Speech production planning affects variation in external sandhi

Oriana Kilbourn-Ceron

Linguistics Department

McGill University, Montréal

May 2017

A thesis submitted to McGill University in partial fulfillment for the requirements of the degree of Doctor of Philosophy

Abstract

Phonological variation is common in many alternations, especially in processes where the target and the trigger of the alternation are in different words—*external sandhi* processes. Much previous work on external sandhi has addressed the morpho-syntactic locality conditions that restrict these cross-word processes, but they are also often sensitive to phonetic and usage factors like pauses, speech rate, lexical frequency, and speech style. What has not been explored in previous work is *why* these factors are consistently associated with external sandhi.

This thesis pursues the hypothesis that patterns of external sandhi variation are shaped by online speech production planning constraints, which can mediate the effect of both grammatical and non-grammatical factors. I investigate the **Production Planning Hypothesis** (PPH) proposal that the narrow window of phonological encoding can block application of external sandhi processes—if the triggering context is not within the same planning window as the target of a process, it cannot apply. The size of the speech planning window is variable, and has been shown to be influenced by many of the factors associated with phonological variation. The predictions of the PPH are tested in case studies of three different external sandhi processes. These studies also contribute to a general understanding of the relationship between variability and syntactic, prosodic, and lexical factors.

The first study investigates the effect of word boundaries, prosodic position, and phonetic pauses on variation of high vowel devoicing (HVD) in Tokyo Japanese. Statistical modeling of HVD patterns in a corpus of spontaneous speech suggests that these factors jointly affect HVD, and that position within a prosodic phrase modulates the effect of speech rate and lexical frequency on HVD. It is proposed that two distinct processes underlie HVD, one that is sensitive to the segmental content of the upcoming word (interconsonantal HVD), and another that is triggered by strong prosodic boundaries (phrase-final HVD). Under this view, part of the variation can be explained under the PPH as production planning effects on interconsonantal HVD.

In a second case study, two factors previously associated with external sandhi variation are tested directly: syntactic structure, and lexical frequency. Both of these factors have also been shown to affect speech planning—and therefore, according to the PPH, should affect external sandhi.

In a production experiment, we examine the effect of a clause boundary on realization of word-final coronal stops in North American English. Clause boundaries are found to have a gradient inhibitory effect on flapping, beyond the effect of associated final lengthening. The PPH explanation for this effect is that a clause boundary induces a delay while high-level planning of the next clause takes place, so segmental details of a potentially flap-triggering word will rarely be available. In contrast, a word that is within the same clause is much more like to be planned within the same window.

Lexical frequency can also affect the time course of speech planning. Higher lexical frequency is associated with faster retrieval, so the PPH predicts that the realization of a word-final coronal stop will be related to the frequency of the word that follows it. A higher frequency following word will be retrieved more quickly, and be more likely to trigger flapping. The relationship between lexical frequency and coronal stop realization is examined in a corpus of American English, and a positive correlation is found between frequency and flapping, as predicted.

The third case study extends testing of PPH predictions to a non-reductive external sandhi process: liaison in French. Frequency and also predictability are used as proxies for upcoming word availability. The PPH predicts that the correlations should be positive, just as for flapping, since both flapping and liaison rely on the knowledge that the upcoming word starts with a vowel.

Examination of two syntactic contexts suggests that increased following word predictability, measured by both local (conditional probability) and global (lexical frequency), increase the likelihood of liaison application. Finding the same effect for the qualitatively distinct processes of flapping and liaison lends support to the PPH proposal that accessibility of the word containing the triggering information for sandhi constrains application of the process.

The PPH offers a unified account of variation in external sandhi related to both grammatical and non-grammatical factors. In addition, the PPH makes many new, testable predictions for future work: the size of the window for phonological encoding should be correlated with the application of external sandhi, with less sandhi applying when the window is more narrow.

Resumé

La variation est un trait fréquent des alternances phonologiques, notamment chez les processus où la cible et le déclencheur ne sont pas dans le même mot—un type de processus qui s'appelle le *sandhi externe*. La recherche portant sur le sandhi externe aborde surtout les contraintes sur la localité morphosyntaxique qui restreignent ces processus, mais ces processus sont généralement aussi sensibles à des facteurs phonétiques et contextuels tels que la présence de pauses, le débit de la parole, la fréquence lexicale et le style du discours. Ce qui n'a pas été exploré jusqu'à présent est *pourquoi* ces facteurs sont constamment associés aux processus de sandhi externe.

Cette thèse poursuit l'hypothèse que la variabilité du sandhi externe est en partie le résultat des contraintes sur la planification de la production de la parole en temps réel, ce qui ont un effet médiateur sur les facteurs grammaticaux et non-grammaticaux. Je teste l'*Hypothèse de la planification langagière* (HPL), une hypothèse qui propose que le fait d'avoir une fenêtre limitée du codage phonologique peut bloquer l'application des processus de sandhi externe; lorsque le déclencheur du processus ne fait pas partie de la même fenêtre de codage que la cible du processus, le processus n'aura pas lieu. La taille de la fenêtre de planification de la parole est variable et est sensible à plusieurs des facteurs associés à la variation phonologique. Je teste ce que la HPL prédit dans trois études de cas de processus de sandhi externe distincts. Ces études contribuent également à une meilleure compréhension de la relation entre la variabilité et des facteurs syntaxiques, prosodiques et lexicaux.

La première étude dans cet ouvrage examine les effets des frontières lexicales, de la position prosodique et des pauses phonétiques sur la variation dans le dévoisement des voyelles hautes (DVH) en japonais de Tokyo. Les résultats de modèles statistiques créés à partir d'un corpus de parole spontanée suggèrent que l'on devrait considérer les effets de ces variables conjointement et que la position dans une phrase prosodique module les effets du débit de la parole et de la fréquence lexicale sur le DVH. Ces résultats mènent également à la proposition qu'il existe deux processus de dévoisement distincts: un premier qui est sensible au contenu segmental du mot suivant (DVH interconsonantique) et un deuxième qui est plutôt déclenché par la présence de frontières prosodiques fortes (DVH en fin de phrase). Grâce à cette constatation que le phénomène de départ est véritablement deux phénomènes, une partie de la variation s'explique par les effets de production de la parole (le DVH interconsonantique) que prédit la HPL.

Ensuite, deux facteurs liés à la variation chez les processus de sandhi externe sont testés de façon directe: la structure syntaxique et la fréquence lexicale. L'effet de ces deux

facteurs sur la planification de la parole a déjà été démontré — et, selon la HPL, on s'attend donc à ce que ces facteurs soient également liés au sandhi externe.

À l'aide d'une expérience de production, nous étudions l'effet de la présence d'une frontière de proposition sur la réalisation des plosives coronales en anglais nord-américain. Les frontières de proposition ont un effet inhibiteur gradient sur le battement de ces plosives au-delà de l'allongement auquel on s'attend en fin de phrase. La HPL réussit à expliquer cet effet: les frontières de proposition induisent un délai pendant la planification de haut niveau de la proposition qui suit, ce qui implique que les détails segmentaux du mot suivant—qui pourrait déclencher le battement de la plosive en finale de mot—ne seront que rarement planifiés. En contraste, un mot dans la même proposition est bien plus propice à être planifié dans la même fenêtre.

La fréquence lexicale peut également avoir un effet sur la planification de la parole en temps réelle. Les mots à fréquence lexicale élevée sont liées à une recherche lexicale plus rapide, donc la HPL prédit que la réalisation d'une plosive coronale en finale de mot sera liée à la fréquence lexicale du mot suivant. Si le mot suivant a une fréquence lexicale élevée, la HPL prédit que la plosive sera réalisée comme battue plus souvent. Cette association entre la fréquence lexicale du mot suivant et la réalisation des plosives coronales est confirmée grâce à une étude de corpus.

La troisième étude de case vise tester les prédictions de la HPL dans un cas de sandhi externe qui n'est pas un processus de réduction: la liaison en français. La disponibilité d'un mot subséquent est estimée selon la fréquence lexicale et la prévisibilité lexicale. La HPL prédit qu'il devrait y avoir une corrélation positive dans les deux cas tout comme pour le battement étant donné que le battement des plosives et la liaison requièrent tous les deux savoir que le mot suivant commence par une voyelle.

Notre étude des deux contextes syntaxiques suggère qu'une augmentation de la prévisibilité lexicale—que ce soit de façon locale (la probabilité conditionnelle) ou de façon globale (la fréquence lexicale)—est liée à une augmentation de la probabilité de liaison. Le fait de trouver le même effet pour des phénomènes aussi différents que le battement et la liaison appuie la proposition de la HPL que la capacité de planifier le mot ayant le déclencheur d'un processus contraint l'application du processus.

La HPL offre une explication unifiée de la variation que l'on trouve chez les processus de sandhi externe pour ce qui est des facteurs grammaticaux et non-grammaticaux. De plus, la HPL offre de nouvelles prédictions testables: tout facteur qui affecte la vitesse du codage phonologique des mots contenant un déclencheur et des mots contenant une cible sera lié à la variation de la réalisation des phénomènes de sandhi.

Contents

List of Figures

List of Tables

Acknowledgements

I am deeply thankful to my supervisors, Morgan Sonderegger and Michael Wagner. Their guidance, mentoring, support, and encouragement has been invaluable, and will stay with me always.

Thank you to all the faculty at the McGill Linguistics Department, especially to Luis Alonso-Ovalle and Bernhard Schwarz for supervising my first Evaluation Paper, and supporting me all the way through to its eventual publication. Thank you to Heather Goad for supervising my second Evaluation Paper, and for offering discussion, ideas and advice throughout my years at McGill. Thank you to Lisa Travis for encouraging me to be a part of the Word Structure Research Group, and giving me the opportunity to pursue my interest in interface-y research. Thank you to Meghan Clayards for being a part of my thesis committee, and sharing her expertise with me.

Thank you to my undergraduate mentors at Concordia University. Mark Hale set me down the path of worrying about the relationship between phonology and syntax, and graciously accepted when I asked him to supervise my undergraduate thesis. Charles Reiss and Dana Isac not only introduced me to the joy of linguistic analysis and cognitive science, but also shaped my philosophy of how linguists and academics can be positive influences in their communities.

Thank you to my fellow students Alanah McKillen, Sepideh Mortazavinia, Gretchen McCullough, Dan Goodhue, Hye-Young Bang, Nina Umont, Henrison Hsieh, Liz Smeets, Colin Brown, Jeffrey Lamontagne, and James Tanner, for the conversation, commiseration, intellectual discussions, and outright silliness that has made these years so much the better.

I am so thankful to my parents for always encouraging me to pursue my interests, and working hard to give me every opportunity they could offer. Thank you to all my family members who have supported and encouraged me in so many ways. Thank you Yuliya Manyakina for the unwavering support through the years and the distances that have separated us and brought us together again. Thank you to Eva Best for understanding me, and being my cheerleader. Thank you to Joshua Rosner for relentlessly believing in me and holding my hand across the finish line.

I would also like to acknowledge the financial support of SSHRC CGS Doctoral Scholarship (767-2012-1089) and the CRBLM Graduate Scholar Stipend. A huge thank you to Jeffrey Lamontagne for translating my thesis abstract into French.

Contribution of Authors

The studies presented in this thesis were prepared as manuscripts for publication elsewhere. I am the primary author of each manuscript. Chapter 2 has been published in *Natural Language and Linguistic Theory* as a co-authored article with Prof. Morgan Sonderegger. I was responsible for the original conception of the research questions and data preparation. The manuscript and statistical analysis were prepared in collaboration with Prof. Sonderegger.

Chapter 3 has been submitted for publication as a manuscript co-authored with Prof. Meghan Clayards and Prof. Michael Wagner. I prepared the manuscript in consultation with Profs. Clayards and Wagner, and we collaborated on revisions and theoretical interpretation. For the corpus analysis, data collection and preparation was carried out by myself and Prof. Wagner, and I prepared the statistical analysis. The experimental results are from an unpublished data set collected by Profs. Clayards and Wagner. I performed the statistical analysis with their input. A version of this manuscript is in press in the *Proceedings of the 52nd Meeting of the Chicago Linguistic Society*.

Chapter 4 is in preparation for submission as manuscript with myself as the sole author. A version of the manuscript appears in *The Proceedings of the 2016 Annual Meeting on Phonology*.

Chapter 1

Introduction

What do speakers learn about the sound patterns of their language, and how do they use that knowledge to produce sounds in context? The assumption of generative phonology is that speakers learn generalizations that abstract away from non-phonological information. But when sound patterns are *variable*, their application is systematically shaped by many factors other than phonological context. Speech rate, syntactic structure, lexical frequency, morphological information, speech style, and social situation can all modulate phonological patterns (Anttila, 2007; Coetzee and Pater, 2011; Guy, 2011). Should the structure of this variation be accounted for within phonological grammar, and if so how?

This thesis focuses on patterns of variation in *external sandhi*, a type of phonological alternation in which the target and the trigger can be in separate words. An example is the well known coronal flapping alternation in North American English: the coronal stops /t,d/ are pronounced as a flap [ɾ] when they appear between two vowels and are not in the onset of a stressed syllable (Kahn, 1976). This alternation can be seen in derived words, like *write* ~ *wri*[ɾ]*er*, but can also apply when the following vowel is in another word, as in *wri*[ɾ]*a poem*, making flapping an external sandhi process[1]. The phonological

[1] Some authors like Kaisse (1985) reserve the term 'external sandhi' for cross-word processes that are (or appear to be) sensitive to morpho-syntactic information. Throughout this thesis, the term is used more broadly to refer to any process that can apply across word boundaries.

environment for flapping could be described with the following rule, where # represents a word boundary, and parenthesis indicate optionality.

(1.1) /t,d/ → ɾ /V___(#)V

However, flapping is considered variable or 'optional,' since the rule does not *always* apply (Randolph, 1989; Patterson and Connine, 2001; Fukaya and Byrd, 2005). Whether flapping applies is affected by the type of boundary following the coronal stop: When the target and trigger are within the same morpheme, flapping is almost categorical (Randolph, 1989), but flapping is less consistent across morpheme and compound boundaries (Patterson and Connine, 2001). Across word boundaries, flapping is even more variable (Gregory et al., 1999), and is affected by the constituency relation between the two words. Flapping is more likely to apply when both words are within a single clause as in (1.2a) than in (1.2b) where they are separated by a clause boundary (Scott and Cutler, 1984).

(1.2) a. The last time we **met Anne** was horrible.

b. The last time we **met, Anne** was horrible.

This pattern is common to most external sandhi phenomena: the 'further apart' the target and trigger of the alternation are, the less likely the process is to apply. This is the puzzle of *locality* of external sandhi, which has been framed in terms of how directly or indirectly syntactic information can affect phonological processes (Cooper and Paccia-Cooper, 1980; Kaisse, 1985; Nespor and Vogel, 1986).

Many external sandhi processes are also variable beyond what can be explained by grammatically-defined locality constraints. For example, flapping is sensitive to speech rate (Kaisse, 1985) and predictability between the target and trigger words (Gregory et al., 1999). Speech rate and lexical frequency are generally considered non-grammatical information, yet they influence variation in many external sandhi processes (Gregory et

al., 1999; Bybee, 2001; Jurafsky et al., 2001; Côté, 2013). What are the mechanisms by which grammatical and non-grammatical variables influence external sandhi variation?

This thesis investigates the hypothesis that online speech production planning constrains external sandhi application, and mediates the effect of both grammatical and non-grammatical factors on variability. Three case studies are presented that test and develop the recently proposed *Production Planning Hypothesis* (PPH; Wagner, 2012; Tanner, Sonderegger, and Wagner, 2017).

The PPH rests on the idea that the size of planning 'chunks' for phonological encoding, being relatively small (Levelt, Roelofs, and Meyer, 1999; Wheeldon and Lahiri, 2002; Wheeldon, 2012), may not always encompass two adjacent words. Hence some words may be phonologically encoded in the absence of information about segments in an upcoming word, preventing interaction between the two words, and therefore blocking external sandhi processes from applying. The details of this hypothesis will be elaborated in Section 1.3, after a summary of previous work on grammatical constraints on external sandhi in Section 1.1 and on quantitative approaches to phonological variation in Section 1.2.

1.1 Locality constraints in grammar

External sandhi processes are often constrained by the locality between the target and trigger containing words. For example, the words *met Anne* are intuitively closer in (1.2a) than in (1.2b), where they are separated by a clause boundary. The question of what counts as 'local enough' for external sandhi to apply has been investigated since early work in generative phonology. Chomsky and Halle (1968) captured these effects by translating syntactic structure into boundary symbols inserted into the phonological string, which could be referred to directly by phonological rules.

The influential framework of Lexical Phonology (Kiparsky, 1982; Mohanan, 1986) posited that cross-word processes happen at a separate, later stage of computation than word-internal processes: the post-lexical level. This architecture marks cross-word processes as exceptional in being influenced by pauses and speech rate. Kaisse (1985) further categorizes cross-word processes according to whether they are sensitive to syntactic categories and structure, dividing them into two classes, P1 and P2. P1 are syntactically-sensitive processes, under Kaisse's model, are restricted by syntax but not pauses, while syntactically insensitive P2 processes apply across the board but are sensitive to pauses.

Direct reference One approach to capturing locality conditions on external sandhi has been to reference syntactic structure directly. Kaisse (1985) proposes that external sandhi processes may be sensitive to syntactic c-command relations between the two words that are potentially interacting phonologically.

More recent proposals by Seidl (2001), Pak (2008), Newell (2008), and Newell and Piggott (2014) tie the possibility of phonological interaction to syntactic cycles: external sandhi only applies between morphemes which are spelled out within the same syntactic phase. These proposals all have in common that there should be a one-to-one correspondence between syntactic structure and external sandhi realization, and variability beyond these restrictions is not accounted for. This leaves open the question, what is the place of syntactic effects that are gradient rather than categorically blocking or licensing?

Prosodic Phonology Another influential theory which addresses locality effects is Prosodic Phonology (Selkirk, 1974; Nespor and Vogel, 1986; Inkelas and Zec, 1990; Inkelas and Zec, 1995; Selkirk, 2011). In this theory, utterances are organized into hierarchical constituents which are derived from syntactic structure, but not isomorphic to it. Locality conditions are captured by restricting the application of a phonological process to words which are within a particular prosodic domain. For example, Nespor and Vogel (1986)

assign the flapping rule to apply within the *phonological utterance* domain, since flapping is possible between any two words in an utterance. On the other hand, a rule like liaison in French applies only within the smaller domain of *phonological phrase*.

Translating locality restrictions from syntactic to prosodic terms can explain certain mismatches between syntactic domains and sandhi application. But assigning domains to phonological rules only restricts the upper bound at which an alternation could apply, and does not address patterns of variability within the domain of application. Processes like flapping and high vowel devoicing, categorized as P2 by Kaisse (1985) for their 'insensitivity' to syntax, are able in principle to apply across any type of syntactic boundary. However, boundaries do influence the likelihood of their application: Scott and Cutler (1984) and Fukaya and Byrd (2005) found that clause boundaries block flapping, for example.

Nespor and Vogel (1986) suggest a possible mechanism for explaining certain kinds of variability within Prosodic Phonology: restructuring. Once prosodic constituents are built in accordance to syntactic structure, the constituents may be merged or split depending on factors like speech rate. Hence, realization of sandhi between two words may vary even when their syntactic relationship is held constant—if the prosodic constituents are later restructured. This way of accounting for variability implies a tight correspondence between surface prosodic structure and external sandhi application. To support this type of account, it would be ideal to find independent evidence of surface prosodic phrasing, and verify whether it consistently matches up with the application of liaison. Pak and Friesner (2006) investigate this issue in French by comparing phrasal accent assignment and liaison realization. They find that the domains signaled by phrasal accents are mismatched and incompatible with those suggested by the realization of liaison. Post (2000) similarly finds that the phrasal domains implied by clash resolution in French do not line up with those of liaison application.

These accounts of external sandhi locality do not address the variation that remains even when locality conditions are met. The following section turns to a different class of theories which addresses variability more broadly, and how they relate to the external sandhi.

1.2 Optionality, variation, and usage

External sandhi processes are associated with variation beyond what can be accounted for by locality constraints. For example, flapping can apply between the words *cat attack*, but it is optional. The realization of flapping is also modulated by speech rate, speech style, or lexical frequency (Gregory et al., 1999; Fukaya and Byrd, 2005). As early as Kiparsky (1982) and Mohanan (1982), in the development of Lexical Phonology, there was recognition of an association between variation/optionality and processes that apply across word boundaries: 'lexical' processes only apply within words and are obligatory, while processes that apply across words must be 'post-lexical' and post-lexical processes "are subject to variation" (Kaisse and Shaw, 1985, p. 6).

As variation has come to the forefront of phonological research programmes in recent years (Coetzee and Pater, 2011; Guy, 2011), there has been substantial development of formal models that incorporate quantitative generalizations about variation. We turn to a brief survey of this work and how it relates to the variability of external sandhi processes in particular.

Variation in sociolinguistics Research in the sociolinguistic tradition has produced decades of work on how linguistic and social factors influence phonological variation (Guy, 2011). The formal framework associated with this research is the Variable Rule model (Labov, 1969; Cedergren and Sankoff, 1974), which allows generalizations about particular pronunciations to be modulated by social and linguistic factors. For example, a rule for the process of t/d deletion in English would have a base probability of application

for each speaker, with this likelihood modulated by contextual factors like surrounding segments, presence of pauses, and morphological class (Guy, 1991). Analyses also can include between-speaker variables like age and gender, which reveal patterns across a speech community rather than within individuals.

A Variable Rule analysis can capture quantitative generalizations about different types of conditioning factors. But this approach doesn't offer an explanation for the directionality of the effects.

Constraint-based probabilistic models Variation has also been formally modeled in constraint-based frameworks, mainly based on Optimality Theory (OT). Variable OT (Reynolds, 1994; Anttila, 1997) derives free variation between forms by allowing competing constraints to be unranked with respect to each other, creating parallel possible rankings with different optimal outputs. Since these rankings are all equally probable, this type of model has difficulty predicting patterns where one variant is highly favoured.

Stochastic OT (Boersma, 1997) adds a numerical component to this notion of different potential rankings, which allows more fine-grained modeling of quantitative patterns. Constraints are assigned numerical values which represent their ranking, so the distance between two constraints can be quantified, unlike in classic OT. During evaluation, 'noise' is added to the value of each constraint. Consequently, constraints which are numerically close may end up having a flipped ranking in a given evaluation. Modulating the values of constraints therefore allows many different distributions of variants to be modeled. Weighted constraint systems have become an important part of phonological research, with more recent variants like Noisy Harmonic Grammar (Smolensky and Legendre, 2006; Potts et al., 2010) and Maximum Entropy OT (Goldwater and Johnson, 2003; Jäger, 2007) being actively developed.

What these probabilistic implementations of OT do not capture are patterns which are conditioned by non-grammatical factors. Coetzee and Kawahara (2013), for example,

point out that there is no way to model the difference between high and low frequency words in likelihood of undergoing a variable process. They propose an extension to noisy Harmonic Grammar in which the weight of faithfulness constraints can be scaled by lexical frequency, with higher frequency words being less faithful and therefore less marked than low frequency words. However, this account remains at the level of the individual word, and leaves open the question of how the frequencies of different words in an utterance may influence phonological interactions between them.

Exemplar Theory Exemplar Theory (Bybee, 2001; Pierrehumbert, 2001; Pierrehumbert, 2006) is another strand of quantitative theories of phonology, focusing on the role of usage in shaping phonological patterns. In exemplar-based theories, detailed phonetic representations are stored for each lexical item, recording phonetic detail for all tokens encountered by a speaker. Lexical items that have variable pronunciations will therefore include in their representations 'exemplars' for both alternants in a contextually-conditioned alternation. Variation is captured by directly representing the distributions of possible pronunciations.

For example, a word like *cat* would be associated with exemplars where the final segment is pronounced [t] and where it is pronounced [ɾ], since it will appear in different contexts. On the other hand, the representation for a word like *catapult* will consist of mostly of [ɾ] pronunciations since the /t/ is always intervocalic. Although this approach can model the observed difference in variability between word-internal and cross-word processes, it does not provide any underlying explanation as to why. In principle, an exemplar-based approach could equally easily model a distribution where the realization of a flap is more variable within a word than when the trigger of flapping is across a word boundary, given the right input distribution. This leaves open the question of why no such distributions are observed across languages. It is proposed in this thesis that considering how the PPH shapes language data is a step towards answering that question.

Gestural overlap Phonological processes are traditionally discussed as alternations between two (or more) discrete phonetic forms. Some external sandhi processes are clearly categorical changes, like liaison in French (Post, 2000), but cross-word assimilatory and reductive processes may involve a range of phonetic outcomes. The Articulatory Phonology (AP) framework proposes that both gradient and categorical processes can be modeled by taking *gestures* to be the primitives of phonological representation (Browman and Goldstein, 1992). The gestures are invariant, abstract representations of physical events in the vocal tract.

Contextual variation arises due to the fact that actual articulatory trajectories for each gesture must be interpolated with those of surrounding gestures, and acoustic realizations can differ significantly with small changes in timing, overlap, or magnitude of gestures. For example, the flap [ɾ] realization of /t/ between vowels can be understood as a consequence of the vowel gestures encroaching on /t/ gestures, resulting in acoustic shortening and loss of voicelessness. Hence, flapping and other overlap-based alternations should be subject to influence from factors which affect timing. Speech rate is a clear candidate, with higher speech rates forcing more gestures to be compressed into less time. Indeed, many cross-word processes including flapping are more likely to apply at faster speech rates. Browman and Goldstein (1992) suggest that prosodic boundaries can also affect gestural overlap by modulating gestural timing and magnitude (see also Byrd and Saltzman, 2003). Under this view of contextual variation, the effects of speech rate and boundaries are qualitatively the same: they do not affect any aspect of the phonological representation, but rather the compilation of these representations into a concrete articulatory motor plan. Gestural overlap as a source of variation is not incompatible with the PPH. However, the PPH assumes that at least some contextual variation (i.e. external sandhi) is explicitly planned, and therefore constrained by the scope of the speech planning window. Hence, the PPH predicts variation due to planning effects that is above and beyond the overlap-induced variation predicted by AP.

1.3 Locality of Production Planning Hypothesis

What is it about word boundaries that causes external sandhi to apply more variably than processes which apply within words? The locality of production planning hypothesis (PPH) (Wagner, 2012; Tanner, Sonderegger, and Wagner, 2017) proposes that the constraints on speech production planning shape variability in external sandhi. This dissertation develops and tests the predictions of the PPH. The following sections present evidence from the speech production planning literature that motivates this idea, which will be explored and tested in Chapters 2, 3 and 4 of this thesis.

1.3.1 Speech production planning

According to influential models of speech production (Levelt, Roelofs, and Meyer, 1999; Dell and O'Seaghdha, 1992), planning an utterance involves several distinct, ordered stages. The planning process for an utterance begins with the formulation of the message to be conveyed, and retrieval of the lexical information associated with the concepts in the message. The next step is retrieval of the *lemma* of each lexical concept, which contains the grammatical information necessary to build the syntactic frame for the utterance, but *not* any phonological information.

Lemmas have diacritics which mark either inherent or contextual grammatical properties (e.g. gender, number, tense). Once a lemma's grammatical information is fully specified by construction of sufficient syntactic structure, the step of *word-form encoding* can proceed. This involves retrieval of the metrical and segmental information associated with the lemma so that it can be phonologically encoded, and later used to produce a phonetic motor plan for eventual articulation.

Under this view of speech planning, any facilitation or difficulty at earlier stages of

planning will affect the time course of later stages. Hence, the timing of phonological encoding and subsequent articulation is dependent on lexical selection and morpho-syntactic encoding. This precedence relation between the planning stages is the first crucial assumption of the PPH.

1.3.2 Scope of phonological planning

The stage of word-form encoding is incremental: speakers do not encode all the details of each word in an utterance before they start speaking (Levelt, 1989; Levelt, Roelofs, and Meyer, 1999). If retrieval of segmental information is a sub-part of the word-form encoding process, then 'look-ahead' to segments of upcoming words must be restricted to within the word-form encoding window. The size of this window is therefore crucial to external sandhi: cross-word phonological processes can only apply if the relevant words (and their segmental material) are both active within the same planning window.

There is evidence that the window is quite narrow, potentially as small as a single prosodic word (see Wheeldon, 2012, for an overview). The advance planning of *segmental* details may be restricted to an even smaller window, and there is evidence that encoding of segmental information must be preceded by planning of higher-level prosodic frames, which may further inhibit 'look-ahead' to segmental content of upcoming words.

In the model of Levelt, Roelofs, and Meyer (1999), prosodic word frames must be built first, then segmental content is associated with the word frame. There is evidence that these word frames, independent of segmental content, can be planned relatively far in advance. Sternberg et al. (1978) showed that the production latency for articulating a previously prepared list of words increases as a function of the number of words in the list. This suggests that the number of word-sized units is planned in advance of the start of articulation, assuming each word incurs an additional processing cost.

Wheeldon and Lahiri (1997) and Wheeldon and Lahiri (2002) show that the time it takes to initiate articulation of a prepared utterance depends on the number of prosodic

(rather than lexical) words it constrains, arguing that the prosodic word is the minimal unit of phonological encoding. Wheeldon and Lahiri (1997) showed that the production latencies differed significantly between utterances with one, two, and three prosodic words, and that these differences were independent of stress placement, intonation patterns and phonological phrase structure. Wheeldon and Lahiri (2002) provided further confirmation that the details of stress patterns do not affect production latencies by showing that compound words, and mono-morphemic words of the same length do not differ significantly from each other, but they all have significantly shorter latencies than an adjective-noun sequence of the same length.

Ferreira (1993) showed that the timing slot associated with a phrase-final position is independent of the segmental content of the word that appears in that position. The study investigated whether the total duration of a phrase-final word plus following pause was consistent for inherently short versus long words. Experiments 3 and 4 showed that the durations of the pauses were inversely proportional to the duration of the phrase-final words, causing the total duration of the phrase-final timing slot (word plus pause) to be consistent across word types, even though there was a significant difference in duration between the short versus long words. This supports Ferreira's view that sentence-final timing slots are planned out in advance of specific knowledge of word length, again showing that segmental details are not planned early on in an utterance.

