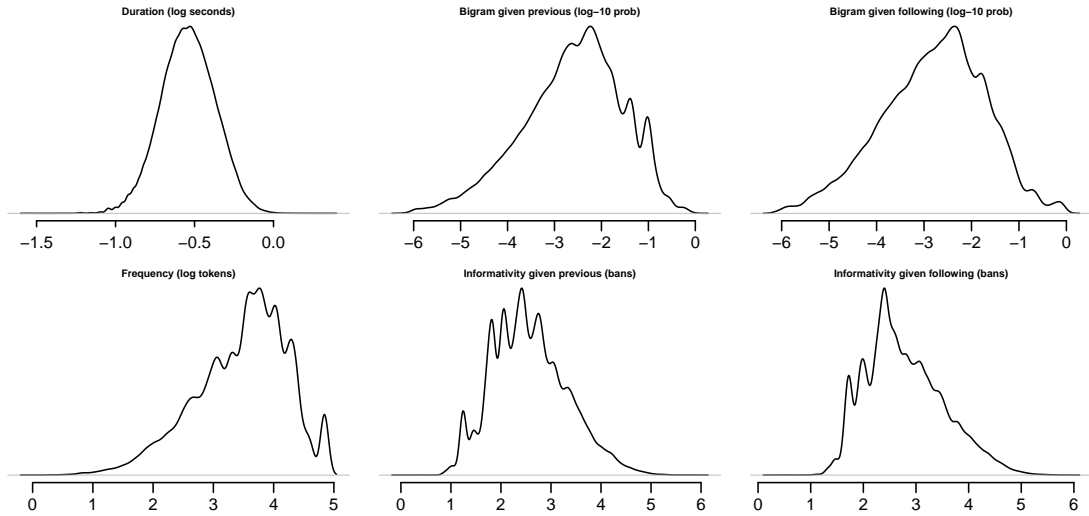
Figure 2.6: Density plots showing the distribution of the probabilistic variables in the Switchboard Corpus. Variables were centered before being entered into the model.

Table 2.7: Spearman correlations between variables in Buckeye data. SEG = segment count, DUR = empirical duration, MARY = baseline duration, SYL = syllable count, RATE = speech rate, BG-P = bigram probability given previous word, BG-F = bigram probability given the following word, FREQ = word frequency, ORTH = orthographic length, INF-P = informativity given the previous word, INF-F = informativity given the following word.

	SEG	DUR	MARY	SYL	RATE	BG-P	BG-F	FREQ	ORTH	INF-P	INF-F
SEG	1.00	0.57	0.86	0.76	0.02	−0.26	−0.25	−0.37	0.87	0.35	0.36
DUR	0.57	1.00	0.63	0.48	−0.21	−0.32	−0.43	−0.44	0.55	0.37	0.50
MARY	0.86	0.63	1.00	0.71	0.01	−0.21	−0.25	−0.31	0.83	0.28	0.35
SYL	0.76	0.48	0.71	1.00	0.06	−0.23	−0.19	−0.29	0.76	0.30	0.29
RATE	0.02	−0.21	0.01	0.06	1.00	0.03	0.02	0.02	0.02	−0.01	−0.02
BG-P	−0.26	−0.32	−0.21	−0.23	0.03	1.00	0.38	0.64	−0.22	−0.73	−0.54
BG-F	−0.25	−0.43	−0.25	−0.19	0.02	0.38	1.00	0.58	−0.23	−0.49	−0.70
FREQ	−0.37	−0.44	−0.31	−0.29	0.02	0.64	0.58	1.00	−0.31	−0.86	−0.84
ORTH	0.87	0.55	0.83	0.76	0.02	−0.22	−0.23	−0.31	1.00	0.28	0.33
INF-P	0.35	0.37	0.28	0.30	−0.01	−0.73	−0.49	−0.86	0.28	1.00	0.71
INF-F	0.36	0.50	0.35	0.29	−0.02	−0.54	−0.70	−0.84	0.33	0.71	1.00

Table 2.8: Predictors removed at $\alpha = 0.15$ in each Buckeye model by segment count. Abbreviations given in caption to Figure 2.7.

# Seg	Non-significant predictors removed
2	SYL ($p > 0.8$), ORTH ($p > 0.6$), BG-P ($p > 0.15$)
3	ORTH ($p > 0.5$), SYL ($p > 0.2$), FREQ ($p > 0.2$), INF-P ($p > 0.4$)
4	ORTH ($p > 0.4$), FREQ ($p > 0.4$), INF-P ($p > 0.2$), SYL ($p > 0.15$)
5	INF-P ($p > 0.8$), ORTH ($p > 0.7$), FREQ ($p > 0.4$)
6	—
7	FREQ ($p > 0.7$), INF-P ($p > 0.5$)

Table 2.9: Spearman correlations between variables in Switchboard data. Abbreviations given in caption to figure 2.7.

	Seg	Dur	Mary	Syl	Rate	Bg-P	Bg-F	Freq	Orth	Inf-P	Inf-F
Seg	1.00	0.60	0.86	0.67	0.00	−0.27	−0.24	−0.38	0.87	0.35	0.36
Dur	0.60	1.00	0.65	0.46	−0.20	−0.34	−0.44	−0.46	0.56	0.40	0.50
Mary	0.86	0.65	1.00	0.64	0.01	−0.24	−0.26	−0.35	0.82	0.31	0.37
Syl	0.67	0.46	0.64	1.00	0.07	−0.26	−0.22	−0.32	0.67	0.34	0.33
Rate	0.00	−0.20	0.01	0.07	1.00	0.02	0.01	0.03	0.01	−0.02	−0.02
Bg-P	−0.27	−0.34	−0.24	−0.26	0.02	1.00	0.39	0.65	−0.23	−0.74	−0.55
Bg-F	−0.24	−0.44	−0.26	−0.22	0.01	0.39	1.00	0.60	−0.23	−0.50	−0.71
Freq	−0.38	−0.46	−0.35	−0.32	0.03	0.65	0.60	1.00	−0.32	−0.86	−0.84
Orth	0.87	0.56	0.82	0.67	0.01	−0.23	−0.23	−0.32	1.00	0.29	0.35
Inf-P	0.35	0.40	0.31	0.34	−0.02	−0.74	−0.50	−0.86	0.29	1.00	0.71
Inf-F	0.36	0.50	0.37	0.33	−0.02	−0.55	−0.71	−0.84	0.35	0.71	1.00

Table 2.10: Predictors removed at $\alpha = 0.15$ in each Switchboard model by segment count.

# Seg	Non-significant predictors removed
2	Inf-P ($p > 0.7$), Freq ($p > 0.8$), Syl ($p > 0.3$)
3	Orth ($p > 0.9$), Freq ($p > 0.8$), Inf-P ($p > 0.6$)
4	Freq ($p > 0.7$), Syl ($p > 0.3$)
5	Freq ($p > 0.3$)
6	Freq ($p > 0.9$), Orth ($p > 0.3$), Syl ($p > 0.2$), Inf-P ($p > 0.15$)
7	Freq ($p > 0.9$), Orth ($p > 0.6$)

2.7 Bibliography

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823.

Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Aylett, M. & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Aylett, M. & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*(5), 3048–3058.

Baese-Berk, M. & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554.

Baker, R. E. & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, *52*(4), 391–413.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1–22.

Bates, D., Maechler, M., & Bolker, B. (2013). lme4.0: Linear mixed-effects models using s4 classes. R package version 0.999999-4.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language*, *60*(1), 92–111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in english conversation. *The Journal of the Acoustical Society of America*, *113*(2), 1001.

Browman, C. P. & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, 341–376.

Brysbaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of google books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology*, *2*.

Brysbaert, M. & New, B. (2009). Moving beyond kucera and francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods*, *41*(4), 977–990.

Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one "fenêtre" in the production lexicon? on-line evidence on the nature of phonological representations of pronunciation variants for french schwa words. *Journal of Memory and Language*, *62*(4), 421–437.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Bybee, J. (2003). Mechanisms of change in grammaticization: the role of frequency. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 602–623). Oxford: Blackwell.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711–733.

Byrd, D. (1996). A phase window model framework for articulatory timing. *Phonology*, *13*, 139–169.

Calhoun, S., Carletta, J., Brenier, J. M., Mayo, N., Jurafsky, D., Steedman, M., & Beaver, D. (2010). The NXT-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, *44*(4), 387–419.

Calhoun, S., Carletta, J., Jurafsky, D., Nissim, M., Ostendorf, M., & Zaenen, A. (2009). NXT switchboard annotations. Technical report, Linguistic Data Consortium, Philadelphia.

Chen, S. & Goodman, J. (1998). An empirical study of smoothing techniques for language modeling. Technical Report TR-10-98, Harvard University.

Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. (2005). *Fisher English Training Part 2*. Philadelphia: Linguistic Data Consortium.

Clopper, C. G. & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America*, *124*(3), 1682.

Cohen Priva, U. (2008). Using information content to predict phone deletion. In Abner, N. & Bishop, J. (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, (pp. 90–98).

Cohen Priva, U. (2012). *Sign and signal: deriving linguistic generalizations from information utility.* PhD thesis, Stanford University.

Connine, C. M. & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review, 23*(3).

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics, 70*(3), 403–411.

Davies, M. (2008). The corpus of contemporary american english: 450 million words, 1990-present.

Demberg, V., Sayeed, A. B., Gorinski, P. J., & Engonopoulos, N. (2012). Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, (pp. 356—-367).

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua, 142*, 27–41.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*(1-3), 162–173.

Ernestus, M. & Baayen, R. H. (2011). Corpora and exemplars in phonology. In J. A. Goldsmith, J. Riggle, & A. C. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 374–400). Blackwell.

Everett, C., Miller, Z., Nelson, K., Soare, V., & Vinson, J. (2011). Reduction of brazilian portuguese vowels in semantically predictable contexts. In *Proceedings of the 17th International Congress of Phonetic Sciences*, (pp. 651–654)., Hong Kong.

Friedman, L. & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician, 59*(2), 127–136.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*(3), 474–496.

Gahl, S. & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language, 80*, 748–775.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*(4), 789–806.

German, J. S., Carlson, K., & Pierrehumbert, J. B. (2013). Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics*, *41*(3-4), 228–248.

Godfrey, J. J. & Holliman, E. (1997). Switchboard-1 release-2. Technical report, Linguistic Data Consortium, Philadelphia.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(5), 1166–1182.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, *195*(2), 251–279.

Goldinger, S. D. (2000). The role of perceptual episodes in lexical processing. In *Proceedings of the Workshop on Spoken Word Access Processes*, (pp. 155–158)., Nijmegen, The Netherlands. Max-Planck Institute for Psycholinguistics.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 49–54).

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, *3*(1), 1–22.

Hooper, J. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. J. Christie (Ed.), *Current Progress in Historical Linguistics* (pp. 95–105). Amsterdam: North-Holland.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, *4*.

Jaeger, T. F. & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, *36*(4), 31–32.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. Mullenix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2004). Massive reduction in conversational american english. In Yoneyama, K. & Maekawa, K. (Eds.), *Proceedings of the 1st Session of the 10th International Symposium on Spontaneous Speech: Data and Analysis*, (pp. 29—54)., Tokyo, Japan. The National International Institute for Japanese Language.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In M. Sole, P. Beddor, & M. Ohala (Eds.), *Experimental Approaches to Phonology. In Honor of John Ohala.* (pp. 25–40). Oxford University Press.

Jurafsky, D., Bell, A., & Girand, C. (2002). The role of the lemma in form variation. *Laboratory Phonology*, *7*, 3–34.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamins.

Kahn, J. M. & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, *67*(3), 311–325.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*(1-3), 117–127.

Komineck, J. & Black, A. W. (2003). CMU ARCTIC databases for speech synthesis. Technical Report CMU-LTI-03-177, Carnegie Mellon University.

Kuperman, V. & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, *66*(4), 588–611.

Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). A study of the non-standard english of negro and puerto rican speakers in new york city. Technical Report Coorperative Research Project 3288, Vol. I, Columbia University.

Lavoie, L. (2002). Some influences on the realization of for and four in american english. *Journal of the International Phonetic Association*, *32*(2), 175–202.

Lee, Y. & Goldrick, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language*, *59*(2), 155–168.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Levy, R. & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems.*

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*(3), 172–187.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the h & h theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403–439). Dordrecht: Kluwer.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, *7*(1), 5–37.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318.

McDonald, S. A. & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295—-322.

Mitterer, H. & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 244–263.

Moore-Cantwell, C. (2013). Syntactic predictability influences duration. In *Proceedings of Meetings on Acoustics*, volume 19.

Mowrey, R. & Pagliuca, W. (1995). The reductive character of articulatory evolution. *Rivista di Linguistica*, *7*(1), 37–124.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. *Laboratory Phonology*, *9*, 201–228.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526–3529.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure: studies in language* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology*, *7*, 101–139.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release). Technical report, Columbus, OH.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? a test of generalization to newly learned words. *Journal of Memory and Language*, *61*(1), 19–36.

Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, *39*(3), 304–311.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*(2-4), 146–159.

R Core Team (2013). R: A language and environment for statistical computing.

Racine, I. & Grosjean, F. (2005). Le coût de l'effacement du schwa lors de la reconnaissance des mots en français. *Canadian Journal of Experimental Psychology*, *59*, 240–254.

Ranbom, L. & Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298.

Ranbom, L. J. & Connine, C. M. (2011). Silent letters are activated in spoken word recognition. *Language and Cognitive Processes*, *26*(2), 236–261.

Schröder, M. & Trouvain, J. (2003). The german text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, *6*(4), 365–377.

Stolcke, A. (2002). SRILM — an extensible language modeling toolkit. In *Proceedings of the International Conference on Spoken Language Processing*, Denver, CO.

Stolcke, A., Zheng, J., Wang, W., & Abrash, V. (2011). SRILM at sixteen: Update and outlook. In *Proc. IEEE Automatic Speech Recognition and Understanding Workshop*, Waikoloa, HI.

Sugahara, M. & Turk, A. (2009). Durational correlates of english sublexical constituent structure. *Phonology*, *26*(3), 477.

Szostak, C. M. & Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics*, *75*(7), 1533–1546.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2), 147–165.

van Son, R. J. J. H. & Pols, L. C. W. (2003). How efficient is speech? In *Proceedings of the 25th Institute of Phonetic Sciences*, volume 25, (pp. 171–184).

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from dutch. *Journal of Phonetics*, *32*(2), 251–276.

Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America*, *90*(4), 2311–2311.

Witten, I. H. & Bell, T. C. (1991). The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory*, *37*(4), 1085–1094.

Wurm, L. H. & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48.

Yao, Y. (2009). Understanding VOT variation in spontaneous speech. In *Proc. 18th International Congress of Linguists (CIL XVIII)*, (pp. 29–43).

Zipf, G. K. (1935). *The Psychobiology of Language.* Boston: Houghton-Mifflin.

Chapter 3

Dynamic hyperarticulation of coda voicing

contrasts

ABSTRACT

This study investigates the capacity for targeted hyperarticulation of contextually-relevant contrasts. Participants communicated target words with final /s/ or /z/ when a voicing minimal-pair (e.g., target *dose*, minimal-pair *doze*) either was or was not available as an alternative in the context. The results indicate that talkers enhance the durational cues associated with the word-final voicing contrast based on whether the context requires it, and that this can involve both elongation as well as shortening, depending on what enhances the contextually-relevant contrast. This suggests that talkers are capable of targeted, context-sensitive temporal enhancements.

3.1 Introduction

If a spoken word is likely to be misunderstood, a talker may enhance it to make it easier to identify. What strategies do speakers use, and what kinds of enhancements are available to them? Here, we study whether and how talkers make enhancements that target contextually-relevant contrasts.

In particular, we ask two questions. First, if an intended word (e.g., *dose*) is likely to be misunderstood in a particular speech context (e.g., because its minimal-pair *doze* is another word available in the context), do talkers selectively enhance those aspects of the signal that increase the contrast between the two words? Second, how are these enhancements realized phonetically? For example, can talkers dynamically elongate and shorten parts of the signal in order to enhance contrasts? Or, does targeted hyperarticulation involve only proportional elongation, as is typical of more global hyperarticulation (Smiljanić and Bradlow, 2008, 2009)? These questions inform broader debates about whether and how communicative goals influence articulation (Lindblom, 1990; Jaeger, 2013), and whether targeted hyperarticulation differs from more general modes of clear speech (Ohala, 1994).

3.1.1 Targeted enhancement of contextually-relevant contrasts

Previous research has shown that when only part of an utterance is misunderstood, talkers focus their hyperarticulation on the misunderstood word (Oviatt et al., 1998; Stent et al., 2008). Further, if only a single segment has been misunderstood (or is likely to be misunderstood), talkers may limit their enhancements to that segment. When a speaker needs to be clear that the intended word is

aspirated *pat* and not unaspirated *bat*, they lengthen only the aspiration of the /p/ (Baese-Berk and Goldrick, 2009; Kirov and Wilson, 2012; Schertz, 2013; Buz et al., 2014). Thus, it has been argued that talkers selectively enhance parts of the signal, based on the context and the needs of the communicative situation (Jaeger, 2013; Schertz, 2013; Buz et al., 2014).

Yet not all studies have found explicit support for targeted enhancements. While there is strong evidence for context-sensitive elongation of word-initial aspiration and prevoicing (see references above), and of vowel length contrasts (de Jong, 2004; Schertz, 2013), findings have been less straightforward for coda voicing contrasts. For example, de Jong (2004) finds that the difference between voiced and voiceless codas increases (as cued by vowel duration) when talkers put focus on the voicing contrast. However, in the same study, the voicing contrast is enhanced even more when talkers put focus on a vowel quality contrast instead. One interpretation of this result is that talkers are limited in their ability to dynamically adjust temporal relations in the rime to enhance relevant contrasts (see de Jong, 2001). Alternatively, when talkers focused on the coda voicing contrast, they may have been primarily targeting different cues for enhancement, as opposed to vowel duration (see Stevens et al., 1992). This could explain why the vowel-duration effect was unexpectedly small during enhancement of the coda voicing contrast.