Shattuck-Hufnagel (2000) and Keating and Shattuck-Hufnagel (2002) argue that higher-level prosodic information like intonational phrasing and the associated tunes is also planned early on, before word-specific metrical and segmental information is available.

Evidence from speech errors suggests that syllabic structure is planned prior to and independently of segmental information. Overwhelmingly, segments that are erroneously duplicated or displaced appear in positions that are structurally similar to the intended target (Shattuck-Hufnagel, 2015). Target syllable onsets are moved to or duplicated in

onsets, as in *The **c**ack of the **b**ar* for *the back of the car*, rather than into coda position. This supports the idea that during speech planning, speakers generate a syllabic structural frame prior to encoding the segmental content of the utterance.

In sum, these studies show that the scope of advance speech planning depends on the level of detail, with higher-level prosodic information planned in advance of detailed, segmental information. The scope of segmental phonological encoding may be as small as a single prosodic word in some cases. The proposal of the PPH is that external sandhi processes can be *blocked* if the size of the planning window is too small to encompass both of the words participating in the alternation. According to the PPH, the application of any external sandhi process that has to reference information in upcoming words will necessarily be constrained by the *locality* of phonological encoding. Therefore, any factor which affects the planning window should affect the application of external sandhi: the wider the planning window, the more likely cross-word processes should be to apply.

The size of the planning window is dependent on a number of factors. Previous work has shown that planning scope is modulated by language-specific factors such as syntactic structure (Ferreira, 1991; Wheeldon, 2012), lexical frequency (Konopka, 2012), and upcoming word predictability (Gahl and Garnsey, 2004). Planning scope can also be affected by other cognitive factors such as working memory load and cognitive load (Ferreira and Swets, 2002; Wagner, Jescheniak, and Schriefers, 2010). These are all predicted under the PPH to also modulate the application of external sandhi processes. In the following chapters of this thesis, it will be shown that effects of prosodic, syntactic, and lexical factors on external sandhi variability can be explained under the PPH.

1.4 Research questions and case studies

The Production Planning Hypothesis relates the locality and variability of external sandhi processes to the size of the planning window for phonological encoding. The following

chapters present case studies investigating patterns of variability in external sandhi, and whether they are consistent with the predictions of the PPH.

The first domain of investigation is the relationship between variability and boundaries. The research on external sandhi's locality conditions, reviewed in Section 1.1, shows that syntactic boundaries affect sandhi application. But the mechanism by which syntax influences sandhi is unclear: is syntactic information referred to directly by phonological processes, is the influence mediated by prosodic structure, or is the variation induced by boundaries a by-product of phonetic boundary marking, like pauses and final lengthening? Chapter 2 investigates how prosodic boundaries and their phonetic cues influence external sandhi, while the first experiment in Chapter 3 tests the effect of syntactic clause boundaries. The planning delays associated with phrase and clause boundaries are predicted by the PPH to inhibit external sandhi application.

The second strand of research tests PPH predictions about ease of retrieval/encoding of the words involved in external sandhi. According to the PPH, sandhi should be more likely to apply if the word that contains the trigger of external sandhi is planned sooner relative to the word that contains the alternating segment. This prediction is investigated using measures of word probability which have been associated with ease of retrieval and planning (Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994). Chapters 3 and 4 examine the effects of lexical frequency and contextual predictability on patterns of external sandhi. These are discussed in the context of PPH predictions, and of previous work on probability and predictability effects in variability (Gregory et al., 1999; Jurafsky et al., 2001; Bybee, 2001; Côté, 2013).

1.4.1 High vowel devoicing in Japanese

The first case study, Chapter 2, investigates high vowel devoicing (HVD), an external sandhi process in Tokyo Japanese. The high vowels /i,u/ are realized without voicing

when they are preceded by a voiceless consonant, and followed by either another voiceless consonant (potentially in another word) or a 'pause' (Fujimoto, 2015).

The latter devoicing environment is considered variable (Vance, 2008), but it also not clear what aspect of 'pauses' trigger HVD. This study investigates the pattern of HVD in relation to a number of boundary-related factors which could be associated with a 'pause' to better understand whether they block HVD, license HVD, and how they interact with other factors like speech rate and lexical frequency. Results show that pauses and prosodic position play a joint role in determining likelihood of HVD, and that the effects of speech rate and lexical frequency look qualitatively different depending on the type of boundary that intervenes in the HVD environment. This pattern can be explained by positing two separate processes of devoicing, leading to two different patterns of variability. The devoicing process which is dependent on the following voiceless consonant is constrained by the PPH, explaining why it becomes less likely across word boundaries, and is less likely across prosodic phrases than within them.

1.4.2 Flapping and glottalization in English

The HVD study identified the PPH as a possible explanation for gradient differences in application between boundaries, but did not test the PPH explicitly. The second case study, Chapter 3 investigates the predictions of the PPH more directly, testing the effect of two factors hypothesized to affect the scope of planning: syntactic junctures, and lexical frequency. The pattern of word-final coronal stop realization in English is analyzed in a production experiment and in a spontaneous speech corpus. Two of these outcomes are compared: flapping, a realization which is only possible when the following word begins with a vowel, and glottalization, which is not strictly dependent on the identity of the following segment, but rather may serve to demarcate prosodic boundaries. The production experiment examined how the realization of a word-final /t/ varied depending on whether there was a clause boundary after it or not. Previous work on flapping

found a blocking effect for clause boundaries (Fukaya and Byrd, 2005), and a more gradient blocking effect for word-internal morpheme and compound boundaries (Patterson and Connine, 2001). The PPH predicts that the flap realization of /t/, which crucially depends on there being a vowel in the following word, should be inhibited by the presence of a clause boundary. Conversely, glottalization is predicted to be more prevalent at a clause boundary, and to be more likely when the prosodic boundary associated with that juncture is stronger. Clause boundaries are found to have a gradient inhibitory effect on flapping, beyond the effect of associated final lengthening. In the corpus study, the effect of lexical frequency is tested for both coronal-final and following vowel-initial words that could trigger flapping. The PPH predicts that the facilitatory effect of lexical frequency should lead to higher rates of external sandhi when the word that contains the *trigger* for flapping is more frequent, since it is more likely to be planned together with the coronal-final word. The likelihood of flapping is shown to indeed have a positive correlation with the frequency of the flap-triggering word. On the other hand, glottalization is not predicted to be modulated by lexical frequency of the following word, since it is not crucially dependent on following context. This is borne out in the results: only frequency of the target word had a significant effect on glottalization.

The PPH predictions are compared and contrasted with other accounts of variability in terms of gestural overlap, prosodic hierarchy, and probabilistic reduction. Although these theories have different scopes, some of their predictions overlap with those of the PPH in the case of reductive processes like flapping. Hence, it is suggested that the same factors should be tested for a non-reductive external sandhi processes, which would uniquely distinguish PPH predictions. This is the objective of Chapter 4, which examines variability in French liaison.

1.4.3 Liaison in French

The effect of lexical frequency of the following word on English flapping presented in Chapter 3 was consistent with PPH predictions, but also with predictions of probabilistic reduction accounts. Chapter 4 examines a non-reductive process, French liaison, again testing lexical frequency of the trigger-containing word, and an additional variable potentially correlated with planning scope: the conditional probability of the trigger-containing word. The PPH predicts that similar planning effects should be found regardless of whether the alternation is reductive or not, since they both rely on information in an upcoming word.

The liaison alternation in French is one in which a latent word-final consonant surfaces when the following word begins with a vowel. For example, the adjective *gros* 'big' is pronounced [gʁo] in isolation or before a consonant, but with a final [z] when it modifies a vowel-initial noun as in *gros enjeu* 'big stake/issue'. Since both liaison and flapping are dependent on the information that the following word begins with a vowel, the PPH predicts the same pattern of variability for the two processes: the harder the triggering word is to plan, the lower the likelihood of applying the process should be. However, since liaison is the realization of an *extra* segment, the predictions of gestural overlap and probabilistic reduction based accounts no longer make overlapping predictions. Examination of two syntactic contexts suggests that increased following word predictability, both local (conditional probability) and global (lexical frequency), increase likelihood of liaison. The results of this analysis are discussed in relation to PPH predictions, as well as from the perspective of other probability-related theories.

Chapter 2

Boundary phenomena and variability in Japanese high vowel devoicing[1]

2.1 Introduction

2.1.1 Categorical and variable devoicing

A highly salient feature of Standard Japanese is the devoicing of the high vowels *i* and *u* (henceforth *high vowel devoicing*: HVD). Since the earliest descriptions of this alternation, the environment has been described disjunctively: high vowels are devoiced when they appear between two voiceless consonants, as in *sika* 'deer,' or when preceded by a voiceless consonant and followed by a pause, as at the end of *ikimasu* '(I will) go' (Han, 1962; McCawley, 1968; Vance, 2008). McCawley (1968) gives the following rule:

Rule 1: $V_{[+high]} \rightarrow V_{[-voice]} / C_{[-voice]}\_\_ \{C_{[-voice]}, \#\}$

Although the generalization captured by this rule remains the starting point for standard descriptions of HVD (Vance, 2008; Labrune, 2012; Fujimoto, 2015), a distinction is normally made between the two environments. Labrune (2012) states that for a high vowel

[1]This chapter appears as Kilbourn-Ceron and Sonderegger (2017) in *Natural Language and Linguistic Theory*, and is reproduced here as part of this dissertation with permission from the publisher.

between voiceless consonants, hereafter referred to as "C̥_C̥", devoicing is "almost compulsory." Nielsen (2015) similarly describes HVD as "almost obligatory in the Tokyo dialect, except in some environments where complete devoicing is often blocked." By contrast, Vance (2008, p.210) notes that devoicing before a pause, hereafter referred to as "C̥_#", is "much less consistent" than in the C̥_C̥ environment. Hence, we are faced with the puzzle that HVD is compulsory, yet sometimes variable.[2] The cause of this difference in variability, and more generally what conditions how often HVD applies in a particular context, remains an open question in the literature. Addressing this question is one goal of this paper.

This question connects to a broader issue: by what mechanism can the "same [phonological] process" be categorical (or nearly so) in some environments, and variable in others? This puzzle has been of interest for HVD in particular, where previous work has ascribed categorical versus variable application to "phonological" versus "phonetic" devoicing (see Sec. 2.2.1); many other cases intuitively involve 'more variability' at some sort of boundary, e.g. Hungarian vowel harmony, English and Navajo phonotactic restrictions (Hayes and Londe, 2006; Martin, 2011). Addressing this issue in the case of HVD is relevant for understanding other such cases cross-linguistically, and how to account for them in a formal analysis.

2.1.2 Devoicing and overlapping environments

The distribution of devoiced vowels in Japanese follows several constraints that are attested cross-linguistically, as shown in typological reviews of non-modal vowels in general (Gordon, 1998) and devoicing in domain-final positions in particular (Barnes, 2006, Section 3.6.1). Many languages show a pattern like Japanese, where high vowels but not non-high vowels undergo devoicing; however, the inverse pattern is unattested (Gordon,

[2]Note that the literature refers to cases of non-application of Rule 1 either as "blocking" or "variability". We use the term "variability" for any non-categorical application of Rule 1.

1998). The environments for vowel devoicing in a given language can make reference to adjacent voiceless consonants, to position within a prosodic domain, or both. For example, Turkish (Jannedy, 1995) and Montreal French (Cedergren and Simoneau, 1985) devoice high vowels only in the C̥_C̥ environment, a subset of Japanese HVD environments. In other languages, vowel devoicing is conditioned by final position without reference to segmental context: Ainu (Crothers et al., 1979) and Woleaian (Sohn, 1975) have devoicing in word-final position, while languages like European French (Smith, 2003), Oneida (Michelson, 1999), and Greek (Dauer, 1980; Kaimaki, 2015) devoice vowels at the end of larger, phrasal domains. In the surveys of both Gordon (1998) and Barnes (2006), devoicing at the end of "smaller" and "larger" domains are always in an implicational relationship within a language: devoicing at a smaller domain edge (e.g. word) implies devoicing at larger domain edges (e.g. utterance).

For Japanese, the environment for HVD takes into account both segmental conditions and domain position. But while the segmental conditions on HVD are clear, the role of domain position (i.e. the meaning of "#" in C̥_#) is not well-understood. In a recent review article, Fujimoto (2015) uses the terms *pre-pausal*, *word-final* and *phrase-final* to describe the C̥_# context, though she notes that "[f]urther investigation is essential in order to clarify the details of word/phrase-final devoicing" (p. 186).

Describing two separate environments for this alternation obscures the fact that the environments can and often do overlap, as schematized in Figure 2.1. On the left is the C̥_C̥ environment, where devoicing seems to be obligatory if all the relevant segments are within the same word and no factors blocking devoicing are present (e.g. a pitch accent; see Section 2.2.1). The right hand side gives an example of utterance-final devoicing which is variable for most words, though obligatory for a small set of high frequency verb particles (Maekawa and Kikuchi, 2005; Vance, 2008; Oi, 2013). The overlap between the two environments is shown in the middle, tentatively labelled as variable. But it is not totally clear what is expected in a case where, for example, a word ending in a voiceless

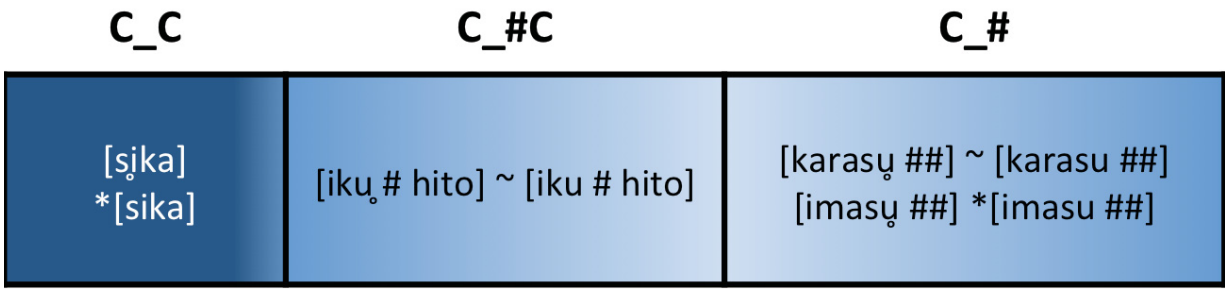

FIGURE 2.1: Schematic representation of the high vowel devoicing environments in Japanese. Darker shade represents more categorical application of HVD. Glosses: *shika* 'deer', *iku hito* 'person who is going', *karasu* '(it's a) crow', *imasu* 'be (animate, formal)'.

consonant and high vowel is followed by a short pause and then another word that begins with a voiceless consonant.

Would such vowels show categorical devoicing, since they are inter-consonantal, or would the devoicing be variable, since the pause precedes the following consonant? Surprisingly little of the substantial literature on Japanese devoicing has directly addressed this issue. A central goal of this paper is to understand what happens when these environments overlap, and more generally, what the relationship is between the two devoicing environments.

Addressing these questions in the case of HVD connects to the broader issue of how to analyze (variable) phonological processes that can apply in *overlapping* environments. Many devoicing processes cross-linguistically fit this description, as do many sandhi phenomena, which can often apply both across or within words (Kaisse, 1985; e.g. North American English flapping). Should such cases be analyzed as two distinct processes with overlapping environments, or one process with complex conditioning factors?

2.1.3 Boundary phenomena

Crucial to understanding the relationship between the two environments for HVD is a definition of what exactly constitutes the C̥_# environment. This question also has rarely been addressed, and as far as we know has not been investigated empirically. Previous work suggests that "physical silent pause" is not sufficient to characterize C̥_# devoicing, although high vowels do become devoiced before some pauses (see Section 2.2.4). Setting aside disfluencies, all cases of pre-pausal devoicing in natural speech are at a *word boundary*. A number of factors come into play at word boundaries which would not affect word-internal vowels; any of these could be responsible for C̥_# devoicing. A number of candidates for such boundary phenomena affecting devoicing rate are raised in the literature—prosodic boundaries, pauses, word boundaries—and will be reviewed further below. The relative role of these boundary phenomena is not clear. Hence, another goal of this paper is to clarify how these boundary phenomena affect devoicing rate, and in doing so to help characterize the C̥_# devoicing environment.

Addressing this goal for the HVD case is relevant for the more general issue of how boundary phenomena affect variable (phonological) processes, and how to account for these effects formally. Many variable processes are said to be conditioned by prosodic boundaries (e.g. Nespor and Vogel, 1986), a physical pause (Stevens, 2012), or word boundaries (Kiparsky, 1985)—but it is often difficult to tease these effects apart given how frequently different kinds of boundary phenomena co-occur (see Sec. 2.2.3, 2.2.4). Clarifying the empirical picture of how different boundary phenomena affect a variable process crucially informs theoretical accounts. If it turns out that only a single kind of boundary is relevant, this can be accommodated in existing theoretical treatments using a formal object that indexes the boundary, e.g. Optimality Theory constraints referring to faithfulness in "pre-pause position" (Coetzee and Pater, 2011) or alignment at prosodic word boundaries (Nagy and Reynolds, 1997). If different boundary phenomena have distinct or interacting

effects, a theoretical treatment becomes more complicated.

2.1.4 Summary

We address the study's three goals related to Japanese vowel devoicing—the source of variability, the relationship between the two devoicing environments, and the role of boundary phenomena—by conducting a multivariate statistical analysis of devoicing in a large corpus of spontaneous speech (Maekawa et al., 2000). The analysis models how different possible correlates of "pause" affect devoicing rate, while controlling for a number of other factors which condition HVD. To address the relationship between the two devoicing environments, the analysis also examines how the effect of other factors depend on the position of the devoiceable vowel.

The results show that, in accordance with native speaker intuitions, devoicing is nearly categorical, but only under certain conditions. HVD is most consistent word-internally, and also at sufficiently "large" domain edges: phrase boundaries that are followed by longer pauses. In other conditions, HVD applies variably. We find that how other factors affect the rate of application of HVD is modulated by the position of the vowel within prosodic phrases: in particular, speech rate and frequency have qualitatively different effects for vowels at the edge of sufficiently "small" domains versus larger domains. This finding leads us to suggest that devoiced vowels in Japanese may be best understood as the result of two different devoicing processes which apply in different, but sometimes overlapping environments. We suggest that some of the variability in these processes can be understood by reference to two sources in phonetic implementation and processing: gestural overlap, which has been previously discussed in the context of HVD (Jun and Beckman, 1993; Beckman, 1996), and the *locality of production planning* (Wagner, 2012), which has not.

In the remainder of this paper, we first present a review of previous findings on variability in high vowel devoicing in Japanese, and outline specific research questions (Section 2.2). We then describe the data (Section 2.3) and methods (Section 2.4) of our corpus study addressing these questions, and present the results (Section 4.4.2). We conclude with interpretation of these results and discussion (Section 4.5), including with reference to the broader issues discussed above beyond the Japanese HVD case.

2.2 Background

Vowel devoicing in Japanese is the subject of a long literature, which comes from many different perspectives (e.g. Han, 1962; McCawley, 1968; Hasegawa, 1979; Yoshida and Sagisaka, 1990; Vance, 1992; Jun and Beckman, 1993; Beckman, 1996; Kondo, 1997; Tsuchida, 1997; Varden, 1998; Maekawa and Kikuchi, 2005; Hirayama, 2009; Varden, 2010; Ogasawara, 2013; Nielsen, 2015, see Fujimoto, 2015 for a recent review). This section gives a brief summary of previous work on high vowel devoicing, focusing on aspects of importance for this paper: variability and the factors affecting variability, especially the role of word boundaries, prosodic information, and pauses.

2.2.1 High vowel devoicing

In Japanese, it is generally assumed that the high vowels /i/ and /u/ have devoiced allophonic variants, [i̥] and [u̥].[3] Textbook descriptions (e.g. Vance, 2008; Fujimoto, 2015) and pronunciation manuals (NHK, 1991, Japanese Pronunciation Accent Dictionary) give the generalization that the high vowels should be devoiced when they are preceded by

[3]There is some debate over whether HVD should be described as "devoicing" or "deletion", or whether both occur (see Fujimoto, 2015, Sec. 4.4). This paper is agnostic to this issue, as we abstract away from the phonetic realization of the vowel and focus on the factors conditioning the probability of application of HVD. We follow Fujimoto (2015) and most previous work in referring to HVD as "devoicing", for convenience.

a voiceless consonant and followed either by another voiceless consonant or by a pause. Examples of typically devoiced vowels are given in Table 2.1.

TABLE 2.1: Examples of words typically pronounced with devoiced vowels in Standard Japanese from Vance (2008).

Preceded & followed by voiceless consonant			$V \rightarrow \underset{\circ}{V}/\underset{\circ}{C}\_\underset{\circ}{C}$
a.	*sika*	[ʃi̥ka]	'deer'
b.	*kusa*	[ku̥sa]	'grass'
Voiceless consonant followed by pause			$V \rightarrow \underset{\circ}{V}/\underset{\circ}{C}\_\#$
c.	*ikimasu*	[ikimasu̥]	'(I will) go'
d.	*karasu*	[karasu̥] ~[karasu]	'(It's a) crow'

However, not all high vowels in the $\underset{\circ}{C}\_\underset{\circ}{C}$ and $\underset{\circ}{C}\_\#$ environments are devoiced. The most important factor is the restriction on devoiced vowels in adjacent syllables: if vowels in consecutive syllables are both in an HVD environment, generally only one of the vowels is devoiced. Also, for some speakers, the presence of a pitch accent or high tone may block HVD. But modulo these blocking factors, HVD is considered compulsory in standard (Tokyo area) Japanese (e.g. Hirayama, 1985, cited in Fujimoto, 2015). This assumption underlies phonological analyses of HVD in the literature, where devoicing is analyzed as categorical assimilation of laryngeal features from surrounding consonants, either [-voice] (e.g. Han, 1962; McCawley, 1968) or [+spread glottis] (Tsuchida, 1997; Tsuchida, 2001). The blocking effects can also be handled in a phonological analysis treating HVD as a categorical phenomenon, for example as proposed in Tsuchida (2001) and Kondo (2005).

However, other work argues that a number of factors gradiently affect devoicing rates in a way that is not easily captured in a categorical phonological analysis. Phonetically-oriented studies of devoicing argue that categorical phonological accounts are belied both by the gradient influence of phonetic context on the rate of devoicing, and by the range of possible realizations of devoiced vowels, including partial devoicing and total deletion (Jun and Beckman, 1993; Beckman, 1996).

Beckman (1996) proposes that devoicing of high vowels is due to gestural overlap—the encroachment of the glottal gestures for surrounding voiceless consonant—rather than a phonological change within the vowel itself (e.g. to [-voice]). In this account, varying articulatory conditions are naturally predicted to affect the likelihood of vowels being produced without voicing as the competing glottal gestures are compressed or change in magnitude for independent reasons. For example, it has consistently been found that vowels preceded by fricatives are devoiced at higher rates than those preceded by stops (see Section 2.2.5). Beckman (1996) suggests that this pattern is predicted by the articulatory differences between stops and fricatives.

This tension between the obligatoriness of HVD in many cases and its variability in others is at the core of much debate over the extent to which HVD is 'phonological' or 'phonetic', and has led to proposals that both phonological and phonetic mechanisms are necessary to account for HVD (Tsuchida, 1997; Varden, 1998; Nielsen, 2015). Tsuchida (1997) proposed that HVD is phonological in environments where it is categorical, but due to gestural overlap in variable cases. Nielsen (2015) showed that both phonetic and phonological factors must be taken into account to predict the realization of HVD in consecutive devoicing environments, arguing that HVD is driven by both types of factors.

Distinguishing between phonological and phonetic vowel devoicing is a challenge in many different languages (Gordon, 1998). In the Japanese case, this debate is complicated by the ambiguous meanings of 'phonological' and 'phonetic': in previous work, these are often used as shorthand for 'categorical' and 'variable', following one longstanding criterion, but variable processes are now routinely addressed in phonological theory (Coetzee and Pater, 2011; Coetzee and Kawahara, 2013), notably for Japanese (e.g. Kawahara, 2011). In this study, we do not directly address the question of which mechanisms underlie HVD in Japanese, but we do take into account both phonological and phonetic factors which have not previously been investigated, and delimit some conditions under which HVD is categorical versus variable, potentially offering some new insights for this debate.

2.2.2 Word boundaries

The literature on high vowel devoicing offers evidence that word boundaries affect variability, although their role has not often been the focus of direct investigation. Vance (1992) argues that one of the factors which disfavors devoicing is the presence of a morphological boundary between a potential target of HVD and the following voiceless consonant: in compound words containing consecutive devoicing environments in the NHK (1991) pronunciation dictionary, if one of the target vowels is followed by a morphological boundary, it is the other vowel that devoices. Varden (1998) reported a similar result from a production experiment. In words containing a consecutive devoicing environment, speakers devoiced the word-final vowel less often than the penultimate vowel in the same word. For example, in the first word in the sentence *Tsuki to hoshi ga kakureta*, the first vowel in *tsuki* was devoiced more often than the second.[4]

As part of a larger study on sociolinguistic effects in HVD, Imai (2004) investigated the effect of different morpheme boundary types, distinguishing between five possible cases: morpheme internal, pause, bound morpheme boundary, compound word boundary, and word boundary. A logistic regression analysis (using Goldvarb software) showed that the morpheme-internal and bound morpheme cases were most likely to devoice, followed by pause and then compound and word boundaries. However, Imai's Table 4.20 shows that when vowels in consecutive devoicing environments are excluded, devoicing rates are more similar for morpheme internal (78%) and word boundary (71%) cases than for bound morphemes (66%) and compound boundaries (35%).

In sum, these results from consecutive devoicing studies suggest that morpheme and word boundaries have some inhibitory effect on HVD, relative to presumably categorical application within a morpheme. That being said, previous work agrees that HVD is possible across both compound-internal morpheme boundaries and word boundaries of all

[4]Note that Varden (1998) interprets this result as a linear order effect, but due to his stimuli construction, linear order is not distinguishable from word boundaries.

types, regardless of syntactic constituency (Kaisse, 1985; Vance, 1992; Kondo, 1997).

Turning to the C̥_# environment in particular: word boundaries are closely tied to this "pre-pausal" environment, since examples given in the literature almost always involve pauses that follow word boundaries. This means that the effect of pause is confounded with the effect of a word boundary (e.g. (c) and (d) in Table 2.1). One study where this is not the case is Vance (1992), who gives the example of syllable-by-syllable pronunciations of words with devoicing environments. He states that if words are pronounced in this way, devoicing of word-internal vowels is blocked. If this is so, it constitutes evidence that word boundaries are at least a necessary condition for C̥_# devoicing to apply. Whatever the most accurate characterization of the C̥_# devoicing environment turns out to be, it will likely be a subset of vowels at word boundaries.

Word boundaries may also be important for devoicing in that they modulate the effect of other factors. While consonant manner and speech rate effects have been reported in many studies (see Section 2.2.5), Kondo (1997) found that consonant manner and speech rate effects were not statistically significant when considering only word-internal single devoicing environments. Hence, these types of effects may be dependent on the presence of a word boundary. More broadly, there is a running question throughout the literature on HVD as to the 'level' at which devoicing applies, closely corresponding to the debate on the 'phonetic' versus 'phonological' nature of HVD discussed above. Vance (1992), in the context of Lexical Phonology, discusses a possible distinction between lexical and post-lexical applications of high vowel devoicing. Within this framework, only post-lexical process/rules should be affected by speech rate and pauses (Mohanan, 1982; Kaisse, 1985).[5] In this study, we compare how devoicing rates are affected by pauses and speech rate in different prosodic positions, including a direct comparison between word-internal and word-final vowels. If it is the case that the effect of pauses and speech rate

[5]Note that only post-lexical applications could apply across words; hence C̥_# devoicing must result from postlexical rule application, while C̥_C̥ devoicing could result from lexical or postlexical rule application.

differs between these two environments, it would lend support to the view that there are two qualitatively different processes that underlie the pattern of high vowel devoicing in Japanese.

In sum, previous work suggests that word boundaries are related to variability in two ways: inhibiting C̥_C̥ devoicing, and as a necessary condition for the variable C̥_# devoicing. A focus of this paper is devoicing variability in those cases where C̥_C̥ and C̥_# environments overlap, a perspective which has not generally been considered in previous work.

2.2.3 Prosodic organization

We begin with a brief review of the prosodic organization of utterances in Japanese, with reference to the X-JToBI system of prosodic annotation (Maekawa et al., 2002) which will be relevant for our corpus study. We then review findings and comments from the literature on how prosodic information might influence HVD.

Above the level of the word, it is commonly argued that Japanese utterances are organized into two hierarchical groupings, although theoretical treatments differ as to the relationships between these levels (e.g. Beckman and Pierrehumbert, 1988; Ito and Mester, 2012). Here we call these levels the accentual phrase (AP) and the intonation phrase (IP), following Beckman and Pierrehumbert (1988) and Venditti, Maekawa, and Beckman (2008). These groupings reflect the syntactic constituency of the utterance, but are not necessarily isomorphic to it. For example, the utterance in Table 2.2 is organized into four APs, which are in turn grouped into three IPs.

These groupings are reflected in both the Tone and Break Index annotations in the X-JToBI system. The Break Index annotations are marks of "degree of perceived disjuncture between words," which listeners judge on the basis of several cues such as pausing, segmental lengthening, F0 lowering or resetting, and creaky voice quality (Venditti, 2005, p.

TABLE 2.2: Prosodic constituency and corresponding break index annotation for *Sankaku no yane no mannaka ni okimasu* 'I will place it right in the centre of the triangle roof' (Venditti, 2005, p. 176).

Accentual phrase	{ }	{ }	{ }	{ }
Intonation phrase	[ ]	[ ]	[ ]	[ ]
Tones	%L H*L L%	H*L L%	H- L%	L%
Break Indices	1 2	1 3	1 3	3
	sa'Nkaku no	*ya'ne no*	*maNnaka ni*	*okima'su*
	triangle-GEN	roof-GEN	middle-LOC	put

184–85). In X-JToBI, each word boundary is assigned a number from 1 to 3, with 3 indicating the highest degree of disjuncture. As shown in Table 2.2, the Break 2 and 3 annotations are typically associated with AP and IP boundaries, respectively.

As for the Tone annotations, the location of tonal targets is constrained by prosodic phrasing, hence these annotations offer some information about the prosodic organization of the utterance. The typical contour of an AP is an initial rise, marked in Table 2.2 by the *%L H-* annotation, followed by a gradual decline to a final low target, *L%* (Beckman and Pierrehumbert, 1988; Venditti, 2005). The AP also constrains the placement of lexical pitch accents, so that a single AP may contain at most one pitch accent. The IP is the domain to which boundary pitch movements (BPMs) are anchored, for example to signal a question or surprise (see Venditti, Maeda, and Santen, 1998, for detailed description of BPMs). The IP is also the domain of F0 downstep, so that each AP within a single IP becomes lower in pitch range, until F0 is "reset" at the beginning of a new IP (Beckman and Pierrehumbert, 1988).

While the effect of tones on HVD, especially pitch accents, has been investigated in several studies (Kuriyagawa and Sawashima, 1989; Hirayama, 2009; Oi, 2013), the effect of phrasal boundaries *per se* on HVD has not been systematically tested. The mentions of prosodic boundary effects on HVD in the literature relate to the definition of the C̥_# environment. For example, Kondo (1997), based on evidence from production experiments,

suggests that the C̥_# environment should instead be characterized as "utterance-final."

In the current paper, we will focus on Break Indices as the operationalization of prosodic phrase boundaries. However, information from Tone annotations will also be included in the model as a control, since previous literature suggests that high tones, particularly pitch accents, may block devoicing for some speakers. With this in mind, we now discuss prosodic phrase boundaries in particular.

Phrase boundaries

Little previous work has addressed the effect of phrase boundaries on HVD *per se*, but phrase boundaries could plausibly decrease or increase devoicing rate.

The idea that stronger phrase boundaries may have an inhibitory effect on HVD seems plausible from a gestural overlap perspective, since phrase boundaries in Japanese (and many other languages) are associated with final lengthening (e.g. Takeda, Sagisaka, and Kuwabara, 1989; Wightman et al., 1992; Den, 2015), in line with articulatory strengthening at phrase boundaries cross-linguistically (Fougeron and Keating, 1997). If HVD is due to overlap of adjacent laryngeal gestures, producing a longer vowel should make it more likely that the vowel's voicing gesture will have time to be realized. Using this logic, Den and Koiso (2015) attribute the negative relationship between devoicing rate and mora duration in utterance-final position to final lengthening. This same logic would apply to *any* possible HVD site—the less gestural overlap obtains, the less likely the voicing gesture will be fully realized. Thus, we include a rough measure of gestural overlap among the variables in our model: *Mora deviation*, defined as the difference between the current Mora's duration and its average duration in the corpus (Wightman et al., 1992). ("Mora" is capitalized for reasons explained below.)

However, there is also good reason to think that phrase boundaries would *increase* devoicing rate. Domain-final vowel devoicing is very common cross-linguistically (e.g. the Greek, French, and Oneida cases discussed above), and has clear phonetic motivation in

the drop of subglottal pressure at utterance/phrase endings (Gordon, 1998; Barnes, 2006). In Japanese in particular, it has been suggested that IP boundaries are the triggers for C̥_# devoicing (Kondo, 1997; Hirayama, 2009; Fujimoto, 2015). Also, prosodic phrasing is well-established as a unit for tonal organization in Japanese, so it seems plausible that segmental processes such as HVD would also be triggered by prosodic phrase boundaries.

To our knowledge, whether prosodic boundaries (e.g. IP) affect HVD has not been *empirically* tested. It is particularly difficult to assess whether it is a prosodic boundary *per se* which affects devoicing rate, or another boundary phenomenon. Phrase boundaries always coincide with word boundaries, and often with pauses, which are a major cue to intonation phrase boundaries (Venditti, 2005). The occurrence of prosodic phrase boundaries are highly correlated with the occurrence and length of pauses, making it difficult to distinguish their relative contributions to devoicing rate. With the large corpus of spontaneous speech used in the present study, we are able to investigate the effect of a prosodic boundary, which we operationalize as Break Indices, while also controlling for pauses and other possible confounding factors (e.g. final lengthening, as assessed by Mora deviation). Given that we restrict our data to tokens which are followed (and preceded) by voiceless consonants, we are also able to investigate the interaction between prosodic boundaries and C̥_C̥ devoicing, a novel empirical contribution to the HVD literature.

2.2.4 Pause

The term "pause" is traditional and often used in the description of C̥_# . Taking this description at face value, how does an actual physical pause affect devoicing? On the one hand, the very use of the term "pause" to define an environment for HVD suggests that a pause may promote devoicing. On the other hand, the few studies addressing the effect of a pause reach the opposite conclusion: Vance (1992) states that pauses *block* devoicing from applying where it otherwise would have, as in a syllable-by-syllable pronunciation of a word containing a C̥_C̥ environment. Kondo (1997), comparing between repetitions

of the same item in a production experiment, also found a negative effect: repetitions in which a pause was present after the devoiceable vowel had *lower* devoicing rates. Den and Koiso (2015), examining a subset of the spontaneous speech dataset used in this paper (Corpus of Spontaneous Japanese), found that devoicing occurs frequently before pauses (defined as silence of at least 200 msec), but that pause *length* does not significantly affect devoicing rate. In sum, the role of pauses in promoting or blocking HVD is unclear.

However, as noted above, word boundaries, phrase boundaries and especially utterance edges are highly correlated with pauses—especially in laboratory experiments, due to the short length of test items (single words or sentences). By investigating HVD in a large corpus of spontaneous speech, we will be able to tease apart the influence of boundaries (of words and prosodic units) and pauses, and delineate their respective roles in HVD.

2.2.5 Other factors

We now turn to some major factors that affect the rate of HVD rate: surrounding consonant articulation, speech rate and style, and lexical frequency and idiosyncrasy. These will be used in our model both as controls, and to investigate the relationship between the C̥_C̥ and C̥_# environments.

Consonantal context

At a basic level, consonantal context is the most important factor in high vowel devoicing, in that the presence of voiceless consonants defines the C̥_C̥ and C̥_# environments.

The *manner* of the surrounding voiceless consonants also influences HVD. In terms of the preceding consonant, there is less devoicing after plosives than after fricatives, both in single-word productions (Kondo, 1997) and in spontaneous speech (i.e. the Corpus of Spontaneous Japanese (CSJ): Maekawa and Kikuchi, 2005). The effect of the following consonant is the reverse, with less devoicing before fricatives than before plosives

(Nielsen, 2015; Maekawa and Kikuchi, 2005; Kuwabara and Takeda, 1988; Lovins, 1976; c.f. Han, 1962). The effect of a preceding or following affricate is inconsistent across studies, but generally patterns with either plosives or fricatives. The preceding and following consonant effects are not independent: a high vowel flanked by voiceless fricatives is generally *less* likely to devoice than other combinations of obstruents, in both laboratory experiments (Kondo, 1997; Tsuchida, 1997; Hirayama, 2009) and in the CSJ (Maekawa and Kikuchi, 2005). Given the important effects of consonant manner on devoicing rate, we include in our model the manner of the preceding and following consonant.

Speech rate and style

Speech rate and speaking style have intuitively opposite effects on HVD: Hasegawa (1979) observed that devoicing is more likely to occur in faster speech, but *less* likely to occur in casual speech. This observation was confirmed by Martin, Utsugi, and Mazuka (2014), a recent corpus study comparing child-oriented, adult-oriented and read speech: high vowels devoiced significantly less in adult-oriented (i.e. conversational) speech than in read speech, but significantly more in faster speech compared to slower speech.

In contrast, Kondo (1997) found no significant effect of speech rate effect when it was tested explicitly in a production experiment, where subjects read test words embedded in paragraphs at slow, normal and fast speaking tempi. However, devoicing rates were very high for all three conditions (81–97%), as expected for a formal speech style. It may be that speaking rate effects are relatively small and are more easily observable in spontaneous speech (as in Martin, Utsugi, and Mazuka, 2014), in which devoicing is more variable, rather than read speech. The current dataset is expected to show a small positive speech rate effect, given that it examines spontaneous speech. While we do track the effect of speech rate, we do not explicitly control for speech style, as the speech contained in the CSJ is almost all from formal settings (academic presentations, simulated public speaking).

Lexical frequency and idiosyncrasy

To our knowledge, the only examination of frequency effects is in Maekawa and Kikuchi (2005, p. 218), who found a small *negative* correlation between devoicing rate and word frequency in the CSJ (empirical correlation, without controlling for other factors). This effect was found for high vowels which were preceded by a voiceless consonant, but with any kind of following segment (or lack thereof). The directionality of this frequency effect is surprising if HVD is seen as a reductive process resulting from gestural overlap, which is expected to be greater for higher-frequency words (Jurafsky et al., 2001; Pluymaekers, Ernestus, and Baayen, 2005); frequency and devoicing rate would then be expected to have a positive correlation.

One aspect of Maekawa and Kikuchi's data points to a positive trend: they highlight two morphemes which stood out as outliers from the negative trend, the verbal particles *desu* (polite form of copula *da*) and *masu* (an auxiliary verb of politeness). These items were among those with the highest frequency, and they also showed extremely high devoicing rates. This pattern accords with native speaker intuitions about these morphemes, as well as the findings of Oi (2013), who specifically tested utterance-final devoicing for lexical words, and found that lexical words were devoiced about 80% of the time, while the particle *masu* was always devoiced for all 10 speakers in the study. One suggested explanation for *desu* and *masu* in particular is that these functional morphemes appear almost exclusively sentence-finally. Hence, they could be much more affected by C̥_# devoicing than other types of words which rarely appear at the ends of utterances.

The case of these two morphemes means that lexical identity is another factor confounded with the boundary phenomena discussed above (e.g. IP boundary, pause). Analyzing HVD in spontaneous speech allows us to address our research questions while controlling for the high devoicing rates of certain words. In addition, by including word frequency in our multivariate model of HVD in the CSJ, we can assess the existence and

directionality of a frequency effect, when other factors (such as lexical identity) are controlled for.

2.2.6 Summary & research questions

We have seen that many factors have been found to affect HVD rate, including consonant manner, high tones, speech rate and style, word boundaries and pauses; prosodic domain edges may also play a role. This paper focuses on three of these factors, which are confounded—word boundaries, prosodic position, and pauses—to address three research questions, in a corpus of spontaneous speech consisting of tokens in C̥_# and C̥_C̥ environments and their intersection.

First, *how do word boundaries affect devoicing rate, and modulate the effect of other factors*? Second, *how do prosodic phrase boundaries affect devoicing rate, and modulate the effect of other factors*? Previous work predicts an inhibitory effect of a word boundary on devoicing rate, and gives reasons to think that phrase boundaries (especially IP boundaries) could either increase or decrease devoicing rate. Whether and how word and phrase boundaries modulate the effects of other factors on devoicing rate will help to understand the relationship between C̥_# and C̥_C̥ devoicing; we consider speech rate, word frequency, Mora deviation, and pauses in particular. Third, *how does a physical pause (presence and duration) affect devoicing rate*? Previous work does not give a consistent prediction on how pauses should affect devoicing rate.

We address these three research questions in a dataset which was selected to best address them, and complements previous work. First, because the three 'boundary phenomena' are highly correlated, we examine HVD variability in a very large dataset of spontaneous speech (Maekawa et al., 2000), where the high degree of variation allows us to tease their effects apart, while controlling for other factors affecting devoicing rate (consonantal context, etc.), in a single statistical model.

Second, in order to understand the relationship between the C̥_C̥ and C̥_# environments, we considered only high vowel tokens which were preceded and followed by voiceless consonants (where the following consonant may occur following a word boundary or pause, in the C̥_# environment). That is, we excluded tokens in the C̥_# environment followed by a voiced segment. This exclusion allows us to understand what happens when the environments overlap, and to delimit the role of boundary phenomena by eliminating a confounding variable (following segment voicing) which could account for any observed difference between HVD application across versus within words. This restriction also means our conclusions about C̥_# position are in fact only based on a subset of the relevant data. We discuss the implications of this in Sec. 2.6.4.

Third, in order to focus on the effects of boundary phenomena, we only consider tokens from single-devoicing environments. Previous work on HVD variability has largely focused on consecutive devoicing environments and lab-elicited speech—precisely *because* speakers seem to apply HVD near-categorically in single devoicing environments in laboratory speech—and it remains unclear how much variability there is in natural speech in single devoicing environments.

Thus, our study contributes a new perspective on HVD variability by examining spontaneous speech, vowels preceded and followed (eventually) by voiceless consonants, and (only) single devoicing environments.

2.3 Data

The source of data for this study is the Corpus of Spontaneous Japanese (Maekawa et al., 2000), a corpus of audio recordings primarily from two styles of spontaneous speech monologues: academic presentation speech and simulated public speaking. We draw

from the "Core" subset of the data which, in addition to being orthographically transcribed and morphologically tagged, includes segmentally-aligned manual phonetic transcription and X-JToBI labels (Maekawa et al., 2002) to mark prosodic information. This subset contains about 44 hours of speech from 201 speakers.[6]

From the XML annotation files, we extracted all tokens of short high vowels[7] and information about whether the vowel was devoiced, immediately adjacent segments, prosody, and other factors expected to affect devoicing rate.

In the segmental phonetic transcription, vowels are transcribed as either voiced or devoiced; we used this manual annotation as our binary measure of *devoicing*. Devoicing was determined by the human labellers preparing the corpus by using information from "the wide-band spectrogram, speech waveform, extracted speech fundamental frequency, peak value of the autocorrelation function, in addition to audio playback" (Maekawa and Kikuchi, 2005, p. 208).

Word and phrase boundaries were derived from the Break Index (BI) annotations in the CSJ. These annotations involve information about the strength of a break (None/1/2/3), as well as other information (e.g. the occurrence of a pause or a "boundary pitch movement"). We collapsed BI annotations into four categories, which closely correspond to word and prosodic phrase boundaries: *None* tokens had no BI marked at the right edge of the vowel, so they are within the same word as the consonants that precede and follow them. Tokens with BI *1*, *2* or *3* are at the right edge of a word. BI *1* tokens are word-final, but not final in their accentual or intonation phrase. Tokens with BI *2* are accentual phrase

[6] A small part of the "Core" subset ($\sim$ 5%) consists of (spontaneous) dialogues and read speech. We found that all results reported in this paper are qualitatively the same if the read speech data (3.3% of our dataset) is excluded. Thus, we report results without excluding this data, and interpret our findings as representative of spontaneous Standard Japanese.

[7] Japanese has a phonological length distinction in vowels, and only phonologically short vowels are said to be affected by devoicing. This is corroborated by Maekawa and Kikuchi (2005) who found less than 0.5% of long high vowels and 1.2% of short non-high vowels were devoiced in the CSJ, compared to 24.3% of short high vowels.

TABLE 2.3: Summary of Break Index annotations in relation to word/phrase position of vowel token.

Break Index	Position of vowel token	Number of tokens
None	word-internal	15355
1	word-final, phrase-internal	23811
2	final in accentual phrase	2361
3	final in intonation phrase	3120

but not intonation phrase final, while BI *3* tokens are final in their intonation phrase.[8] Table 2.3 shows the definition and number of tokens for each BI category.

Tone annotations were also extracted in order to control for effects of high tones. Annotations for lexical pitch accents and phrasally-assigned tones in the CSJ are aligned with "the corresponding F0 event" (Venditti, 2005). Hence, the annotations track the surface realization of tones, potentially differing from underlying pitch accents or usual alignment of phrasal high tones (second mora by default). We considered tone labels to be part of a token if their timestamps were within the start and end times of the token vowel.

Pause duration following the token was defined as the time difference between the end of the CV Mora and the beginning of the next segment. This interval sometimes included a manually annotated "pause" in the CSJ (200 msec or longer), and sometimes did not, i.e. for brief silences or other non-speech. 2634 tokens (5.9%) were followed by a pause.

As a measure of final lengthening, which is associated with larger prosodic phrase boundaries, we used a measure based on the duration of the CV sequence containing the target vowel. The duration of the vowel itself was not used because the left boundaries of devoiced vowels are often unclear, and are indicated as such in the CSJ annotations. (For example, in many [sụ̥] tokens there is no clear acoustic landmark differentiating the fricative and (devoiced) vowel portion.) Our use of the duration of a larger unit than the

[8]Note that Intonation Phrase boundaries (BI 3) in this dataset include "utterance" boundaries as well as "intermediate phrase" boundaries, in the terminology of Beckman and Pierrehumbert (1988) (Igarashi, Kikuchi, and Maekawa, 2006, p. 348).

vowel itself which can be more reliably measured follows other work examining vowel devoicing (e.g. Torreira and Ernestus, 2011 for French). In the CSJ XML annotations, segments are hierarchically organized into Mora units, which include a vowel segment and its onset consonant for all tokens where HVD can apply.[9] (To avoid confusion of "mora" as referring to physical duration with the abstract weight unit used in phonological theory, we capitalize Mora throughout this paper to emphasize that it is the physical duration of a CV sequence that is referred to.) For each token, we recorded the duration of the Mora containing it. From this value we subtracted the average duration of that particular CV Mora across the CSJ corpus. This gave a measure of *Mora deviation*, a positive value if the Mora was longer than average and negative it was shorter. For example, a token of /u/ preceded by /k/ would be in a /ku/ Mora, and the difference between the duration of that Mora and the average duration of all /ku/ Moras would yield its value for Mora deviation.

We extracted two measures of *speech rate* to be included in the model. We first calculated raw speech rate as the number of phones per second in the inter-pausal unit according to the CSJ annotation (where pauses of >200 msec are manually annotated). Raw speech rate was used to calculate *speaker speech rate*, the average rate over all the speaker's utterances, and *local speech rate*, the difference between an utterance's speech rate and the speaker's average. Using separate speaker-level and observation-level speech rate predictors, following Snijders and Bosker (1999), allows us to differentiate between devoicing occurring more often for faster speakers, versus faster utterances (within a speaker). Both variables are in units of phones per second, such that an increase in the variable corresponds to faster speech.

The data was restricted to tokens of high vowels that were preceded and followed by voiceless obstruents (n = 52809). To focus on the single devoicing environment, we

[9]Note that Japanese has moras which are not CV units (Labrune, 2012), but only CV-type moras contain vowels in C̥_# and C̥_C̥ environments.

excluded tokens that were adjacent to other potential devoicing sites (i.e. "consecutive devoicing environments," see Section 2.2.1; n = 7102, 13.45% of tokens). Remaining tokens that were part of disfluencies were also excluded (n = 984, 5.36 % of tokens). Finally, 76 tokens were excluded whose prosodic annotations reflected pathological cases or probable coding errors.[10] The final dataset contains 44647 tokens for analysis, of which 91.17% were devoiced.

2.4 Methods

The data was analyzed using mixed-effects logistic regression, a type of multivariate statistical model, which predicts the outcome (whether a vowel was devoiced) as a function of a number of variables (e.g. Gelman and Hill, 2007; Baayen, 2008). Mixed-effects logistic regression has been applied to HVD data in particular by Nielsen (2015). The advantage of using a multivariate model is that it allows the comparison of several effects at once, and the possibility of comparing their relative effect size. A mixed-effects model in particular also allows the inclusion of both fixed effects, which are the factors of interest discussed above, and random effects, which account for differences in baseline HVD rates and effect sizes within different speakers or words. The dependent variable for this study is the binary outcome of devoicing (1) or no devoicing (0) based on the phonetic transcription in the corpus. Hence, positive coefficient estimates indicate an increase in the likelihood of devoicing. More precisely, each coefficient gives the estimated effect of a factor of interest on the *log-odds* of devoicing.

2.4.1 Model terms

We now turn to the variables which are included in the statistical model as fixed or random effects, and how they are related to our research questions.

[10] These were: all remaining tokens whose (collapsed) *Break Index* was not 1, 2, 3, or None followed by no physical pause.

Word and phrase boundaries The four-level Break Index (BI) is the independent variable of primary interest, as it lets us examine the effect of word and phrase boundaries. This variable was included in the model as a four-level categorical variable with Helmert contrast coding. With this type of coding, the estimated coefficients have interpretations that directly address our first and second research questions. The first coefficient will compare the devoicing rate in word-internal tokens (BI *None*) versus word-final tokens (BI *1/2/3*). The second coefficient estimates the difference in devoicing rate among word-final tokens which are phrase-internal (BI *1*) versus phrase-final (BI *2/3*). The final coefficient compares tokens at accentual phrase versus intonation phrase edges (BI *2* v BI *3*). *Break Index* is included as a main effect in the model, as well as in a number of interaction terms, discussed below.

Pauses To address our third research question, how *pause* affects the rate of devoicing, pause duration was included in the model. Because the distribution of pause duration is highly skewed, with the vast majority of tokens showing no pause or a short pause, it was not possible to include pause duration as a continuous variable.[11] Instead, pause duration was discretized into a four-level factor, called *Pause*, which allowed comparison between tokens with and without following pauses, and allowed for non-linear effects of pause duration. The first level corresponded to tokens with no pause. Tokens that did have a following pause were categorized in to three bins (levels 2–4) of roughly equal size (using the `cut2` function in R; Harrell Jr, 2015) according to pause duration: less than 85 msec, 85-463 msec, and over 463 msec. The four-level factor was coded such that the intercept corresponded to no pause, and the three contrasts corresponded to Helmert contrasts: no pause vs. pause, short vs. medium/long pause, medium vs. long pause.

[11]The distribution is highly skewed because within-word environments always show no pause, and are much more frequent than cross-word environments. Thus, discretizing pause duration is necessitated by our focus on *both* devoicing environments and the intersection between them.

An interaction of pause duration with break index was included in the model, to allow for the possibility of different pause effects at different boundaries. However, because there were almost no word-internal tokens that were followed by a pause,[12] *Pause* and *Break Index* are not independent, and the model structure must somehow take into account that there can be no *Pause* effect for word-internal tokens. We did this by excluding the main effect of pause duration. Intuitively, the interaction terms describe the *Pause* effect when *Break Index* is 1, 2, or 3.

Mora deviation Mora deviation was included in the model to control for final lengthening as a confound for phrase boundaries, and to capture the effects of gestural overlap. Exploratory plots suggested a nonlinear effect of mora deviation on devoicing rate, of a roughly quadratic shape (in log-odds space). *Mora deviation* was thus coded as a nonlinear spline with three knots (using `rcs` in the `rms` R package; Harrell, 2014), which corresponds to a curve with a single "bend", and included in the model as a main effect and in interactions (see below). The two spline components correspond approximately to linear and nonlinear effects, of a continuous variable. Before coding as a spline, *Mora deviation* was centered and divided by two standard deviations (Gelman and Hill, 2007).

Interactions Our first and second research questions address how word and phrase boundaries modulate the effect of other variables. The model includes interactions of *Break Index* (corresponding to phrase boundaries) with four variables in particular: *local speech rate*, *lexical frequency*, *Mora deviation*, and *Pause*. Interactions with speech rate, frequency, and Mora deviation are of interest in that differences in their qualitative effects

[12] Such tokens exist in the corpus, but were excluded from analysis as they were determined to be mostly disfluencies.

depending on *Break Index* would bear on the relationship between C̥_C̥ and C̥_# devoicing.[13] The interaction with *Pause* is partially necessitated by the structure of the data (pauses do not occur for BI=*None*, as discussed above). The possibility of the effect of *Pause* differing at different boundary types (BI=*1*, *2*, *3*) emerged in exploratory data analysis, and will turn out to be crucial for interpreting our results.

Controls A number of other variables expected to affect devoicing rate based on previous work (Section 2.2.5) were included in the model as controls, as main effect terms. Terms were included for *Preceding consonant manner* and *following consonant manner*, coded using sum contrasts as factors with the levels *stop*, *affricate* and *fricative*, with *stop* as the base level. Based on previous findings that vowels between two fricatives have very low devoicing rates, we also included a term for the interaction between these two factors. The presence of a high tone associated with the vowel was included as a binary predictor, which was converted to a numerical variable and centered.[14]

A continuous *lexical frequency* measure was included in the model: frequency was defined as a word's count in the CSJ divided by the total number of words in the CSJ; this measure was then log-transformed.

Finally, the model includes both measures of speech rate described above, *speaker speech rate* and *local speech rate*. Frequency and speech rate predictors were centered and divided by two standard deviations (Gelman and Hill, 2007).

Coding and model interpretation The coding of variables included in the model results in a straightforward interpretation of model coefficients, which will be important in interpreting our results. Holding the *Pause* contrasts at zero corresponds to a token with

[13] We included only *local speech rate* in interactions, and not *speaker speech rate*, to limit model complexity, and since *local speech rate* corresponds more closely to measures of speech rate used in previous work on HVD (e.g. Kondo, 1997).

[14] Exploratory analysis suggested possible differences in the effect of pitch accents (H*), phrasal (H-) and boundary tone-associated H tones on devoicing rate, but due to the low number of tokens bearing a high tone in the dataset, these differences were collapsed into a single binary predictor of high tone presence.

no pause, while all other variables have been centered, or coded using Helmert or sum contrasts, where the intercept corresponds to averaging across factor levels. Hence, the interpretation of the intercept in the statistical model reflects the estimated devoicing rate for word-internal cases with no pause, with all other variables held at their mean values. All fixed effect coefficients can be interpreted as the estimated effect of one or more predictors, holding other variables at their mean values.

Random effects Previous research has reported differences in devoicing rates across both speakers and lexical items, and any spontaneous speech corpus is inherently unbalanced, such that certain words and speakers have much more data than others. These facts must be controlled for in the statistical model, or the effects of interest will be unduly influenced by a small group of speakers or words. For example, high-frequency verbal particles (e.g. *desu*) are highly prone to devoicing (potentially skewing the estimate of overall devoicing rate), and occur disproportionately often in phrase-final position (potentially skewing the estimate of e.g. the *Break Index* effect). In a mixed-effects model, these issues are mitigated by the inclusion of random-effect terms. The model reported here includes by-speaker and by-word intercept terms, which directly account for differences between speakers and words in overall devoicing rate. We also included by-speaker random slope terms, which account for differences between speakers in effect size, for all fixed-effect terms of interest for our research questions: all terms involving *Break Index* or *Pause*, as well as main effects of variables involved in any interactions with *BreakType* (i.e. *local speech rate*, *lexical frequency*). These terms result in more accurate p-values and coefficient estimates for the fixed-effect terms of interest (Barr et al., 2013).[15] The model does not include random slopes corresponding to fixed-effect terms *not* of interest for our research questions (such as surrounding consonant manner), in order to limit model complexity.

[15]It would have also been preferable to include by-word random effect terms corresponding to the fixed effects of interest for our research questions, e.g. for *Break Index*. Adding these terms resulted in unstable models, presumably due to the high number of word types relative to the size of the dataset; we thus did not include them in the final model.

The coefficients and p-values for these terms are thus less reliable (Barr et al., 2013). Finally, correlation terms between random effects were excluded, to aid model convergence.

2.4.2 Model construction

A mixed-effects logistic regression was fit using the `glmer` function in the `lme4` package (Bates et al., 2013) package in R (R Core Team, 2013). The inclusion of the full random effect structure described above led to non-convergent models. Analysis of the distribution of the data, guided by `glmer` warning messages, suggested that convergence issues were due to sparsity in certain parts of the data, reflecting collinearity between the presence of medium and long pauses and the type of *Break Index*. In particular, longer pauses are relatively rare at BI *1* or *2*, occurring mostly at BI *3* (94%, n = 1756).

In order to arrive at a convergent model, random-effect and fixed-effect terms flagged by `glmer` as unstable were iteratively removed, until a well-conditioned model was achieved. The fixed and random effect terms removed for the final model were two of those comparing medium versus long pauses: one estimating the difference between BI *1* and BI *2* and *3* (in the effect of medium vs. long pauses on devoicing rate), and the other estimating the difference between BI *2* and BI *3* (same). Hence, in the final model, the difference between medium and long pauses (in devoicing rate) was only estimated as a single effect across all word-final tokens (*Break Index* = *1*, *2*, and *3*), which will be important for interpreting the results.

2.5 Results

Here we report the results of the statistical model of devoicing rate. The model's estimates for the fixed-effect terms are shown in Table 2.4. We first discuss the results for control predictors, then turn to predictors relevant for our research questions: *Break Index*, *Pause*, and interactions between *Break Index* and *Mora deviation*, *lexical frequency* and *speech rate*.

To aid in interpretation of the model's results, we use partial effect plots (in addition to reporting model coefficients): these show the predicted effect of varying one or more predictors, while others are held constant, with predictions transformed into probability space (instead of log-odds). Model predictions in these plots were computed using the fixed effect coefficient estimates. Errorbars on model predictions correspond to two standard errors.

We do not discuss the model's random effect terms, which are shown in Appendix 2.7.

2.5.1 Control predictors

The estimates for the effect of consonant manner are consistent with previous findings. Compared to the mean devoicing rate, a fricative preceding the token increases the likelihood of devoicing ($\hat{\beta} = 0.75$, $p < 0.001$), while a fricative following decreases the likelihood ($\hat{\beta} = -0.99$, $p < 0.001$). The effects of affricates are not as clear, with a preceding affricate slightly decreasing odds of devoicing relative to the mean rate, and a following affricate being not reliably different ($\hat{\beta} = -0.34$, p 0.033; $\hat{\beta} = 0.16$, $p = 0.297$). There is also a significant interaction between preceding and following consonant manners. We do not discuss these terms in detail, but note that the negative coefficient for the interaction between terms for a preceding and following fricative ($\hat{\beta} = -0.49$, $p < 0.001$) suggests that vowels flanked by fricatives on both sides have particularly low devoicing rates, as expected (Tsuchida, 1997).

The presence of a high tone strongly decreases the likelihood of devoicing ($\hat{\beta} = -4.35$, $p < 0.001$), again consistent with previous findings discussed in section 2.2.3. The large effect of tone confirms that devoicing of vowels associated with a high tone is indeed highly dispreferred, but due to the small number of H-associated tokens in our data set, it was not possible to distinguish between pitch accents, phrasal high tones, and other high tones.

TABLE 2.4: Fixed effects for the statistical model: coefficient estimates, standard errors, z-values, and significances (assessed using a Wald test). Main-effect terms are shown above the middle line, and interaction terms below the line.