Thus, in the current study, we investigate enhancement of coda voicing contrasts. We measure a number of cues that may be targeted for enhancement. In particular, we evaluate temporal enhancement of the English word-final /s, z/ contrast (e.g., *dose* versus *doze*), which potentially involves at least three differences in the rime. Voiceless /s/ is longer than voiced /z/, but the nucleus vowel before coda /s/ is much shorter than before coda /z/. In addition, although voiced obstruents are typically partially or fully devoiced in this position, voicing may be

maintained longer into /z/ than into /s/ (Derr and Massaro, 1980; Smith, 1997; see also Stevens et al., 1992). Besides these temporal differences, spectral differences between /s/ and /z/ are illustrated in Maniwa et al. 2009, and spectral enhancement of sibilants is investigated by Silbert and de Jong (2008); Maniwa et al. (2009); Julien and Munson (2012); Clayards and Knowles (2015).

3.1.2 Realization of enhancements

There are several ways that talkers might enhance final /s/ and /z/ to increase the contrast between the two. What strategies do they actually use for targeted enhancements? For global temporal enhancements—clear speech that is not targeted to a particular lexical competitor—talkers elongate segments in proportion to their durations in unenhanced, conversational speech, as well as inserting more and longer pauses (de Jong, 2001; Smiljanić and Bradlow, 2008, 2009). For targeted temporal enhancements, previous work has found mainly elongation of particular segments or acoustic cues (Oviatt et al., 1998; Silbert and de Jong, 2008; Baese-Berk and Goldrick, 2009; Julien and Munson, 2012; Schertz, 2013).

This might suggest that talkers generally slow their speech rate when making enhancements, but just do so over a longer or shorter stretch of speech, depending on how much is relevant to the context. For example, Kirov and Wilson (2012) find that English /p/ aspiration is lengthened not only when contrasting target *peak* with *beak*—where longer aspiration is a direct cue to the difference—but also when contrasting target *peak* with *teak*, where longer aspiration is not a direct cue to the difference. While it is plausible that lengthening improves other cues to the place contrast (see also Kirov and Wilson, 2012), or that lengthening is a side-effect of a more direct enhancement of the place contrast, it remains possible that general elongation of a crucial segment is the default strategy for enhancing

a contrast. Under this account, talkers would be predicted to elongate both coda /s/ and /z/ in our experiment when the coda contrast is contextually-relevant.

Alternatively, talkers might be able to make more dynamic temporal adjustments, beyond elongation of the intended segment. For example, they might shorten the short vowel before /s/ and lengthen the long /s/ coda in *dose*, but lengthen the vowel before /z/ and shorten the coda in *doze*. In this way, they would adjust the overall duration of voicing in the signal when the voicing contrast is relevant (Keyser and Stevens, 2006).

3.1.3   The current study

We asked experimental participants to produce /s, z/-final words in two contexts. In one context, participants had to be sure that a listener would not confuse a target word with its voicing minimal-pair (e.g., target *dose* must be differentiated from *doze*, or vice-versa). In the other context, there was no potential to confuse the target word with its voicing minimal-pair.

We measured vowel duration, coda duration, and the duration of voicing in the coda. Because coda voicelessness is cued by a short vowel but a long coda (and vice versa for voicedness; Derr and Massaro, 1980), these stimuli allowed us to evaluate whether talkers invariably use elongation to make words easier to identify, or whether they might use a different strategy to enhance a contrast where across-the-board elongation would be less useful.

3.2   Methods

Forty participants were recruited through Amazon Mechanical Turk to play an online communication game with a partner. A browser-based Flash application

captured audio at 22.05 kHz with 16-bit depth from participants' microphones. All participants reported being native English speakers.

The task was modeled after Baese-Berk and Goldrick (2009), using the software and paradigm developed in Buz et al. (2014). On each trial, a participant would see three words on their screen. After 1.5 seconds, one of the three words was highlighted by the computer. The participant was asked to verbally produce this word (the target) for a partner, who could also see the three words but did not know which of the three was the target. They were told that their partner would listen to the word that they said, and try to choose it from the three possibilities. The partner was simulated by the computer, and always chose the target word.

The critical targets were /s/- or /z/-final words. In *contrastive* trials, one of the two alternative words was the voicing-final minimal pair of the target. In *control* trials, neither of the two alternatives was the voicing-final minimal pair. For example, when the target was /s/-final *dose*, the control and contrastive conditions had the following words (with the target word outlined):

| *Control* | bade | dose | maul |
| *Contrastive* | doze | dose | maul |

When the target was /z/-final *doze*, the control and contrastive conditions had these words:

| *Control* | mode | doze | maul |
| *Contrastive* | dose | doze | maul |

Thus, in the contrastive condition, participants were aware that their partner must identify whether the target is voiceless-final *dose* or voiced-final *doze*. Based on previous work using this method (Baese-Berk and Goldrick, 2009; Kirov and Wilson, 2012; Buz et al., 2014), we expected that they would target the voicing contrast for enhancement. On the other hand, in the control condition, while

Table 3.1: Critical target words.

/s/	base	bus	cease	deuce	dice	dose	face	fuss	goose
/z/	bays	buzz	seize	dues	dies	doze	phase	fuzz	goos

/s/	hiss	mace	moose	niece	noose	pace	piece	spice	vice
/z/	his	maze	moos	knees	news	pays	peas	spies	vies

participants may speak clearly, the coda voicing contrast does not need to be enhanced.

3.2.1 Items

The critical targets were 18 word-final /s, z/ minimal pairs, listed in Table 3.1. Half of the participants saw only the /s/-final words, and half saw only the /z/-final words. Matched pairs were chosen so that, with the exception of any contrastive enhancements, the vowel and coda durations would be as similar as possible.

Participants saw each target word in only one trial. Each participant saw half of their targets in a contrastive trial, and half in a control trial. In addition to the critical trials, participants saw 39 filler trials. In 9 of these trials, the filler target was a minimal pair with one of the other words in the trial, to avoid drawing special attention to the /s, z/ voicing contrast. Fillers and critical trials were presented according to a pseudo-randomized list, with half of the participants seeing the list in backwards order (following Buz et al., 2014). Target presentation was balanced so that the highlighted targets appeared roughly equally often in all three positions on the screen (left, center, or right).

3.3 Analysis

Because participants recorded themselves with their own laptop or peripheral microphone, the recording quality was variable. Of the forty included participants, 20 used a built-in laptop or desktop microphone, 15 used a head-mounted microphone, 4 used a peripheral desk-mounted microphone, and one participant declined to provide information. Because the experimental condition (control versus contrastive) was manipulated within-participant, uneven recording quality across participants could not have biased the results.

Prior to annotation, seven participants (12.5%) were excluded because of a technical issue (such as failing to upload audio), excessively noisy recordings, or failing to follow directions. Nine additional participants (16.1%) were excluded after they wrote on a debriefing questionnaire that they suspected their partner was simulated. Discussion of the questionnaire, and of the believability of the simulated partner in this paradigm, is provided in Buz et al. (2014).

The experiment was run until reaching 20 included /s/ participants and 20 included /z/ participants, totaling 720 productions of critical targets (40 subjects ∗ 18 items). All exclusion criteria were identical to Buz et al. (2014).

3.3.1 Annotation

Four annotators marked the vowel and coda segment boundaries. Annotators were naïve to the trial condition. Vowel onsets were marked at the onset of periodicity, or at the onset of dark formant bands if the preceding segment was voiced. Vowel offsets were marked at the onset of sibilant noise in the range above 3500 Hz. Coda segments were marked from the onset to the offset of sibilant noise

above 3500 Hz immediately following the vowel.

To assess inter-annotator agreement, all annotators segmented productions from a test set of two /s/ participants and two /z/ participants. Pearson's r was calculated between each pair of annotators within each participant. For vowel durations, the mean pairwise r values were 0.87, 0.95, 0.99, and 0.99 for the four test participants. For coda durations, the mean pairwise r values were 0.65, 0.72, 0.89, 0.96. Because of the lower agreement rates for coda durations, the coda segment results should be interpreted cautiously, as the noisier annotations inflate the Type II error rate.

Thirteen productions (1.8%) were removed from all analyses because of heavy audio clipping or another recording issue that prevented segmentation (such as a loud background noise at a segment boundary), or because the word was cut off in the recording. Twelve productions (1.7%) were removed from all analyses because the participant said the wrong word, no word, or more than one word. Five more productions (0.1%) were excluded from the vowel analyses because the vowel duration was more than 2.5 standard deviations from the participant's mean. Eleven productions (1.5%) were excluded from the codas analyses because the coda duration was more than 2.5 standard deviations from the participant's mean.

After annotating the vowel and coda boundaries, Praat was used to count the total number of voiced 10 ms frames in each coda fricative (Boersma, 1993; Boersma and Weenink, 2014). Figure 3.1 shows vowel durations, coda durations, and coda voicing proportions for the /s/- and /z/-final target words.

3.3.2 Models and results

Vowel durations, coda durations, and coda voicing durations were analyzed in separate linear mixed-effects models. Fixed effects in all three models were
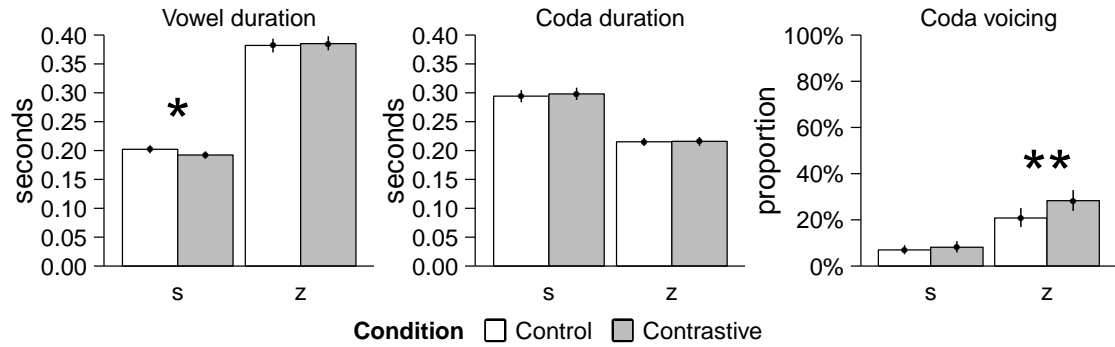
Figure 3.1: Mean segment durations and coda voicing proportions, by experimental condition. Error bars show bootstrapped 95% confidence intervals of the means.

phonological coda voicing (voiceless or voiced) and critical trial type (control or contrastive), plus the interaction. Models also included by-participant intercepts and slopes for critical trial type, and by-item intercepts and slopes for all three fixed effects. For the random groupings, a minimal pair (e.g., *dose* and *doze*) was treated as a single item, but all results were the same if /s/ and /z/ stimuli were modeled separately (without parameters for phonological voicing), or if they were modeled together but treated as different items. *p*-values were calculated using the Satterthwaite approximation for degrees of freedom. Results were qualitatively the same with and without log-transformation. Segment duration models were planned analyses; the coda voicing duration analysis was added post-hoc in response to the low agreement rates for the coda offsets.

Vowel durations: Vowels were significantly shorter overall in /s/-final words compared to /z/-final words ($\hat{\beta} = -187$ms, $t = 11.5$, $p < 0.0001$). Crucially, for /s/-final words, vowels were also significantly shorter in the contrastive condition, where the target word's voicing-final minimal pair was present, compared to the control condition ($\hat{\beta} = -9$ms, $t = 2.3$, $p < 0.05$). For /z/-final words, vowel durations were not significantly different between the control and contrastive conditions ($p > 0.7$).

Coda durations: Coda /s/ was significantly longer overall than coda /z/ ($\hat{\beta}$ = 79ms, t = 4.9, p < 0.0001). There was no significant difference in coda durations between conditions for either /s/-final words (p > 0.2) or /z/-final words (p > 0.9).

Coda voicing durations: Coda voicing was maintained significantly longer in /z/ than in /s/ ($\hat{\beta}$ = 2.6 frames, t = 2.8, p < 0.01). Crucially, for /z/ words, voicing was also maintained significantly longer in the contrastive condition compared to the control condition ($\hat{\beta}$ = 1.6 frames, t = 4.0, p < 0.001; also significant after family-wise error correction). This model was fit without by-item slopes for phonological voicing, since it did not converge with slopes. For /s/-final words, coda voicing durations were not significantly different between the control and contrastive conditions (p > 0.7).

3.4   Discussion

Talkers produced relatively shorter vowels before voiceless /s/ and maintained voicing longer into voiced /z/ when the voicing contrast was contextually-relevant. Our first question was whether talkers selectively enhance aspects of the signal that increase a relevant contrast. Both of the timing changes that we found increase the contrast between the target word and its voicing-final minimal pair. This indicates that talkers make selective, context-specific enhancements in our contrastive condition, when targeting the coda voicing contrast. This extends similar findings on other contrasts, such as voice-onset time in word-initial plosives, the tense-lax English vowel distinction, and some spectral measures that distinguish fricatives (Maniwa et al., 2009; Kirov and Wilson, 2012; Schertz, 2013; Buz et al., 2014; Clayards and Knowles, 2015). Our second question was whether targeted

hyperarticulation invariably uses the elongation processes that are typical of more global hyperarticulation. The results suggest that this is not the case: talkers are capable of dynamic temporal enhancements in particular contexts where across-the-board or proportional elongation of a word or segment may be less helpful.

Both effect sizes are comparable to the enhancement of prevoicing and aspiration durations in tasks where talkers explicitly clarify and repeat words that have been misidentified by a listener (Schertz, 2013). This suggests that the effects observed here were plausibly intended to improve lexical identification. Buz et al. (2016) argue that such contrastive enhancement needs to be understood with regard to the avoidance of perceptually-ambiguous productions near phonetic category boundaries, where even smaller durational changes are known to affect comprehension (McMurray et al., 2002). It is a question of ongoing research as to what kinds of enhancements serve to facilitate comprehension (Uchanski, 2008; Smiljanić and Bradlow, 2009).

One view of the two effects is that talkers may use different strategies to enhance coda voicelessness (vowel shortening) versus voicedness (longer phonetic voicing). Alternatively, it may be that the enhancement target is the overall voicing duration or the relative timing of the voicing offset within the word, rather than the duration of the vowel or coda individually (Keyser and Stevens, 2006; Choi et al., 2015; see Massaro and Cohen, 1977; Stevens et al., 1992 on perception).[1] This would help explain why previous work that investigates enhancement of the coda-voicing contrast has not found the expected effect on vowel duration alone.

---

[1]To evaluate whether talkers initiate devoicing earlier before voiceless codas when they are contrasted with voiced ones (cf. Clayards and Knowles, 2015, on prominence effects), we conducted a post-hoc analysis of voicing duration during the vowel. Voicing duration was shorter before /s/ in the contrastive condition relative to the control condition ($\hat{\beta} = -9$ms, $p < 0.05$). However, most vowel productions in our data were fully voiced, and the duration of vowel voicing was highly correlated with the duration of the vowel ($r = 0.96$), making it impossible to distinguish this effect from the effect on vowel duration reported in the text.

For example, Goldrick et al. (2013) found no lexically-mediated enhancement of vowel duration (after controlling for onset aspiration), de Jong (2004) found that the vowel duration contrast between voiced- and voiceless-final words was enhanced more weakly in the context of voicing-relevant competitors (contrasting *bat–bad*, *bet–bed*) than voicing-irrelevant ones (contrasting *bat–bet*, *bed–bad*), and Choi et al. (2015) found no vowel duration enhancement in the context of voicing-relevant competitors.

More generally, our findings provide evidence that selective, context-specific enhancement is not limited to onsets, and point to ways in which talkers may use different phonetic strategies for targeted enhancements as compared to more global hyperarticulation.

3.5   Acknowledgments

We thank Anna MacDonald, Aishwarya Krishnamoorthy, Brian Leonard, and Lindsey Harris for help with segmentation; and Meghan Clayards, Marc Garellek, Matthew Goldrick, the phonetics and phonology group at UCSD, and audiences at the University of Geneva and CUNY 2015 for helpful discussion and comments. This work was supported by an NSF Graduate Research Fellowship to SJS, an NRSA Predoctoral Fellowship (F31HD083020) to EB, and an NSF CA-REER (IIS-1150028) to TFJ. The views expressed here do not necessarily reflect those of the funding agencies.

<div align="center">* * *</div>

Chapter 3, in full, is a reprint of the material as it appears in Seyfarth, Buz, and Jaeger (2016) [Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustical Society of America*, 139 (2), EL31–37. doi:10.1121/1.4942544]. This work was reported, in part, at the 28th Annual CUNY Conference on Human Sentence Processing, and at the 15th Conference on Laboratory Phonology. This chapter reports co-authored work, which is used here with permission from the co-authors. The dissertation author was the primary investigator and author of this paper.

3.6 Bibliography

Baese-Berk, M. and Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4):527–554.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, volume 17, pages 97–110. University of Amsterdam. Available online at http://www.fon.hum.uva.nl/Proceedings/Proc_17_Postscript/Proc_Boersma_1993.ps

Boersma, P. and Weenink, D. (2014). Praat: Doing phonetics by computer. Available online at http://www.fon.hum.uva.nl/praat/

Buz, E., Jaeger, T. F., and Tanenhaus, M. K. (2014). Contextual confusability leads to targeted hyperarticulation. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Buz, E., Tanenhaus, M. K., and Jaeger, T. F. (in press). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89: 68–86.

Choi, J., Cho, T., Kim, S., Baek, Y., and Jang, J. (2015). Phonetic encoding of coda voicing contrast and its interaction with information structure in L1 and L2 speech. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Clayards, M., and Knowles, T. (2015). Prominence enhances voicelessness and not place distinction in English voiced sibilants. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32(4):493–516.

de Jong, K. J. (2001). Effects of syllable affiliation and consonant voicing on temporal adjustment in a repetitive speech-production task. *Journal of Speech, Language, and Hearing Research*, 44(4):826–840.

Derr, M. A. and Massaro, D. W. (1980). The contribution of vowel duration, f0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics*, 27(1):51–59.