Fixed effects	β	se(β)	z	Pr(z)
(Intercept)	5.88	0.3	19.62	< 0.001
Break Index				
1, 2, 3 − None	−2.45	0.29	−8.41	< 0.001
2, 3 − 1	−1.97	0.24	−8.05	< 0.001
3 − 2	−1.21	0.32	−3.77	< 0.001
Mora deviation				
linear	0.48	0.27	1.75	0.081
nonlinear	−2.7	0.3	−8.85	< 0.001
Lexical frequency	0.28	0.33	0.84	0.402
Speech rate within utterance	−0.04	0.12	−0.37	0.714
Speech rate average by speaker	0.45	0.16	2.74	0.006
High tone *non-high − high*	−4.35	0.29	−15.02	< 0.001
Manner of previous phone				
fricative	0.75	0.14	5.39	< 0.001
affricate	−0.34	0.16	−2.13	0.033
Manner of following phone				
fricative	−0.99	0.1	−9.96	< 0.001
affricate	0.16	0.15	1.04	0.297
Pause : Break Index				
No Pause − Pause : *1, 2, 3 − None*	−0.01	0.37	−0.02	0.986
No Pause − Pause : *2, 3 − 1*	−2.69	0.81	−3.32	< 0.001
No Pause − Pause : *3 − 2*	−0.47	0.75	−0.63	0.529
Short Pause − Medium/Long Pause : *1, 2, 3 − None*	−0.39	0.55	−0.7	0.481
Short Pause − Medium/Long Pause : *2, 3 − 1*	−3.96	1.26	−3.14	0.002
Short Pause − Medium/Long Pause : *3 − 2*	0.08	1.12	0.07	0.944
Medium Pause − Long Pause : *1, 2, 3 − None*	−2.01	0.37	−5.43	< 0.001
Break Index : Lexical Frequency				
1, 2, 3 − None : *Frequency*	0.17	0.37	0.45	0.652
2, 3 − 1 : *Frequency*	1.25	0.32	3.93	< 0.001
3 − 2 : *Frequency*	1.49	0.52	2.85	0.004
Break Index : Speech Rate within utterance				
1, 2, 3 − None : *Speech Rate*	0.2	0.18	1.1	0.27
2, 3 − 1 : *Speech Rate*	0.51	0.24	2.17	0.03
3 − 2 : *Speech Rate*	−0.05	0.34	−0.15	0.884
Break Index : Mora deviation				
1, 2, 3 − None : *linear*	−2.18	0.44	−4.94	< 0.001
2, 3 − 1 : *linear*	−0.11	0.61	−0.18	0.857
3 − 2 : *linear*	−1.97	0.94	−2.09	0.037
1, 2, 3 − None : *nonlinear*	−0.03	0.44	−0.07	0.946
2, 3 − 1 : *nonlinear*	−0.29	0.57	−0.51	0.611
3 − 2 : *nonlinear*	3.12	0.86	3.61	< 0.001
Previous phone manner : Following phone manner				
Preceding fricative: Following fricative	−0.49	0.13	−3.84	< 0.001
Preceding fricative: Following affricate	0.75	0.2	3.77	< 0.001
Preceding affricate: Following fricative	−0.08	0.16	−0.5	0.615
Preceding affricate: Following affricate	−0.84	0.25	−3.4	< 0.001

For the speech rate predictors, among main-effect terms, only the main effect of the speaker's mean speech rate reaches statistical significance, with a higher likelihood of devoicing for faster-talking speakers ($\hat{\beta} = 0.45$, $p = 0.006$). Neither local speech rate ($\hat{\beta} = -0.04$, $p = 0.714$) nor lexical frequency ($\hat{\beta} = 0.28$, $p = 0.402$) reached significance as main effects. However, terms in the interactions between *Break Index* and these two variables do reach significance. These interactions will be discussed below.

2.5.2 Break indices

The coefficients for this predictor address our first two research questions, comparing word-internal, word-final and phrase-final (AP or IP-final) vowels. Figure 2.2 shows the predicted probabilities for each value of *Break Index* with no pause following, and all other variables held constant at average values.

First of all, the rate of devoicing for word-internal vowels is very high, essentially at ceiling (Intercept: $\hat{\beta} = 5.88$, predicted probability: 99.72%). Regarding the effect of word boundaries, the model confirms that, all else being equal, vowels followed by a word boundary (in any phrasal position) are significantly *less* likely to devoice than vowels that are within the same word as their following consonant (*Break Index* 1/2/3 - None: $\hat{\beta} = -2.45$, $p < 0.001$). This finding, on the effect of word boundaries for single devoicing environments, is consistent with the results of Varden (1998), who found that in *consecutive* devoicing environments, a word-internal vowel was more likely to be devoiced than a word-final one.

Among word-final vowels, the model finds a reliable difference between devoicing rates for phrase-internal vowels compared to vowels at the edge of an accentual phrase or intonation phrase (*Break Index* {2, 3} - 1: $\hat{\beta} = -1.97$, $p < 0.001$). Among vowels at prosodic phrase edges, vowels at the edge of an intonation phrase are *less* likely to devoice than vowels at the edge of an accentual phrase (*Break Index* 3 - 2: $\hat{\beta} = -1.21$, $p < 0.001$).

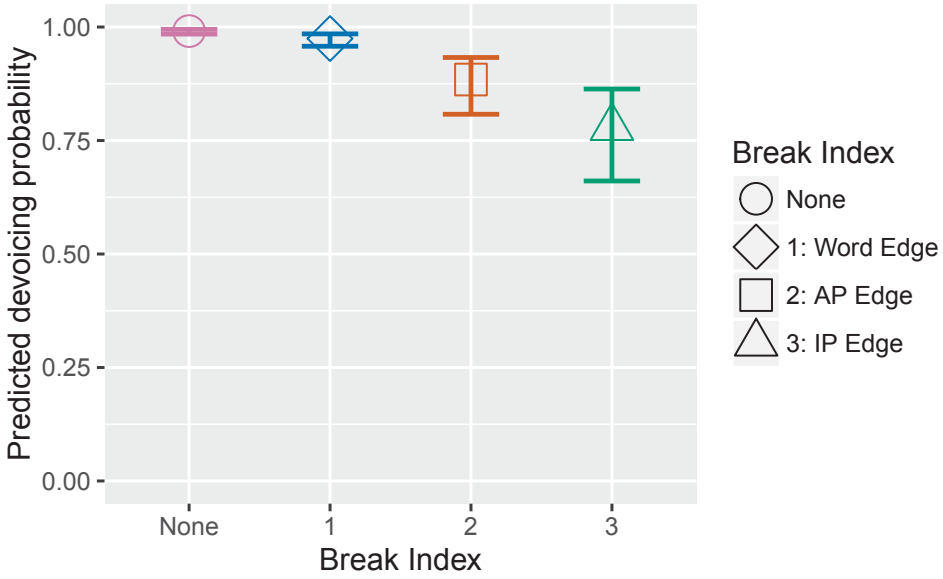

FIGURE 2.2: Predicted probability of devoicing for a high vowel that is (a) word-internal, (b) at a word boundary, but phrase-internal, (c) at an accentual phrase (AP) boundary, (d) at an intonation phrase boundary; in all cases, the prediction assumes no following pause, and others variables held constant at mean values. Shapes represent the predicted probabilities, and bars show the 95% confidence intervals.

In sum, the main effect of the *Break Index* predictor confirms that, when no pause follows and other predictors are controlled, the 'higher' the boundary (greater *Break Index* value: None $<$ word boundary $<$ AP $<$ IP), the less likely devoicing becomes.

2.5.3 Pause

The effect of *Pause* was included in the model only as an interaction with *Break Index*, since there are no word-internal tokens that are followed by a pause. When considering all word-final tokens jointly, the model does not find a significant difference in devoicing rate depending on the presence/absence of a pause ($\hat{\beta} = -0.01$, $p = 0.986$), or on the difference between a short/longer pause ($\hat{\beta} = -0.39$, $p = 0.481$). There is a significant difference between tokens followed by medium and long pauses, with long pauses associated with

higher rates of devoicing ($\hat{\beta} = -2.01$, $p < 0.001$). Since the model only compares medium and long pauses across all values of *Break Index* jointly (see Section 2.4.2), it is not possible to say whether this effect is similar at all types of boundaries, but examination of the empirical data for each *Break Index* value suggests that it is driven by tokens at IP boundaries (*Break Index*=3, which contains the most data for medium–long pauses).

The model also compares the differences in the effect of *Pause* among vowels in different prosodic positions. The presence of a pause has a *smaller* effect on the probability of devoicing following phrase-internal word-final vowels, relative to following phrase-final vowels ($\hat{\beta} = -2.69$, $p < 0.001$). There is also a difference in the effect of short pauses (<85 msec) and longer pauses (>85 msec): tokens followed by short pauses have a higher devoicing rate than tokens followed by long pauses, for *Break Index* 1 (phrase-internal word boundary); but if the token is at a phrase boundary (*Break Index* 2 or 3) then it is *longer* pauses that have higher devoicing rates than short pauses ($\hat{\beta} = -3.96$, $p = 0.002$).

The larger pattern expressed by these coefficients can be seen in the prediction plots in Figure 2.3. The effect of a pause is strikingly different between the phrase-internal and phrase-final vowels. At phrase-internal vowels (left panel), an increase in pause duration has a consistently negative effect on devoicing rate, at least for null/short/medium pauses.[16] For phrase-final vowels, an increase in pause duration is associated with an *increase* in the probability of devoicing.

In sum, the relationship of pause length to devoicing rate looks qualitatively different in different prosodic positions. Pauses have an inhibitory effect on devoicing for phrase-internal vowels, but a facilitatory effect for phrase-final vowels.

[16]The high standard errors of the <463ms and >463ms points, presumably due to the small number of phrase-internal tokens followed by appreciable pauses, prevent us from concluding there is an effect of increasing pause duration from medium to long pauses, in either direction.

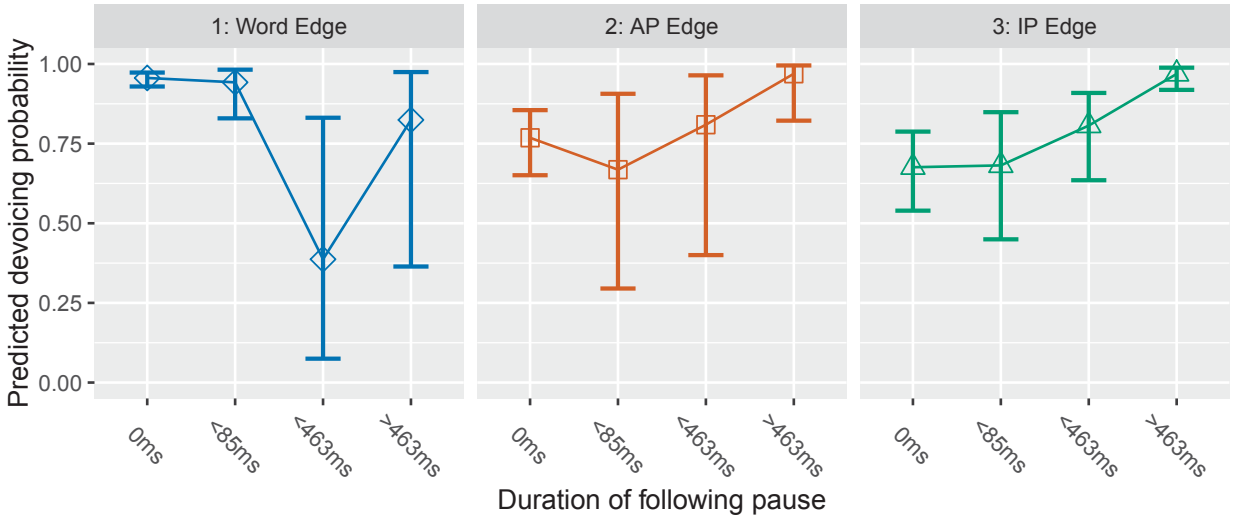

FIGURE 2.3: Predicted probability of devoicing for a high vowel at a word boundary as duration of the following pause increases, by prosodic position (*Break Index*), with other predictors held constant at mean values. Shapes represent the estimated probabilities, and bars show the 95% confidence intervals.

2.5.4 Mora deviation

Mora deviation strongly affects the likelihood of devoicing. As shown in Figure 2.4, devoicing is progressively less likely for longer Moras, and this holds across prosodic positions (when other variables are held constant). The regression terms are difficult to interpret directly, but their significance can be evaluated jointly: a likelihood ratio test (comparing the full model with one where all terms involving *Mora deviation* are excluded) shows that information about *Mora deviation* contributes significantly to explaining the variation in the data ($\chi^2(8) = 2325$, $p < 0.0001$). To visualize the predicted effect of *Mora deviation*, the model-predicted probabilities of devoicing as a function of *Mora deviation* for each prosodic position, with other variables held constant, are shown in Figure 2.4.

For word-internal vowels, devoicing is predicted to be at ceiling until the duration of the Mora is around the mean value (represented by 0 on the x axis in Figure 2.4), and from

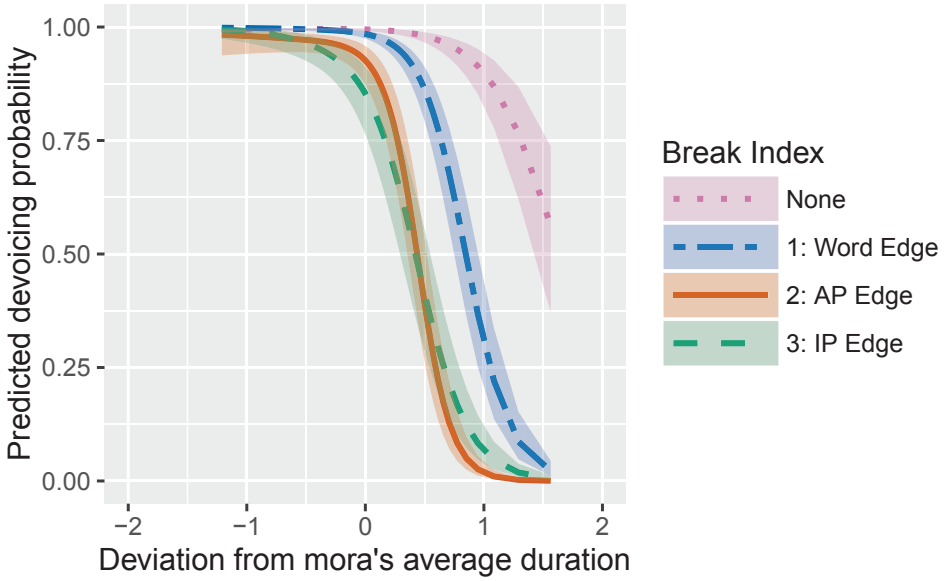

FIGURE 2.4: Predicted probability of devoicing for a high vowel by the degree of Mora deviation, by prosodic position (*Break Index*), with other predictors held constant at mean values. Lines represent the estimated probability, and shading shows 95% confidence intervals.

there ranges to about 50% at its lowest value. This agrees with previous work on consecutive devoicing environments which found that a Mora is significantly shorter when it is produced with a devoiced vowel (Kondo, 2005). For word-final vowels (*Break Index* = 1, 2, 3), the probability of devoicing ranges from almost 100% to almost 0% across the range of observed *Mora deviation* values, as shown in Figure 2.4.

Some of the interaction terms with *Break Index* were statistically significant. The slope of the estimated linear effect was significantly different between word-internal and word-final position, with devoicing probability being less affected by *Mora deviation* in word-final position ($\beta = -2.18$, $p < 0.001$). In addition, the effect of Mora deviation differs between IP-final vowels and AP-final vowels ($\beta = -1.97$, $p = 0.037$), such that Mora deviation has a stronger effect on AP-final vowels (steeper slope in Figure 2.4). However, none of the interaction terms change the qualitative shape of the effect of *Mora deviation*, which

is similarly negative across prosodic positions.

In sum, the duration of the Mora has a significant negative correlation with probability of devoicing. Interpreting higher *Mora deviation* as a proxy for more final lengthening and less gestural overlap, this pattern suggests that devoicing is less likely when there is more final lengthening, and more likely when there is more gestural overlap. The effect is qualitatively similar across all prosodic positions, in contrast with the effect of *Pause* described above.

2.5.5 Lexical frequency

The main effect of *lexical frequency* does not reach significance ($\hat{\beta} = 0.28$, $p = 0.402$), suggesting that word frequency does not play an important role in determining devoicing rates, averaging across prosodic positions. This is in contrast to an empirical plot of word frequency by devoicing rate of our data, which suggested a slightly negative effect, similar to the negative effect found by Maekawa and Kikuchi (2005) for the same corpus (although their analysis was for high vowels preceded by a voiceless consonant, but with any following environment). The fact that the model does not find a significant effect, in contrast to plots of the empirical data, suggests that the trend is primarily an artefact of other factors (variables which may be confounded with frequency, or lexical idiosyncrasies).

However, there are significant terms for the interaction of *lexical frequency* with *Break Index*, suggesting that word frequency does affect devoicing rate for some prosodic positions. Figure 2.5 shows the predicted frequency effect for each prosodic position, illustrating the pattern captured by these interaction terms. For word-internal vowels, the devoicing rate is at ceiling. Among word-final tokens, the frequency effect is slightly negative at phrase-internal word boundaries, versus positive at phrase-final word boundaries ($\hat{\beta} = 1.25$, $p < 0.001$): thus, we again see a qualitative difference among word-final vowels depending on whether they are phrase-internal or phrase-final. The frequency effect is significantly larger (= more positive) at IP boundaries than at AP boundaries ($\hat{\beta}$ = 1.49,

p = 0.004). Both of these terms point to the broader pattern in Figure 2.5: the effect of frequency is essentially restricted to IP-final vowels, where there is a strong *positive* effect: devoicing is more frequent for more frequent words.

A frequency effect in phrase-final position is expected under our account of phrase-final devoicing as a phonetically-motivated reduction process, discussed further below. However, we do not have a good explanation for why the frequency effect is essentially restricted to IP-final vowels. This may be due in part to high-frequency words which devoice near-categorically and occur disproportionately in IP-final position (e.g. *desu*, *masu*), though the by-word random intercept term should mitigate such effects of individual words.

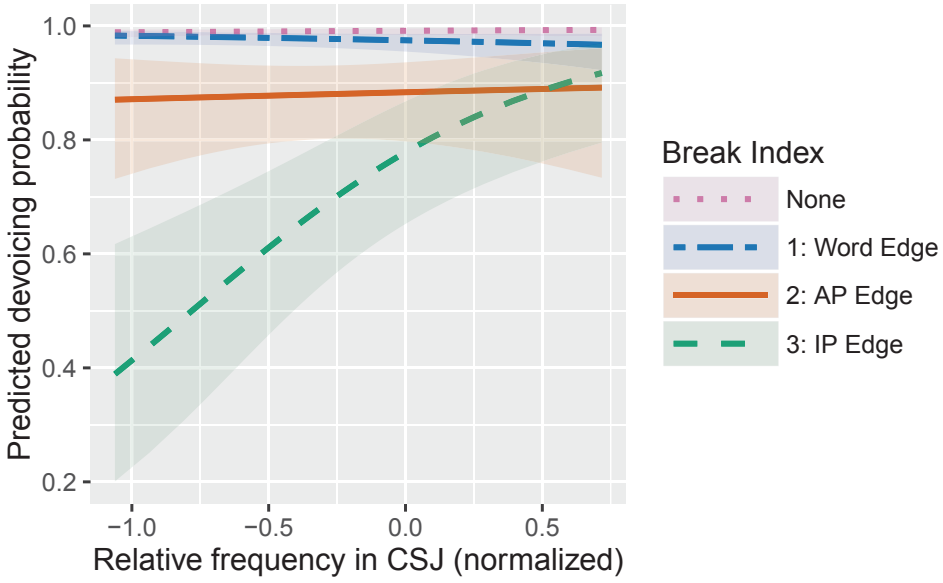

FIGURE 2.5: Predicted probability of devoicing for a high vowel by relative lexical frequency (log-transformed and normalized), by prosodic position (*Break Index*), with other predictors held constant at mean values. Lines represent the estimated probability, and shading shows 95% confidence intervals.

2.5.6 Speech rate

Two measures of speech rate were included in the model, average talker speech rate and local deviation from the talker's average speech rate. The *average speech rate* significantly

increases the probability of devoicing ($\hat{\beta} = 0.45$, p = 0.006), suggesting that faster talkers devoice vowels more readily. The main effect of *local speech rate* has a fairly small coefficient estimate, and does not reach significance ($\hat{\beta} = -0.04$, $p = 0.714$), suggesting that how fast a speaker is talking relative to their norm has little effect on devoicing rate, averaging across prosodic positions.

However, as with the lexical frequency effect, there is a significant interaction of *local speech rate* with *Break Index*, suggesting that the speech rate effect differs qualitatively by prosodic position. Figure 2.6 shows the predicted rate effect for each prosodic position, illustrating the pattern captured by these interaction terms.

For word-internal vowels, the devoicing rate is at ceiling regardless of speech rate. Among word-final vowels, the speech rate effect is significantly greater for phrase-final vowels than for phrase-internal vowels ($\hat{\beta} = 0.51$, $p = 0.03$). This results in the pattern apparent in Figure 2.6: phrase-final vowels tend to devoice more in faster speech, while phrase-internal vowels are not greatly affected by speech rate, if anything showing a tendency to devoice *less* in faster speech. Thus, we again see a qualitative split by prosodic position, depending on whether the vowel is internal or at the edge of a prosodic phrase.

2.6 Discussion

The results of the mixed-effects regression show that in the single devoicing environment, the devoicing rate for high vowels surrounded by voiceless consonants is affected by a number of factors—notably prosodic position, which both directly affects devoicing rate and modulates other variables, in a way which suggests a qualitative split between phrase-internal and phrase-final environments. These results bear on the three questions raised at the outset about Japanese vowel devoicing: the role of boundary phenomena, the relationship and characterization of the two environments in which devoicing applies, and the sources of variability. We first discuss our findings with respect to our research

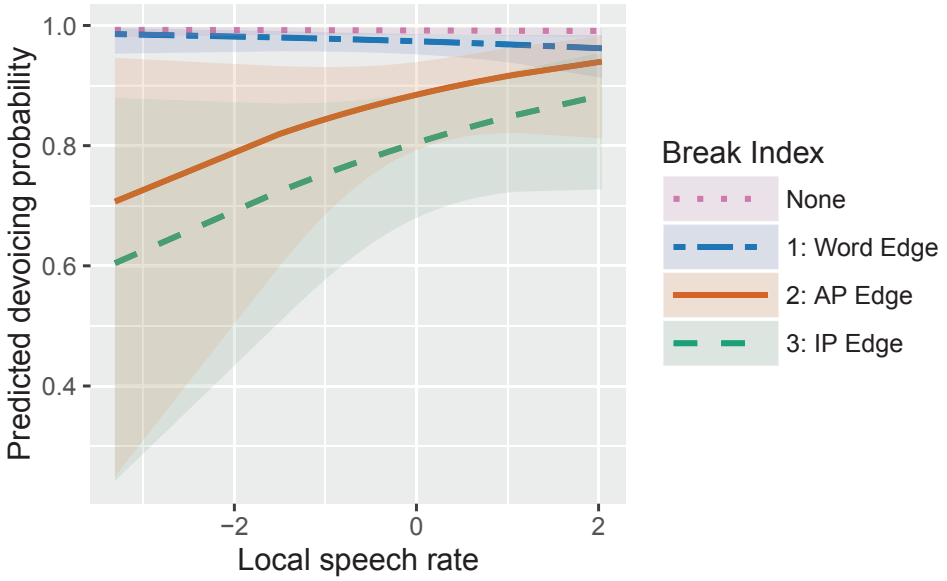

FIGURE 2.6: Predicted probability of devoicing for a high vowel by local speech rate (phones/second, normalized), by prosodic position (*Break Index*), with other predictors held constant at mean values. Lines represent the estimated probability, and shading shows 95% confidence intervals.

questions, which focused on boundary phenomena: how do word and prosodic boundaries affect devoicing rate (including modulating other factors), and what is the role of physical pauses? We then turn to the broader issues of how to characterize HVD, and sources of variability in its application.

2.6.1 Boundary phenomena

Word boundaries

Our first research question was how word boundaries affected the rate of high vowel devoicing and modulated the effects of other factors. The results confirmed a difference in devoicing rate between word-internal and word-final vowels, with a significantly lower probability of devoicing expected for vowels followed by word boundaries. This finding may seem unsurprising, given that external sandhi processes cross-linguistically usually

apply more consistently within words than across words, but to our knowledge this study is the first to demonstrate and quantify this effect for Japanese high vowel devoicing. Furthermore, the word boundary effect exists after controlling for confounding factors that could be correlated with word boundaries, such as domain-final lengthening and high tones, by including appropriate terms in the statistical model. This seemingly intuitive and simple inhibitory effect of word boundaries points to a new question for HVD, and external sandhi processes more generally: why should word boundaries *per se* have an inhibitory effect on process application, when other factors are held constant?

Another interesting result is that the estimated baseline rate for word-internal vowel devoicing is so high that it is not appreciably lowered by most inhibiting factors.[17] For example, Figures 2.5 and 2.6 show that when other factors are held constant, the probability of devoicing a word-internal vowel stays more or less at ceiling for any value of speech rate or lexical frequency (99.72% at mean values, 98.32% with frequency and speech rate at -2.5 standard deviations away from their mean value). Hence, these subtle effects are predicted to be masked for word-internal vowels. Even the relatively large effect size of *Mora deviation*, illustrated in Figure 2.4, is not predicted to completely block word-internal vowel devoicing at its most extreme value, with the lowest predicted probability reaching only about 50%. This is in striking contrast to word-final vowels, where devoicing is predicted to be almost totally absent for the most lengthened Moras. These results confirm textbook statements (e.g. NHK, 1991; Vance, 2008) and native speaker intuitions that high vowel devoicing is obligatory, but with the qualification that this holds for word-internal devoicing environments.

On the other hand, for word-final vowels the devoicing rate is estimated to be reliably slightly lower (Figure 2.2). This makes the devoicing rate at word boundaries more susceptible to the influence of even relatively small effects like *local speech rate* and *lexical*

[17] The one exception is the presence of a high tone, confirming the intuition that this blocks devoicing for some speakers (Han, 1962; Lovins, 1976; Hirayama, 2009; Nielsen, 2015).

frequency, as well as large effects like *Mora deviation*. Importantly, this difference in susceptibility is not due to the effects of word frequency, Mora duration, or speech rate actually differing between word-final and word-internal vowels—the relevant model terms are not significant (frequency, speech rate) or do not change the effect's direction (*Mora deviation*). Rather it is simply due to the much higher baseline devoicing rate for word-internal vowels.

On the whole, the results show that the presence of a word boundary is correlated with a decrease in devoicing rate, all else being equal. The effects of *Mora deviation*, *local speech rate*, and *lexical frequency* do not differ qualitatively if we compare their effects on word-internal versus word-final vowels.

Prosodic phrase boundaries

Our second research question was how prosodic phrase boundaries affected the rate of high vowel devoicing and modulated the effects of other factors. In examining the effect of prosodic phrase boundaries, we discuss both the presence/absence of a phrase boundary (either accentual or intonation phrase) at a word edge, and the difference between AP-final and IP-final tokens.

The statistical analysis shows that word-final devoicing rates differ significantly depending on whether a phrase boundary follows. In the absence of a pause, the presence of an accentual or intonation phrase boundary significantly decreases the probability of devoicing compared to a word-final vowel that is not followed by a phrase boundary. Among vowels that are followed by a phrase boundary, the stronger intonation phrase boundary is associated with significantly less devoicing than a weaker accentual phrase boundary. The overall pattern (Figure 2.2) is that as Break Index increases, devoicing rate decreases. What is driving this effect, and how does it fit in with current accounts of HVD?

Consider first the C̥_C̥ environment. Taking the view of HVD as a reductive process, the decrease in devoicing at stronger boundaries fits in with the cross-linguistic

tendency to see less reduction at stronger prosodic boundaries (Wightman et al., 1992; Keating, 2006). Since phrase boundaries are associated with segmental lengthening, this would also fit nicely with a gestural overlap account of HVD: the phrase-final vowel is lengthened, so the gestures of the surrounding consonants are less likely to overwhelm the vowel's voicing gesture. However, our model included *Mora deviation* as a separate factor, which accounts for this kind of temporal overlap. Indeed, our model estimates that as the Mora becomes longer (relative to its expected duration), the rate of devoicing declines sharply, so gestural overlap may play a role, but the effect of prosodic boundaries cannot be simply attributed to the temporal alignment of gestures. For example, if we consider two identical word-final vowels, both surrounded by the same consonants and *of the same duration*, the vowel followed by an accentual phrase boundary is more likely to be devoiced than the one followed by an intonation phrase boundary. A gestural overlap analysis of devoicing would have to be augmented to account for these effects. One possibility is that higher level prosodic boundaries are associated with some increase in magnitude (rather than timing) of the voicing gesture for the vowel, which leads to devoicing rates even lower than would be expected from simply articulating the vowel more slowly. In sum, our results show that prosodic boundaries have an effect on HVD above and beyond the potential confound of final lengthening, but overall it makes sense that a stronger boundary would disrupt the interaction between a word-final vowel and following voiceless consonant.

If we now consider the C̥_# environment, we run into a different puzzle. It has been suggested that "#" should be interpreted as the end of an intonation phrase or an utterance (Kondo, 1997; Hirayama, 2009; Fujimoto, 2015). If we take IP boundary as the definition of "#" in the C̥_# environment, then a feature-based analysis such as Rule 1 would not immediately explain the difference in variability between C̥_C̥ and C̥_# devoicing environments – if the process is the same in both environments, it should apply equally often in both cases. In fact, HVD at intonation phrase boundaries was much less consistent

overall, with the model estimating rates between 56% and 93% depending on the manner of surrounding consonants, all other variables held constant. Again, these differences between phrase positions are found even after controlling for presence of pause and Mora deviation, so the inhibitory effect is above and beyond these correlates of prosodic boundaries. On the other hand, defining "#" as a physical pause is also clearly not right, since the effect of a pause is inhibitory for phrase-internal vowels. The environment in which we find categorical HVD, other than word-internally, is defined by the *joint* effect of a phrase boundary and a longer pause, so both factors must somehow be incorporated into the definition of C̥_# .

The model also shows that prosodic phrase boundaries strongly modulate the effects of other variables. Most strikingly, the effects of *pause duration*, *lexical frequency*, and *local speech rate* are significantly different for phrase-internal vowels and phrase-final vowels, and that the effects of these variables is qualitatively different depending on prosodic position. We return to the pause effect below (Section 2.6.1), and here discuss the frequency and speech rate effects.

Overall, *lexical frequency* has little effect on devoicing rate. However, phrase-internal vowels and phrase-final vowels show a qualitatively different frequency effect. As Figure 2.5 shows,this effect is driven mostly by IP-final vowels, for which there is a strong positive frequency effect. This is the direction predicted for a reductive, phonetically-motivated process (e.g. Jurafsky et al., 2001; Pluymaekers, Ernestus, and Baayen, 2005), and consistent with a gestural overlap account of HVD.

A similar pattern emerges for the effect of *local speech rate* on HVD. For phrase-internal vowels at a word-boundary the effect is slightly negative, meaning that devoicing becomes *less* probable as speech rate increases. This is the opposite of what is expected for a reductive process, since reductions typically become more common at faster speech rates (e.g. Fosler-Lussier and Morgan, 1999). Phrase-final vowels, on the other hand, show the expected pattern (for a reductive process) of higher likelihood of devoicing at faster

speech rates. The positive speech rate effect is consistent with previous findings, such as the study in Martin, Utsugi, and Mazuka (2014), although other studies have failed to find speech rate effects in the single devoicing environment (e.g. Kondo, 1997).

It is striking that in both of these cases—as well as for the case of *Pause*, discussed below—there is a clear qualitative split between phrase-internal (*Break Index* None and 1) and phrase-final (*Break Index* 2 and 3) vowels. It would have been possible for these differences in effects to be only differences in magnitude, but still going in the same direction, as is the case for the *Mora deviation* effect. It could also have been the case that presence/absence of word boundaries modulated the frequency and rate effects, rather than phrase boundaries. The fact that the interaction terms involving phrase-internal/phrase-final differences in Table 2.4 are consistently significant (for *Pause*, *frequency*, and *local speech rate*) suggests that something about higher-level prosodic groupings must be invoked to explain this pattern of variability.

Physical pause

Part of the puzzle of high vowel devoicing we seek to address in this paper was the effect of a "pause", which ostensibly triggers devoicing, is associated with variable devoicing, and blocks devoicing. Our results on the effect of a physical pause, our third research question, show that these claims are all valid, but depend on context.

First of all, for word-final vowels that are not at any larger phrase boundary (Figure 2.3, left panel), pauses inhibit devoicing: devoicing is less probable if there is a pause following, of any duration. This effect is consistent with Vance's 1992 observation that in syllable-by-syllable pronunciations of words with potential devoicing sites, the pauses between the syllables block devoicing. It also supports the intuition expressed by some authors that C̥_# devoicing is not exactly conditioned by the pause itself, but by finality in an intonation phrase or utterance (Kondo, 1997; Hirayama, 2009; Fujimoto, 2015). The exact duration of the pause also affects devoicing rate, in a similar way: devoicing is more

likely before a short pause than before a medium pause. Thus, pauses gradiently and negatively affect the likelihood of devoicing for phrase-internal word-final vowels.

For phrase-final vowels, pauses have the opposite effect (Figure 2.3, middle–right panels): vowel devoicing becomes *more* probable before a pause, and more probable as pause duration increases. Thus, pauses gradiently and *positively* affect the likelihood of devoicing. In fact, with all other predictors held constant, devoicing is predicted to reach almost 100% probability for vowels which are followed by a long pause ($>$ 463 msec), but only if they are at an accentual or intonation phrase boundary.

The differences in the effect of pause once again mirrors the split seen for lexical frequency and speech rate effects: phrase-internal and phrase-final vowels are affected differently by these variables.

The role of boundary phenomena

Our findings on how boundary phenomena condition HVD is relevant for the more general issue of how boundary phenomena affect variable (phonological) processes. We found that prosodic boundaries and physical pauses have distinct and interacting effects: notably, the direction of one effect (*Pause*) flips depending on the value of the other (*Break Type*). Thus, the correct characterization of the 'pre-pausal' environment is more complicated than just one boundary phenomenon (e.g. 'utterance boundary') or another (e.g. 'long pause'). An interesting question for future work is whether this empirical pattern holds for other cases where variable processes apply in 'final' or 'pre-pausal' position, especially for vowel devoicing processes, where this description is common (Gordon, 1998; Barnes, 2006). How to capture the observed patterns in a formal analysis is a non-trivial question, which depends on how one assumes HVD is characterized. We return to this issue below.

2.6.2 High vowel devoicing as two overlapping processes

This study has shown that in a large corpus of spontaneous speech, it is possible to tease apart the effect of several (often correlated) variables on HVD. The results of our analysis suggest a complex relationship between HVD variability and word and phrase boundaries, pauses, lexical frequency, and speech rate measures.

It was confirmed that, in line with native speaker intuitions, HVD is "almost compulsory" when the following voiceless consonant is within the same word or across a word boundary (phrase-internal, no intervening pause). But HVD is also nearly categorical in basically the opposite context, when the vowel is followed by a prosodic phrase boundary *and* a relatively long pause. The results also confirmed the seemingly contradictory claims that pauses trigger devoicing (cf. the traditional description of C̥_# : Han, 1962; McCawley, 1968) and block devoicing (Vance, 1992; Kondo, 1997): in fact, pauses have opposite effects on devoicing depending on whether the vowel is at the edge of a prosodic phrase or not. Prosodic position also modulates the effect of lexical frequency and speech rate in a similar way. We now discuss possible interpretations of this complex pattern of variability within existing analyses of HVD, and follow with our own proposal. We suggest that by breaking down the source of devoiced vowels into two separate processes, we can describe two sub-patterns in the distribution of devoiced high vowels, which can help explain the overall pattern of variability.

The traditional description of Japanese high vowel devoicing exemplified in Rule 1 implies that the alternation between voiced and voiceless vowels is the same qualitative process, independent of which environment is the trigger of the change.

This assumption is difficult to reconcile with the patterns of high vowel devoicing variability observed in this study. Even allowing a rule such as Rule 1 to apply variably would not go very far toward explaining why devoicing is categorical or variable in a particular prosodic context. Furthermore, our results show that the position of the vowel

within the prosodic phrase affects not only the amount of variability in devoicing, but also the manner in which pauses, speech rate, and lexical frequency influence variability. In our view, this pattern suggests that two different processes underlie the alternations between voiced and voiceless vowels in Japanese.

The idea that devoiced high vowels may have different underlying sources has already been suggested in the literature, but with a different motivation. Tsuchida (1997), focusing on variability in consecutive devoicing environments, proposed that devoiced vowels in Japanese have two different underlying mechanisms depending on context. In the C̥_C̥ environment, it is argued, devoicing is categorical and due to a phonological rule. This classification is motivated by its categorical rate of application in the C̥_C single devoicing environment. The variability of devoicing in consecutive devoicing environments, and for vowels flanked by fricatives, suggests a phonetically-driven process in those cases. This dual-mechanism account is also defended by Varden (1998) and Nielsen (2015). Under these accounts, two processes are needed to account for variable and categorical application: variation within a consistent phonological context implies a phonetic process, while a phonological process should be categorical within a given context.

While the results of the present study agree with the claim that devoicing is near categorical within a word, the pattern of variability becomes more complicated as we investigate what happens to vowels at word boundaries and in different positions in a prosodic phrase. Surprisingly, we also see near categorical devoicing when there is the most disjuncture between a vowel and following consonant, namely at a prosodic phrase boundary with a long pause. Intuitively, if we think of C̥_C̥ devoicing as applying categorically word-internally, and C̥_# devoicing as applying categorically at a very strong boundary, all cases where there is variability lie in between these two extremes, where the two environments *overlap*. The picture of devoicing that emerges is not easily interpreted within a dichotomy of categorical/phonological versus continuous/phonetic.

However, we agree with the intuition that two different processes underlie devoiced

vowels in Japanese. Recall that cross-linguistically, there are two attested parameters that define the environments for vowel devoicing: the segmental context, and position with a (prosodic) domain. We propose that Japanese has two separate devoicing processes that differ precisely along these parameters, corresponding intuitively to C̥_C̥ and C̥_# . The C̥_C̥ process, which we call **interconsonantal devoicing**, is triggered by the voicelessness of the surrounding segments, but not by finality within a domain. This process parallels devoicing in Turkish (Jannedy, 1995) and Montréal French (Cedergren and Simoneau, 1985) in which vowels are only devoiced between voiceless consonants.

On the other hand, what has been described as C̥_# devoicing is a separate process, which we call **phrase-final devoicing**, which *does not* make reference to the following segment's properties, but rather the position of the vowel within a larger domain—tentatively, finality in an accentual phrase (and thus also in intonation phrases or utterances). This process parallels devoicing in languages like Greek (Dauer, 1980; Kaimaki, 2015), in which vowels are only devoiced phrase or utterance-finally. Note that an important caveat to our characterization of phrase-final devoicing is that our data only contains vowels followed by voiceless consonants. We assume in the following discussion that the "phrase-final" characterization is correct, but come back to this caveat in Sec. 2.6.4.

Overlapping environments

Our two-process proposal for HVD connects to the broader issue of how to analyze (variable) processes that apply in overlapping environments. We argued for two overlapping processes based on their distinct phonetic sources, cross-linguistic typology (where both kinds of devoicing processes are attested), and qualitatively different effects of non-grammatical factors (frequency, speech rate) by prosodic position. If our two-process proposal is correct, a formal analysis would be fairly straightforward: intervocalic devoicing and phrase-final devoicing could each be analyzed similarly to other cases of intervocalic devoicing or phrase-final devoicing (respectively) (e.g. Tsuchida, 2001). Devoiced vowels

between voiceless consonants in Japanese would then result from two different processes, analogously to other such cases, like word-final underlyingly-voiced obstruents in languages with both final devoicing and regressive voicing assimilation for obstruents (e.g. Polish, Catalan: Lombardi, 1991).

In contrast, a formal analysis of our HVD data as a single process would need to account for the complex effects of boundary phenomena, in particular the fact that the effect of one variable (pause duration) on devoicing rate *reverses direction* depending on the value of another variable (Break type). In a standard constraint-based analysis of a variable process (e.g. Maximum Entropy harmonic grammar: Hayes and Wilson, 2008; Coetzee and Pater, 2011), the effects of these two variables would be captured by two (sets of) constraints, each of which always assumes the same directionality of an effect. For example, one constraint could penalize devoicing before shorter pauses (accounting for the pattern in phrase-final position), but the opposite effect of a pause in phrase-internal tokens would be unaccounted for. In order for the effect of one variable to 'flip' depending on the value of another variable, additional mechanisms would need to be invoked, such as weighted constraint conjunction (e.g. Shih, 2016; Hayes, Wilson, and Shisko, 2012). While such an analysis is certainly possible, it would leave unexplained *why* the effect of pause differs by prosodic position, which falls out naturally from the two-process proposal: pauses interrupt the HVD environment phrase-internally, but not at phrase edges. We elaborate how the pattern of variability can be captured in the following sections.

2.6.3 Sources of variability

Our account of HVD in terms of two processes, which differ in sensitivity to following context versus prosodic position, helps elucidate the overall pattern of variability, shown in Figure 2.7, and the differing effects of pause, lexical frequency, and speech rate for phrase-internal and phrase-final vowels. The pattern of variability can be further explained by reference to two aspects of phonetic implementation and processing: gestural

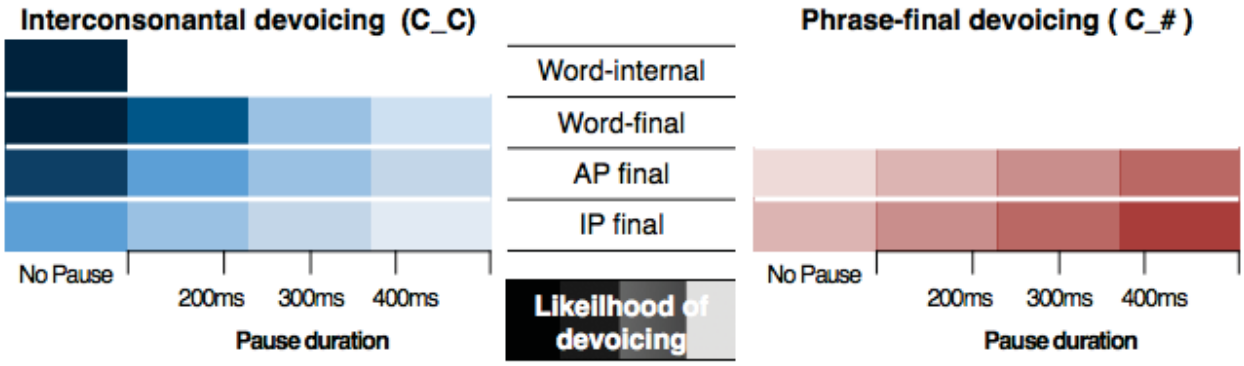

FIGURE 2.7: Schema of the pattern of variability for our two proposed processes of devoicing in Japanese. Darker colors represent higher likelihood of a devoiced vowel. Each row represents one of the prosodic conditions investigated in this study, and each column represents an interval of pause durations, with the first column being the case where there is no pause at all.

overlap, and the locality of production planning.

Across prosodic positions, *Mora deviation* is an important predictor of devoicing rate: devoicing becomes much less likely as Mora duration increases (as found by Den and Koiso, 2015 for utterance-final vowels), even for word-internal vowels. We suggest that the strong Mora duration effect reflects gestural overlap as a major source of variability in HVD, in addition to the effect of final lengthening on gestural overlap (Byrd and Saltzman, 2003). In all prosodic positions, shorter Mora duration will correlate with more gestural overlap with adjacent voiceless segment(s) making devoicing more likely. For word or phrase-final vowels, final lengthening will correlate with less gestural overlap and a longer vowel, either of which make devoicing less likely (Gordon, 1998; Barnes, 2006).

Turning to phrase-final devoicing in particular: the profile of variability we observe for phrase-final tokens is mostly consistent with phrase-final devoicing being a phonetically-motivated process (e.g. a postlexical, or 'late' phonological process; Coetzee and Pater,

2011), in particular reduction due to gestural overlap and aerodynamic factors. Phrase-finally, two kinds of phonetic factors promote devoicing: gestural overlap with the preceding voiceless segment, and decreased subglottal pressure over the course of an utterance (Gordon, 1998, p. 100). The positive effects of lexical frequency and speech rate for phrase-final vowels are consistent with the first of these sources: there should be more gestural overlap for higher-frequency words, or in faster speech, making the duration and magnitude of the voicing gesture shorter, both of which make it less likely that the aerodynamic conditions for voicing are met. The effect of pause duration makes sense assuming that a longer pause correlates with decreased subglottal pressure; there is then less likely to be sufficient pressure across the glottis to initiate voicing. Thus, the directions of the frequency, speech rate, and pause effects are consistent with a phonetically-motivated devoicing process which applies phrase-finally.

How to explain variability in application of interconsonantal devoicing, on the other hand, is a more challenging question. Interconsonantal devoicing does not show significant effects of lexical frequency or speech rate (when *Break Index* = None, 1), and is generally very consistent as long as no pause follows. However, the presence of a word boundary and the strength of the prosodic juncture between the vowel and the following consonant have gradient inhibitory effects on devoicing rates: devoicing is progressively less likely for higher *Break Index* values (Figure 2.2), for longer pauses (for *Break Index* = None, 1), and for higher *Mora deviation*, which we assume in part reflects the final lengthening expected for stronger prosodic boundaries. These patterns cannot be explained solely by reference to gestural overlap, which would lead us to expect the same patterns as for phrase-final devoicing: positive frequency and speech rate effects, and positive effects of pause and boundary strength. Thus, another explanation is needed for variability in interconsonantal devoicing: why is there variability *at all*, why do higher break indices condition less devoicing, and why do pauses condition less devoicing for phrase-internal word-final vowels?

We suggest that the locality of production planning can help explain these aspects of variability in the dataset, while allowing us to maintain a simple description of the environment for both processes. We offer a brief overview of the locality of production planning hypothesis before discussing how it may explain some of the patterns of variability found in this study, complementing those patterns which are well-explained by reference to gestural overlap.

The locality of production planning hypothesis (PPH) is a proposal developed in Wagner (2012), Tanner, Sonderegger, and Wagner (2017) which relates prosodic boundaries to phonological variability. It is proposed that the scope of speech planning constrains the application of phonological processes across word boundaries.

This hypothesis is based on findings in the psycholinguistics literature on speech production that at the phonological level, speech is planned hierarchically and incrementally (Sternberg et al., 1978; Ferreira, 1988; Ferreira, 1991; Dell and O'Seaghdha, 1992; Levelt, Roelofs, and Meyer, 1999). Higher-level information, such as number of words in an utterance, is planned before all lower-level information, such as number of syllables or segmental content, is retrieved or encoded. For example, Sternberg et al. (1978) found an asymmetry in the type of information that induced delays in initiating an utterance: the overall number of words in the utterance always increased the delay, but the number of syllables in a word only had an effect for the first word in the utterance. Wheeldon and Lahiri (1997) and Wheeldon and Lahiri (2002) similarly found the overall number of words in an utterance affected latency, but the number of syllables only had an effect when considering the first word (i.e. the number of syllables in the second word did not have an effect). They furthermore showed that prosodic organization plays a significant role in production planning, with production latencies crucially depending on the number of *prosodic* rather than lexical words. In Levelt's influential model of speech production (Levelt, Roelofs, and Meyer, 1999), segmental information is retrieved only incrementally, in word-sized planning units. Although there is an ongoing debate in the literature as to the

size of the window for phonological encoding (see Wheeldon, 2012, for an overview), it is agreed that in some cases, especially in spontaneous speech, the window is fairly limited, possibly as small as a single prosodic word. Hence, it must be the case that segments early in an utterance are planned in the absence of detailed information about later segments. The PPH is premised on the idea that even the segmental details of the *very next segment* may not be always be available. This situation is predicted to be more likely if the following segment is in a separate planning unit, and should be made even more likely by any other factors which delay the retrieval and encoding of phonological material.

How does this help explain the variability of interconsonantal devoicing? The PPH predicts that any alternation that is dependent on information from a following word (i.e. a separate planning unit) should be subject to variability. Applications of interconsonantal devoicing across a word boundary fall under this category: a word-final high vowel may have to be planned without the information that the upcoming word begins with a voiceless consonant, and hence there would be no motivation to plan a devoiced vowel. This would not be the case for word-internal applications of HVD, where the following consonant is always in the same planning unit and therefore always known at the moment of planning the vowel. Hence, the PPH explains the consistent difference in variability between word-internal and word-final vowels.

Our results also showed that as prosodic boundary strength increases, there is a gradient decrease in the probability of devoicing beyond what could be attributed to temporal overlap of voicing gestures. Wagner (2012) suggests that the strength of the boundary between two words is correlated with the likelihood of their being planned within the same window. Hence, under the PPH, it is predicted that stronger prosodic boundaries are associated with less availability of the segment following the boundary. For interconsonantal devoicing, this would lead to a decreased probability of application for higher level prosodic boundaries.

The inhibitory effect of pauses (for word-final, phrase-internal vowels) can be explained along similar lines. Pauses are associated with complexity of the upcoming phrase being planned (Sternberg et al., 1978; Ferreira, 1991; Wheeldon and Lahiri, 2002), so they may also track availability of the segment following the pause. Again, decreased availability of the following segment would lead to decreased application of interconsonantal devoicing.

In sum, the PPH offers an explanatory mechanism as to why a seemingly planned, phonological process may show variability in spontaneous speech. When an alternation depends on information in an upcoming word, many factors may interfere with online phonological encoding during the course of speech planning, leading to an "opaque" output from the perspective of the ultimate pronunciation (e.g. a voiced high vowel between two voiceless consonants).

Production planning and formal analysis

We have invoked the PPH to describe variability observed in our empirical data, without providing a formal analysis. But, related to the broader theoretical issue of how to account for processes which show near-categorical behaviour in some environments and variability in others, it is worth discussing how the mechanism of PPH could be incorporated into a formal analysis of HVD, and of other such processes. We propose two options. As suggested by Wagner (2012), explaining variability in a process' application in terms of production planning could be used to maintain a *non*-probabilistic account: interconsonantal HVD could be a categorical process described in purely segmental terms (i.e. devoice high vowels between voiceless consonants), as in traditional descriptions, while the variability observed across word boundaries and as a function of various factors (prosodic boundary, pause, speech rate, frequency) would be 'factored out' to production planning. Alternatively, PPH could be incorporated into the structure of phonological grammar, as

a factor restricting which phonological patterns are possible—similarly to projecting constraint scales based on perceptibility in an Optimality Theoretic analysis which restrict possible neutralization patterns (Steriade, 2008). For example, it is predicted to be impossible to have a process which is "more variable" across words than within words. The choice between these options is beyond the scope of this dissertation.

2.6.4 Other issues

An important caveat to our characterization of phrase-final devoicing is that we have characterized it without considering a large subset of cases where it could apply: phrase-final short high vowels followed by a *voiced* segment. Recall that the dataset was restricted to tokens followed by a voiceless consonant, for reasons discussed above. Thus, strictly speaking we cannot show that "phrase-final devoicing" shows particular behavior without showing that all aspects of phrase-final devoicing (e.g. frequency effect, devoicing rate) do not depend on the following segment's voicing. As a basic check of this, Figure 2.8 shows the empirical devoicing rates for all phrase-final tokens (BI = 2, 3) in the CSJ ("Core" subset), broken down by following pause duration, as a function of voicing of the following segment. When any pause is present—the positions where phrase-final devoicing is most likely to apply—there is no apparent effect of following segment voicing on devoicing rate, suggesting that our characterization of "phrase-final" devoicing is on the right track.[18] A more detailed check would need to consider the role of following segment voicing more generally, in all positions (and report a statistical analysis): within-word and across-word. In fact, the facts are complex: Maekawa and Kikuchi (2005) showed that in the CSJ, devoiced short vowels actually do occur in C̥_[+voi] context *within words*, not infrequently (10%, in the current dataset). By any treatment of HVD, devoicing in this

[18] When no pause is present, in the left panel, there is an effect of voicing in the direction expected if both C̥_C̥ and C̥_# devoicing can apply before a voiceless segment but only C̥_# can apply before a voiced segment.

context should be impossible, suggesting that the role of following segment voicing is an important but complex direction for future work.

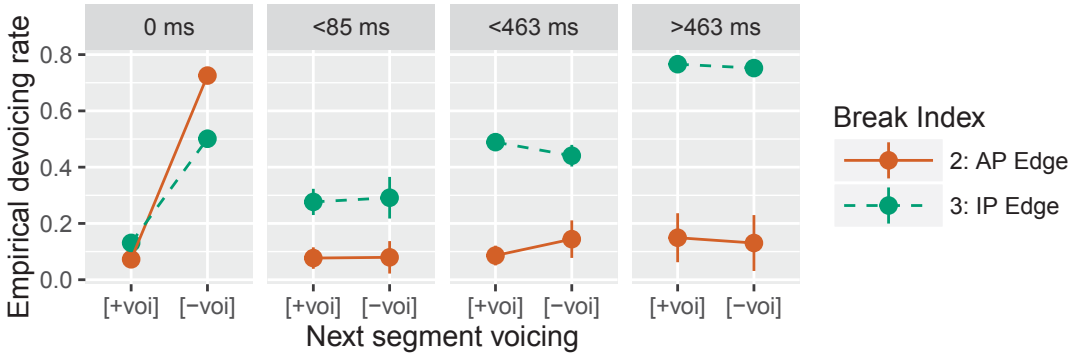

FIGURE 2.8: Percentage of devoiced high short vowels in phrase-final position as a function of following segment voicing, by phrase type (Break Index) and duration of following pause (panel labels). Errorbars indicate ±2 standard errors.

Another direction for future work is the relationship between the categorical measure (constituent boundaries: AP, IP) and continuous phonetic measures (pause duration, Mora deviation) of boundary strength, and how they affect the two hypothesized devoicing processes. These three measures are strongly correlated, and a more thorough examination of their relationship could give a better understanding of the articulatory characteristics of "final" position. While we have shown that the three measures all independently affect devoicing rate in a dataset where the C̥_C̥ and C̥_# environments are pooled, their relative effects on each kind of devoicing remains unclear. If the phrase-final devoicing process is phonetically driven and interconsonantal devoicing is an "early" phonological rule, then one might expect these processes to be more/less affected by "phonetic" variables (Mora deviation, pause duration) than by categorical constituent boundaries (Break Type), respectively. A reviewer notes that it is difficult to examine this issue without considering Pause Duration as a continuous variable, rather than discretizing it into bins—as

was done in this dissertation to better address our research questions (see Sec. 2.4.1). Future work examining the effect of pause duration (as a continuous measure) on phrase-final devoicing rate could reveal a more nuanced relationship, potentially related to a finer prosodic hierarchy than AP/IP.

2.7 Conclusion

This paper investigated the role that boundary information plays in the variability of Japanese HVD. We focused on teasing apart the effect of highly correlated boundary phenomena—including prosodic phrase boundaries, pauses, and final lengthening—and how these might interact with C̥_C̥ devoicing and define the C̥_# devoicing environment. By examining these factors in a large corpus of spontaneous speech and controlling for other factors known to influence HVD, we were able to pinpoint different sources of variability for HVD depending on the particular context the vowel appears in.

Our results showed that the correlated boundary phenomena have a *joint* influence on variability in HVD. All else being equal, a larger prosodic phrase boundary following the vowel was correlated with a decrease in devoicing rate. Also, the duration of a particular Mora relative to its average duration in the corpus was negatively correlated with the likelihood of HVD, which likely reflects gestural overlap and final lengthening. But the effect of a physical pause was dependent on whether the target vowel was phrase-internal, where pause inhibited HVD, or phrase-final, where pause promoted HVD. The joint effect of a phrase boundary and a long pause led to almost categorical devoicing rates. Phrase-internal and phrase-final vowels were also influenced in qualitatively different ways by speech rate and word frequency: phrase-final vowels showed a positive effect, typical of reductive processes in general, while phrase-internal vowels showed no such effects.

We proposed that there are two separate devoicing processes: interconsonantal and phrase-final devoicing, which show different patterns of variability.

Phrase-final devoicing shows telltale signs of a reductive process, namely the positive effect of speech rate and lexical frequency. This pattern could be accounted for under existing proposals of devoicing as gestural overlap and reduction. Phrase-final devoicing is also promoted by a long pause, which could also receive an articulatory explanation if it is assumed that long pauses at the end of a phrase or utterance are associated with a decrease in subglottal pressure, making it harder to initiate voicing.

Interconsonantal devoicing, on the other hand, shows a pattern of variability that is less easily explained by gestural overlap and reduction. We suggest that its variability can be better understood by reference to the *locality of production planning* hypothesis, which explains part of the variability as a consequence of limitations imposed by online speech production. The inhibitory effect of larger prosodic phrase boundaries, and negative effect of pause for phrase-internal word-final vowels, are due to these two factors correlating with later planning of an upcoming voiceless obstruent, which interferes with the planning of a devoiced vowel variant in the interconsonantal environment.

Acknowledgments

A preliminary version of this work was reported in Kilbourn-Ceron (2015). We thank audiences at LabPhon 14 and ICPhS 2015, Kuniko Nielsen, Hisako Noguchi, and James Tanner for feedback on this project; Michael Wagner, Heather Goad, three anonymous reviewers, and Rachel Walker for useful comments on manuscript drafts; and Michael McAuliffe for translation help. This work was supported by a SSHRC CGS Doctoral Scholarship (767-2012-1089) and CRBLM Graduate Scholar Stipend to OKC, and research grants from SSHRC (#430-2014-00018) and FRQSC (#183356) to MS.

Appendix: Random effects

TABLE 2.5: Summary of random-effect terms for the statistical model of HVD: variances and corresponding standard deviations.

Predictor	Variance	Standard Deviation
Word		
(Intercept)	5.139	2.267
Speaker		
(Intercept)	0.769	0.877
Break Index		
1, 2, 3 − None	0.719	0.848
2, 3 − 1	1.447	1.203
3 − 2	1.467	1.211
Lexical frequency	0.042	0.205
Speech rate within utterance	0.000	0.000
Pause : Break Index		
No Pause − Pause : 1, 2, 3 − None	3.441	1.855
No Pause − Pause : 2, 3 − 1	8.697	2.949
No Pause − Pause : 3 − 2	5.076	2.253
Short Pause − Medium/Long Pause : 1, 2, 3 − None	2.924	1.710
Short Pause − Medium/Long Pause : 2, 3 − 1	16.638	4.079
Short Pause − Medium/Long Pause : 3 − 2	7.409	2.722
Medium Pause − Long Pause : 1, 2, 3 − None	3.549	1.884
Break Index : Lexical Frequency		
1, 2, 3 − None : Lexical frequency	1.332	1.154
2, 3 − 1 : Lexical frequency	2.412	1.553
3 − 2 : Lexical frequency	10.890	3.300
Break Index : Speech rate within utterance		
1, 2, 3 − None : Speech Rate	0.623	0.789
2, 3 − 1 : Speech Rate	1.817	1.348
3 − 2 : Speech Rate	1.281	1.132

Preface to Chapter 3

Chapter 2 proposed that the distribution of devoiced vowels in Japanese is the result of two distinct high vowel devoicing (HVD) processes: interconsonantal devoicing, which depends on adjacent voiceless consonants, and phrase-final devoicing, which does not depend knowledge of upcoming segments. This was motivated by the finding that there are *qualitative* differences in the effects of lexical frequency and speech rate on HVD depending on prosodic position. The results also showed that after controlling for pauses, relative duration, frequency and speech rate, *prosodic boundaries* had a significant inhibitory effect on HVD. We proposed that this effect could be understood under the PPH as inhibition of the interconsonantal devoicing process by online speech production planning constraints. In order to find further evidence for production planning effects, the next step was to test an external sandhi process whose outcome is clearly dependent on the following segmental context. This is tested in Chapter 3 with flapping in North American English.

In North American English, a coronal stop is realized as a flap [ɾ] if it is preceded and followed by vowels, and it is not in the onset of a stressed syllable. This process can apply even if the following vowel is in a separate word, so flapping is a type of external sandhi. In Chapter 3, flapping is used to test predictions of the PPH about the effect of *syntactic structure* and *lexical frequency* on external sandhi realization.