Goldrick, M., Vaughn, C., and Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*, 134(2):EL172–EL177.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, 4:1–4.

Julien, H. M. and Munson, B. (2012). Modifying speech to children based on their perceived phonetic accuracy. *Journal of Speech, Language, and Hearing Research*, 55(6):1836–1849.

Keyser, S. J. and Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language*, 82(1):33–63.

Kirov, C. and Wilson, C. (2012). The specificity of online variation in speech production. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, Sapporo, Japan.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling*, pages 403–439. Springer.

Maniwa, K., Jongman, A., and Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America*, 125(6):3962–3973.

Massaro, D. W. and Cohen, M. M. (1977). Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Perception & Psychophysics*, 22(4):373–382.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86:B33–B42.

Ohala, J. J. (1994). Acoustic study of clear speech: A test of the contrastive hypothesis. In *Proceedings of the International Symposium on Prosody*, pages 75–89, Yokohama, Japan.

Oviatt, S., Levow, G.-A., Moreton, E., and MacEachern, M. (1998). Modeling global and focal hyperarticulation during human–computer error resolution. *Journal of the Acoustical Society of America*, 104(5):3080–3098.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, 41(3-4):249–263.

Silbert, N. and de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America*, 123(5):2769–2779.

Smiljanić, R. and Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, 36(1):91–113.

Smiljanić, R. and Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, 3(1):236–264.

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics*, 25(4):471–500.

Stent, A. J., Huffman, M. K., and Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, 50(3):163–178.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America*, 91(5):2979–3000.

Uchanski, R. M. (2008). Clear speech. In D. Pisoni & R. Remez (Eds.), *The Handbook of Speech Perception*, pages 207–235, Blackwell.

Chapter 4

Acoustic differences in morphologically-distinct homophones

ABSTRACT

Previous work demonstrates that a word's status as morphologically simple or complex may be reflected in its phonetic realization. One possible source for these effects is phonetic paradigm uniformity, in which an intended word's phonetic realization is influenced by the articulatory plans of its morphological relatives. For example, the realization of the English inflected word *frees* should be influenced by the plan for *free*, and thus be non-homophonous with the morphologically-simple word *freeze*. We test this prediction by analyzing productions of forty such inflected/simple word pairs, embedded in pseudo-conversational speech structured to avoid metalinguistic task effects, and balanced for frequency, orthography, as well as segmental and prosodic context. We find that stem and suffix durations are significantly longer by about 4–7% in fricative-final inflected words (*frees*, *laps*) compared to their simple counterparts (*freeze*, *lapse*). This suggests that such words are influenced by the articulatory plans of their phonologically-lighter stems. The result provides new evidence for a mechanism of interaction between morphology and phonetic realization.

4.1   Introduction

When a language-user produces a spoken word, its exact articulation is influenced by a wide range of linguistic and psycholinguistic variables, such as the word's position in a phrase (Oller, 1973), overall frequency in the language (Gahl, 2008), and its predictability in context (Lieberman, 1963; Bell, Brenier, Gregory, Girand & Jurafsky, 2009). How do the morphological properties of a word influence its phonetic realization? Modular, sequential processing architectures (e.g., Levelt, Roelofs & Meyer, 1999) and other non-interactive models of language production (e.g., Kiparsky, 1982) propose that when phonetic attributes such as duration and pitch are encoded from a phonological representation, the word's morphological status is inaccessible. However, a growing body of work demonstrates that morphology does interact with phonetic characteristics such as formant trajectory alignment (Scobbie, Turk & Hewlett, 1999), /l/-darkening (Sproat & Fujimura, 1993; Hayes, 2000; Lee-Kim, Davidson & Hwang, 2013; Strycharczuk & Scobbie, 2015), and segment duration (Pluymaekers, Ernestus, Baayen & Booij, 2010; Smith, Baker & Hawkins, 2012; Plag, Homann & Kunter, 2015). For example, the /t/ is aspirated in the derived word *mistime*, but not in the morphologically-simple word *mistake*, even though it occurs in the same phonological environment in both words (Baker, Smith & Hawkins, 2007; Smith et al., 2012; Zuraw & Peperkamp, 2015).

What causes these effects? One possible mechanism is *phonetic paradigm uniformity*: the articulatory influence of an intended word's morphological relatives on the articulatory plan of that word (Hayes, 2000; Steriade, 2000; Frazier, 2006; Ernestus & Baayen, 2006; Roettger, Winter, Grawunder, Kirby & Grice, 2014). There is some existing evidence for morphological family effects on speech

production latencies (Hay & Baayen, 2005; Baayen, Levelt, Schreuder & Ernestus, 2007). However, these effects are nonetheless compatible with a model that segregates morphology and phonetics, in that they arguably involve competition during word or formative planning processes, rather than during articulation itself (see Goldrick, Baker, Murphy & Baese-Berk, 2011; Goldrick, Keshet, Gustafson, Heller & Needle, 2016). In this paper, we argue that an intended word's morphological relatives also interact with that word's phonetic realization, and test this hypothesis by looking at the durational influence of freestanding English stems on the wordforms in their inflectional paradigms. This work addresses broader questions about interaction among different components of the linguistic signal, and the role of analogy between wordforms in phonetic and phonological representation.

4.1.1 Paradigm uniformity

A morphological paradigm is the set of wordforms that have a lexeme in common. For example, the inflectional paradigm of the English verb FREE is *free*, *frees*, *freeing* and *freed*. Paradigm uniformity is a pressure for invariance among the phonological forms of an inflectional or derivational paradigm (Hayes, 2000; Steriade, 2000). This phenomenon occurs in the pronunciation of the American English words *capitalistic* and *militaristic*. The unstressed syllable /tə/ in the word *capitalistic* is normally produced with an alveolar tap [ˌkæpɪɾəˈlɪstɪk]. This follows the phonological pattern in which intervocalic /t/ is flapped when it is unstressed. However, the same syllable /tə/ in the word *militaristic*—which is unstressed, just as in *capitalistic*—can be pronounced with an aspirated [t] ([ˌmɪlɪtʰəˈɹɪstɪk]), even though this violates that phonological pattern (Withgott, 1982). This can be accounted for by uniformity pressures within the two words' derivational paradigms (Steriade, 2000). The syllable that corresponds to /tə/ is unstressed in *capital*

/ˈkæpɪtl̩/, but is stressed in *military* /ˈmɪlɪˌtɛɹi/. Even though /tə/ is unstressed in the derived *militaristic*, the pressure for paradigmatic uniformity with *military* prevents it from being realized as a tap in *militaristic*. On the other hand, there is no such influence on *capitalistic*, because the /t/ is also realized as a tap in *capital* (see also Davis, 2005, for a different uniformity-based analysis).

While paradigm uniformity has been formalized in several symbolic phonological theories (e.g., Benua, 1997; McCarthy, 2005, see Steriade, 2000 for a summary), it has also been argued to influence more fine-grained production patterns (Hayes, 2000; Steriade, 2000; Frazier, 2006). As one instance, paradigm uniformity may account for incomplete voicing neutralization patterns in Germanic languages (Ernestus & Baayen, 2006, 2007; Winter & Röttger, 2011; Roettger et al., 2014; Kaplan, 2016). For example, the German words *Rad* 'wheel' and *Rat* 'council' are typically considered to be homophones, ending in a final voiceless segment: both are pronounced [ʁaːt]. *Rad* is morphologically related to *Räder* 'wheels', in which the corresponding segment is voiced [d] ([ʁæːdɐ]). However, *Rat* has no such voiced relative. A body of research demonstrates that there are fine-grained phonetic differences between *Rad* and *Rat* such that *Rad*, but not *Rat*, is produced with some of the phonetic cues associated with a final voiced segment (see Winter & Röttger, 2011 and Roettger et al., 2014 for recent reviews and experimental data). One account for these results is that incomplete neutralization is the result of paradigm uniformity effects (e.g., operationalized in terms of spreading activations or lexical analogy, Ernestus & Baayen, 2006, 2007; Winter & Röttger, 2011; Roettger et al., 2014). When a speaker produces the form *Rad*, the morphologically-related voiced form *Räder* is also activated, and influences how voicing cues are realized in *Rad*.

4.1.2 Paradigm effects on English inflected forms

One measurable type of uniformity effect that has been proposed is word and segment duration (Steriade, 2000; Frazier, 2006). There is evidence that durational targets are specified in a wordform's phonological plan (Tauberer & Evanini, 2009; Katz, 2010, 2012; Seyfarth, 2014), and duration has previously been used as a test case for competition among articulatory plans (Goldrick et al., 2011). Here, we investigate the effects of monosyllabic English stems such as *free* on an inflected paradigm member with a heavier coda, such as *frees*. The paradigm uniformity account predicts that the timing of the segments in *frees* should be influenced by the durational targets of *free* (see Frazier, 2006 for a similar proposal based on moraic structure). As a baseline for what the timing of *frees* should be if there were no interference from paradigm members, we compare each inflected word to a segmentally-identical but morphologically-simple homophone, such as *freeze*.

The inflected word *frees* should show the following uniformity effects from the influence of *free*. First, the word *free* [fɹi] has no coda, and therefore the nucleus is longer than if it were in a closed syllable (Munhall, Fowler, Hawkins & Saltzman, 1992; Shaiman, 2001; Katz, 2010, 2012). If the phonetic plan for the stem *free* influences the plan for *frees*, the nucleus should be relatively longer in *frees* compared to *freeze*, where it is not influenced by a longer stem word (Frazier, 2006).

Second, *free* constitutes a full prosodic constituent when it occurs as a freestanding word. As a consequence, there should be some degree of lengthening in *free* adjacent to the stem-final prosodic boundary (Oller, 1973; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992). This should influence the production of its morphological relative *frees*, so that the corresponding string [fɹi] will be

overall longer in *frees* compared to *freeze* (cf. Sugahara & Turk, 2009). In particular, we expect that final lengthening should be greatest on segments immediately adjacent to the boundary (Shattuck-Hufnagel & Turk, 1998; Byrd & Saltzman, 2003; Byrd, Krivokapić & Lee, 2006). Thus, the vowel should be lengthened more in *free* than in *freeze*, and uniformity effects should produce the same distinction between *frees* and *freeze*.

The prosodic boundary at the end of the freestanding stem *free* may also influence the production of the suffix in the related form *frees*. If there is a gradient internal boundary, the post-boundary suffix [z] should be lengthened relative to the corresponding [z] suffix in *freeze*, due to domain-edge strengthening effects (Fougeron & Keating, 1997; Turk & Shattuck-Hufnagel, 2000; Keating, 2006). In particular, Cho, Lee & Kim (2014) show domain-initial durational lengthening of English [s]; however, we also note that there is language-specific variation as to what segment classes undergo lengthening in initial position (Oller, 1973; Fougeron, 2001; Keating, 2006; Hualde & Prieto, 2014; Strycharczuk & Kohlberger, 2016), and it is somewhat less clear whether there should be lengthening for the English [t, d] inflectional suffixes in this study. We might also expect lengthening of the post-boundary inflectional suffixes because they are immediately following a prosodic boundary, as domain-final lengthening should extend rightward from the boundary, regardless of initial strengthening effects (Shattuck-Hufnagel & Turk, 1998; Byrd et al., 2006).

4.1.3   Previous evidence

Prior work has compared the durations of simple and inflected English homophones (*tax/tacks*), but with unclear results. Two laboratory studies report that suffix durations are longer in inflected words than their simple homophones

(Walsh & Parker, 1983; Losiewicz, 1992), and two report the same pattern for vowel or stem durations (Frazier, 2006; Sugahara & Turk, 2009, as well as mixed results discussed in Sugahara & Turk, 2004). However, a major concern with the interpretation of these findings is that the word productions were elicited in short lists of homophones and in short phrases intentionally designed to highlight contrasts between the target words. It has been shown that phonetic variation between orthographically-distinct homophones increases when the target homophones are dictated in an isolated-word list or in contrastive sentences, as compared to when the target words are disguised in longer contexts (Fourakis & Iverson, 1984; Port & Crawford, 1989; Kharlamov, 2014, see also Winter & Röttger, 2011; Roettger et al., 2014). Thus, while the participants in these studies may have been encouraged by the experimental design to produce phonetic distinctions, those distinctions may have been motivated by orthography or metalinguistic knowledge as much as by the words' morphological properties (Fourakis & Iverson, 1984; Jassem & Richter, 1989; Kharlamov, 2014; Mousikou, Strycharczuk, Turk, Rastle & Scobbie, 2015, see also Warner, Jongman, Sereno & Kemps, 2004; Warner, Good, Jongman & Sereno, 2006; Ernestus & Baayen, 2006).

More broadly, the generalizability of previous reports has also been criticized (see Bermúdez-Otero, 2010; Hanique & Ernestus, 2012; Plag, 2014; Mousikou et al., 2015; Plag et al., 2015), including the findings of one corpus study that reports longer suffix durations for inflected words (Song, Demuth, Evans & Shattuck-Hufnagel, 2013). These studies have often tested very few items (3 homophone pairs in Walsh & Parker, 1983; 6 pairs in Losiewicz, 1992; 9 non-homophonous words in Song et al., 2013), found the effect only at a slow speech rate (Sugahara & Turk, 2009), only utterance-finally (Song et al., 2013), or were not robust to current statistical practices (Plag, p.c., on Losiewicz, 1992). Additionally, the

inflected and simple words in the prior laboratory work were not balanced for frequency, which is well-known to influence acoustic duration.[1] Several authors also raise a concern about orthographic differences (Fourakis & Iverson, 1984; Winter & Röttger, 2011 on incomplete neutralization; Sugahara & Turk, 2004, 2009; Mousikou et al., 2015 on duration), which might affect production independently of morphological status (Warner et al., 2004, 2006; Ernestus & Baayen, 2006; Brewer, 2008; Bürki, Spinelli & Gaskell, 2012).

In addition to these concerns, a recent study of a larger number of non-homophonous inflected and simple words in conversational speech reports the *opposite* pattern for English [s] suffix durations: final [s] is *shorter* when it signals an inflectional suffix (Plag et al., 2015). Since this study found the opposite pattern than found in prior laboratory experiments, one interpretation is that the experimental work may have been confounded by task effects or other methodological issues. At the same time, a corpus-based analysis raises a different set of analytical and interpretability challenges due to the heterogenous word types in the data, as well as the unbalanced prosodic contexts that English inflected and uninflected words tend to appear in. We return to these questions in the discussion §4.4.3.

4.1.4 The current study

In the current study, we analyze the stem and suffix durations in forty pairs of monosyllabic English homophones, in which one member of the pair is inflected (*frees*) and the other is morphologically-simple (*freeze*). Under the paradigm uni-

---

[1]For example, Frazier (2006) found that vowel durations were longer in inflected words than in morphologically-simple homophones. However, the log wordform frequency in the SUBTLEX-US corpus (Brysbaert & New, 2009; Brysbaert, New & Keuleers, 2012) was significantly greater for the morphologically-simple words ($\mu = 2.90$) compared to the inflected words in that study ($\mu = 2.02$; unpaired $t(32) = 2.48$, $p < 0.05$; excluding two inflected words *brayed* and *rued* which have zero frequency in SUBTLEX). Losiewicz (1992) had the same confound (Hanique & Ernestus, 2012); and see also discussion in Sugahara & Turk (2004, 2009).

formity account, the prediction is that stem durations should be relatively longer in inflected words like *frees*, compared to simple *freeze*, due to the prosodic influence of a lighter stem word (*free*) on the inflected but not simple words. This account also predicts that the suffix duration in *frees* should be lengthened relative to the same segment in *freeze* as a result of domain-edge strengthening effects. Because the theory predicts no differences between different morphological suffixes, we include a variety of [s, z, t, d] English suffixes in our stimuli set. Besides these planned tests, we also use the data to explore the influence of probabilistic variables on inflected words (Hay, 2003; Cohen Priva, 2012; Schuppler, van Dommelen, Koreman & Ernestus, 2012; Cohen, 2014; Rose, Hume & Hay, 2015). In particular, we evaluate whether the predicted influence of morphological relatives is stronger if these relatives are more frequent, in either absolute or relative terms.

In order to elicit more natural speech and avoid metalinguistic task effects, yet still maintain a phonetically-controlled context, we adapt a method used by Port & Crawford (1989), Baker et al. (2007), and Smith et al. (2012) in which the homophone pairs are embedded in conversational dialogues that are matched for prosodic and segmental context. These dialogues are read by pairs of naïve participants who are already friends, and who are familiarized with the dialogues prior to participation in the experimental task (see Warner, 2012).

This method has several crucial advantages over previous work. First, by using controlled dialogues rather than completely spontaneous speech, we are able to collect productions of a large number of homophone pairs which are matched for frequency and orthographic length, factors that have been potential confounds in previous work (Hanique & Ernestus, 2012; Plag, 2014; Mousikou et al., 2015; Plag et al., 2015). The use of homophone pairs allows us to compare matched stems, and to be sure that durational differences are truly independent of segmental content.

Second, by concealing the task and target words within a meaningful conversation, speakers are unlikely to explicitly attend to orthographic or morphological differences in the targets. In particular, because Plag et al. (2015)—who looked at spontaneous speech—found a durational effect in the opposite direction as previous experimental work, this method allows us to evaluate whether that difference can be ascribed to metalinguistic task effects in the lab reading experiments. While our hypothesis does not predict the result in Plag et al. (2015), if we do replicate their effect in an experiment that uses conversational speech styles, it would suggest that task effects did in fact confound prior experimental work, and thus help reconcile that study's findings with earlier work.

Third, our hypothesis requires that the target words be parsed as having both prosodic and morphological structure. By embedding the words in a meaningful conversational dialogue, speakers are much more likely to generate an appropriate prosody and morphological parse, as compared to if they produce items from a word list or in a fixed carrier phrase.

4.2 Methods

4.2.1 Participants

Forty participants were recruited from the UC San Diego community. They each brought a friend to the experiment, and together read through a list of short conversational dialogues that included the target words. All participants and their friends reported that they had started learning to speak English before age 6. They gave informed consent, and received course credit in exchange for participation.