The PPH predicts that the planning of upcoming syntactic structure can affect the planning window for phonological encoding. Words that are at the beginning of complex

syntactic constituents will take longer to retrieve and encode, since planning of higher-level structure must be completed prior to word-form encoding. This prediction is tested in a production experiment, which tests the realization of word-final /t/ followed by a vowel-initial word. Syntactic structure is varied so that the vowel-initial word is either in the same constituent as the target word, or at the beginning of a new clause. The PPH predicts that flapping should be less likely in the latter case, where a clause boundary separates the /t/-final and the vowel-initial word, since it is less likely that the vowel will be retrieved and phonologically encoded within the same window.

Another factor which could affect the planning window, and therefore external sandhi, is lexical frequency. Previous work has shown that more frequent words are retrieved more quickly. Therefore, the frequency of the word following a word-final coronal stop should modulate whether it is in the same planning window, with higher frequency words being more likely to affect the realization of the previous word. This prediction is tested in a corpus of North American English.

Chapter 3

External sandhi and the locality of production planning: The case of flapping[1]

3.1 Introduction

Listeners trying to decode the message of an utterance face a complex task: They need to segment the incoming acoustic stream into the words that were used and identify them (speech segmentation), and at the same time they need to figure out how these words relate to each other and combine to form the overall meaning of the message (syntactic parsing).

One strategy to segment the incoming continuous signal into words is to identify possible words and find the best parse that assigns all of the signal to words (e.g. McQueen et al., 1995; McClelland and Elman, 1986; Norris, 1994). This is the strategy that we have to use when given an orthographic string without spaces, like *Theyweretalkingveryunclearly*. But in the acoustic signal, there are also cues that give us more direct clues about the location of word boundaries (e.g. Lehiste, 1960). This makes this task much easier, and makes listening to speech more like the task of recognizing the words in an orthographic string

[1] A version of this chapter will appear in *The Proceedings of the 52nd Meeting of the Chicago Linguistics Society.*

in which words are separated by spaces, as in *They were talking very unclearly* (see Mattys, White, and Melhorn, 2005; Mattys and Melhorn, 2007; Mattys, Melhorn, and White, 2007, for evidence that both strategies play a role)

In speech, having a 'space' or a pause between words is the exception rather than the rule, but many other cues that have been shown to influence parsing. There are consistent cues to prosodic junctures at phrase boundaries such as final lengthening and boundary tones (Wightman et al., 1992; Price et al., 1991), and also acoustic cues to word boundaries (Davis, Marslen-Wilson, and Gaskell, 2002; Salverda, Dahan, and McQueen, 2003). Metrical structure, such as the position of stressed syllables, also plays a role (e.g. Cutler and Norris, 1988; Cutler and Butterfield, 1992; Banel and Bacri, 1994; see Cutler, Dahan, and Van Donselaar, 1997 for a review). Another important type of cue comes from positional variants of certain sounds (e.g. Nakatani and Dukes, 1977; see also Gaskell and Marslen-Wilson, 2002; Mattys, White, and Melhorn, 2005, for a review. Such 'allophonic' variation often helps encode word boundaries. An example is that /t/s at the ends of words in English can be pronounced as a glottal stop [ʔ], a process we will refer to as 'glottalization' (Kahn, 1976). Realizing an underlying /t/ as a glottal stop is a cue for a word juncture, and, as we will see in the present study, a cue that a prosodic boundary follows.

Some allophonic processes differ from glottalization in that they rely on phonological information in a following or preceding word. Such processes are often referred to as 'external sandhi'. These processes can play a role in encoding which words in an utterance form syntactic-semantic units. An example of a process that can apply across word boundaries is 'flapping' in North American English. The coronal stops /t,d/ are often realized as an alveolar flap when they appear between vowels.[2] When they occur intervocalically within a word, the flap realization is nearly categorical, unless the following vowel is stressed (Kahn, 1976; Eddington and Elzinga, 2008). For example, the stops in *atom* and *Adam* are highly likely to be realized as a flap (except maybe in very slow speech,

[2] And also after certain sonorants, e.g. in /ntV/ and /ndV/ sequences, which we will not discuss here.

cf. Gussenhoven, 1986), but the stops in *atomic* or *adorable* are not. When a word boundary separates the occurrence of /t,d/ and the following vowel, as in *at Olivia's*, this is an instance of external sandhi, and these instances of flapping across word boundaries are our main focus here.

If glottalization helps demarcate word boundaries, flapping has the opposite effect: It makes sequences of words more similar to single words. But it might serve a different function, helping demarcate which strings of words form syntactically and/or semantically coherent sequences. In other words, it helps with the challenge of syntactic parsing. Scott and Cutler (1984), for example, looked at the perception of structural ambiguities such as the following:

(3.1) a. For those of you who'd like to eat, early lunch will be served.

b. For those of you who'd like to eat early, lunch will be served.

In their perception experiment, Scott and Cutler (1984) found that once other acoustic cues to syntactic boundaries are controlled for, flapping is a highly reliable cue to distinguish such attachment ambiguities. If a flap is realized in *ea*[ɾ] *early*, then the sentence must receive the (3.1b) parse, with *early* modifying *eat*. Allophonic cues likely play a bigger role in spontaneous speech, where clause boundaries are less likely to be marked by other means, e.g. by pauses (Kowal, Bassett, and O'Connell, 1985). The locality conditions on external sandhi have the effect that the application/non-application of a process across word boundaries can effectively encode information about which words belong together syntactically/semantically. Other cross-word processes, such as assimilation, have been shown to correlate with locality relations between adjacent words (Holst and Nolan, 1995; Nolan, Holst, and Kühnert, 1996), and can serve as cues to structural relations between words.

Apart from being subject to **locality**, external sandhi has a second characteristic property: It is more likely than other types of phonological processes to be 'optional' or **variable**. Across word boundaries, the application of flapping is much more variable than word-internally. A number of factors have been hypothesized in previous literature to affect whether flapping will occur, including the syntactic relation between the two words (Kahn, 1976; Nespor and Vogel, 1986) and whether they form part of the same prosodic domain (Nespor and Vogel, 1986).[3] The pattern of variability observed in flapping is common in sandhi processes cross-linguistically (see e.g. Kaisse, 1985). But *why* should segmental processes that span word edges be more variable?

While both the locality and variability of external sandhi processes have each been extensively discussed in the literature, the *link* between these two properties has not. The locality of external sandhi processes, for example, has received various accounts in the literature on prosodic phonology (Cooper and Paccia-Cooper, 1980; Selkirk, 1984; Kaisse, 1985; Odden, 1987; Nespor and Vogel, 1986; Pak, 2008; Selkirk, 2011; Šurkalović, 2016). But these accounts do not link the locality conditions on sandhi to its variability, which is usually noted, but not taken to be part of what the theory of phono-syntactic locality should explain. Conversely, recent phonological work trying to gain a better understanding of phonological variability (see Anttila, 2007; Coetzee and Pater, 2011, for overviews) has not tried to answer the question of how the variability of phonological processes relates to their locality conditions, and also has not explored the question of why cross-word processes specifically are more likely to be variable than word-internal ones.

In this paper, we draw on speech production planning research in order to connect these two properties of external sandhi and relate them to a single underlying source: the

[3] Across word boundaries, flapping is possible even when the following vowel is stressed, as in *at Olive's*, while aspiration is often said to be impossible in the same context. This could be evidence that, even in fast speech, the /t/ never occupies the onset position of the syllable, or at least it *also* has to be syllabified into the coda. We will not discuss issues of syllabification in this paper, see Kahn (1976) and Gussenhoven (1986) for discussion.

locality of speech production planning (see also Kilbourn-Ceron, 2015; Tanner, Sonderegger, and Wagner, 2017; Wagner, 2011; Wagner, 2012; Kilbourn-Ceron and Sonderegger, 2017).[4] By relating results from current research on speech production to external sandhi phenomena, we aim to provide a new perspective on the patterns of locality and variability in connected speech.

In the following, we will first present the **Production Planning Hypothesis** (PPH) in more detail, outline its predictions for external sandhi phenomena, and how the PPH differs from alternative accounts. In particular, we compare the PPH to an alternative account of flapping purely in terms of gestural overlap. We then present a case study on English intervocalic /t,d/ realization, in order to test these predictions. We report on a production experiment which looks at the effects of syntactic juncture on /t/ realization, and on a corpus study looking at the effects of lexical frequency of the target and following words. Both experiments provide new insights into the nature of flapping and glottalization, and provide evidence for different predictions of the PPH. In concluding, we compare the PPH with two other alternative accounts: an account in terms of the prosodic hierarchy, and an account in terms of probabilistic reduction.

The Locality of Production Planning

The PPH proposes that the realization of external sandhi is constrained by the availability of phonological information during speech production planning. Many aspects of an utterance can be planned incrementally, with articulation happening in parallel with planning of upcoming words and sentences. The scope of advance planning for detailed segmental and metrical information in particular is generally considered quite narrow, possible as small as a single word (Levelt, 1989). Sternberg et al. (1978) found that utterance

[4] See also MacKenzie (2013) for similar effects of production planning locality on contraction of function words, and Tamminga, MacKenzie, and Embick (2016) for a more general discussion of the role of production planning effects in accounting for individual variation.

initiation time, the amount of time is takes to start speaking after a start signal is given, correlates with the overall number of words in a word list the participant has prepared before the signal. This suggests that the number of items in the list is planned early on, with each additional word incurring additional planning time before articulation of the utterance can start. Utterance initiation time was also correlated with the number of syllables in the first word, suggesting that additional syllables require additional planning time (see also Wheeldon and Lahiri, 1997; Wheeldon and Lahiri, 2002). In contrast, increasing the number of syllables in the second word or later had no effect. In other words, the planning of lower-level detail including syllables is effected only very locally, within a small window, while higher level information like number of (prosodic) words is planned over a large planning window. Shattuck-Hufnagel (2000) and Keating and Shattuck-Hufnagel (2002) argue that higher-level prosodic information—for example intonational phrasing and the associated tunes—is also planned early on, before lower level prosodic and segmental information is available. This pattern of higher level information being planned over a larger window than lower level information is compatible with evidence from speech errors (Fromkin, 1971; Shattuck-Hufnagel, 1979; Garrett, 1988).

According to Levelt's influential model of speech production (Levelt, 1989; Levelt, Roelofs, and Meyer, 1999), segmental phonological information is encoded in roughly word-sized planning chunks, i.e. within a very local window. Of course, the application of an external sandhi process necessarily requires planning a chunk that encompasses the current word and at least the beginning of an upcoming word. Levelt (1989, p. 377) discusses such processes, and acknowledges that cross-word phonological phenomena imply that "a lookahead of no more than one word is required."

The size of the speech planning window has in fact been shown to vary. Wheeldon and Lahiri (1997) and Wheeldon and Lahiri (2002) found that depending on the task, utterance initiation can be driven more by the number of upcoming prosodic words (with more planning time), or the internal complexity of the first upcoming prosodic word (with

shorter planning time). In other words, how far a speaker plans ahead is task-dependent. We submit that the variability in the application of sandhi rules can be linked to this variability in the size of the planning chunks involved in speech planning (cf. Wagner, 2012; Tanner, Sonderegger, and Wagner, 2017).

Prior research has identified a number of factors affecting planning scope. The PPH predicts that those same factors should affect the rate at which external sandhi processes apply. The size of the planning window has been found to depend on syntactic constituency and semantic coherence (Wheeldon, 2012) and on the lexical frequency of the words involved (Konopka, 2012). An increase in cognitive load has been shown to reduce speech rate (Mitchell, Hoit, and Watson, 1996), and been argued to decrease planning scope (Ferreira and Swets, 2002; Wagner, Jescheniak, and Schriefers, 2010). Also, individual differences in working memory correlate with planning scope (Swets, Jacovina, and Gerrig, 2014).

Prosodic groupings appear to be closely correlated with the size of the planning window. Just as planning scope is affected by syntactic complexity, so is prosodic phrasing. For example, Breen, Watson, and Gibson (2011) found that following a verb, an upcoming subject of a following clause would be much more likely to be set off by a prosodic boundary than an upcoming direct object that the verb takes as an argument. More generally, if the upcoming word is syntactically and semantically closely linked to the interpretation of the verb, and more predictable, the prosodic boundary separating the words tends to be weaker. Gahl and Garnsey (2004) found that verbs tend to be shorter if the complement of the verb is of a more predictable syntactic category for that verb.

Krivokapić (2007) showed that post-boundary pauses—an important cue to boundary strength—are affected by the complexity of upcoming prosodic constituents both in terms of their length (in syllables) and how many sub-constituents they contain. This suggests that both factors influence the course of phonological encoding. This result can be explained if more complex constituents take a longer time to plan, and if the strength

of prosodic boundaries correlates with the degree to which an upcoming constituent has already been planned at that point. It is hard to tell whether the strength of a prosodic boundary itself affects the likelihood of the following word being planned, or conversely whether the likelihood of upcoming material being planned determines the strength of the prosodic boundary produced—what's crucial here is that prosodic boundary strength *correlates* with the amount of look-ahead in phonological planning (see also Ferreira, 1991).

The PPH makes the strong prediction that any phonological alternation which relies on phonological information from an upcoming word *must* be variable, since phonological processes cannot apply if the conditioning phonological environment in the next word has not yet been retrieved, and we know that speakers do not reliably plan the phonological detail of more than one word ahead of time. It also makes predictions about the sensitivity to locality that a process should show. Finer grained information is planned in a more narrow window, so if finer grained information about an upcoming word is needed (e.g. does it begin with a vowel?), the process will be more variable than one which relies on higher level information (e.g., is there another following word at all?), which is planned earlier. We return to these predictions about locality differences in our final discussion.

Flapping and gestural overlap

In the phonological literature (Kahn, 1976; Kiparsky, 1979; Gussenhoven, 1986; Nespor and Vogel, 1986) flapping is usually treated as a choice of allophone for an underlying /t/ or /d/. The assumption is that flapping involves a categorical change from a stop to an flap. There is another perspective on how to characterize the nature of this alternation, under which flaps involve similar gestures as a regular [t] and its surrounding vowels, but these are blended in such a way that the acoustic result is a much shorter (voiced) consonant. Herd, Jongman, and Sereno (2010), for example, characterize flaps as a result of blending the gestural requirements of the stop with the surrounding vowels. Fukaya and Byrd (2005) explore several possible articulatory explanations on why word-final [t]s

might be realized in a way that is perceived as a flap, including differences in the start and end point of the gestures, a greater velocity of the gesture, or a truncation of the gestures associated with the [t] due to greater overlap with adjacent vowels. While there was substantial variation in their data in how often speakers produced word-final stops that were perceived as flaps, and the articulatory means by which they achieved a flap-like realization, there was no evidence that flaps involve qualitatively different gestures. This is consistent with earlier articulatory measures suggesting that flapping is gradient rather than categorical (e.g. Fox and Terbeek, 1977), although the acoustic consequence may be more categorical (De Jong, 1998). The gestural account also receives support from the observation that consonants other than /t,d/ are subject to similar temporal reductions in flapping environments (Browman and Goldstein, 1992; Turk, 1992).

A gestural account is also supported by findings that flapping does not neutralize the distinction between an underlying /t/ and /d/, which remains detectable in small but consistent phonetic differences in the length of the preceding vowel (Malécot and Lloyd, 1968; Herd, Jongman, and Sereno, 2010; Braver, 2011). This pattern is unexpected if flapping involves a categorical phonological change (though see Bermudez-Otero, 2011).

If flapping is due to gestural overlap, acceleration, or reduction, this may already explain part of the locality and variability of this process. Under this view, we would expect flapping to be local and variable if it is a direct consequence of temporal compression of the word ending in the stop and the following word. For example, syntactic and prosodic boundaries could reduce the flapping rate by way of inducing a temporal slow-down at the boundary due to the effect of final lengthening (Byrd and Saltzman, 2003). This view also accounts for why other temporal modulations, such as changes in speech rate, affect the overall likelihood of flapping (cf. Browman and Goldstein, 1992).

An articulatory overlap account of flapping variability therefore makes overlapping predictions with the PPH. For example, a stronger prosodic boundary should lead to a lower rate of transcribed flaps, since stronger prosodic boundaries lead to less gestural

overlap and compression (Browman and Goldstein, 1992). Similarly, if increased frequency of a following word leads to shorter duration of the current word (Jurafsky et al., 2001), then we expect a higher rate of flapping with higher frequency of the following word, again mirroring the prediction of the PPH.

We will see, however, that temporal compression alone will not be sufficient to explain the observed patterns of variability. As pointed out in Whalen (1990), gestural overlap accounts assume that coarticulation is not planned, but rather automatically emerges from the temporal overlap of articulatory gestures associated with different segments. In contrast, Whalen (1990) presents compelling evidence that anticipatory coarticulation *is* planned. In a production experiment, Whalen tested whether speakers would show coarticulation in the initial vowel of a nonsense string like *api, apu, abi, abu* and the following consonant and vowel if they had to initiate speaking before the entire string was revealed. If the second vowel was not known in advance, then vowel-to-vowel coarticulation was absent, while the effect of consonant voicing on the initial vowel duration remained. If the consonant was initially hidden, then vowel-to-vowel coarticulation was present, but the durational effect of the consonant disappeared. In other words, anticipatory coarticulation was only possible when the triggering segment was known in advance, suggesting that coarticulation is not just an emergent property of temporal compression. Articulating a flap in particular requires close coordination of the gestures of the flap and the following vowel. It stands to reason that the realization of a flap requires fairly detailed planning of the following vowel, and if so the locality of production planning should severely constrain flapping rate.

We can distinguish the PPH from an account relying only on gestural overlap by looking for evidence that speakers make choices about the articulatory plan rather than simply

compressing the existing articulatory plan depending on speech rate and related temporal factors. More specifically, an account purely in terms of temporal compression predicts that once we have controlled for the durational compression of the segments involved, there should not be any further effect of syntax, frequency, or other factors that affect production planning. In our experiments, we will try to control for duration, in order to test whether factors affecting production planning still have an effect, and to test whether there is a straightforward relationship between duration and flapping rate. Finding an effect of factors affecting production planning, such as syntax and frequency, *after* controlling for temporal compression would provide evidence that production planning constraints do play a role in explaining the variability of flapping.

A second potential source of evidence for the PPH comes from examining the overall pattern of alveolar stop realizations under different planning conditions. Flapping requires segmental information about and gestural coordination with the following phone—unlike released/unreleased stops, glottal stops and deletions which can be realized in clear absence of a following context, like before a pause (though of course context can modulate their likelihoods, see Randolph, 1989). The PPH predicts that the less a variant is dependent on following context, the less its variability should be correlated with difficulty of planning upcoming material. To test this secondary prediction, we focus on the glottalization, a common sentence-final realization of /t/. We investigate whether syntax and lexical frequency influence glottalization in the same way they do flapping. A glottal stop realization is more likely at larger prosodic breaks, so we expect a trade-off between glottalization and flapping as boundary strength increases. As for lexical frequency, the PPH predicts that in the case of glottalization there should be little to no effect of the following word's frequency, since glottalization does not require knowledge about the phonological make-up of the following word, and hence it does not have to be to be planned out in order to realize a /t/ as a glottal stop.

3.2 Boundary strength: production experiment

A production experiment was conducted to test the effect of phonological, prosodic and syntactic context on the realization of word-final /t/. We manipulated whether the target word was followed by a direct object within the same clause, or by the subject of the following clause, creating a large syntactic break in that the two relevant words are separated by a clause boundary. Syntactic constituency has an effect on the likelihood of two words being planned in the same planning window (cf. Wheeldon, Smith, and Apperly, 2011; Lee, Brown-Schmidt, and Watson, 2013). The presence of a clause juncture is therefore predicted to make it less likely that the upcoming word is planned together with the target word, making it less likely that the relevant phonological information about the following segment is available, and hence decrease the rate of flapping.

Similarly, a stronger prosodic boundary plausibly reduces the likelihood that an upcoming word is planned within the same window as the target word—or maybe strong boundaries are strong partly *because* the upcoming word was not planned yet. While we will not try to tease apart the directionality of this effect, we expect a correlation such that the flapping rate is expected to decrease as prosodic boundary strength increases.

To distinguish the PPH from theories relying purely on gestural overlap, acceleration or reduction, we will include temporal measures of boundary strength in our analysis. We predict an effect of syntax even after controlling for duration.

We will also examine realizations of /t/ other than flapping. One alternative outcome is glottalization, where the oral gesture of /t/ is not realized, or at least reduced to the point where an annotator no longer detects it.[5] Speakers can also completely delete a segment, an outcome that is more likely when a consonant follows (and also appears to

[5] Note that stops in North American English can also be accompanied by glottalization during the preceding vowel while maintaining the oral closure gesture. These stops involve a glottal closure as well as an oral closure. We will here only look at glottalized versions of /t/ in which the oral closure was completely perceptually absent. Flaps are generally considered incompatible with glottalization, and our data is compatible with this view.

be modulated by production planning constraints, see Tanner, Sonderegger, and Wagner, 2017). Finally, while stops in North American English that occur word finally are often unreleased, a stop can also be released, and this outcome is more likely phrase finally. A speaker therefore has many choices in how to articulate a word-final /t/, and our interest is how the first segment of an upcoming word affects the choice a speaker makes.

3.2.1 Methods

Participants Twenty-three participants were recruited from the McGill University community. All were native speakers of North American English with limited exposure to French.

Materials The materials for the production study consisted of eight sets of sentences in four conditions, varying two factors: *Phonology* (did the upcoming word start with a vowel or a consonant) and *Syntax* (was there a clause boundary before the upcoming word or not). Each item was a sentence with two clauses, as in Table 3.1, where the verb in the initial embedded clause was a nonce word that contained the target word-final /t/. The target /t/ was always preceded by a vowel, and followed by either a vowel- or consonant-initial proper noun depending on *Phonology* condition. In the following analysis, we will focus only on the items where a vowel followed, since flapping before a consonant was extremely rare.

The *Syntax* manipulation varied whether the nonce verb was followed by a clause boundary. In one condition, the following word was the object of the nonce verb (*No Clause Boundary* condition), forming a close syntactic relationship, while in the other, the following word was the subject of the main clause (*Clause Boundary* condition), creating a large syntactic break after the target word.

TABLE 3.1: A sample item from the production experiment, showing the four conditions.

Phonology	**Syntax**	
	Clause Boundary	*No Clause Boundary*
Consonant	If you **plit**, Alice will be mad.	If you **plit** Alice, John will be mad.
Vowel	If you **plit**, Penny will be mad.	If you **plit** Penny, John will be mad.

Procedure Participants saw the target sentence on a visual display, and were given time to familiarize themselves with the sentence. Then, they were asked to read the sentence aloud. Participants recorded each item in each condition once at normal speaking tempo, and once at a fast tempo (*Speech Rate* manipulation). After each trial, they were prompted to press a key when ready for the next trial. Every speaker saw each sentence from each item set. The trials were randomized into four blocks, each block consisting of an equal number of trials from each of the four conditions (*Clause Boundary/No Clause Boundary* and *Vowel/Consonant* following). Each block only contained one condition from each item set, like in a Latin-square design. Within each block, a condition was not repeated more than once, otherwise the order was completely random. The order of the blocks was also randomized between participants. This randomization scheme minimized the number of repetitions of trials from the same condition or the same item set. Each trial involving a /t/-final word was followed by a filler trial involving a similar sentence which did not involve a /t/-final word. Sample items are shown in 3.1.

Analysis The recorded utterances were force-aligned using the prosodylab-aligner (Gorman, Howell, and Wagner, 2011), and annotated manually by a research assistant. The realization of the /t/ was a choice between the following categories: released, unreleased,

glottalized, deleted or flapped (Randolph, 1989), tokens which were problematic or unclear were also noted and these were excluded from analysis (n = 40). Articulatory research has shown that instances that are transcribed as deletion often actually still involve the relevant gestures, but they are undershot and thus do not lead to acoustic results that are audible enough for an annotator (Purse and Turk, 2016). This means that our annotation will likely contain many instances of glottalization or deletion where a reduced oral gesture was still present.

Acoustic measures were extracted using Praat scripts: duration of the /t/, of the preceding vowel, and of the following segment. We used the duration of the preceding vowel as a proxy measure for the strength of the boundary separating the two words of interest, coded as the continuous variable *Vowel Duration*.

We report here only on the condition in which a vowel follows the target word and a flap-realization was possible in principle, but point out that this was only the case for half the stimuli that participants were asked to produce, so they could not fall into a strategy where they simply flap every coronal stop. Furthermore, we report only the data elicited at the fast speech tempo, since flapping was very rare at a normal tempo, and excluded participants who did not flap on any trial.[6] Finally, we excluded tokens that were followed by a pause, since none of these tokens were ever flapped, and flapping might actually be physiologically impossible in this case. This left 235 tokens for analysis.

3.2.2 Results

First, we present the overall pattern of /t/ realizations in the data. Figure 3.1 shows the rate of flapping (blue), glottalization (red), deletion (green), stops with release (black) and without release (orange). Across all the conditions, the most common realization is glottalization, i.e. realization of /t/ as a glottal stop. Focusing only on the *Clause Boundary* condition of Syntax, we see that in the *Fast Speech Rate* condition, there appears to be a

[6] Results were similar even when the non-flapping participants were included.

trade-off between glottalization and flapping depending on the strength of the prosodic boundary (as indexed by vowel duration, a proxy for final lengthening; see e.g. Wightman et al., 1992). There is more flapping at shorter vowel durations, decreasing as the vowel gets longer, and glottalization shows the opposite pattern of becoming more likely as the vowel gets longer (i.e. when the prosodic boundary is stronger).

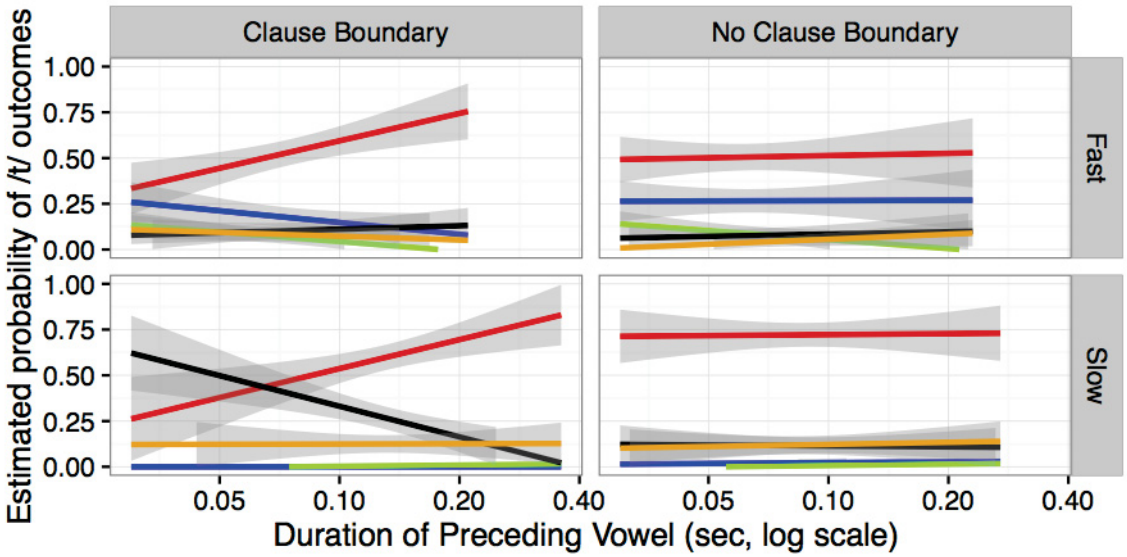

FIGURE 3.1: Empirical plots of the correlation between different /t/ realizations for the target word (flapping (blue), glottalization (red), deletion (green), stops with release (black) and without release (orange)) and the duration of the vowel preceding the word-final /t/, plotted by condition of *Syntax* and *Speaking Rate* for production experiment data.

If /t/ realization was modulated simply by the degree of temporal compression (which could require gestural reduction, acceleration, or overlap), we would expect to see the same pattern for how it is affected by vowel duration in the *No Clause Boundary* condition. It is clear, however, that vowel duration affects /t/ realization qualitatively differently depending on the syntactic condition. The rates for the different realizations appear constant in the *No Clause Boundary* condition, but seem to correlate with prosodic boundary strength (as measured by vowel duration) in the *Clause Boundary* condition. This overall interaction is predicted by the PPH, since syntax affects planning scope (Ferreira, 1991;

Lee, Brown-Schmidt, and Watson, 2013), but is not expected under the gestural overlap approach. In the following sections, we will look more closely at flapping rate and glottalization respectively, and report regression models to assess the significance of the observed patterns.

Flapping The overall rate of flapping annotated in our data was 34.47% (n = 235, again counting only participants who flapped on at least one trial). This is in contrast to the flapping rate of 93.9% found by Patterson and Connine (2001) for *word-medial* /t/ in the SWITCHBOARD corpus of conversational speech, highlighting the contrast between word-internal and cross-word applications of flapping. In another corpus study, Randolph (1989) found a more comparable rate of flapping (67%, n = 1398) for intervocalic alveolar stops, which included both word-medial and word-final stops from the TIMIT corpus of read speech.

The rate of flapping was lower when a clause boundary followed the target word with a rate of 27.83%, compared to 40.83% when no clause boundary followed. As for the effect of duration of the preceding vowel, which indexes temporal compression due to speech rate and prosodic boundary strength, empirical examination suggests that there is a negative correlation with flapping rate. This is shown in Figure 3.2.

Furthermore, as noted above, the empirical plots show that the correlation between vowel duration and flapping rate may only hold if there is a clause boundary following the /t/—the error bars in the plot in the *No Clause Boundary* condition indicate that the observed fluctuations are probably not meaningful. In other words, the rate of flapping is essentially 'flat' in this condition, and does not depend on the strength of the boundary separating the words, although there does appear to be a downtrend in flapping rate with increased vowel duration in the *Clause Boundary* condition.

To test these patterns statistically, we fitted a logistic mixed-effects model using the `glmer` function in the `lme4` package (Bates et al., 2013) package in R (R Core Team, 2013).

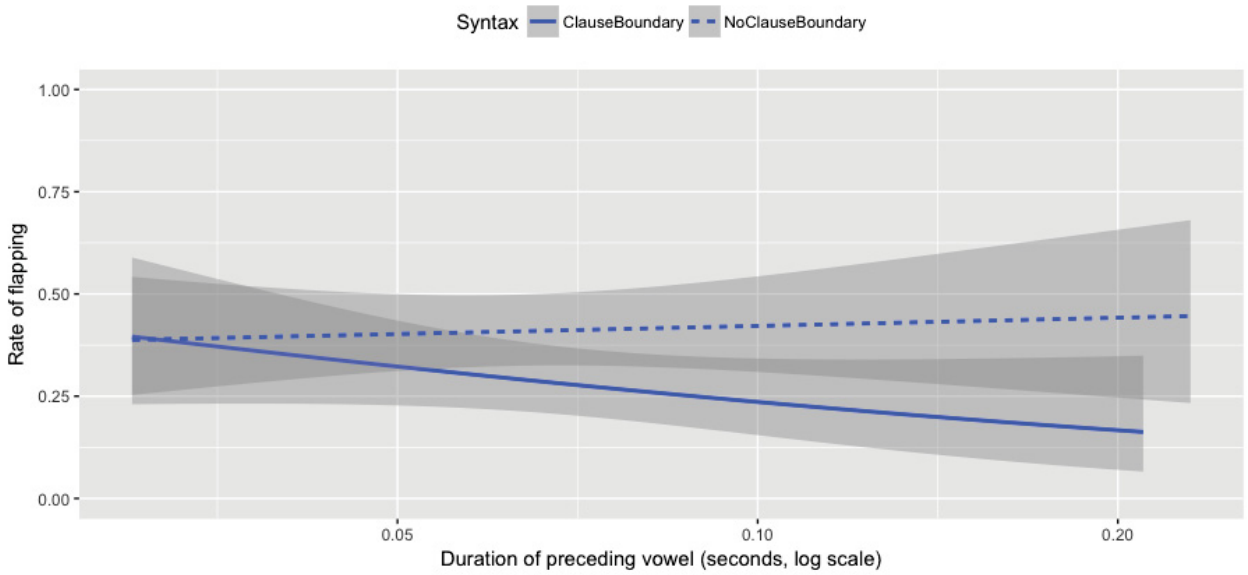

FIGURE 3.2: Empirical plots of the correlation between the rate of flapping for the target word and the duration of the vowel preceding the word-final /t/ for production experiment data.

The model included *Syntax* and *Vowel Duration* and their interaction as fixed effects. The dependent variables were standardized: both were centered around zero by subtracting the mean, and the continuous variable *Vowel Duration* was divided by two standard deviations. The model also included full random effects structure by participant and by item, which controls for possible differences in baseline flapping rates and in effect size for each variable across individuals (Barr et al., 2013). This model is reported in Table 3.2.

Syntax significantly affected flapping rate ($\hat{\beta} = 0.87$, $p = 0.03$), with flapping being about 2.6 times more likely in the *No Clause Boundary* condition. *Vowel Duration*, our proxy measure for the prosodic boundary separating the words, correlated with *Syntax*, as expected given the influence of syntax on prosodic boundary strength ($r = -0.16$, $p = 0.002$). Its effect on flapping rate was in the expected negative direction, but it was not possible to confirm an independent effect of *Vowel Duration* with a 95% confidence level ($\hat{\beta} = -0.82$, $p = 0.2$). The interaction between vowel length and syntax suggested by the empirical data

TABLE 3.2: Fixed effects for the statistical model of **flapping** in the production experiment: coefficient estimates, standard errors, z-scores, and significances (assessed using a Wald test)

Fixed effects	$\hat{\beta}$	se($\hat{\beta}$)	z	Pr(z)
Intercept	-1.037	0.557	-1.860	0.063
Syntax	0.866	0.400	2.166	0.030
Vowel Duration	-0.824	0.649	-1.269	0.204
Syntax:Vowel Duration	0.856	0.771	1.111	0.267

was not significant ($\hat{\beta}$ = 0.86, p = 0.27). In other words, we cannot conclude from this data that there is a reliable difference between the effect of *Vowel Duration* in *Clause Boundary* vs *No Clause Boundary* conditions. This may be a statistical power issue, due to the low number of tokens realized as flap. We will test this interaction for another, more common realization of /t/, namely glottalization.

Glottalization The tokens that were realized as glottal stops show a similar pattern to flapping, but in the opposite direction: The rate of glottalization has a positive correlation with *Vowel Duration*, suggesting that glottal stop realizations of /t/ are more likely at stronger junctures. This relationship between *Vowel Duration* and glottalization seems to be absent in the *No Clause Boundary* condition, where glottalization applies at a constant rate. This apparent interaction is similar to the empirical trend we observed for flapping. Since glottalization is the most common outcome in this experiment, and occurs in both conditions of *Speech Rate*, there are a higher number of tokens which allows us to estimate the effects of the predictors more reliably. Since glottalization does not require the upcoming word to be planned, the PPH does not predict that syntactic clause boundaries should have an inhibitory effect on this outcome, though a higher rate of glottalization in the *Clause Boundary* condition would be compatible with the idea that glottalization marks prosodic boundaries.

The overall rate of glottalization in the data was 58.91% (n = 735). We fitted a mixed-effects logistic regression to this data similar to the one for flapping with *Syntax* and *Vowel Duration* as predictors, but also including a predictor for *Speech Rate*. All two-way interactions and the three-way interactions were also included in the model, both as predictors and in the random effect structure. Factors were again standardized.

TABLE 3.3: Fixed effects for statistical model of **glottalization** in the production experiment: coefficient estimates, standard errors, z-scores, and significances (assessed using a Wald test)

Fixed effects	$\hat{\beta}$	se($\hat{\beta}$)	z	Pr(z)
Intercept	0.375	0.368	1.018	0.309
Syntax	0.445	0.312	1.423	0.155
Vowel Duration	0.903	0.372	2.424	0.015
Speech Rate	-0.285	0.253	-1.127	0.260
Syntax:Vowel Duration	-1.078	0.511	-2.111	0.035
Syntax:Speech Rate	-0.810	0.236	-3.433	0.001
Vowel Duration:Speech Rate	-0.058	0.306	-0.190	0.849
Syntax:Vowel Duration:Speech Rate	-0.281	0.451	-0.622	0.534

The model does not show a significant effect of *Syntax* ($\hat{\beta}$ = 0.44, p = 0.15), in contrast to flapping. There was a significant effect of *Vowel Duration* ($\hat{\beta}$ = 0.9, p = 0.02), with longer vowels correlating with more glottalization, reversing the pattern we observed in the case of flapping. Glottalization does not depend on phonological information about the upcoming word, and might therefore simply be a cue to the strength of the upcoming prosodic boundary, with greater lengthening correlating with a stronger boundary and hence a greater likelihood of glottalization.

There was also an effect of speech rate, with more glottalization occurring at the *Slow Speech Rate* ($\hat{\beta} = -0.28$, p = 0.26).

The interaction between *Syntax* and *Speech Rate* is also statistically significant, suggesting that the rate of glottalization is attenuated in the *Slow*, *Clause Boundary* condition compared to what would be predicted by the main effects alone ($\hat{\beta} = -0.81$, $p < 0.01$).

Finally, the interaction between *Syntax* and *Vowel Duration* was significant in this model ($\hat{\beta} = -1.08$, $p = 0.035$), showing that the effect of prosodic boundary strength was indeed modulated by *Syntax*.

3.2.3 Discussion

These results show that the presence of a clause boundary after a word-final /t/ has a significant influence on its likelihood of being flapped. Since the model controls for *Vowel Duration* (i.e. final lengthening) as an independent, continuous measure of prosodic boundary strength, we conclude that the effect of the syntactic manipulation is not completely reducible to durational effects associated with clause boundaries, and supports the existence of production planning effects. Previous studies have shown that greater syntactic complexity can delay production latencies, suggesting an increased planning load for more complex upcoming constituents (Ferreira, 1991; Krivokapić and Byrd, 2012). Hence, an upcoming noun phrase (e.g. *Alice* in *plit Alice*) should incur less of a processing load if it is the final noun in the embedded clause, where it would be a small one-word constituent, rather than the first noun in a completely new clause, which would initiate planning of the entire new sentence.

Interestingly, the effect of *Syntax* is not categorical—a clause boundary does not completely block flapping, but rather decreases the probability in a gradient way. For example, holding *Vowel Duration* at its mean value, the probability of flapping is estimated to be 19% if a clause boundary follows, but increases to 35% if no clause boundary follows. An account purely in terms of gestural reduction would predict that our durational measures should account for the variability, and that syntax should not affect the flapping rate above and beyond durational effects. The gradient effect of *Syntax* is compatible with the PPH view that whether or not an upcoming word forms part of the same clause influences both the prosodic realization of the boundary and concomitantly how early or late the upcoming word is retrieved relative to the word containing the word-final /t/.

Empirically, glottalization generally showed the mirror image pattern of that of flapping. This may not be surprising: the two are not independent variables, since a glottalized /t/ cannot be a flap and vice versa. When analyzing the glottalization rates we found an interaction between *Syntax* and *Vowel Duration*, which suggests that prosody has a different effect on /t/ realization depending on whether there is a clause boundary. This is compatible with the PPH and unpredicted by an account in terms of gestural overlap alone. The significance of this interaction is another piece of evidence that the realization of /t/ is not a pure function of the degree of articulatory closeness of the two words. We suggest that the reason there is apparently little or no effect of vowel duration in the *No Clause Boundary* condition is that the rate of planning the two words together is basically at ceiling. The verb-direct object unit could be a minimal planning unit (at least under these particular experimental conditions), preventing any modulation from prosodic distance effects. However, there is some evidence in the literature that sometimes the subject and verb are planned to the exclusion of the complement of the verb (Lindsley, 1975). It would be interesting in a follow up study to manipulate the difficulty of planning ahead, and see whether in a slightly different task we find a correlation between flapping rate and prosody also in the *No Clause Boundary* case.[7] On the other hand, in the *Clause Boundary* condition the two words are a priori at a lower probability of being planned together, so that probability can also be influenced by factors such as prosodic boundary strength, as measured here by *Vowel Duration*.

Our glottalization results also bear on hypotheses about the origin of glottalization patterns in English. Eddington and Channer (2010) argue that glottalization of word-final /t/ originally happens when a consonant-initial word follows. They show evidence that words are generally more often followed by consonant-initial words, which they take to be the true glottalization environment for word-final /t/, and argue that glottalization in

[7] See Kilbourn-Ceron and Sonderegger (2017) for related findings about ceiling effects when looking at speech rate and lexical frequency in the absence of the temporal slow down induced by prosodic boundaries.

those cases where a vowel follows involves reuse of an exemplar of a previous occurrence of that word where a consonant followed. The fact that speakers in our experiment use the glottal stop in both environments even with nonce words casts doubt on this explanation: Glottalization of /t/ is an option even in cases when speakers are not familiar with the word, and hence cannot draw on a previous, glottalized realization when it was followed by a consonant-initial word. Since each speaker pronounced each nonce word 4 times (for the 4 conditions in the experiment), we can check if it mattered whether a speaker had produced the nonce word before. The rate of glottalization when a vowel followed was overall 58.9%; the rate on the first use of a word was 57%; the rate on the last use was 61%. This upward trend might be evidence that speakers are more likely to glottalize in words they are familiar with, and it could be that a prior glottalized realization makes glottalization in future productions more likely. The high rate of glottalization in vowel-following contexts on first occurrences clearly shows that word-final glottalization cannot generally be 'imported' from prior uses of the word where a consonant followed.

The production data show evidence for an effect of clause boundaries even after controlling for temporal modulation, as predicted by the PPH since syntax modulates planning scope. The PPH makes the prediction that *any* factor which affects or modulates planning will interact with the application of external sandhi processes. In the following, we turn to frequency effects as observed in a corpus of conversational speech to test further predictions.

3.3 Flapping and lexical frequency

Word frequency interacts with production planning: retrieving and planning frequent words takes less time (Oldfield and Wingfield, 1964; Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994). With respect to our hypothesis, this means that we expect that a highly frequent following word should be planned earlier relative to the target word, and

sandhi should be more likely to apply. Crucially, this effect should be present even after controlling for durational measures. The predictions of the PPH for the frequency of the word containing the final /t/ are less straightforward. Depending on the level at which frequency effects apply (at the lemma, or at the level of phonological form), one might expect a higher or a lower rate of the sandhi process.[8] Gregory et al. (1999) investigated cross-word flapping in a corpus, and found an effect of the mutual information between two words, but no effect of the frequency of the first word containing the word-final stop. This finding is consistent with the PPH, which predicts an effect of the predictability of the following word on flapping rate. Furthermore, the PPH predicts that predictability-related factors should have effects above and beyond their modulation of duration, which we control for in the present study. The PPH furthermore predicts that in the case of glottalization, the frequency of the following word should not matter because it is not necessary for glottalization. We tested these predictions with data from a corpus of spontaneous North American English.

3.3.1 Data set

The data source for this study was the Buckeye Corpus of conversational speech (Pitt, Dilley, Johnson, Kiesling, Raymond, Hume, and Fosler-Lussier, 2007). The corpus includes word- and phone-level time-aligned annotations, which were prepared with automatic phonetic transcription and subsequently hand-corrected by phonetically trained research assistants (Kiesling, Dilley, and Raymond, 2006). Transcribers were instructed to label all /t,d/ phones which show glottalization with the label [tq], a glottal stop. Flaps were identified by transcribers based on listening cues and spectrogram inspection, with voicing throughout the closure, and labelled as [dx].

[8] See Tanner, Sonderegger, and Wagner (2017) for an in-depth discussion of potential effects of the target word's lexical frequency on planning two word sequences.

Using the Montreal Corpus Tools software, we extracted[9] 11863 tokens of words which end in a vowel followed by /t/ or /d/ and were followed by a vowel-initial word (46.24% were transcribed as flaps). Of these, we excluded tokens where the following word was a disfluency marker[10] (18.26% of tokens), and where the following word was reduced to a syllabic consonant on the surface (0.07% of tokens). This left 9208 tokens for analysis.

Word frequencies were retrieved from SUBTLEX-US, a database of word frequencies based on film and television subtitles (Brysbaert and New, 2009). Our temporal measure was observed/expected word duration, where expected duration was the mean duration for that word in the entire Buckeye corpus. This measure was meant to control for temporal compression due to either speech rate, or boundary-induced final lengthening (Wightman et al., 1992), since we did not have syntactic boundary information for this corpus. This measure relates to the predictions of gestural overlap, since a greater O/E ratio would reflect less temporal overlap between the adjacent segment gestures. Number of syllables was calculated for both the token and following word, with each syllabic segment in the Buckeye surface transcription counting as one syllable.

3.3.2 Results

Flapping We first illustrate the distribution of flapping in terms of *Observed/Expected (O/E) Duration* in Fig. 3.3. The flapping rate inversely correlates with the (normalized) duration of the target, as expected if flapping is more likely across weaker prosodic boundaries, or if flapping is a consequence of gestures being temporally compressed and overlapped.

Fig. 3.4 illustrates that in the case of both /d/ and /t/, the rate of flapping correlates with the frequency of the following word.

[9] We gratefully acknowledge the assistance of Michael McAuliffe in extracting these data.

[10] These words were 'uh','um','okay','yes','yeah','oh','heh','yknow','um-huh','uh-uh','uh-huh','uh-hum','mm-hmm', and 'and', all of which were associated with flapping rates well below the mean by word-type.

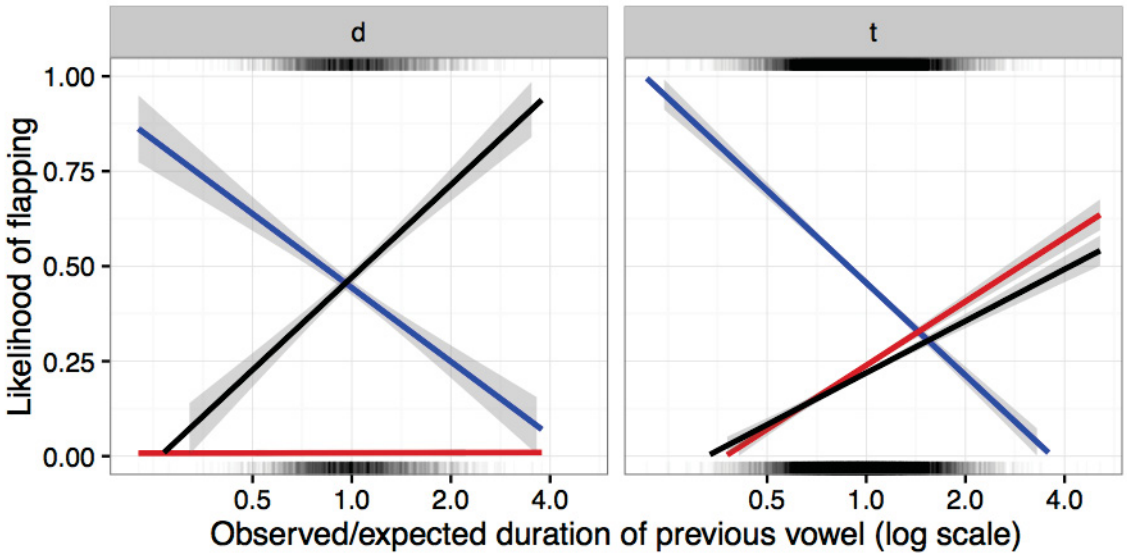

FIGURE 3.3: Relationship between Observed/Expected word duration and rate of flapping (blue), glottalization (red) and alveolar closure (black) in the Buckeye corpus.

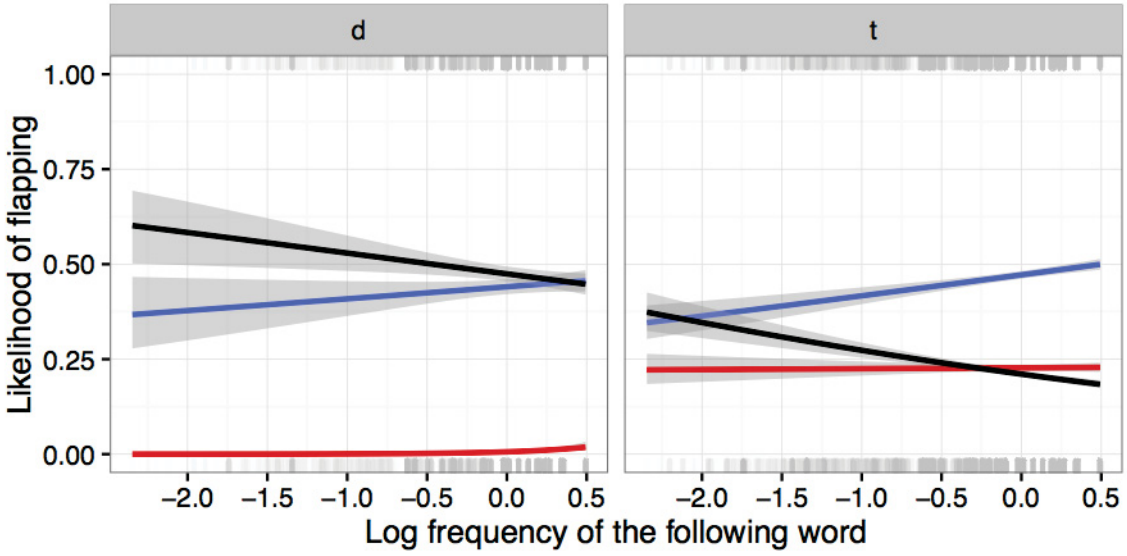

FIGURE 3.4: Relationship between Following Word Frequency and rate of flapping (blue), glottalization (red) and alveolar closure (black) in the Buckeye corpus.

TABLE 3.4: Fixed effects for the statistical model of **flapping** in the Buckeye corpus: coefficient estimates, standard errors, z-values, and significances (assessed using a Wald test)

Fixed effects	$\hat{\beta}$	se($\hat{\beta}$)	z	Pr(z)
Intercept	0.654	0.115	5.681	<0.001
Underlying /t,d/	0.338	0.117	2.888	0.004
Target Word Frequency	0.213	0.124	1.712	0.087
Following Word Frequency	0.290	0.096	3.010	0.003
Observed/Expected Duration	0.047	0.133	0.356	0.722
Pause	-4.796	0.239	-20.042	<0.001
Interactions				
# of Syllables: Target Word	-0.182	0.142	-1.284	0.199
# of Syllables: Following Word	-0.021	0.091	-0.231	0.818
Target:Following Word Freq	0.228	0.136	1.677	0.094

Given our hypothesis, we were particularly interested in whether there is an effect of frequency after controlling for our prosodic lengthening proxy measure, in other words, whether after controlling for *O/E Duration*, we still see that the flapping rate goes up if the following word is easier to plan. In order to test this, we again analyzed the data using logistic mixed-effects models. The log-transformed lexical frequency of the token word and the following word were standardized and included as fixed effects. Control predictors included presence of *Pause*, a binary variable, underlying voicing of the word-final segment (*Underlying /t,d/*), *O/E Duration*, log-transformed and standardized, and binary variables tracking whether the token and following words were monosyllabic or not. Random effect structure included by-speaker and by-word intercepts, and by-speaker and by-word slopes for following word frequency. We fit a model with the dependent measure of whether or not the underlyingly /t,d/-final word was annotated as a flap in the surface transcription. Table 3.4 shows the model estimates for the fixed effects coefficients. Each coefficient represents the estimated change in log-odds of flapping when other predictors are held at their mean observed values, except *Pause* which is held at 0 (no pause).

The model finds a reliable difference between flapping rates for *Underlying /t,d/* once the effects of other variables are taken into account, with /t/-final words more likely to be flapped ($\hat{\beta}$ = 0.34, p = 0.004). The estimate for *Pause* is negative and of very large magnitude compared to other effects ($\hat{\beta} = -4.8$, $p < 0.001$), confirming that flapping in the presence of a pause is very rare (just under 1% of tokens followed by pause in the subset under analysis are annotated as flaps). The effect of *O/E Duration* was not statistically significant ($\hat{\beta}$ = 0.05, p =0.722). Nor did the number of syllables in the target or following word have a statistically significant effect.

As for our crucial variable, the model confirms that the lexical frequency of the second, vowel-initial word in the sandhi pair has a reliable effect on the likelihood of flapping. Higher frequency vowel-initial words are more likely to trigger flapping on a preceding coronal stop ($\hat{\beta}$ = 0.29, p =0.003). The frequency of the coronal-final word itself showed a positive trend, but the effect was not statistically reliable in the full model ($\hat{\beta}$ = 0.21, p =0.087). There was also a positive interaction between these predictors: Increasing the frequency of both words in the sandhi pair increases the likelihood of flapping even more than would be expected from the sum of the independent effects of each word's frequency ($\hat{\beta}$ = 0.23, p =0.094), though again the interaction was not significant at a $p < 0.05$ level in the model with full random effect structure.

Glottalization We also examined the pattern of glottalization for /t/ in the Buckeye corpus. Although glottal stop is an alternative realization of a /t/ in a flapping environment, the conditions on glottalization are qualitatively different, since a /t/ can easily be realized as a glottal stop regardless of which type of segment follows, and even if the word occurs at the end of an utterance. Therefore, according to the PPH, there should not necessarily be an effect of the following word's frequency on glottalization, since the glottalized realization does not depend on the following environment. Given the results of the production experiment, it seemed that glottalization is a cue for boundary strength,

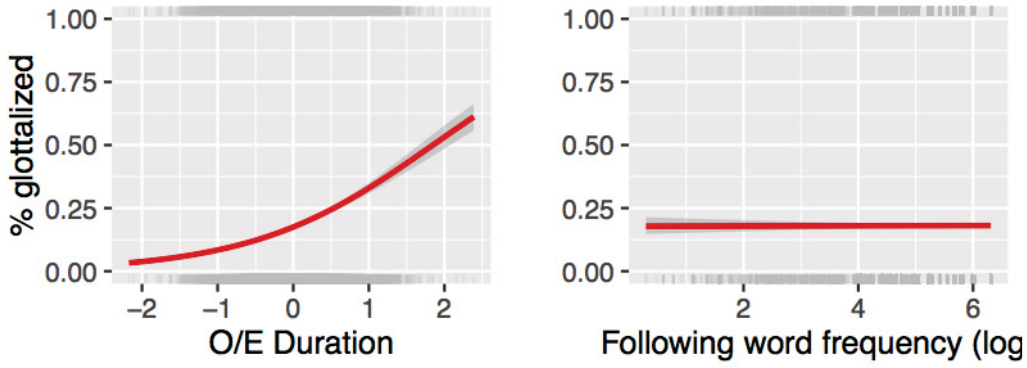

FIGURE 3.5: Relationship between rate of /t/-glottalization and *O/E Duration* (left) and *Following word frequency* (right)

with glottalization being more likely before stronger boundaries, so we expect a positive effect of *O/E Duration* similar to the *Vowel Duration* effect in the production experiment.

The overall rate of glottalization in the corpus was 18.31% (n = 11863), slightly lower than the 24% (n = 1101) reported by Eddington and Channer (2010) using data from another corpus of spontaneous speech. The rate of glottalization before a pause is much higher than when no pause follows, 43.8% versus 12.1% (n = 2301 with pause, 9562 without pause). This is not surprising, since we have already seen that flapping occurs almost exclusively in the absence of pauses, probably for articulatory reasons. Only when there is no pause does flapping compete with glottalization as one of the possible realizations of /t/. The rate of glottalization, shown in Figure 3.5, seems to show a correlation with our measure of final lengthening, *O/E Duration*, suggesting that stronger boundaries are indeed associated with higher rates of glottalization. As for the effect of the following word's frequency, Figure 3.5 shows no strong correlation with the rate of glottalization, suggesting that the realization of /t/ as a glottal stop is not particularly sensitive to the planning of upcoming information.

The data for glottalization was analyzed with a logistic mixed effects model similar

TABLE 3.5: Fixed effects for the statistical model of **glottalization** in the Buckeye corpus: coefficient estimates, standard errors, z-values, and significances (assessed using a Wald test)

Fixed effects	$\hat{\beta}$	se($\hat{\beta}$)	z	Pr(z)
Intercept	-1.591	0.151	-10.540	<0.001
Target Word Frequency	-0.395	0.173	-2.282	0.022
Following Word Frequency	-0.008	0.143	-0.053	0.958
Observed/Expected Duration	0.455	0.131	3.461	0.001
# of Syllables: Target Word	0.145	0.181	0.801	0.423
# of Syllables: Following Word	-0.003	0.104	-0.026	0.980
Pause	0.656	0.028	23.820	<0.001
Interactions				
Target Word Freq:Following Word Freq	-0.147	0.204	-0.720	0.471
Target Word Freq:O/E Duration	-0.117	0.194	-0.606	0.544
Following Word Freq:O/E Duration	0.002	0.143	0.011	0.991
Target Word Freq:Following Word Freq: O/E Duration	0.079	0.340	0.232	0.817

to the one presented for flapping. The fixed effects included target word frequency, following word frequency, and *O/E Duration*. *Pause* was included as a categorical variable (pause or no pause), and, as controls, whether the target and following words were mono- or polysyllabic. We also included as controls the interactions between the two word frequency measures and O/E Duration, including their three-way interaction. All duration and frequency measures were log transformed and all variables were standardized. The results of this analysis are shown in Table 3.5.

There was a significant effect of *O/E Duration* ($\hat{\beta}$ = 0.46, $p = 0.001$), with a higher rate of glottalization after words that were relatively long, replicating the effect of *Vowel Duration* observed in the production experiment. There was also a significant effect of *Pause* ($\hat{\beta}$ = 0.66, p <0.001), with a higher rate of glottalization before pauses. Both results point to glottalization being a cue for boundary strength.

There was no effect of the frequency of the following word ($\hat{\beta} = -0.01$, $p = 0.958$). Since

glottalization does not necessarily depend on phones in the upcoming word, the PPH does not predict such an effect. There was, however, an effect of word frequency for the word containing the /t/ ($\hat{\beta} = -0.4$, $p = 0.022$), with more frequent words being less likely to end with a glottalized /t/. None of the interactions reached statistical significance.

3.3.3 Discussion

The results of the corpus study show that there is a strong correlation between the *following word frequency* and the likelihood of flapping. This is consistent with the prediction of the PPH that flapping should be more likely when the following word is easier to plan. Lexical frequency is known to have a facilitatory effect on word form retrieval (Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994). This has the consequence that phonological encoding of the following vowel-initial word may begin sooner for more frequent words, thus making the vowel more likely to be available to trigger flapping on the target coronal-final word, according to the PPH.[11]

The measure of duration compression that we included in this analysis, *O/E Duration*, looked to be negatively correlated with flapping when we examined the empirical trends. However, the effect was not statistically significant in the model, once other factors like the presence/absence of a pause were controlled for. This is slightly surprising, since syntactic and prosodic boundaries are correlated with longer durations, as we saw in the previous study, and we did not separately control for syntax in the corpus study. One reason we may not have detected an effect of lengthening in this analysis is that our measure of temporal compression was over the entire word instead of targeting the final syllable, which is the main locus of boundary-associated lengthening (Wightman et al., 1992; Keating and Shattuck-Hufnagel, 2002).

[11] We believe that a potentially better measure of the predictability of an upcoming word would be an estimate of its conditional probability given the first word, but we have not yet explored this.

In contrast to flapping, the frequency of the following word did not have a detectable effect on the glottalization rate. The absence of an effect of the frequency of the following word for glottalization is compatible with PPH predictions, since glottalization as a process does not depend on the phonological form of the upcoming word. Glottalization seems to be used as a cue for strong boundaries, with a greater rate of glottalization at stronger boundaries (as measured by *O/E Duration*) and before pauses.

3.4 General discussion

The results of the two studies show that both spontaneous and lab-elicited speech exhibit patterns of variability compatible with the predictions of the locality of production planning hypothesis.

The results of the production experiment show that syntactic boundaries have a gradient blocking effect on flapping likelihood, and that this effect does not appear to be entirely due to the temporal slow-down effects at the clause boundary, since it was significant when also controlling for pre-boundary lengthening.

As discussed, syntactic locality effects and variability in general could receive an alternative explanation if temporal compression automatically leads to gestural reduction, acceleration, or overlap. Edges of larger syntactic constituents are known to be associated with longer duration (Wightman et al., 1992), as is lower lexical frequency (Jurafsky et al., 2001). Both factors could interfere with the gestures involved in realizing [t]. However, the results of our production experiment suggest that the inhibitory effect of a clause boundary goes above and beyond that expected based on durational effects. Final lengthening, measured as the duration of the vowel preceding the target /t/, did not have a statistically significant effect in a model that also included syntax.