4.2.2 Stimuli

The target words were 40 pairs of English homophones in which one member of the pair was uninflected, and the other had an inflectional suffix. 26 pairs had fricative [s] or [z] suffixes (e.g., plural *lapse/laps*, third-person singular *freeze/frees*) and 14 had stop [t] or [d] suffixes (past *duct/ducked*, participle *tide/tied*).

4.2.2.1 Dialogues

The two homophones in each pair were embedded in phonetically-matched dialogues (Baker et al., 2007; Smith et al., 2012). Each dialogue was a short conversation between two people, and was preceded by a one-sentence description of the scenario in which the conversation took place. For example, the descriptions and dialogues for the target words *freeze/frees* were the following:

Two housemates are wrapping up a surprise birthday party that they put on for a friend.

B: It looks like most people are leaving now. I guess I'm going to start cleaning up a little bit.

A: There's so much cake leftover. I don't want it to go bad.

B: If we freeze it, it should be fine.

Two rural neighbors are talking about a friend, Rich, who is an avid hiker and animal-lover.

B: Rich decided to take care of the injured hawk that he found yesterday.

A: They don't do well in captivity. Wouldn't it be better to let it go?

B: If he frees it, it won't survive.

The complete set of 40 dialogue pairs is given in 4.6. All of the target words received nuclear accent in their phrase. Within each pair of dialogues, each of the two target homophones were preceded by the same number of syllables and stresses in the phrase. If the target homophones were not in the first phrase of the speaker's turn, there were also the same number of syllables, stresses, and phrases between the beginning of the turn and each of the two target homophones. To manage the possibility that a suffix could be resyllabified, the targets were followed by the same segment, or by a phrase boundary.[2] To control for the spread of phrase-final lengthening, each pair of target homophones was followed by the same number of syllables in the phrase and turn. In addition, the targets bore the same type of focus, occurred on the same conversational turn (e.g., the third turn

---

[2] In some cases in which both words were followed by a vowel or semivowel, they had different qualities. For two pairs, the target words were followed by a different segment, but excluding these from the suffix duration analysis did not qualitatively affect the results. Additionally, the target words in 33 of 40 pairs were preceded by the same segment, or else by vowels or semivowels with a different quality.

in the dialogue), and where it was possible, the target words (or their phrases) had the same discourse relation with the preceding utterance.

4.2.2.2 Frequency and orthography

Across pairs, the morphologically-simple and inflected words were not significantly different on log SUBTLEX wordform frequencies (mean of differences $= 0.21$, paired $t(39) = 1.03$, $p > 0.3$), log SUBTLEX word frequency specific to the words' part-of-speech ($\mu_d = 0.75$, $t(39) = 1.46$, $p > 0.15$), or on orthographic length ($\mu_d = -0.33$ letters, $t(39) = 1.65$, $p > 0.1$).

In addition to being matched across the stimuli pairs overall, both frequency measures were matched across the 26 fricative-final pairs alone (frequency: $\mu_d = 0.40$, $t(25) = 1.39$, $p > 0.17$; part-of-speech-specific frequency: $\mu_d = 1.19$, $t(25) = 1.46$, $p > 0.11$) and across the 14 stop-final pairs alone (frequency: $\mu_d = -0.13$, $t(13) = 0.48$, $p > 0.64$; part-of-speech-specific frequency: $\mu_d = -0.05$, $t(13) = 0.08$, $p > 0.93$). Orthographic length was matched overall, as well as across the fricative-final pairs alone ($\mu_d = 0.19$, $t(25) = 0.93$, $p > 0.36$). However, orthographic length was not matched across the stop-final pairs ($\mu_d = -1.29$, $t(13) = -4.84$, $p < 0.001$); we discuss this issue further in §4.4.1.

4.2.2.3 Predictability norming experiment

Beyond the effects of frequency on word and segment durations, it is well-known that words that are predictable in the discourse context are shortened (e.g., van Son & Pols, 2003; Bell et al., 2009). We estimated the contextual predictability of each word by recruiting 40 different participants for a cloze norming task via Amazon Mechanical Turk, using the JavaScript library `jsPsych` (de Leeuw, 2015). On each trial, each cloze participant saw the first part of one dialogue

(including the one-sentence description), which was truncated immediately before the target word. They were asked to complete the partial dialogue with the first word, phrase, or sentence(s) that came to mind. Cloze participants saw half of the 80 dialogues (i.e., only one member of each dialogue pair). We collected 20 individual completion judgments per dialogue.

The predictability of each target word was considered to be the proportion of individuals who wrote down that word immediately following the partial context ($\mu = 0.09$, $\sigma = 0.18$, range $= 0.00$–1.00). On this measure, there was no significant difference between inflected and morphologically-simple words, either across pairs overall or across fricative pairs or stop pairs, by either paired t-test or paired Wilcoxon test (since the distribution of predictability was highly non-normal; all $p > 0.15$). We also used this experiment to estimate the probability of inflectional agreement in the dialogues containing inflected words. For each dialogue containing an inflected word, the probability of inflectional agreement was considered to be the proportion of individuals who wrote down a word with the same inflection immediately following the partial context. For example, in the *frees* dialogue above, the probability of third-person singular agreement was the proportion of participants who completed the truncated phrase in the third turn of the dialogue *"If he . . . "* with any third-person singular verb. We explore these data further in §4.3.3.3 (see also Cohen, 2014; Rose et al., 2015).

4.2.3   Procedure

4.2.3.1   Lists

Each participant pair in the primary experiment read through one of four lists containing half of the dialogues. Each list had 20 inflected targets and 20

simple targets, and included one member of each homophone pair. The first list was constructed by randomly selecting one member of each dialogue pair, and sorting them in a random order. The second list was the mirror-image of the first list (i.e., the first list began with the dialogues containing *prize*, while the second list began with *pries*). To control for possible trial order effects, the third and fourth lists were reversed versions of the first two lists. The order of the dialogues is provided in 4.7.

Lists were randomly assigned to participants so that each list was seen by 10 participant pairs. Participants were given their experimental list at least one day in advance. They were instructed to familiarize themselves with the dialogues and to share the list with their friend before arriving for the experiment. They were asked to try to read the dialogues as conversationally and as naturally as possible. During the recording session, participants were given additional time before each item to silently review each dialogue before reading it out loud. To avoid clear speech styles, participants were told not to worry if they stumbled or misspoke, and just to start over where they left off as they would normally do. This resulted in some excluded data, described below in §4.3.1.

4.2.3.2   Recording

Participants were given the same role (speaker A or speaker B—see example in §4.2.2.1) for all of the dialogues in their list. The target words were always produced by speaker B. Each participant pair sat together in a sound-attenuated booth in a quiet room. Both participants wore head-mounted microphones, and the person given the speaker B role was recorded at a 44.1 kHz sampling rate with 16-bit depth. Although both microphones were set up in the same way, the person assigned to the speaker A role was not recorded.

Table 4.1: Criteria used to mark the onset of the stem region in the target words.

Word-initial segment	Example	Onset boundary
[p, t, k, tʃ, b, d, g, m, n]		beginning of closure
[f]	we **f**reeze	onset of broadband frication noise
[s]	we **s**eize	onset of sibilant noise > 3500 Hz
[h]	the **h**ose	intensity drop following a vowel
[l]	the **l**aps	onset of low intensity trough
[ɹ]	alread**y wr**apped	onset of intensity rise
[oʊ]	a**n o**de	end of preceding nasal closure

4.2.4 Segmentation

Each target word was extracted from the dialogues recorded by the participant pairs, and segmented into two regions. The stem region was the word onset to onset of the final [s, z, t, d] suffix segment. For example, for the words *freeze/frees* [fɹiz], the stem was [fɹi]. For the words *mist/missed* [mɪst], the stem was [mɪs]. The suffix region was the final segment [s, z, t, d].

Segmentation was performed using the waveform and broadband spectrogram view in Praat (Boersma & Weenink, 2016). The acoustic criteria that were used to mark the onset boundary for the stem region are given in Table 4.1, with the following additional procedures. For five pairs, an onset plosive followed another plosive segment (e.g., in the phrase *bad **b**ruise*). If the preceding segment was unreleased, the midpoint of the two-segment closure was used as the stem onset boundary (e.g., the midpoint of the [db] closure in *bad bruise*). For [l]-initial pairs, if the intensity contour was flat, the onset of a low F2 or high F3 plateau was used as a boundary instead.

The [s, z] fricative suffixes were segmented from the onset to the offset of sibilant noise in the range above 3500 Hz. If there was broadband aspiration noise following the sibilant noise, it was not included in the suffix duration. If a plosive

preceded the sibilant (e.g., *lax/lacks* [læks]), the plosive release burst (if any) was not included in the suffix duration.

The [t, d] stop suffixes were segmented from the onset of the closure to the offset of a release burst (if present), or to the end of the closure, if a release burst was not visible. Closure durations were also segmented; all results for [t, d] suffixes were the same when closure durations were analyzed alone. If there was no burst, no closure (complete or incomplete), and a relatively small drop in intensity, the segment was marked as an approximant. If there was also no drop in intensity, no audible percept of a coronal stop, and no visible F2 transition (when adjacent to non-front vowels), it was marked as deleted. Plosives that were part of a coda cluster (e.g., *duct/ducked* [dʌkt]) were segmented beginning after of the first segment's release burst. If no release burst was visible, the midpoint of the two-segment closure was used as the suffix onset boundary.

4.3   Results

The experiment was run until reaching 40 included participant pairs, with a total of 1600 tokens of the target words (1 of 2 words in each of 40 homophone pairs $*$ 40 participants). Data from one additional participant was excluded without being annotated because of a lisp.

4.3.1   Exclusions

65 tokens (4.1%) were excluded from all analyses because the target word was disfluent, which was defined as a hesitation immediately before the word, a mispronunciation or speech error on the target word (whether or not the speaker corrected it), or laughter during the word. 40 additional tokens (2.5%) were ex-

cluded because the speaker misread the target phrase (e.g., they said *had packed it* instead of *had it packed*). 54 tokens (3.4%) were excluded from the duration analyses because the suffix segment was judged to be deleted (see criteria given in §4.2.4; an analysis of deletion rates appears below in §4.3.3.1), as well as 5 other tokens (0.3%) which had no visible landmarks on the spectrogram that could be used for segmentation.

For the stem duration analysis only, 22 tokens (1.4%) were excluded because they were 2.5 standard deviations or more from the mean stem duration of their respective items. For the suffix duration analysis only, 33 (2.1%) [t, d] tokens were excluded because they were approximated, 29 (1.8%) because they were spirantized, and 6 (0.4%) because they were phrase-final but unreleased, which made it impossible to identify the suffix offset. Additionally, 25 tokens (1.6%) were excluded from the suffix duration analysis because they were 2.5 standard deviations or more from the mean suffix duration of their respective items.

4.3.2 Models

Stem and suffix durations were analyzed in separate linear mixed-effects models. Fixed effects were word type (simple or inflected) and suffix manner (fricative [s, z] or stop [t, d]), plus the interaction. These analyses were planned, designed to replicate a significant interaction found with 20 different participants and variant dialogues in a pilot experiment.[3] Models also included by-item intercepts and slopes for word type, and by-participant intercepts and slopes for all three fixed effects. Each homophone pair was treated as a single item for the purpose of random groupings. Mixed-effects models were fit using the R package

---

[3]For the suffix durations, pilot results were the same as those reported here; stem durations and other measures were not analyzed. Pilot results are reported by Seyfarth, Garellek, Malouf, and Ackerman (2015, oral presentation).

`lme4` (R Core Team, 2015; Bates, Maechler, Bolker & Walker, 2015).

Effect sizes were estimated with the R package `lsmeans` (Lenth, 2016) by predicting the appropriate marginal means from each model, and then calculating the difference and p-value for the contrast with the default Kenward-Roger approximation for degrees of freedom (Halekoh & Højsgaard, 2014). The multivariate t distribution was used for multiplicity correction within each family of tests (stem durations and suffix durations).

4.3.2.1 Stem durations

The left panel of Figure 4.1 shows a summary of stem durations, by word type and manner of the suffix. Crucially, for fricative-final words, stem durations were significantly longer in inflected words (*frees*) compared to simple words (*freeze*) ($\hat{\beta} = 18$ms, $t(39.73) = 2.91$, $p < 0.02$). However, for stop-final words, stem durations were not significantly different between word types ($\hat{\beta} = -16$ms, $t(47.43) = 1.82$, $p > 0.14$), with a non-significant effect in the reverse direction.

4.3.2.2 Suffix durations

The right panel of Figure 4.1 shows a summary of suffix durations, by word type and manner of the suffix. Crucially, for fricative suffixes, suffix durations were significantly longer in inflected words compared to simple words ($\hat{\beta} = 6$ms, $t(32.51) = 2.73$, $p < 0.03$). However, for stop suffixes, suffix durations were not significantly different between word types ($\hat{\beta} = -2$ms, $t(33.12) = 0.54$, $p > 0.8$). Results were qualitatively the same if release bursts were excluded from the suffix region, and stop suffix durations were considered to be the closure only.

Figure 4.1: Stem and suffix durations for morphologically-simple and inflected words, by manner of the suffix.

4.3.3 Additional analyses

4.3.3.1 Deletion rates

Final [t, d]-deletion is a well-attested process in American English. Further, a variety of studies have found that final [t, d] are deleted more often when they represent an inflectional suffix (as in *paced*) than when they do not (*paste*) (Labov, Cohen, Robins & Lewis, 1968; Guy, 1980; Neu, 1980; Guy, 1991; Bybee, 2000; Guy, Hay & Walker, 2008, among others). In our data, deletion rates were roughly the same regardless of inflectional status. Excluding disfluent or misread tokens (§4.3.1), $27/249 = 10.8\%$ of inflected [t, d] suffixes were deleted, and $26/251 = 10.4\%$ of morphologically-simple [t, d] suffixes were deleted (plus 1 token that was both misread and deleted).

To evaluate whether the non-effect was biased by particular items or subjects, we fit a logistic mixed-effects model (using the [t, d] data only) to predict

deletion, with a fixed effect of word type (simple or inflected), plus by-item and by-participant intercepts and slopes. The effect of word type was marginally non-significant in the expected direction ($\hat{\beta} = -2.22$, $z = 1.70$, $p < 0.09$). While it is difficult to interpret a null result, the balanced design of the dialogues suggests that the robust differences in deletion rates that have previously been reported may have been partially driven by frequency effects (Guy et al., 2008) or by the different segmental contexts in which simple and inflected words tend to appear (cf. Bybee, 2002).

4.3.3.2 Frequency effects

Besides wordform frequency, several other probabilistic measures may influence the realization of the inflected words in our study. For example, in a dual-route model of morphological processing, morphologically-complex words are accessed through both whole-word representations, and through decomposed forms (e.g., Caramazza, Laudanna & Romani, 1988; Baayen, 1992; Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 1995; Hay, 2003). If the complex word form has a high frequency relative to its components, it is predicted to behave more like a morphologically-simple form, potentially including stem reduction (Losiewicz, 1992; Hay, 2003; Cohen, 2014, though see Hanique & Ernestus, 2012). Additionally, two studies have also found that suffixes are lengthened (Cohen, 2014) or less likely to delete (Schuppler et al., 2012) with higher relative frequency.

This processing model potentially has implications for the paradigm-uniformity account. In order to explain incomplete neutralization effects (§4.1.1), Winter & Röttger (2011) and Roettger et al. (2014) propose a version of the paradigm-uniformity account that involves spreading activation among morphological paradigms. In particular, if a paradigm member is highly frequent, it is

predicted to have a stronger influence on its morphological relatives. We explored this prediction by examining whether inflected words with high-frequency stems (such as *guys*) show different effects. Following Hay (2001) (and others), we also tested whether inflected words with a high frequency relative to their stem (such as *bored*) show different effects. These results should be interpreted with caution, in particular because the stimuli were not selected to include a broad range of either frequency measure.

Following the procedure in §4.3.2, we fit a separate linear mixed-effects model to predict stem durations. The model included stem frequency as an additional fixed effect, as well as all interactions with word type and manner, plus by-participant random slopes. We also fit three models which replaced the stem frequency parameters with relative frequency, in order to predict stem durations, suffix durations, and stop deletion rates. Stem frequency was the log wordform frequency of the freestanding stem in SUBTLEX. Relative frequency was the log ratio of the inflected word frequency to the freestanding stem frequency.

There was no effect of stem frequency on stem durations ($p > 0.3$), or of relative frequency on stem or suffix durations ($p > 0.4$). However, there was a marginally non-significant effect in which inflectional stop suffixes were less likely to delete as relative frequency increased ($\hat{\beta} = 0.99$ per SD, $z = 1.95$, $p < 0.051$). This follows Schuppler et al. (2012), who had a similar finding for Dutch /t/ suffixes in a corpus of spontaneous speech.

4.3.3.3 Predictability effects

In addition to frequency, we explored a possible effect of the probability of an inflection in context (Cohen, 2014; Rose et al., 2015). Inflectional probability was estimated using the cloze norming experiment, as described in §4.2.2.3. We fit

additional models, following the procedures in §4.3.2 and §4.3.3.2, to predict suffix durations and stop deletion rates. As before, because the stimuli were not selected to include a broad range of the predictability measure, these results should be interpreted with caution. There was a marginal effect of inflectional probability, such that inflectional fricative suffixes (but not stop suffixes) were non-significantly longer with lower inflectional probability ($\hat{\beta} = 12$ms per SD, $t(39.40) = 2.054$, $p < 0.10$). There was no significant effect of inflectional probability on stop deletion rates ($p > 0.17$).

4.4   Discussion

English words with English [s, z] inflectional suffixes (e.g., *frees*) had significantly longer stems and suffixes than morphologically-simple homophones (*freeze*). This result supports the phonetic paradigm uniformity account, which predicted that inflected words such as *frees* should be influenced by the articulatory plan for their freestanding stems, such as *free*. In particular, we predicted (following Frazier, 2006) that the stem in an inflected word should be lengthened, because the same stem has a lighter coda and a longer duration when it occurs as a word on its own. Further, the freestanding stem word *free* is subject to prosodic domain-final lengthening. This should gradiently influence the realization of the inflected form *frees*, such that the stem is lengthened when it occurs within the inflected word *frees*, and the suffix is lengthened as a result of domain-edge strengthening effects.

4.4.1   Results for stop suffixes

In addition to the positive results for [s, z] suffixes, we also found a null effect for stem and suffix durations when the final suffixes were [t] or [d]. While it

is difficult to interpret a null result, the paradigm uniformity hypothesis predicts that these words should also be lengthened. Only one prior study has investigated [t, d] durations (Losiewicz, 1992), but also found a null result under a mixed-effects analysis (Plag, p.c.).

While orthographic length was unbalanced across stop-final pairs (see §4.2.2.2), complex words had more letters than simple words, and the stem duration effect for stop-final pairs is in the opposite direction: complex words are non-significantly shorter than simple words. Therefore, it is unlikely that orthographic length contributed to the null result.

However, one likely explanation for the null result comes from different parts-of-speech between the simple and inflected words. Several corpus studies report significant differences in word duration as a function of part-of-speech (Gahl, 2008; Gahl, Yao & Johnson, 2012; Seyfarth, 2014). In particular, nouns tend to be longer in duration than verbs. Because these differences have so far been attributed to systematic differences in phrase position and accent in spontaneous English speech (Gahl, 2008; Gahl et al., 2012), and because the items in the current study were matched for phrase position and accent (see §4.2.2), part-of-speech was not intentionally balanced across pairs. Impressionistically, we found that participants were very reliable at accenting the expected word. During segmentation of the data, we noted only 21 tokens (1.3% of the data) in which an unexpected word was accented; exclusion of these data did not qualitatively affect the results. Nevertheless, it is plausible that nouns are more likely to attract a stronger prominence than verbs, even with all else held equal, which might result in longer durations. It is also possible that there are fundamental differences in the phonological representation or processing of different parts of speech (e.g., Farmer, Christiansen & Monaghan, 2006).

Table 4.2: Distribution of part-of-speech for simple and inflected words, by manner of the suffix.

Word type	Noun	Verb	Other
Fricatives			
Simple	17	7	2
Inflected	16	10	0
Stops			
Simple	11	2	1
Inflected	0	11	3

Table 4.2 shows the distribution of part-of-speech for simple and inflected words (see also 4.6), within fricative-final pairs (*freeze/frees*) and within stop-final pairs (*tide/tied*). While the fricative-final pairs were balanced for part-of-speech ($\chi^2 = 2.56$, $p > 0.27$), the stop-final pairs included mainly verbs among the inflected words, but mainly nouns among the simple words ($\chi^2 = 18.23$, $p < 0.001$). This confound could thus have led to a null result for the stop suffixes (with a non-significant trend such that the inflected words were shorter). One future direction is therefore to investigate potential systematic differences in inherent duration, or in the realization of accent, across different parts of speech.

4.4.2 Other accounts

4.4.2.1 Internal hierarchical prosodic structure

Although we argue that the lengthening of inflected words derives from the influence of their morphological relatives, there are other accounts that may accommodate this result. One proposal is that an English inflected word like *tacks* has a hierarchical prosodic structure, such that the inflectional suffix is adjoined to an internal prosodic-word constituent corresponding to the stem *tack* (Goad, White & Steele, 2003; Sugahara & Turk, 2009, contra Hall, 2001; Raffelsiefen, 2005).

Because syllable rhymes are lengthened before prosodic boundaries at various levels of prosodic constituency (Wightman et al., 1992), the stem within an inflected word should be lengthened. This account is similar to the one described in §4.1.2. It differs in that in this alternative proposal, the word-internal prosodic boundary is part of the inflected word's phonological representation that is derived from the morphological parse of the word, rather than arising via uniformity with the inflected word's stem.

While this analysis is possible, it entails that the final [z] of *frees* either comprise its own syllable, or else be extra-syllabic (not part of any syllable) (Goad et al., 2003; Sugahara & Turk, 2009). However, we are unaware of psycholinguistic evidence that would support either possibility, and from a theoretical perspective, there are alternative explanations for the phenomena that extra-syllabicity has been used to account for (Hall, 2002). The exception is the durational lengthening observed here (Sugahara & Turk, 2009, p. 482–485), which we claim can be explained by a more general uniformity mechanism.

Sugahara & Turk (2009, p. 506) argue against the uniformity account on the grounds that it would require including duration—which is highly variable in usage—in phonological representation. However, there is recent evidence which supports this assumption (Tauberer & Evanini, 2009; Katz, 2010, 2012; Seyfarth, 2014). Further, it is not clear that word and segment duration is necessarily any more variable in usage than features such as (for example) English obstruent voicing, which is standardly assumed to be part of a contrastive phonological representation.

4.4.2.2 Relative frequency and the dual-route model

A probabilistic version of the paradigm-uniformity proposal might account for Cohen's (2014) finding that high relative frequency of a complex form is associated with reduced stems (cf. Hay, 2003). High relative frequency means that the full inflected form is relatively more frequent than the stem, and so the influence of the stem's independent phonetic plan may be weaker (cf. Zuraw & Peperkamp, 2015). In §4.1.2, we argued that the stem's influence should produce relatively longer durations. Therefore, when this influence is weaker, the stems in inflected words should be relatively shorter. The same reasoning might apply to the hierarchical structure proposal in §4.4.2.1: if the inflected form is more likely to be retrieved as a whole word, it is less likely to include an internal boundary that would condition stem lengthening. However, our study did not replicate the effect of relative frequency on stem durations found in Cohen 2014, nor did we find an effect of stem frequency (§4.3.3.2, with caveats noted in that section). There was a marginal effect of relative frequency on inflectional suffix deletion (supporting the findings of Schuppler et al., 2012), but it is not clear that this particular effect bears on the predictions made by the paradigm-uniformity account.

4.4.2.3 Communicative enhancement

One reason to expect that inflectional suffixes might be lengthened is because a suffix like the [z] in *frees* signals a morphosyntactic property (third-person singular agreement, in our materials), whereas the same [z] suffix in a word like *freeze* carries no additional information beyond that conveyed by any other word-final segment (Pluymaekers et al., 2010; Cohen Priva, 2012; Hanique & Ernestus, 2012; Rose et al., 2015). This suggests an alternate explanation for our finding

that inflectional [s, z] suffixes were longer: speakers may use the details of phonetic implementation to enhance the intelligibility of morphological distinctions. While previous work has found suggestive effects (Cohen, 2014; Rose et al., 2015, though see Hanique, Ernestus & Schuppler, 2013), we did not find an effect of inflectional probability on suffix duration in our analysis.

At the same time, it is not necessarily true that lengthening a suffix improves its discriminability. In particular, if duration is not used by listeners as a perceptual cue, elongating a particular segment may not necessarily efficiently enhance its contribution to the linguistic signal. For example, Maniwa, Jongman & Wade (2009) find that although talkers enhance a variety of spectral cues when they hyperaticulate fricatives, the actual fricative durations are lengthened roughly proportionately, even though duration can be used to acoustically discriminate fricative voicedness.

4.4.3   Comparison with previous work

Our finding that the English words inflected with [s, z] had longer stems and suffixes than uninflected words agrees with some existing experimental work (Walsh & Parker, 1983; Frazier, 2006; Sugahara & Turk, 2009). However, it does not reconcile the differences between that work and Plag et al. (2015), who analyzed suffix durations in a spontaneous speech corpus. In particular, Plag et al. (2015) found that voiceless [s] suffixes were *shorter* in inflected words compared to uninflected ones; as well as a more complicated pattern of differences within several kinds of voiced [z] suffixes (e.g., plural [z] was longer than third-person singular [z]). These corpus results are not predicted by the paradigm uniformity account, or by any current production theories (Plag et al., 2015, p. 29–32).

Why did our suffix duration results pattern in the opposite direction as Plag

et al. (2015)? In an exploratory analysis, we tested interactions between voicing, word type, and manner, but found no significant interactions (all $p > 0.3$).[4] Two methodological differences were the use of read-aloud versus truly-spontaneous conversational speech, and the analysis of homophones versus non-homophones. However, it is unclear whether either consideration would cause the effect to reverse direction.

It is also possible that the unbalanced nature of the corpus data in Plag et al. (2015) influenced the analysis. For example, Hsieh, Leonard & Swanson (1999) find that in natural speech, plural nouns appear in final position much more often than third-person verbs—contributing to lengthening—and other work has pointed out that different parts-of-speech (likely correlated with inflectional status) may systematically occur in different prosodic and segmental contexts (see §4.4.1; Bybee, 2002). Plag et al. (2015) take this into account and include an appropriate variety of lexical and contextual control variables in their models. However, in order to accurately estimate parameters for correlated variables (e.g., suffix type and syntactic position), it is necessary to have many observations in most cells of the design, which may not have been the case (the analysis selected about 650 tokens at random from the corpus). Further, as the authors acknowledge (p. 14), it is challenging to code and statistically control for the effects of diverse prosodic contexts.

4.4.4   Conclusions

We found that English inflected words with [s, z] suffixes had significantly longer stems and suffixes than uninflected words that were segmentally-identical:

---

[4]A main effect of voicing was significant ($\hat{\beta} = 31$ms, $t(36.31) = 3.104$, $p < 0.01$), but did not qualitatively affect the results reported in §4.3.2.2.

frees is not homophonous with *freeze.* This supports predictions based on a model of phonetic paradigm uniformity, in which the durational targets of a target word's morphological relatives influence the realization of that word (Hayes, 2000; Steriade, 2000; Frazier, 2006). We found this result based on a large and diverse set of word types, which were balanced for frequency and orthography, and elicited in phonetically-matched conversational speech designed to avoid metalinguistic task effects that have challenged the interpretability of previous work (see e.g. Bermúdez-Otero, 2010; Hanique & Ernestus, 2012; Plag, 2014; Mousikou et al., 2015; Plag et al., 2015).

This finding challenges modular accounts of language production in which morphological information does not interact with phonetic realization once a phonological plan is specified (Kiparsky, 1982; Levelt et al., 1999). In particular, the phonetic paradigm uniformity account suggests one specific mechanism involving the cross-influence of articulatory plans among morphological relatives (cf. Kuperman, Pluymaekers, Ernestus & Baayen, 2007), and makes straightforward, testable predictions about phonetic realization (Frazier, 2006; Kaplan, 2016, and see ongoing work by Abby Kaplan). Future work might investigate especially the cross-linguistic validity of these predictions, and evaluate the interaction of probabilistic variables with paradigm uniformity effects.

4.5   Acknowledgments

<center>* * *</center>

Chapter 4, in full, has been submitted for publication as Seyfarth, Garellek, Gillingham, Ackerman, and Malouf [Acoustic differences in morphologically-distinct homophones]. This work was reported, in part, at the 3rd American International Morphology Meeting. This chapter reports co-authored work, which is used here with permission from the co-authors. The dissertation author was the primary investigator and author of this paper.

4.6   Items

board, bored

board (Noun)

Two restaurant executives are talking about expanding to a new location downtown.

B: Getting this construction permit is going to be more of a headache than I thought. I was hoping we could just get the city planning committee chairman to sign off on it, so we wouldn't have to bring it before the whole committee. But the chairman said we need to go to the next meeting.

A: Do you think we can still just ask the committee chairman at the meeting?

B: No... now we'll have to ask the board at the meeting.

bored (Adjective, participle)

Two neighbors are talking about volunteering with their city government.

B: The planning committee released their meeting schedule this week. They're not discussing my proposal for the new library exhibit until January.

A: I'm sorry to hear that. Are you still excited about volunteering at the meetings?

B: No... now I think I'll just be bored at the meetings.

booze, boos

booze (Noun)

Two friends are talking about a football game that B went to with their mutual friend, Tim.

A: How was the football game yesterday? I heard Tim got really drunk and started a fight!

B: It was great—we won in double overtime!

A: How could he get angry about winning?

B: Well... I think Tim just had too much booze.

boos (Noun, plural noun)

Two friends are talking about a football game that B went to with their mutual friend, Tim.

A: How was the football game? I'm impressed that you drove all the way up to Portland just to see your team play.

B: It was fine—we won, but Tim got pretty upset that he was the only one cheering for our team.

A: I bet that must have been frustrating.

B: Yeah. . . all his cheers were drowned out by boos.

brood, brewed

brood (Verb)

Two high school students are talking about their college applications.

A: Are you still upset about your SAT scores?

B: A little bit. I'm really worrying about my chances. My grades haven't been so great this year either.

A: Oh, I'm sorry. You've been stressing about this a lot. Do you want to talk about it?

B: I just brood sometimes. Thanks anyway, though.

brewed (Verb, past)

Two accountants are working together in an office in the late afternoon.

A: I'm heading out for lunch. Do you want anything?

B: I brought my lunch today. I'm feeling really tired anyway.

A: Oh, maybe you need some caffeine! Do you want me to pick up some tea while I'm out?

B: I just brewed some tea. Thanks anyway, though.

bruise, brews

bruise (Noun)

Two friends are talking about an intramural baseball game.

B: Anderson got hit with a baseball yesterday. The pitcher was trying to throw a fastball but missed.

A: Is Anderson all right?

B: Yeah, but he got a bad bruise.

brews (Noun, plural noun)

A has just arrived at a housewarming party. Most of the guests have been there for a little while.

B: If you're thirsty, Anderson brought all of this beer. He went to that fancy beer place he's always talking about.

A: Oh, I think he spends too much money there.

B: Yeah, but he got these good brews.

chard, charred

chard (Noun)

B is giving A a tour of a newly-planted vegetable garden.

A: You have a huge garden! I can't wait to come over for dinner in a few months.

B: Thanks! I've got some great Mediterranean recipes planned out. Be careful when you're stepping over the fence.

A: Oops, I think I stepped on something! What's growing here?

B: Hm. . . it looks like it's chard. It'll be fine, don't worry.

charred (Adjective, participle)

A is out on some errands, and is calling B on the phone.

A: Hi, I forgot to take the lasagna out of the oven. Could you check on it for me?

B: Sure, I'll go take a look.

A: Does it look okay? Did it burn around the sides of the pan?

B: Hm. . . it looks like it's charred. It's fine in the middle, though.

choose, chews

choose (Verb)

Two college students are walking back to their dorm.

B: We should get back soon, so we don't miss our course enrollment time.

A: Are you worried about getting a math section with a bad TA?

B: If we choose one, we'll be fine. I looked through the course reviews.

chews (Verb, third singular)

Two roommates are talking.

B: I read that lilies can be dangerous to animals.

A: Do you think we should keep them away from the cat?

B: If he chews one, he'll get sick. We should just put them outside.

clause, claws

clause (Noun)

>*Two lawyers are revising a contract that they've been hired to write.*
>**B:** Pat added this one sentence to the contract that I think is going to cause us a lot of trouble.
>**A:** Yeah, and he also gave us a big list of other changes that he wanted us to make.
>**B:** I think those are mostly low priority. We can work on those later.
>**A:** So which of Pat's ideas do you want to start with?
>**B:** Let's try changing his clause.

claws (Noun, plural noun)

>*Two roommates are talking about a cat that they just adopted.*
>**B:** The cat has been scratching the sofa a lot.
>**A:** Should we get another scratching post?
>**B:** Well, we don't have a lot of space in that room.
>**A:** So how should we deal with the cat instead?
>**B:** Let's try trimming his claws.

cruise, crews

cruise (Noun)

>*Two young friends are talking about their families' vacation plans.*
>**B:** My grandparents fired their travel agent last year. They started booking everything online themselves.
>**A:** I didn't know your grandparents could use the internet.
>**B:** Yeah. . . that's how they found their cruise.

crews (Noun, plural noun)

>*A naval historian is giving a lecture about Caribbean piracy.*
>**B:** The pirates would go from island to island, robbing townships and abducting young men.
>**A:** Did they mostly just force people into serving on their ships?
>**B:** Yes. . . that's how they found their crews.

daze, days

daze (Noun)

Two political campaign workers are talking about the election results that were announced the day before.

B: When Kate found out that she won the election, she was just in a state of total shock.

A: That's great! What did she say after the shock had worn off?

B: Once she'd gotten past the daze, she thanked all of her supporters. She was very thoughtful.

days (Noun, plural noun)

A journalist is talking to a co-worker about a political convention in New York City.

B: Every morning, we went to two or three fundraising meetings, then we listened to some of the candidates' speeches during lunch, then we had a five-hour press conference with the candidates immediately afterwards.

A: So it sounds like you were pretty busy! Did you get to go out at night, once all the events were over?

B: Once we'd gotten through the days, we just wanted to go to sleep. We were so exhausted.

duct, ducked

duct (Noun)

A repair person just arrived at an office on a hot summer day.

A: Maintenance sent me over to fix your broken air conditioning system. They said you're not getting any air flow here. Have you checked to see if your vent is blocked?

B: No. . . but this duct is blocked.

A: Okay, we'll take a look.

ducked (Verb, past)

Two friends on a baseball team are talking about the last pitch.

B: The pitcher threw a wild fastball. Grace almost got hit!

A: Did you see what happened? Did she swing in time?

B: No. . . but Grace ducked in time.

A: At least she's okay.

flex, flecks

flex (Verb)

Two construction workers are working on a new house.

B: Two more pallets of lumber just arrived. Could you get the forklift and take one of them to Sheila? She's putting up a frame for the shed in the back.

A: Sure, I'll go do that now.

B: Just be careful when you load them up.

A: Why? They look pretty sturdy.

B: We can't let them flex. If they're bent, we can't use them.

flecks (Noun, plural noun)

An art historian is showing a colleague around a Gothic church.

B: This is the area where the earthquake last month caused most of the damage. But there was a silver lining too—when that stairwell collapsed, we found a mural behind the broken stones.

A: Do you know yet if the mural was originally part of the building?

B: Not yet. A forensics team came in to analyze the paint on the mural. They took some samples last week, but we haven't gotten the results back yet.

A: Oh, I hope they didn't damage it! Did they scrape off any of the paint?

B: They scraped off some flecks. It's the normal procedure.

freeze, frees

freeze (Verb)

Two housemates are wrapping up a surprise birthday party that they put on for a friend.

B: It looks like most people are leaving now. I guess I'm going to start cleaning up a little bit.

A: There's so much cake leftover. I don't want it to go bad.

B: If we freeze it, it should be fine.

frees (Verb, third singular)

Two rural neighbors are talking about a friend, Rich, who is an avid hiker and animal-lover.

B: Rich decided to take care of the injured hawk that he found yesterday.

A: They don't do well in captivity. Wouldn't it be better to let it go?

B: If he frees it, it won't survive.

graze, grays

graze (Verb)

Two farmers are talking about their shared pasture, near a busy highway.

B: I'm worried about my sheep. I think the highway construction is scaring them.

A: Why do you say that?

B: They went back to the barn right away this morning. They usually graze first.

grays (Verb, third singular)

A retired couple is out for lunch together.

B: I keep finding all these white hairs around my temples!

A: I wouldn't worry about it. It makes you look distinguished.

B: I didn't think my hair would turn white right away. Hair usually grays first.

guise, guys

guise (Noun)

Two neighbors are talking on the porch.

A: Nice to chat with you. It looks like the plumbing company just got here, so I should go let them in.

B: Oh, did you call Northern Plumbing? Don't trust them with your money. Remember how Martha and Tony around the corner got scammed? That's them.

A: Are you sure it's the same company? It's a similar van, but I thought they were called Mission Plumbing.

B: It's a different guise, but it's the same company.

A: Oh. . . I'll keep an eye on them. Thanks.

B: No problem. Talk to you later!

guys (Noun, plural noun)

Two bicycle store employees are talking.

A: Do you have a minute to help me unload the delivery truck?

B: I want to keep an eye on those three customers. I think they're with the ring of bike thieves that's been trying to sell us parts.

A: Are you sure they're part of that group? I don't recognize them.

B: They got different guys, but they all work together.

A: How do you know?

B: I saw their van out in the parking lot.

hose, hoes

hose (Noun)

Two neighbors are talking about a community garden.

A: I left some things out yesterday when I was watering the plants. Did you put anything away?

B: Yeah, I put the hose away, back in the shed.

A: Thanks—sorry for leaving a mess everywhere.

hoes (Noun, plural noun)

Two neighbors are talking about a community garden.

A: I left some tools out yesterday when I was weeding the garden. Did you put any of them away?

B: Yeah, I put the hoes away, back in the shed.

A: Thanks—sorry for leaving a mess everywhere.

lapse, laps

lapse (Noun)

Two soccer players just finished a game.

B: Miranda was pretty upset after she let that goal through.

A: She went left and the ball went right. She almost never misjudges it like that.

B: Yeah. . . it was just a lapse. She'll bounce back tomorrow.

laps (Noun, plural noun)

Two athletes just finished a morning workout.

B: I'm exhausted. Let's take it easy this afternoon. My calves are going to be sore.

A: Did we spend too much time on the track?

B: Yeah. . . it was all the laps. I'll bounce back tomorrow.

lax, lacks

lax (Adjective)

Two students are walking to class.

A: The lecture started a couple minutes ago. Do you think the professor is going to mark us late?

B: I heard she doesn't actually care. I think she'll be pretty lax.

A: That's good to know.

lacks (Verb, third singular)

Two students are talking about a mutual friend.

A: Jay just told me that he's going out tonight. But this morning, he said he'd be up all night finishing the final paper for his class.

B: He doesn't have a great work ethic. It's something he really lacks.

A: Well, I'm sure he'll get his paper done eventually. He always manages to pull it off.

lynx, links

lynx (Noun)

A park ranger is giving a presentation on how they deal with injured animals.

B: Right now, we're caring for a dog and a wild cat that were injured by non-native predators. We're hoping that the cat will be well enough to be released back into the area soon.

A: Are you going to release the dog too?

B: No, just the lynx. We'll see if the dog can be put up for adoption.

links (Noun, plural noun)

An IT worker is giving a presentation on parental-control software for browsing the internet.

B: This software will keep your kids from clicking on URLs that go to potentially unsafe sites. It actually hides them, so kids won't even see them on their screen.

A: Will it hide the images I don't want my kids to see too?

B: No, just the links. You'll have to use other software to block images.

mist, missed

mist (Noun)

Two commuters are chatting in a coffee shop.
B: How has the traffic been so far?
A: The traffic isn't bad, but it's always so foggy in the morning. I can never see the bay on my way to work.
B: Well, it's not bad now. It's just mist out there.
A: That's nice to hear! Maybe it will clear up while I'm driving.

missed (Verb, past)

Two tennis players are talking about the match they just watched.
B: Becca fumbled her first serve attempt, so the other player got to serve first.
A: She's been having trouble with her wrist. Is that still bothering her? Maybe that's the reason she fumbled.
B: No, I think she's fine. She just missed out there.
A: That's too bad.

nose, knows

nose (Noun)

A doctor is helping a patient who is having trouble breathing at night.
A: I always wake up feeling short of breath. Do you have any suggestions?
B: Try using this breathing strip for a week first. It should help keep your sinuses open, and relieve congestion while you're sleeping.
A: I'll give it a try—do I put it on my sinuses?
B: Place it on your nose. There's an adhesive that will help it stick.

knows (Verb, third singular)

Two event planners are working on a schedule for an upcoming fundraiser.
A: Did you ask Taylor for her opinion on the fundraiser schedule? I thought we'd start advertising about three weeks before.
B: She said that we aren't going to raise any money if we don't start at least two months in advance.
A: Oh, is that what Taylor thinks?
B: It's what Taylor knows. She's run a lot of fundraisers before.

ode, owed

ode (Noun)

Two poets are talking about their friend, Gordon.

A: Gordon's really excited about the open-mic night next week.

B: I know. He said he just finished writing something to present. Now he's busy practicing his performance.

A: Did he write a poem for it?

B: Yeah... he wrote an ode for it.

owed (Verb, past)

Two parents are talking about their college-age daughter, Jan.

A: Jan said she had a great time on the ski trip. She said it turned out to be pretty expensive, though.

B: Yeah, she asked me if she could borrow five hundred dollars and pay me back when her job starts.

A: Five hundred dollars? Was that Jan's share of the trip?

B: Yeah... that's what Jan owed for it.

pact, packed

pact (Noun)

Two parents are talking about a backpacking excursion that their teenage sons had been planning.

A: It looks like it's going to rain pretty heavily this weekend. Are Peter and Kevin really still going on a trip to the mountains?

B: Yeah, they made a pact for their trip. They've been planning it for months, and they agreed not to put it off.

A: Well, tell them to bring the snow chains. There might be snow up there.

B: I will. I put the chains in the back of the car already.

packed (Verb, participle)

Two parents are talking about a backpacking excursion that their teenage sons had been planning.

A: Peter and Kevin had a ton of stuff in the car yesterday, but it's empty now. I thought they were going to take a trip.

B: Yeah, they had it packed for their trip. Kevin told me they decided to put it off until next week.

A: Why—what happened? They were really looking forward to going.

B: The weather report said it was going to snow all weekend. They didn't want to drive in the weather.

past, passed

past (Noun)

Two congressional representatives are talking about a vote last month.

B: The majority leader wasn't happy that I didn't vote for her bill last month, but I think she understands why. She knows that it would have hurt me in my district's next election, and her bill ended up getting approved anyway, even without my vote.

A: So do you think that she's forgiven you for it?

B: Yeah...it's in the past.

passed (Verb, participle)

Two congressional representatives are talking about a vote the next day.

B: We need to be sure that everybody from our party is there to vote on the conservation bill tomorrow.

A: Are you worried about whether the bill is going to be approved?

B: No...it'll get passed.

paste, paced

paste (Verb)

Two office workers are working on a newsletter.

B: Thanks for helping me with the newsletters. Let's put this photo on the front page.

A: Should we just staple it to the front?

B: I thought we'd paste it there instead. Then we won't have to damage the edges.

A: That's a good idea.

paced (Verb, participle)

Two runners are talking about a trail race that goes through a state park next month.

B: Sam said he's been training to run the trail race with Amy next month. Amy's run it every year for the past five years, but Sam said he's pretty nervous about all the hills.

A: Did he say he'd run through the course together with her?

B: He said they'd paced it together. He hasn't run through it with Amy yet.

A: I'm sure he'll do fine. Amy's a great trainer.

pause, paws

pause (Verb)

Two friends are watching a news program on TV.

A: This show always has really interesting charts, but they never show them for very long.

B: I pause at them sometimes.

A: I usually watch them live. Maybe I should just record them.

paws (Verb, third singular)

Two friends are talking.

A: I like your aquarium! Does your cat bother the fish when she's climbing up on the mantle?

B: She paws at them sometimes.

A: Be careful. My friend lost a goldfish that way.

please, pleas

please (Adverb)

A parent and their child are buying ice cream.

B: Go ahead and ask for your favorite flavor, but be sure to be polite.

A: Do I have to say thank you?

B: You have to say please.

A: Could I have some mint chocolate chip ice cream, please?

B: That's right. Then you should say thank you afterwards.

pleas (Noun, plural noun)

A tour group is being shown around a municipal court room.

B: The next room is where the judge formally presents the charges to people accused of a crime.

A: Does the judge listen to their cases?
B: He listens to pleas.
A: Then what?
B: If someone pleads innocent, their case is assigned to another judge.

praise, prays

praise (Noun)

A is talking to a vet about his dog's behavior issues.
B: If you're having trouble with your dog obeying commands, you need to be firm.
A: How should I reward him? Should I give him a treat for obeying?
B: You should give him praise. You can give him a treat sometimes, but you shouldn't do it very often.
A: Okay, I'll remember that.

prays (Verb, third singular)

Two football fans are talking about how their favorite players celebrate when they score a touchdown.
B: I think John is very professional. He never rubs it in the other team's face when he scores.
A: Really? Isn't John the guy that dances every time he scores a touchdown?
B: John's the guy who prays. Terry is the one that dances. He got a penalty for it last year.
A: I remember when that happened.

pride, pried

pride (Noun)

Two college-age friends are chatting.
A: How did Sarah's blind date with Franklin go?
B: She said it didn't go very well. She thought he was really irritating.
A: What about him annoyed her?
B: His pride annoyed Sarah. He would never admit he was wrong.

pried (Verb, past)

Two neighbors are talking after a burglary at one of their houses.
A: Do you know how the burglar got in? Were any of the windows broken?
B: He probably came in through the front door. He left some mud on the carpet.
A: How was he able to get your door open?
B: He pried our door open. We found the broken crowbar he used.

prize, pries

prize (Noun)

Two neighbors are chatting at a barbecue.
A: Oh, here comes Don. He's got his poodle with him.
B: It looks like that poodle has gotten groomed pretty recently.
A: Didn't Don just enter him in a dog show?
B: Yeah, he won a prize. Don's been doing well now that he's retired.

pries (Verb, third singular)

Two neighbors are chatting at a barbecue.
A: Oh, here comes Don. He looks like he's got something on his mind.
B: Uh oh. I'm going to go see how the grill is doing.
A: Why don't you want to talk to Don?
B: Well, he really pries. He always asks very personal questions.

quartz, quarts

quartz (Noun)

A museum director is talking to an artist.
A: Do you need any more stone to finish the sculpture?
B: I need some more quartz.
A: Sure, we'll order some right away.

quarts (Noun, plural noun)

Two chefs are starting their morning shift.
A: Do we still need two more pints of milk to make the cheese for today?
B: We need two more quarts.
A: Okay, I'll go get some from the fridge.

raise, rays

raise (Noun)

> *Two coworkers are chatting.*
> **B:** I just had my performance review yesterday. Did I tell you about that?
> **A:** No! How did it go? Did you get a good evaluation?
> **B:** Yeah...it went pretty well. I got a good raise.
> **A:** That's great! I scheduled mine next week, so I'm crossing my fingers.

rays (Noun, plural noun)

> *Two friends are chatting.*
> **B:** Daniel looks really tan! Did he go to the beach this weekend?
> **A:** He said he spent the whole day there. He definitely got a good tan.
> **B:** Yeah...it looks like he did. He got some good rays.
> **A:** I guess we missed out. I always get sunburned anyway.

rapt, wrapped

rapt (Adjective)

> *A is running for city mayor.*
> **A:** How do you think that speech went?
> **B:** It went really well! You had great delivery, and you covered all the key issues.
> **A:** Thanks—did it seem like the audience was paying careful attention?
> **B:** They were paying rapt attention.

wrapped (Verb, participle)

> *B just gotten a call from a friend who's getting married soon.*
> **A:** Thanks for RSVPing to our wedding reception! A couple people have asked about what kind of gifts would be appropriate, so I just wanted to call to let you know that we're not expecting attendees to bring anything.
> **B:** Oh, thanks...I need to tell James that.
> **A:** Uh-oh. Did he already buy a present?
> **B:** He's already wrapped a present.

ruse, rues

ruse (Noun)

Two roommates are talking.
A: Who was that at the door?
B: It was someone asking if we needed carpet cleaning services. I said no, but he claimed he needed to come in anyway because city hall had designated our house as the possible source of a mold infestation on this block, and his company was the only one licensed to deal with it.
A: Isn't that kind of a complicated story?
B: It's a complicated ruse. I've read about this scam online.

rues (Verb, third singular)

Two neighbors are chatting about their friend, Ed.
A: Did you hear that Ed broke his leg on his skydiving trip last month? He was really upset! He was even planning to sue the tour company.
B: Yeah, he told me about it. He said that they ended up paying him six hundred dollars to waive his rights to a lawsuit. But then his lawyer told him he could have won a lot more in court if he hadn't signed the waiver.
A: Is that a decision Ed regrets?
B: It's a decision Ed rues. He's been dwelling on it too much.

sax, sacks

sax (Noun)

A is helping B, a musician, pack up the car for an out-of-town gig.
B: Thanks for helping me with all this stuff.
A: No problem. Should I get your instrument first, or the recording equipment?
B: Let's start with the sax.
A: Sure. I'll put it in the front of the car so it doesn't get damaged.

sacks (Noun, plural noun)

A is helping B pick up supplies for a new garden shed that B is building.
B: Thanks for helping me get all this stuff home.
A: No problem. Should we carry the bags of cement first, or the roofing material?
B: Let's start with the sacks.

A: Sure. These are heavy!

seize, sees

seize (Verb)

Two TSA officers have identified a suspicious package in the concourse.
A: That piece of luggage doesn't belong to any passengers on this plane. Should we confiscate it now?
B: If we seize it, there might be a big scene. Let's just tell the baggage handlers.
A: That's a good idea. Let's do that instead.

sees (Verb, third singular)

Two travelers are waiting in an airport security line.
A: Do you think it's okay to take this pocket knife through security? It looks like the guy at the X-ray machine isn't paying very close attention.
B: If he sees it, there might be a big scene. You should just drop it in the mail.
A: That's a good idea. Let's do that instead.

size, sighs

size (Noun)

B is at the customer service desk at a department store.
B: I bought a shirt from your company's online store, but you guys sent me this one instead of what I ordered.
A: Sure, I can exchange that for you. Is it the wrong color?
B: No, it's the wrong size. I ordered a medium.

sighs (Noun, plural noun)

Two political analysts are watching a candidate give a speech on TV.
B: I don't think Allan is going to win the election. He has a good record, but he doesn't relate well to the voters.
A: Do you think it's his body language?
B: No, it's all the sighs. He doesn't seem excited.

suede, swayed

suede (Noun)

B has just walked into a shoe store.
A: Good morning! Just let me know if you need help finding anything.
B: Hello. I was hoping to find a pair of leather boots.
A: Of course. Are you looking for a particular kind of leather boots today? Fancy ones, or boots for when it rains?
B: I was looking for suede ones today... so they won't be for rain.
A: No problem—please follow me.

swayed (Verb, past)

Two neighbors are chatting.
A: That was some storm yesterday! The weather channel said the wind speed got up to fifty miles per hour.
B: Wow! I guess we got pretty lucky. I was watching the willow tree in our yard get blown around for a couple hours.
A: Did your tree end up losing any branches in the wind?
B: It definitely swayed in the wind... but it's fine, thankfully.
A: That's good to hear.

tease, teas

tease (Verb)

Two parents are talking while their son Eugene is playing a game with his cousins.
A: I'm glad Eugene is finally playing Jenga with his little cousins.
B: Maybe... I wish he would be nicer to them.
A: He keeps saying how well they're doing at the game. It seems like he's trying to be nice!
B: He's trying to tease them. They don't realize he's being sarcastic yet.
A: Oh... maybe we should say something to him.

teas (Noun, plural noun)

Two coffee shop customers are waiting at a table for their order to be ready.
A: I don't know why we keep coming back here.
B: I know.... we've been here for twenty minutes and our order still isn't up!

A: Can you see what's on the counter? Are there two coffees there for us?
B: It looks like two teas there. That must be another customer's order.
A: Let's go somewhere else next time. This place takes forever.

tide, tied

tide (Noun)

Two friends are staying at a beach cabin in the summer.
A: Do you want to go fishing this morning?
B: The conditions out there aren't good for fishing right now.
A: Is the wind too high?
B: No. . . the tide is too high.

tied (Adjective, participle)

Two parents are talking before a little league awards ceremony.
A: Did you put together the award yourself?
B: Sort of. I added all these blue ribbons to this medal I bought.
A: What's keeping the ribbons together? Are they all glued underneath?
B: No. . . they're tied underneath.

tract, tracked

tract (Noun)

Two friends are talking.
B: Did you see Russell trying to hand out all those copies of "Our World's Promise" at the book fair?
A: What's that? Is it another religious book that he wrote?
B: Well. . . it's a tract he wrote.

tracked (Verb, past)

Two camp directors are planning for the summer.
B: Russell said that there was a bear last year that had been getting close to the cabins. But he talked to the park rangers, and they think they've scared it off for good.
A: Were they able to find the bear's den?
B: Yeah. . . once they tracked it there.

4.7   Item presentation order

List 1: prize, tide, choose, nose, booze, tracked, lynx, lacks, quartz, praise, bored, sighs, missed, crews, past, pleas, charred, flex, ode, rues, brews, laps, paws, guys, swayed, freeze, days, ducked, hose, raise, paced, seize, brood, pride, rapt, teas, pact, graze, claws, sacks

List 2: pries, tied, chews, knows, boos, tract, links, lax, quarts, prays, board, size, mist, cruise, passed, please, chard, flecks, owed, ruse, bruise, lapse, pause, guise, suede, frees, daze, duct, hoes, rays, paste, sees, brewed, pried, wrapped, tease, packed, grays, clause, sax

List 3: sacks, claws, graze, pact, teas, rapt, pride, brood, seize, paced, raise, hose, ducked, days, freeze, swayed, guys, paws, laps, brews, rues, ode, flex, charred, pleas, past, crews, missed, sighs, bored, praise, quartz, lacks, lynx, tracked, booze, nose, choose, tide, prize

List 4: sax, clause, grays, packed, tease, wrapped, pried, brewed, sees, paste, rays, hoes, duct, daze, frees, suede, guise, pause, lapse, bruise, ruse, owed, flecks, chard, please, passed, cruise, mist, size, board, prays, quarts, lax, links, tract, boos, knows, chews, tied, pries

4.8 Bibliography

Baayen, H. (1992). Quantitative aspects of morphological productivity. In *Yearbook of Morphology* (pp. 109–149).

Baayen, R. H., Levelt, W. J., Schreuder, R., & Ernestus, M. (2007). Paradigmatic structure in speech production. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 43, (pp. 1–29). Chic Ling Society.

Baker, R., Smith, R., & Hawkins, S. (2007). Phonetic differences between mis-and dis-in English prefixed and pseudo-prefixed words. In *Proceedings of the 16h International Congress of Phonetic Sciences.*

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*(1), 92–111.

Benua, L. (1997). *Transderivational identity.* PhD thesis, University of Massachusetts, Amherst, Amherst, MA.

Bermúdez-Otero, R. (2010). Morphologically conditioned phonetics? Not proven. *On Linguistic Interfaces II, Belfast.*

Boersma, P. & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.14, retrieved 6 March 2016 from http://www.praat.org.

Brewer, J. B. (2008). *Phonetic Reflexes of Orthographic Characteristics in Lexical Representation.* ProQuest.

Brysbaert, M. & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–90.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*(4), 991–997.

Bürki, A., Spinelli, E., & Gaskell, M. G. (2012). A written word is worth a thousand spoken words: The influence of spelling on spoken-word production. *Journal of Memory and Language, 67*(4), 449–467.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-Based Models of Language*. CSLI Publications.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America*, *120*(3), 1589.

Byrd, D. & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297–332.

Cho, T., Lee, Y., & Kim, S. (2014). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics*, *46*, 128–146.

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement: Contextual and paradigmatic effects on morpheme pronunciation. *Morphology*.

Cohen Priva, U. (2012). *Sign and signal: deriving linguistic generalizations from information utility*. PhD thesis, Stanford University.

Davis, S. (2005). "Capitalistic" versus "militaristic": the paradigm uniformity effect reconsidered. In L. Downing, T. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory* (pp. 107–121). Oxford: Oxford University Press.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

Ernestus, M. & Baayen, H. (2007). Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes*, *22*(1), 1–24.

Ernestus, M. & Baayen, R. H. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. *Laboratory phonology*, *8*(1), 27–49.

Farmer, T., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 12203–8.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*, 109–135.

Fougeron, C. & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*(6), 3728–3740.

Fourakis, M. & Iverson, G. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica*, *41*, 140–149.

Frauenfelder, U. H. & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: the role of productivity. *Yearbook of Morphology*, 165–183.

Frazier, M. (2006). Output-output faithfulness to moraic structure: Evidence from American English. In Davis, C., Deal, A. R., & Zabbal, Y. (Eds.), *Proceedings of the 36th Annual Meeting of the North East Linguistics Society*, (pp. 1–14).

Gahl, S. (2008). Time and Thyme are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Language*, *84*(3), 474–496.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806.

Goad, H., White, L., & Steele, J. (2003). Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, *48*(2), 243–263.

Goldrick, M., Baker, H. R., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, *121*(1), 58–72.

Goldrick, M., Keshet, J., Gustafson, E., Heller, J., & Needle, J. (2016). Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition*, *149*, 31–39.

Guy, G. R. (1980). Variation in the group and the individual: the case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space*, volume 1 of *Quantitative analyses of linguistic structure* (pp. 1–36). Academic Press.

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, *3*(1), 1–22.

Guy, G. R., Hay, J., & Walker, A. (2008). Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English. *Laboratory Phonology*, *11*.

Halekoh, U. & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30.

Hall, T. A. (2001). The distribution of superheavy syllables in Modern English. *Folia Linguistica*, *35*(3-4), 399–442.

Hall, T. A. (2002). Against extrasyllabic consonants in German and English. *Phonology*, *19*(01).

Hanique, I. & Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e Linguaggio*, *11*(2), 147–164.

Hanique, I., Ernestus, M., & Schuppler, B. (2013). Informal speech processes can be categorical in nature, even if they affect many different words. *Journal of the Acoustical Society of America*, *133*, 1644.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, *39*(6; ISSU 376), 1041–1070.

Hay, J. (2003). *Causes and consequences of word structure*. Routledge.

Hay, J. & Baayen, R. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, *9*(7), 342–348.

Hayes, B. (2000). Gradient well-formedness in Optimality Theory. *Optimality Theory: Phonology, syntax, and acquisition*, *88*, 120.

Hsieh, L., Leonard, L. B., & Swanson, L. (1999). Some differences between English plural noun inflections and third singular verb inflections in the input: The contributions of frequency, sentence position, and duration. *Journal of Child Language*, *26*(03), 531–543.

Hualde, J. I. & Prieto, P. (2014). Lenition of Intervocalic Alveolar Fricatives in Catalan and Spanish. *Phonetica*, *71*(2), 109–127.

Jassem, W. & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, *17*, 317–325.

Kaplan, A. (2016). What phonology knows about the lexicon.

Katz, J. (2012). Compression effects in English. *Journal of Phonetics*, *40*(3), 390–402.

Katz, J. J. I. (2010). *Compression effects, perceptual asymmetries, and the grammar of timing.* PhD thesis, Massachusetts Institute of Technology.

Keating, P. A. (2006). Phonetic encoding of prosodic structure. In *Speech production: models, phonetic processes, and techniques* (pp. 167–186). Psychology Press.

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics*, *43*, 47–56.

Kiparsky, P. (1982). Lexical morphology and phonology. In *Linguistics in the Morning Calm.* Seoul: Hanshin Publishing Company.

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America*, *121*(4), 2261–2271.

Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Technical Report Coorperative Research Project 3288, Vol. I, Columbia University.

Lee-Kim, S.-I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, *4*(2).

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*(3), 172–187.

Losiewicz, B. (1992). *The effect of frequency on linguistic morphology.* PhD thesis, University of Texas at Austin.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America*, *125*(6), 3962–3973.

McCarthy, J. J. (2005). Optimal paradigms. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory*, Oxford Studies in Theoretical Linguistics. New York: Oxford University Press.

Mousikou, P., Strycharczuk, P., Turk, A., Rastle, K., & Scobbie, J. M. (2015). Morphological effects on pronunciation. *Proceedings of the 18th ICPhS, Glasgow.*

Munhall, K. G., Fowler, C. A., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, *20*, 225–239.

Neu, H. (1980). Ranking of constraints on /t, d/ deletion in American English: a statistical analysis. In W. Labov (Ed.), *Locating language in time and space*, volume 1 of *Quantitative analyses of linguistic structure* (pp. 37–54). Academic Press.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, *54*(5), 1235–1247.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics*.

Plag, I., Homann, J., & Kunter, G. (2015). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 1–36.

Pluymaekers, M., Ernestus, M., Baayen, R. H., & Booij, G. (2010). Morphological effects on fine phonetic detail: The case of Dutch –igheid. *Laboratory Phonology*, *10*, 511–532.

Port, R. & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, *17*, 257–282.

R Core Team (2015). R: A language and environment for statistical computing.

Raffelsiefen, R. (2005). Paradigm uniformity effects versus boundary effects. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in Phonological Theory*, Oxford Studies in Theoretical Linguistics. New York: Oxford University Press.

Roettger, T., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics*, *43*, 11–25.

Rose, D., Hume, E., & Hay, J. (2015). Morphological predictability shapes the phonetic realization of morphemes.

Schreuder, R. & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Erlbaum.

Schuppler, B., van Dommelen, W. A., Koreman, J., & Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, *40*(4), 595–607.

Scobbie, J. M., Turk, A., & Hewlett, N. (1999). Morphemes, Phonetics and Lexical Items: The Case of the Scottish Vowel Length Rule. In *Proceedings of the XIVth International Congress of Phonetic Sciences.*, volume 2, (pp. 1617–1620). International Congress of Phonetic Sciences.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*(1), 140–155.

Shaiman, S. (2001). Kinematics of compensatory vowel shortening: the effects of speaking rate and coda composition on intra- and inter-articulatory timing. *Journal of Phonetics, 29*, 89–107.

Shattuck-Hufnagel, S. & Turk, A. (1998). The domain of phrase-final lengthening in English. In *Proceedings of the 16th International Congress on Acoustics*, volume 2, (pp. 1235–1236).

Smith, R., Baker, R., & Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics, 40*(5), 689–705.

Song, J. Y., Demuth, K., Evans, K., & Shattuck-Hufnagel, S. (2013). Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *Journal of the Acoustical Society of America, 133*(5), 2931–2946.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21*, 291–311.

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology 5.*

Strycharczuk, P. & Kohlberger, M. (2016). Resyllabification reconsidered: On the durational properties of word-final /s/ in Spanish. *Laboratory Phonology, 7*(1), 1–24.

Strycharczuk, P. & Scobbie, J. M. (2015). Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English. *Proceedings of the 18th ICPhS, Glasgow*, (0309).

Sugahara, M. & Turk, A. (2004). Phonetic reflexes of morphological boundaries at a normal speech rate. In *Speech Prosody 2004, International Conference.*

Sugahara, M. & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology, 26*(03), 477–524.

Tauberer, J. & Evanini, K. (2009). Intrinsic vowel duration and the post-vocalic voicing effect: some evidence from dialects of north american English. In *INTERSPEECH*, (pp. 2211–2214). Citeseer.

Turk, A. E. & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*, 397–440.

van Son, R. J. J. H. & Pols, L. C. W. (2003). How efficient is speech? In *Proceedings of the 25th Institute of Phonetic Sciences*, volume 25, (pp. 171–184).

Walsh, T. & Parker, F. (1983). The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics*, *11*, 201–206.

Warner, N. (2012). Methods for studying spontaneous speech. In A. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 621–633). Oxford: Oxford University Press.

Warner, N., Good, E., Jongman, A., & Sereno, J. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics*, *34*(2), 285–293.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics*, *32*(2), 251–276.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*(3), 1707–1717.

Winter, B. & Röttger, T. (2011). The nature of incomplete neutralization in German. *Grazer Linguistische Studien*, *76*, 55–76.

Withgott, M. (1982). *Segmental evidence for phonological constituents*. PhD thesis, University of Texas at Austin.

Zuraw, K. & Peperkamp, S. (2015). Aspiration and the gradient structure of English prefixed words. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Chapter 5

Discussion

5.1 Summary of main findings

This research presented in this dissertation investigated how speech production is influenced by the relations that a target word has with other words in the context and in the lexicon. Word and segment durations were used to evaluate three questions about these influences: whether there is a lasting effect of syntagmatic reduction, whether hyperarticulation is tailored to a phonological contrast, and whether morphological paradigms influence phonetic realization. The corpus studies in Chapter 2 focused on probabilistic reduction, a phenomenon in which words are reduced when they are highly predictable given other words in the context. It was argued that the repeated effects of syntagmatic predictability accumulate in the long run (through one or more possible mechanisms; see §2.4.1), and permanently shorten the phonological duration of usually-predictable words. The experiment in Chapter 3 tested how context-specific competition with a minimal-pair word affects hyperarticulation strategies. It was found that talkers can shorten vowels to enhance a context-relevant lexical contrast, which suggests that they have the ability to target specific phonological distinctions that are communicatively important. This shows that speakers can learn to make sophisticated, contrast-specific adjustments to enhance intelligiblity that go beyond across-the-board enhancements in clear speech. Chapter 4 investigated a way in which morphologically-related words might influence the phonetic realization of a target word through uniformity effects. The results of a laboratory production ex-

periment showed that English words with /-z/ inflectional suffixes (e.g., *frees*) are produced with significantly longer duration than uninflected homophones (*freeze*). This supports a model in which the morphological relatives of a word interact with the phonetic encoding of that word, such as through spreading activations or lexical analogy.

5.2   Lexical relations and language change

Although these studies focused on online or synchronic effects, there is broader interest in whether and how lexical interactions affect phonological representation over the long-term. This is also called the *conditioning* question of language change (Garrett & Johnson, 2013; cf. Weinreich, Labov & Herzog, 1968):

(4) What role do lexical and morphological factors play in sound change?

Among the factors that might be involved, Garrett & Johnson (2013, §3.5.3) point out:

(5) a. lexical usage patterns such as WORD FREQUENCY;

b. morphology-phonetics interactions, especially the MORPHOLOGICAL CONDITIONING of a phonological or phonetic pattern;

c. functional concerns including HOMOPHONY AVOIDANCE

In understanding the role of each factor, it is important to examine how it synchronically interacts with articulation and representation. The research in this dissertation is intended to contribute to this question: how do lexical relations and patterns interact with phonetic realization in ways that might provide

the precursors for long-term phonological change? For example, in understanding why apparently-functional homophony avoidance occurs, it is essential to understand what hyperarticulation and clear-speech strategies are available to speakers in normal conversation (cf. Chapter 3). Without such research, questions about language change such as Ohala's (1993) challenge to functionalism are difficult to address (cf. Ohala, 1994):

> "If speakers have such control over their pronunciation as to worry about maintaining the phonetic distinctions between words ... why are they helpless in the face of one phonological change but masters of the situation in the other?"

In this case, a possible answer may arrive through a better understanding of what kinds of phonetic distinctions and lexical contrasts are accessible to (and adjustable by) speakers in production, and crucially in what contexts they choose (or choose not) to maintain phonetic distinctions. The following sections thus address in more detail how the findings in each chapter might contribute to a synchronic understanding of contextual and paradigmatic interactions in speech, with the goal of informing research into the conditioning question of language change.

5.2.1   Cumulative effects of inter-word reduction

With regard to lexical patterns, word and construction frequency has been emphasized in the literature on usage-driven sound change (Philips, 1984; Lindblom, Guion, Hura, Moon & Willerman, 1995; Bybee, 2000; Pierrehumbert, 2001). However, it is not clear whether word frequency conditions sound change more generally (Garrett & Johnson, 2013; Zellou & Tamminga, 2014), and it has been

found that frequency may interact with a sound change in more than one way (Wedel, Jackson & Kaplan, 2013; Hay, Pierrehumbert, Walker & LaShell, 2015; Hay & Foulkes, 2016). The findings in Chapter 2 suggest that it is important to consider the contexts that a word tends to appear in, not just how frequently it appears (see also Bybee, 2002; Cohen Priva, 2008, 2012, 2015)

In fact, there is earlier evidence showing that the diachronic trajectory of a word within a sound change can be shaped by the synchronic distribution of lexical contexts in which that word appears (Bybee, 2002; Brown, 2004; Brown & Raymond, 2012; Raymond & Brown, 2012; Brown, 2014). For example, in Spanish, there is a class of Peninsular Latin words beginning with a singleton onset /f/ (see Brown & Raymond, 2012). Deletion of singleton /f/ is lexically idiosyncratic. In some words, such as in (6), below, the /f/ has undergone deletion in modern Spanish. In others, such as in (7), it has been retained.

(6) a. Latin *fabulari* > Spanish *hablar* 'to talk'

b. Latin *fornus* > Spanish *horno* 'oven'

(7) a. Latin *favor* > Spanish *favor* 'favor'

b. Latin *focus* 'fireplace' > Spanish *foco* 'focus'

Brown & Raymond (2012) demonstrate that these diachronic lexical idiosyncrasies can be captured by looking at the words that tend to occur adjacent to *hablar* and *horno* (in 6), as compared to *favor* and *foco* (in 7). Initial deletion (or debuccalization) is favored following non-high vowels (Raymond & Brown, 2012), and the words in (6) disproportionately occurred in these contexts. Over time, onset /f/ was deleted in words that tended to occur following non-high vowels, but

was retained in words that more often occurred in other contexts. Related work has shown similar cumulative effects of lexical context, in Spanish and in English (Brown, 2004; Raymond & Brown, 2012; Brown, 2014; Barth, 2015; Raymond, Brown & Healy, 2016).

More broadly, this evidence aligns with the findings and evidence discussed in Chapter 2 to suggest that frequent reduction in syntagmatic contexts leads to the lexicalization of that reduction, such that frequently-reduced word forms become reduced out-of-context. With regard to predictability-driven reduction in particular, the result here may help explain why the most predictable words have the fewest number of segments across languages (Piantadosi, Tily & Gibson, 2011): the cumulative effects of predictability-driven reduction were lexicalized in highly predictable words, such that these forms became phonologically short (although this is not the only process that could contribute to this phonological pattern; e.g., Kanwal, Smith, Culbertson & Kirby, 2016).

This raises questions about whether and how inter-word production patterns might condition language change more generally. Within a word, the phonetic precursors associated with a particular sound change are reliably present. For example, phonological vowel nasalization within a word is likely related to coarticulatory nasalization before a nasal consonant (Beddor, 2009). Prior to a categorical change, the non-categorical nasalization would have been relatively consistent in synchronic productions, since the coarticulatory context always exists within a word.

However, between words, the presence of phonetic precursors is substantially more variable. What is the necessary distribution of inter-word effects such that they lead to lasting change? Among other things, this is likely to involve the robustness of an inter-word pattern and the range of variation involved. Chapter

2 looked at the rate of probabilistic reduction, arguing that words which typically undergo probabilistic reduction are shortened. At the same time, this measure was insensitive to the overall frequency with which context-driven reduction occurs for a given word. A low-frequency word that is used very rarely, but is predictable in the one or two contexts in which it does occur, was expected to undergo the same representational change as an extremely common word that occurs over and over in predictable contexts. With regard to the incorporation of these reduction patterns into representation, it may be that speakers are sensitive not only to the proportion of times in which a word appears in a predictable context, but also to the weight of the evidence. Frequency is thus expected to play a more indirect role in conditioning lexical change. However, this depends crucially on what mechanism is responsible for the lexicalization of context-driven reduction, and how speakers adapt their perception and production to accommodate variable inter-word patterns in speech (e.g., §2.4.1).

5.2.2   Morphological conditioning in sound change

The 19th-century model of language change considers morphological conditioning to involve categorical lexical analogy, rather than regular sound change with phonetic precursors.[1] These precursors may have a variety of phonetic sources (e.g., motor planning, gestural mechanics, aerodynamics, and perception; Garrett & Johnson, 2013), but usually do not include morphological or lexical interactions (Garrett, 2015). One example of a precursor pattern, mentioned above, involves coarticulatory vowel nasalization preceding a nasal coda consonant, which may be

---

[1]Although it is not yet clear how exactly phonetic precursors lead to long-term change; for research on this topic, see e.g. Ohala, 1989, 1993; Labov, 1994; Kiparsky, 1995; Blevins, 2006; Beddor, 2009; Baker, Archangeli & Mielke, 2011; Garrett & Johnson, 2013; Yu, 2013b; Fruehwald, 2016 among many others.

phonologized to create nasal vowels (Beddor, 2009).

Nevertheless, Garrett (2015) points out that the influences of morphological relatives on non-contrastive phonetic variation (called *subphonemic analogy*) have been observed since at least Bloomfield (1933). One example is the Scottish Vowel Length Rule, in which /i, ʉ, ai/ are lengthened in open syllables, as well as before voiced fricatives and /r/ (Scobbie, Turk & Hewlett, 1999; Scobbie, Hewlett & Turk, 1999; Ladd, 2005, 2016). Below, (8a) is lengthened in an open syllable, in contrast with (8b). However, lengthening also occurs in (8c), even though it does not have the appropriate phonological environment (Scobbie et al., 1999).

(8) a. *brew* [brʉː]

b. *brood* [brʉd]

c. *brewed* [brʉːd]

This can be explained as an analogical (or paradigm-uniformity) pattern: the long vowel in (8a) is extended to its morphological relative in (8c) (Scobbie et al., 1999; Ladd, 2005, 2016; Scobbie et al., 1999). How did this pattern arise? One possibility is that this analogical pattern developed suddenly and categorically during the change, perhaps on a word-by-word basis. However, an alternative is that it had a gradual development involving a synchronic precursor pattern, which was not conditioned purely by a phonological environment (see Garrett & Johnson, 2013; Garrett, 2015). While lexical interactions have not typically been considered to be a source of phonetic precursors, the data in Chapter 2 provide evidence for a kind of representational change that has a phonetic precursor (online probabilistic reduction), and which applies to all words in the lexicon, yet is not regular in the sense of sound change.

The same argument might be made for subphonemic analogy (or paradigm uniformity) as a precursor pattern (see Garrett & Johnson, 2013; Garrett, 2015). In particular, the finding in Chapter 4 was that some English inflected wordforms may be influenced by the prosodic properties of their stem words, and crucially that this influence can have non-categorical effects on phonetic realization. If gradient coarticulatory interactions can serve as a phonetic precursor, and if lexical interactions lead to representational change (Chapter 2; §5.2.1), this finding offers the possibility that the influence of morphological relatives on articulation can serve as a precursor as well. This might help explain how a pattern such as the Scottish Vowel Length Rule in (8) arose: paradigm-uniformity effects on inflected word duration became phonologized at the same time as the durational differences in open versus closed syllables (see also Strycharczuk & Scobbie, submitted).

Garrett (2015) lists a number of further English allophonic patterns that involve subphonemic analogy. While detailed phonetic analysis of historical or related English varieties would be informative (if possible), each alternation may have been preceded by the low-level phonetic influence of morphological relatives in synchronic speech production. In this vein, a growing body of experimental work highlights a variety of synchronic patterns in which morphological structure or status conditions acoustic or articulatory differences (see references in §4.1, as well as Cho, 2001; Mousikou, Strycharczuk, Turk, Rastle & Scobbie, 2015; Tomaschek, Tucker, Wieling & Baayen, 2014; Hay, 2003, 2007; Plag, 2014; Yu, 2007). In the same way that phonetic context (such as coarticulatory patterns) may provide the precursors of regular sound change, each synchronic subphonemic pattern may provide a precursor to a morphologically-conditioned sound change.

5.2.3 Phonological similarity and homophony avoidance

A third kind of lexical interaction that might play a role in language change is homophony avoidance (Martinet, 1952; Blevins & Wedel, 2009; Kaplan, 2011). In sound change, mergers are less likely to occur when a large number of homophones will be created (Wedel, Kaplan & Jackson, 2013; Bouchard-Cote, Hall, Griffiths & Klein, 2013), and words that are most confusable in the lexicon may provide the most resistance to a threatened merger (Maclagan & Hay, 2007; Wedel et al., 2013; Hay et al., 2015). It is not clear why languages tend to avoid homophones, but a variety of mechanisms have been proposed. Some possibilities include semantic change—in which highly-confusable homophonous words are often avoided by speakers in usage, and gradually replaced with near-synonyms (Kaplan, 2011)—or else more low-level processes in speech perception (Wedel, 2006; Boersma & Hamann, 2008; Żygis & Padgett, 2010; Sonderegger & Yu, 2010; Yu, 2013a).

Accounts involving hyperarticulation or similar production-side mechanisms are generally avoided as an explanation. It seems implausible that speakers might collectively decide to halt a sound change, or to change their articulations, in order to avoid a merger. While speakers do adapt their hyperarticulation strategies in order to improve intelligibility of specific contrasts in limited contexts (as in Chapter 3), it is not clear how this would influence population-level diachrony. For this reason, hyperarticulation (even contrast-specific hyperarticulation) is not generally considered to be a motivation for sound change, or for homophony avoidance (cf. Ohala, 1993).

Nevertheless, it may be premature to dismiss such explanations without better understanding the contexts, limitations, and scope of hyperarticulation. Chapter 3 provides evidence for context- and contrast-specific hyperarticulation.

An immediate question that occurs is whether context-specific enhancement patterns might accumulate (or otherwise be learned) in the same way that context-specific reduction was argued to do in Chapter 2. One proposed mechanism for the accumulation of probabilistic reduction (see §2.4.1.4) is that speakers acquire more precise gestural scores for words that are likely to occur in a confusable context. With regard to homophony avoidance, the expectation might be that if there are a large number of such words containing a particular segment, these words—and that segment—will on average be produced more precisely, and that particular segment is thus more likely to avoid merger with the distribution of a neighboring segment (see Wedel et al., 2013). This might be tested with synchronic data: first, are words that are highly confusable (e.g., high-informativity) produced with less variability? Second, on average, are segments that are involved in relatively few minimal-pair contrasts within a language produced with more variability than segments that are not involved in such contrasts?

5.3 Additional questions and directions

5.3.1 Context-specific reduction and enhancement

The studies here investigated how acoustic word and segment durations are affected by a speaker's knowledge of related words and patterns, such as inter-word trends in usage contexts, and phonological or morphological paradigms. Studies on hyperarticulation strategies (Chapter 3) are informative about how speakers structure the phonological and lexical contrasts among words. For example, it is an open question as to how hyperarticulated productions are related to a speaker's ideal phonetic or phonological target (e.g., Lindblom, 1990; Johnson, Flemming &

Wright, 1993; Whalen, Magen, Pouplier, Min Kang & Iskarous, 2004; Whalen, Magen, Pouplier, Kang & Iskarous, 2004; Kang & Guion, 2008), and to how language-users represent the contrasts among spoken forms. Identifying what perceptual features are enhanced in production, and what strategies speakers use to do so, is likely to provide insight into how speakers perceive and represent phonological and phonetic forms (Stevens & Keyser, 1989; Keyser & Stevens, 2006). The results in Chapter 3 demonstrate, at a minimum, that speakers are able to identify some of the perceptual cues that distinguish a context-relevant lexical contrast, and efficiently use that knowledge to improve the likelihood of communicative success (cf. Buz, Tanenhaus & Jaeger, 2016).

The results in Chapter 2 provides strong support for the incorporation of context-driven production patterns into offline lexical representations (e.g., Bybee, 2002; Brown, 2004; Bybee, 2006). A future area of research is to explore the particular conditions that are necessary for context-specific effects to be incorporated into lexical representations. Beyond the directions discussed in §5.2, one promising direction involves testing the role of internal monitoring processes with regard to how they might guide or limit adaptation (cf. Frank, 2011, and see discussion in §2.4.1.3), as well the long-term persistence of context-specific adaptation in production. More broadly, it may be found that syntagmatic contextual patterns have an increasingly larger role to play in understanding language change and contrast maintenance.

5.3.2 Paradigmatic interactions

A continuing question for future research is to better understand what kinds of paradigmatic lexical interactions are possible, both in terms of the phonetic attributes that may be affected and the kinds of morpho-phonological relations that

are relevant. The work here has focused on acoustic durations, but the hypotheses predict related effects on other auditory or articulatory qualities. For example, inasmuch as open syllables have different f0 or formant trajectories compared to syllables with a coronal coda, the expectation is that these trajectories should be different in *frees* compared to *freeze*, due to uniformity effects. With respect to morphological paradigms, research so far has tested primarily the effect of stem words on more complex inflected or derived forms. Which words within a paradigm can influence the production of which other words (Kaplan, 2016b,a; Hall & Scott, 2007; Albright, 2008)? Further, what are the psycholinguistic mechanisms that are responsible? For example, one such mechanism for this effect might be cascading activation (Peterson & Savoy, 1998; Rapp & Goldrick, 2000), in which an intended word's morphological relatives are co-activated along with the intended word (Goldrick, Baker, Murphy & Baese-Berk, 2011; Winter & Röttger, 2011; Roettger, Winter, Grawunder, Kirby & Grice, 2014). Exploring these questions and testing these hypotheses with synchronic data is likely to shed light on how morphological relations, and other kinds of lexical patterns, might condition sound change.

5.4 Bibliography

Albright, A. (2008). Modeling analogy as probabilistic grammar. In J. Blevins & J. Blevins (Eds.), *Analogy in Grammar: Form and Acquisition.* Oxford: Oxford University Press.

Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, *23*(03), 347–374.

Barth, D. G. (2015). *To HAVE and to BE: Function Word Reduction in Child Speech, Child Directed Speech and Inter-adult Speech.* PhD thesis, University of Oregon.

Beddor, P. S. (2009). A Coarticulatory Path to Sound Change. *Language*, *85*(4), 785–821.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, *32*(2), 117–166.

Blevins, J. & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica*, *26*(2), 143–183.

Bloomfield, L. (1933). *Language.* New York: Henry Holt.

Boersma, P. & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*(02), 217.

Bouchard-Cote, A., Hall, D., Griffiths, T. L., & Klein, D. (2013). Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences*, *110*(11), 4224–4229.

Brown, E. L. (2004). *The reduction of syllable-initial /s/ in the Spanish of New Mexico and southern Colorado: A usage-based approach.* PhD thesis, University of New Mexico, Albuquerque, NM.

Brown, E. L. (2014). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. *International Journal of Bilingualism*, *19*(4).

Brown, E. L. & Raymond, W. D. (2012). How discourse context shapes the lexicon: Explaining the distribution of Spanish f-/h words. *Diachronica*, *29*(2), 139–161.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent productions. *Journal of Memory and Language*.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-Based Models of Language*. CSLI Publications.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711–733.

Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica, 58*, 129–162.

Cohen Priva, U. (2008). Using Information Content to Predict Phone Deletion. In Abner, N. & Bishop, J. (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, (pp. 90–98).

Cohen Priva, U. (2012). *Sign and signal: deriving linguistic generalizations from information utility*. PhD thesis, Stanford University.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, *6*(2).

Frank, A. F. (2011). *Integrating linguistic, motor, and perceptual information in language production*. PhD thesis, University of Rochester.

Fruehwald, J. (2016). The early influence of phonology on a phonetic change. *Language*, *92*(2), 376–410.

Garrett, A. (2015). Sound change. In C. Bowern & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 227–248). Routledge.

Garrett, A. & Johnson, K. (2013). Phonetic bias in sound change. In A. C. Yu (Ed.), *Origins of Sound Change: Approaches to Phonologization*. Oxford: Oxford University Press.

Goldrick, M., Baker, H. R., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition, 121*(1), 58–72.

Hall, T. A. & Scott, J. H. G. (2007). Inflectional paradigms have a base: evidence from s-Dissimilation in Southern German dialects. *Morphology, 17*(1), 151–178.

Hay, J. (2003). *Causes and consequences of word structure*. Routledge.

Hay, J. (2007). The phonetics of 'un'. In J. Munat (Ed.), *Lexical creativity, texts and contexts* (pp. 39–57).

Hay, J. & Foulkes, P. (2016). The evolution of remembered /t/ over real and remembered time. *Language*, *92*(2), 298–330.

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

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 505–528.

Kang, K.-H. & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America*, *124*(6), 3909–3917.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2016). The evolution of Zipf's law of abbreviation. In *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*.

Kaplan, A. (2011). How much homophony is normal? *Journal of Linguistics*, *47*(03), 631–671.

Kaplan, A. (2016a). Incomplete Neutralization and the (A) symmetry of Paradigm Uniformity.

Kaplan, A. (2016b). What phonology knows about the lexicon.

Keyser, S. J. & Stevens, K. N. (2006). Enhancement and Overlap in the Speech Chain. *Language*, *82*(1), 33–63.

Kiparsky, P. (1995). The phonological basis of sound change. In J. Goldsmith (Ed.), *Handbook of Phonological Theory* (pp. 1–31). Oxford: Blackwell.

Labov, W. (1994). *Principles of Linguistic Change, Volume I: Internal Factors*. Blackwell.

Ladd, B. (2005). Vowel length in Scottish English: new data based on the alignment of pitch accent peaks. *Journal of Phonetics*, *37*, 145–161.

Ladd, D. R. (2016). Lexical allophones. In *Perspectives on Marginal Contrasts*, Ithaca, NY.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&h theory. In *Speech production and speech modelling* (pp. 403–439). Springer.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, *7*(1), 5–37.

Maclagan, M. & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English 'short' front vowel shift. *Language Variation and Change*, *19*, 1–25.

Martinet, A. (1952). Function, Structure, and Sound Change. *Word*, *8*(1), 1–32.

Mousikou, P., Strycharczuk, P., Turk, A., Rastle, K., & Scobbie, J. M. (2015). Morphological effects on pronunciation. *Proceedings of the 18th ICPhS, Glasgow*.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (Eds.), *Language change: Contributions to the study of its causes* (pp. 173–198). Berlin: Mouton de Gruyter.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives* (pp. 237–278). London: Longman.

Ohala, J. J. (1994). Acoustic study of clear speech: a test of the contrastive hypothesis. In *Proceedings of the International Symposium on Prosody*, (pp. 75–89)., Yokohama, Japan.

Peterson, R. R. & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 539.

Philips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320–342.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526–3529.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structurestudies in language* (pp. 137–157). Amsterdam: John Benjamins.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics*.

Rapp, B. & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499.

Raymond, W. D. & Brown, E. L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. *Frequency effects in language*, *2*, 35–52.

Raymond, W. D., Brown, E. L., & Healy, A. F. (2016). Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. *Language Variation and Change*, *28*(02), 175–202.

Roettger, T., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics*, *43*, 11–25.

Scobbie, J. M., Hewlett, N., & Turk, A. E. (1999). Standard English in Edinburgh and Glasgow: the Scottish Vowel Length Rule revealed. In P. Foulkes & G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles* (pp. 230–245). London: Arnold.

Scobbie, J. M., Turk, A., & Hewlett, N. (1999). Morphemes, Phonetics and Lexical Items: The Case of the Scottish Vowel Length Rule. In *Proceedings of the XIVth International Congress of Phonetic Sciences.*, volume 2, (pp. 1617–1620). International Congress of Phonetic Sciences.

Sonderegger, M. & Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In *Proceedings of the 32nd annual conference of the cognitive science society*, (pp. 375–380).

Stevens, K. & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language*, *65*(1), 86–106.

Strycharczuk, P. & Scobbie, J. M. Gradual or abrupt? The phonetic development of morphologisation.

Tomaschek, F., Tucker, B. V., Wieling, M., & Baayen, R. H. (2014). Vowel articulation affected by word frequency. In *Proceedings of the 10th International Seminar on Speech Production*, (pp. 425–428).

Wedel, A. (2006). Exemplar models, evolution and language change. *Linguistic review*, 1–40.

Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*, *56*(3), 395–417.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: a corpus study. *Cognition*, *128*, 179–186.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. C. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 95–195). Austin: University of Texas Press.

Whalen, D., Magen, H., Pouplier, M., Kang, A., & Iskarous, K. (2004). Vowel targets without a hyperspace effect. *Language*, *80*, 377–378.

Whalen, D., Magen, H., Pouplier, M., Min Kang, A., & Iskarous, K. (2004). Vowel Production and Perception: Hyperarticulation without a Hyperspace Effect. *Language and Speech*, *47*(2), 155–174.

Winter, B. & Röttger, T. (2011). The nature of incomplete neutralization in German. *Grazer Linguistische Studien*, *76*, 55–76.

Yu, A. C. L. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology*, *24*(01), 187.

Yu, A. C. L. (2013a). Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. Yu (Ed.), *Origins of Sound Change: Approaches to Phonologization*. Oxford: Oxford University Press.

Yu, A. C. L. (2013b). *Origins of Sound Change: Approaches to Phonologization*. Oxford University Press.

Zellou, G. & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, *47*, 18–35.

Żygis, M. & Padgett, J. (2010). A perceptual study of Polish fricatives, and its implications for historical sound change. *Journal of Phonetics*, *38*(2), 207–226.