These results show that accounts of flapping in terms of temporal compression of the

gestures involved will not be entirely sufficient to explain the variability and locality effects in the data. Which is not to say that gestural overlap does not play a role in explaining some variation, but accounting for non-temporally based variability requires additional explanation, which the PPH provides.

The corpus study revealed a positive effect of the following word's frequency on flapping, after controlling for potentially correlating temporal effects. This confirms another prediction of the PPH: Words that are planned more quickly are more likely to influence the phonological encoding of the word that precedes them, which in this case means triggering flapping.

In our complementary analysis of glottalization, we found a different pattern. The crucial phonological difference between the two realizations is that glottalization, although it may be influenced by a following segment, does not *require* a particular following segment the way that flapping requires a following vowel. We investigated the same variables for glottalization as we did for flapping, and found three major results. First, in the production experiment there was a statistically significant interaction between the presence of a clause boundary and the effect of vowel duration. Qualitatively, the rate of glottalization was not greatly affected by the duration of the vowel in the condition where there is no clause boundary, which makes sense under the PPH as there is high cohesion between the target and following words. On the other hand, when the target word was followed by a clause boundary, the duration of the vowel was positively correlated with the rate of glottalization, mirroring the *negative* trend found in the analysis of flapping. Our interpretation of this interaction is that the target verb and following object consistently form a planning unit when they are not separated by a clause boundary, and the /t/ realizations in this condition reflect the rate of variation under "ideal" production planning conditions. In the condition that included a clause boundary on the other hand, we expect the difficulty of planning an upcoming clause to lead to higher variability in whether the target and following word form a planning unit. We expect this in turn to

be correlated with (and perhaps influenced by) degree of final lengthening, possibly as an indicator of prosodic boundary strength. Then, when the two words do not form a unit (more lengthening), we expect higher rates of glottalization and when they do form a unit (less lengthening) we expect higher rates of flapping.

In our analysis of glottalization in the Buckeye corpus, we found that our word-normalized measure of duration was again positively correlated with the probability of realizing a /t/ as a glottal stop. Interpreting this measure as a cue to boundary strength as before, this dovetails with the other factor that significantly increased glottalization rates, namely the presence of a pause. On the other hand, unlike in the case of flapping, we were not able to detect any effect of the following word's frequency. We interpret this result in light of the PPH as being consistent with the idea that variants which do not require information about the following segment should not be influenced by how difficult the upcoming word is to plan.

There are two other theoretical perspectives on external sandhi that we have not addressed so far. The first is the account of external sandhi in Prosodic Phonology, and the second is the probabilistic reduction approach, which has gained importance in recent years. We next discuss how our results bear on these approaches in turn.

Prosodic Phonology

Prosodic Phonology (e.g. Nespor and Vogel, 1986; Selkirk, 1986) captures locality conditions by positing phonological domains of particular types (phonological word, phonological phrase, intonational phrase, phonological utterance), which are organized into a Prosodic Hierarchy. This model was proposed as a way to enrich earlier models of phonology that only made reference to word-prosodic structure, such as Chomsky and Halle (1968), which were found wanting because they seemed unable to capture locality generalizations about sandhi processes.

In this model, phonological processes are tied to a particular domain on this hierarchy. Phonological domains are mapped from syntactic structure (e.g., clauses map to intonational phrases), though eurythmic principles can override syntactic mappings. Nespor and Vogel (1986), for example, argue that flapping can occur throughout the entire domain of a phonological utterance, the evidence being that flapping can occur even across sentence boundaries. However, processes like flapping pose a problem for this model in that they do not seem to have an absolute syntactic upper bound that makes application impossible. It seems that flapping can in principle apply across any two words, even when separated by clause junctures, but is increasingly less likely to apply across bigger boundaries (Scott and Cutler, 1984; Fukaya and Byrd, 2005). It also often does not apply even between words that are part of the same syntactic phrase, and should therefore form part of a much smaller phonological domain than the 'phonological utterance' that Nespor & Vogel propose as the domain for this process. This pattern of variability seems to contradict the kind of distributional pattern that the prosodic hierarchy aims to explain, namely that syntactic locality effects are due to segmental processes being bounded by prosodic domains. Nespor and Vogel (1986) attribute this type of unexplained variation to variability in prosodic phrasing. In their analysis, a 'phonological utterance' (the domain of flapping) is frequently restructured into several such phonological constituents, blocking flapping from applying. However, this raises the question whether it is necessary to tie external sandhi to prosodic domains in the first place (see Scheer, 2012, for a critical review of Prosodic Phonology).

The PPH account of syntactic and prosodic influences on flapping is compatible with a statement of the flapping process without any reference to a prosodic domain: The gradient effect of syntactic boundaries, the greater variability when applying between words, and the correlation between juncture size and application rate can in principle all be a direct consequence of the locality of production planning. The fact that it is possible to capture insights into external sandhi segmental processes without explicitly tying them

to phonological domains and without a rich representational inventory of hierarchical phonological structure does not necessarily refute this account. If the factors that affect planning scope affect the restructuring posited by Nespor and Vogel (1986), then our results may still be compatible with this theory. However, in the absence of independent evidence that all utterances in which flapping fails to occur are prosodically restructured into multiple phonological 'utterances', an interpretation purely in terms of the PPH seems more parsimonious.

At least in the case of flapping, the data seems compatible with a model in which phonological processes cannot 'see' syntactic or higher level prosodic structure at all, and are sensitive only to very local phonological information (e.g., *Is a vowel/sonorant following?*), not unlike the model proposed in Chomsky and Halle (1968). For example, see Scheer (2012) for a current model of phonology, in which phonology can only make reference to segmental and word-prosodic information. The PPH could complement this type of model, with the apparent effects of syntax and prosody on variability mediated by the locality of planning rather than being directly encoded into the process.

Of course, this does not mean that all locality effects can or should be accounted for in this way, since it is certainly possible that phonological processes are tied to certain phonological domains. For example, the evidence that there are phonological processes tied to prosodic words seems quite compelling, since in the case of prosodic words various phonological criteria converge on a particular phonological category playing a role. One type of evidence that could support the prosodic hierarchy account would be to show that independent cues to phonological junctures between words, for example tonal events such as boundary tones, can predict the presence/absence of flapping. Since we have not looked at intonational cues to phrasing, we cannot be sure that there is not a convergence of evidence motivating a categorical phonological 'flapping' domain.

The PPH predicts that different processes might be sensitive to different planning window sizes, while the Prosodic Hierarchy theory predicts convergence of locality domains

with a fixed hierarchy of phonological domains. In order to test this divergence in predictions, one could look at different processes with different locality restrictions, as well as tonal evidence for phrasing. While both theories try to capture that certain sandhi processes only apply when the two words involved are 'close enough', they differ in how they do so. The Prosodic Hierarchy theory 'earmarks' certain processes to apply in certain domains, without giving an explanation as to why that process should show its particular locality pattern. According to the PPH, on the other hand, the locality of a phonological process is predictable from the availability of the the type of information needed to apply a phonological process. At least for some sandhi processes, such as liaison, it has been shown that intonational criteria do not correlate with the domain of liaison (Post, 2000; Pak and Friesner, 2006)—a paradox for the prosodic hierarchy theory, but not unexpected by the PPH.[12]

So while the PPH does not predict a perfect convergence of the locality of different processes and tonal criteria for phrasing on a small number of phonological phrasing categories, it does predict that we should be able to make predictions about locality and variability by looking at the phonetic and phonological substance of a process. French liaison has been reported to show a different pattern of locality compared to flapping, and so has been argued to apply within a 'phonological phrase' (Selkirk, 1986), a unit smaller than the domain suggested for flapping. But this might not be an accident: French liaison is quite different from flapping. For example it seems to be less susceptible to speech rate effects than flapping (Kaisse, 1985). Under the PPH, we can tie this difference to the fact that flapping necessarily requires planning the articulatory gestures of the following vowel at the same time, since the flap has to be planned to be released into the vowel;

[12] A similar dissociation between sandhi domains and prosodic domains has been shown for Taiwanese tone sandhi in (Chen, 1987). Tone sandhi in Taiwanese applies as long as a word is not final within a certain type of syntactic domain. Crucially, it does not require information about the phonological content of the upcoming word. As noted in Wagner (2012), the PPH correctly predicts it to be less local than sandhi processes that rely on tonal information about the upcoming word, such as tone 3 sandhi in Mandarin (Chen, 2000).

liaison, however, only requires knowledge *that* a vowel follows, but does not require coordinating the gestures involved. The articulation of the liaison consonant is not necessarily coarticulated with the following vowel (Côté, 2011), and this might account for why it is less dependent on speech rate.

Many languages show sandhi phenomena that are sensitive to whether or not the *previous* word ends in a vowel. A well-known example is the spirantization of voiced stops in Spanish (Hualde, 2013), which has been described as occurring across word boundaries without regard to syntactic or phonological junctures.[13] This asymmetry is compatible with the PPH, which predicts that processes sensitive to preceding phonological information do not have to show locality and variability effects when applying across word boundaries—the previous word, after all, has already been planned out at the time of planning the current word. Exploring such asymmetries between processes that are sensitive to upcoming and preceding information will be of interest in order to explore and further refine the PPH, and in comparing it with the prosodic hierarchy account.

In sum, we believe that the PPH offers the potential of a deeper explanation for the empirical patterns observed the in locality and variability of external sandhi phenomena. We turn now to discussion of another potential interpretation of our results related to predictability.

Flapping as a means of probabilistic reduction?

Flapping can be viewed as a form of reduction, since flaps are much shorter than fully articulated oral stops (Fukaya and Byrd, 2005, i.a.). Over the past two decades, many studies have shown that highly predictable information tends to be reduced. Jurafsky et al. (2001) found that frequent words tend to be shorter and more prone to final t/d deletion, as well as words that are highly predictable given a following word (in terms of

[13]Similar processes that could be explored are consonant mutation in Corsican and /v/ alternation in Belarussian (Scheer, 2012).

bigram frequency and other probabilistic measures). Pluymaekers, Ernestus, and Baayen (2005) showed that for seven high frequency words in Dutch, mutual information with the following word was predictive of reduction, with fewer segments realized when mutual information was high; Torreira and Ernestus (2009) found an effect of bigram frequency with the following word on the acoustic realization of /t/; Ernestus et al. (2006) showed that a sandhi phenomenon in Dutch, voice assimilation, is more likely to occur within a compound when the two component words have a high co-occurrence frequency.

These patterns fit into a framework in which reduction is used by speakers in a 'rational' way to achieve communicative goals, such that reduction is observed when there is low information in the signal, i.e, highly predictable information, and reduction is rare when information is not predictable (i.a. Aylett and Turk, 2004; Jaeger, 2010; Turk, 2010). Under this approach, the motivation for reduction phenomena can be seen in its effect on the listener side: Less reduction takes place where the listener has to work harder to retrieve the information. These theories have some commonalities with Lindblom's Hypo- & Hyperarticulation theory (Lindblom, 1990; Lindblom, 1995), where in addition to production-oriented constraints, speakers manipulate the acoustic properties of their utterances in order to facilitate message decoding for the listener. Hall et al. (2016) similarly propose a model of phonology which takes phonological processes to be a way to manage predictability within an utterance. They argue that phonological lenition and fortition processes can be rationalized as ways to optimize message transmission, with higher deletion/reduction rate occurring where there is greater redundancy in the signal.

Even if one assumes the general idea that language is optimized for communication, and that phonological patterns can be rationalized as part of this optimization process, there are many ways in which this might be actuated in phonology. For example, Cohen Priva (2015) argues that coronal stops are more likely to delete because they are, on average, more predictable within the lexicon. Gregory et al. (1999) and Jurafsky et al. (2001) find that word-final deletion rate of coronal stops depends on the frequency of the word

containing the stop and on the mutual information between it and the following word, thus deletion in this case relates to contextual predictability factors within an utterance, rather than the static distribution of the sound in the lexicon. The two types of effects are compatible with each other and might very well coexist, they just rely on different types of mechanisms. Different hypotheses within this general framework therefore differ with respect to the mechanism they consider responsible for predictability effects, and the level at which they calculate predictability.

Since flapping is arguably a reductive process, both our syntactic effect and our frequency effect can in principle be rationalized in terms of such models. Syntactic complements are plausibly generally more predictable than the first words of an upcoming sentence, and upcoming frequent words are on average more expected than comparable words with lower frequency. The results are therefore in line with the more general finding that predictability within an utterance correlates with reduction, including the more specific claim that prosody in general and prosodic phrasing in particular is sensitive to predictability and information density (Aylett and Turk, 2004; Turk, 2010).

The PPH can be seen as a potential mechanism responsible for these effects: Reduction in terms of flapping is more likely when a following word is very predictable, since it is planned faster relative to the previous word, and hence the application for the process becomes more likely. But other approaches, including approaches that view reduction as a way to manage predictability for the listener could also rationalize this effect without reference to production planning, so we cannot conclude based on the present study alone that the observed frequency effects are indeed due to the mechanism assumed by the PPH.

We point out, however, that the PPH is not a hypothesis specifically about reductive processes. It actually makes the same predictions for phonological processes that span word boundaries that are *non-reductive*, where probabilistic reduction accounts either make no predictions or opposite predictions. In that regard, the PPH is compatible with the view that sandhi processes serve the function of encoding prosodic structure, rather

than being primarily a tool to reduce the content of predictable information (Kingston, 2008; Katz, 2016). Glottalization can be seen as a cue to the end of a prosodic domain, flapping and liaison as a cue that the domain continues.

The PPH predicts an external sandhi process in which a segment is inserted rather than lenited, e.g., liaison in French, should be affected in similar ways by factors associated with planning scope. The realization of liaison consonants, which depends on an upcoming word starting with a vowel, should increase with a greater predictability of an upcoming word. This prediction will be tested in Chapter 4 of this thesis (see also Kilbourn-Ceron, 2017). For such non-reductive processes, the Probabilistic Reduction Hypothesis would make no prediction, or maybe in fact predict a lower rate of liaison with greater predictability of the upcoming word, since predictability should correlate with more reduction.

A much greater range of processes will have to be looked at closely in order to tease apart which mechanism(s) are responsible for the observed effects. Our main goal here was to show that the PPH makes very concrete predictions in this regard, which differ from the predictions of alternative hypotheses, and that our data support these predictions.

3.5 Conclusion

This paper looked at North American English flapping and tested predictions of the Production Planning Hypothesis, a proposal that relates locality and variability in phonology to the locality of speech production planning. Our production experiment tested how the likelihood of flapping is affected by syntactic clause boundaries, after controlling for prosodic boundary strength. Results showed that clause boundaries make flapping less likely, but do not rule it out completely. Moreover, we found a gradient effect such that prosodic boundary strength correlates with a lower likelihood of flapping. This gradient

effect is compatible both with an explanation in terms of the PPH and with an alternative account in terms of gestural overlap under temporal compression. However, we found that the syntactic effect persisted even after temporal factors were taken into account. Given findings that production planning is constrained by syntactic constituency (cf. Wheeldon, Smith, and Apperly, 2011; Lee, Brown-Schmidt, and Watson, 2013), the PPH provides an explanatory mechanism for this interaction.

In a second study, we looked at a corpus of conversational speech, and tested for effects of frequency of the upcoming word on flapping rate. The more frequent the following word, the more likely the coronal stop was to be flapped, as predicted by the PPH. Again, this effect was present even after controlling for measures of prosodic compression, suggesting that it cannot be reduced to an explanation in terms of greater articulatory overlap.

Our analysis of glottalization in both the experimental and corpus studies showed a contrasting pattern of variability. Unlike flapping, glottalization does not necessarily require coordination with the gestures of the following segment, but it is a marker of strong prosodic boundaries. We saw that while both flapping and glottalization were affected by gradient measures of boundary strength, glottalization was *not* significantly modulated by the difficulty of planning the upcoming word (i.e. lexical frequency of the following word). This result is in agreement with the PPH: Only processes that are dependent on detailed information of the following context should have a pattern of variability that is dependent on the difficulty of retrieving and encoding the following word. Furthermore, we found evidence in both studies which suggests that glottalization is significantly correlated with a gradient measure of prosodic boundary strength. Particularly striking was the fact that the final lengthening measure in our production experiment only exerted its effect if there was a *clause boundary* following the /t/-final word, suggesting that the glottal stop is an important marker of major boundaries, even though it often occurs within smaller phrases.

Flapping, as opposed to glottalization, is a process that depends on the phonological

content of a following word. The source of the locality and some aspects of the variability of flapping, we argued, is that the following word is not reliably planned at the time when the current word is phonologically encoded. The PPH makes the strong prediction that any process that depends on phonological detail of an upcoming word will show a pattern of production planning-induced variability, and that the precise pattern of locality and variability depends on the kinds of information a sandhi process relies on.

Preface to Chapter 4

Chapter 3 tested the predictions of the PPH for /t/-realizations in North American English. The predictions of the PPH for syntactic effects and lexical frequency effects were supported, the former with results from a production experiment, and the latter with results from a corpus analysis.

These results were also discussed from the perspective of other theories on variability, including Prosodic Phonology and probability-based accounts of reduction. Probability-based accounts of reduction were found to have predictions overlapping with those of the PPH for flapping, a reductive process. The PPH explanation for frequency effects is not mutually exclusive with probabilistic reduction accounts, but their domains of explanation are not completely overlapping, so it should be possible to find evidence for production planning effects that are not attributable to probabilistic reduction.

A distinguishing case that would uniquely support the existence of PPH effects would be the finding of a similar pattern for a *non-reductive* external sandhi process. This study is presented in Chapter 4 which tests for probability and predictability effects in liaison in French.

Liaison is the realization of a word-final consonant which is only pronounced when the following word is vowel-initial. Similarly to flapping, the realization of liaison requires knowledge that the following word begins with a vowel to be available at the time that the consonant-final word is being encoded. And just as for flapping, the PPH predicts that the ease of retrieval of the vowel-initial word should correlate with the likelihood

that the phonological information will be available, and that liaison will apply. Chapter 4 tests these predictions in a corpus study, and tests an additional measure of predictability: conditional probability.

Chapter 4

Speech production planning affects phonological variability: a case study in French liaison[1]

4.1 Introduction

Connected speech processes have played a major role in shaping theories about phonological organization, and how phonology interacts with other components of the grammar. In particular, processes that can apply across word boundaries, often called 'external sandhi', have been important to the development of theories of the syntax-phonology interface (Selkirk, 1974; Kiparsky, 1982; Kaisse, 1985; Nespor and Vogel, 1986) due to their **locality** conditions. External sandhi processes are subject to locality restrictions that appear to be syntactic, or correlated with syntactic structure. There are approaches which characterize locality as directly syntactic, or as locality within phonological domains that only indirectly reflect syntax, or even as frequency of co-occurrence.

But external sandhi processes also seem to be variable above and beyond locality restrictions, in that they often do not apply categorically even when locality is held constant. As phonological variability has come to the forefront of phonological research programmes (Coetzee and Pater, 2011), several theories have been and are being developed

[1] A version of this chapter appears in *The Proceedings of the 2016 Annual Meeting on Phonology*.

to understand variable realizations, especially in spontaneous speech. Applying these ideas to external sandhi processes has been part of these developments, since these processes are highly variable and by definition apply only in connected speech. One strand of research has focused on the role of probability in predicting the prevalence of reduction processes such as coronal stop deletion and flapping in English, e.g. Gregory et al. (1999) proposing the Probabilistic Reduction Hypothesis (Jurafsky et al., 2001). Relatedly, the Smooth Signal Redundancy Hypothesis ties prosodic modulations to the information density of an utterance, where duration (for example) is increased for more informative (i.e. less predictable) words (Aylett and Turk, 2004; Turk, 2010).

These approaches may explain some, but not all variability in external sandhi processes. In particular, they do not make clear predictions about sandhi processes that are non-reductive. For example, liaison is an alternation in which a consonant is pronounced between an (etymologically/orthographically) consonant-final word (W1) and a following vowel-initial word (W2), but is not pronounced if W1 is utterance-final or before a consonant-initial W2 (e.g. *peti*[t] *ami* 'little friend' but *peti*[∅] *garçon* 'little boy'). Hence, the context-sensitive variant involves the articulation of an *extra* segment rather than deletion or reduction. This study investigates the effects of probability and predictability on the pattern of variability in liaison, and proposes that these effects can by understood by reference to the **Production Planning Hypothesis**.

4.1.1 Locality of Production Planning

The Production Planning Hypothesis (PPH) (Wagner, 2012; Tanner, Sonderegger, and Wagner, 2017; Kilbourn-Ceron, Wagner, and Clayards, 2016) proposes that the constraints on speech production planning play a role in explaining external sandhi patterns. The core idea of the PPH is that the choice of pronunciation for a word cannot be affected by the following phonological context if that context is not available at the time the word is being encoded, and the following context is only probabilistically available.

According to influential models of speech production (Levelt, Roelofs, and Meyer, 1999; Dell and O'Seaghdha, 1992), planning of connected speech proceeds hierarchically and incrementally. Units like syllables and prosodic words are planned before detailed segmental information, and larger units are planned further in advance than smaller ones (Sternberg et al., 1988; Wheeldon and Lahiri, 2002; Ferreira and Swets, 2002; Keating and Shattuck-Hufnagel, 2002). Consequently, at the moment of phonological encoding for a particular word, more details of the higher-level utterance structure are known than lower-level information, including segmental content of upcoming words.

The PPH rests on the idea that the size of planning 'chunks' for phonological encoding, being relatively small (Wheeldon and Lahiri, 2002; Wheeldon, 2012), may not always encompass two adjacent words. Hence some words may be phonologically encoded in the absence of information about segments in an upcoming word, preventing interaction between the two words, and therefore blocking external sandhi processes from applying. Furthermore, there is evidence that the size of the planning window is not fixed: It expands or contract depending on many factors, including cognitive load (Wagner, Jescheniak, and Schriefers, 2010), complexity of an upcoming syntactic constituent (Ferreira, 1991), and working memory load (Ferreira and Swets, 2002). This leads to a potential explanation for different patterns of variability under different speaking conditions.

The PPH predicts that factors which modulate the difficulty of planning the upcoming word should led to less application of external sandhi, since the planning window is less likely to include the triggering word when the target word is being planned. In Chapter 3 (also reported in Kilbourn-Ceron, Wagner, and Clayards, 2016), this prediction was tested for coronal stop flapping in two ways: by manipulating whether or not the triggering word was in the same syntactic clause, and by varying the lexical frequency of the trigger containing word. The first experiment tested the effect of syntactic constituency on the likelihood of flapping across a word boundary in a production experiment. Subjects read aloud sentences with target nonce verbs in an embedded clause, and varied whether

the target verb was followed by a clause boundary or not. A logistic regression analysis showed significantly lower likelihood of flapping in the presence of a clause boundary. However, the effect of clause boundary present was *gradient*, not blocking flapping completely but decreasing the likelihood by about half. The PPH predicts both aspects of the syntax effect found in this experiment: syntax has a probabilistic and relatively subtle indirect effect through its influence on the course of speech production planning.

Kilbourn-Ceron, Wagner, and Clayards (2016) also reports a corpus study testing the effect of lexical frequency of both the target coronal-final word and following vowel-initial word (e.g. *cat attack*). Many studies have shown that high lexical frequency facilitates word form retrieval (Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994). Accordingly, the PPH predicts that lexical frequency should influence external sandhi in a very specific way: the higher the frequency of the word *following* the target word, the more likely it should be for sandhi to apply, since the words are more likely to be encoded within the same planning window. This prediction was tested for flapping using data from the Buckeye Corpus of Conversational Speech, and it was found that indeed there is a statistically significant increase in the likelihood of flapping as the frequency of the word following the target increases. Under the view that flapping is a reductive process due to gestural overlap with adjacent vowels (e.g. Fukaya and Byrd, 2005), this finding fits in well with the broader research on probabilistic effects on reduction (Bybee and Scheibman, 1999; Jurafsky et al., 2001; Bell et al., 2003). But unlike probabilistic accounts like the Probabilistic Reduction Hypothesis of Bell et al. (2003), the predictions of PPH extend also to non-reductive sandhi processes, predicting similar effects of lexical frequency in both cases.

In the present study, the same prediction is tested for a non-reductive process: liaison in French. The main research question of this study is whether ease of retrieval/encoding of W2 in a potential liaison environment increases the rate of liaison. This question is addressed by presenting an analysis of data from the Phonologie du Français Contemporain

(PFC) corpus (Durand and Lyche, 2003; Durand, Laks, and Lyche, 2009). The measure of retrieval/encoding difficulty is operationalized with two variables: the lexical frequency of W2, which measures its global difficulty across all contexts, and the conditional probability of W2 given W1, which measures the *local* predictability of W2 when W1 is known. The PPH predicts that, to the extent that these measures have independent effects on lexical retrieval and encoding, both could independently influence liaison rates. Both variables are predicted by the PPH to be positively correlated with liaison.

If the locality of production planning does indeed play a role in shaping the pattern of external sandhi variability, then it will be crucial to take these effects into account in future empirical studies of external sandhi processes, especially in spontaneous speech.

Section 4.2 reviews previous research on liaison and its variability, followed by the presentation of the data set used for the present study in Section 4.3. Section 4.4 presents the construction of the statistical model and the results of fitting it to the data. Section 4.5 discusses how the results bear on the PPH and other accounts of variability, and Section 5.3 concludes.

4.2 Liaison

The liaison alternation in French is one in which a latent final consonant on Word 1 (W1) surfaces when the following word (W2) begins with a vowel. For example, the adjective *gros* 'big' is pronounced [gʀo] in isolation or before a consonant, but with a final [z] when it modifies a vowel-initial noun as in *gros enjeu* 'big stake/issue'. There is an extensive literature on liaison (see Côté (2011) for a review), with ongoing debates regarding the lexical affiliation of the liaison consonant (Morin, 2003; Côté, 2005; Smolensky and Goldrick, 2016) as well as the nature of morpho-syntactic restrictions on the process (Selkirk, 1974; Morin and Kaye, 1982; Kaisse, 1985; Nespor and Vogel, 1986; Bybee, 2001; Côté, 2013). The

issue of lexical affiliation will not be addressed in detail in this paper as it is not crucial to our point, but the issue is briefly touched on in the discussion.

4.2.1 Locality conditions

The realization of liaison is dependent on the relationship between W1 and W2, traditionally described in syntactic terms, and divided into three categories: excluded, variable, or categorical. Intuitively, locality seems to depend on degree of 'cohesion' between the words in the liaison pair: the closer the words are to forming a unit, the more likely liaison is to apply. For example, liaison is obligatory between a verb and following subject clitic (*di**t**-il* 'says he'), but variable between a verb and following prepositional phrase (*j'irai**s** à Paris* 'I'd go to Paris'). Different generalizations about liaison's locality conditions have been proposed in syntactic (e.g. Kaisse, 1985; Pak, 2008), prosodic (Selkirk, 1986; Féry, 2004), and usage-based (Bybee, 2001; Côté, 2013) terms, with debate still open as to which of these definitions explains the pattern of locality most closely.

The realization of liaison also varies by region (see Côté, 2017, for a comprehensive review). A small class of liaison contexts have been found to be categorically realized in every variety of French so far examined: determiner + adjective/noun, proclitic + proclitic/verb, verb + enclitic, and *en* "in" + X (Durand and Lyche, 2008; Côté, 2017). For other liaison contexts, the prevalence differs markedly by region, both in overall rates of realization and in pattern. For example, Côté (2017, :19) calculates the liaison rate of *est* "is (3rd person singular)" for four different regions: 39.3% for Île-de-France, 30.6% for Switzerland, 75.5% for Canada, and 12.8% for Africa. Differences between dialects will not be directly addressed in this study, but will be controlled for in the statistical model since the corpus used for this study includes data from several different dialects across the French-speaking world.

Previous work in exemplar-based models has explored the link between probability and liaison. Bybee (2001) proposes that usage frequency accounts for locality conditions.

Under this account, phrases and constructions can be stored in memory. Liaison pairs in higher frequency phrases and constructions are more likely to be stored together, and therefore more likely to retain liaison. This predicts a correlation between bigram frequency and liaison, a prediction consistent with the empirical data she examines.

Similarly, Côté (2013) puts forth that liaison likelihood depends on the probability of a given W1 being followed by a particular syntactic category. Her results also show an empirical correlation between rate of liaison and predictability of W2's category given W1. However, these studies focused on a relatively limited number of W1 types. The current study tries to broaden the range of word types by focusing on two particular syntactic contexts where both W1 and W2 are open-class lexical items. This should also allow investigation of a wider range of frequency values.

The first context tested in this study is an adjective followed by a noun (Adj-Noun), for example *petit* [t] *ami* "little friend". This construction is classified by Delattre (1955) as an obligatory liaison context. However, Durand and Lyche (2008) noted counter-examples in the PFC data they examined (a subset of the data used for this study). For example, the adjective *gros* "big" appears in prenominal liaison context 8 times in their data, but is only realized with a liaison consonant in 6 of 8 occurrences. Côté (2017) reports a rate of 82% for prenominal adjectives in Louisiana French, again not categorical but relatively high.

The second liaison context examined is a plural noun followed by an adjective (PlNoun-Adj), for example *les pas* [z] *enjoués* "the cheerful steps" (Côté, 2013). Between a *singular* noun and adjective, liaison is reportedly impossible. PlNoun-Adj is classified in the literature as a 'variable' liaison context (Côté, 2011).

One important difference to note between the PlNoun-Adj context and the Adj-Noun context is their prosodic organization: pre-nominal adjectives are consistently phrased together with the noun, while post-nominal adjectives can be phrased separately (Post, 2000). Although liaison can *sometimes* be realized across a large prosodic boundary (Côté, 2011), the PlNoun-Adj context may have an overall lower rate of liaison partly due to

cases where a prosodic phrase boundary appears between the noun and adjective. This is a difference which could be interesting to explore from a PPH perspective in future work. Another interesting point is that there may be two different syntactic structures for post-nominal adjectives, a difference which could be relevant to the realization of liaison (Post, 2000). These differences would be interesting and relevant to investigate with controlled tests, but for the purpose of the present study, they will be set aside. In this study, the crucial question is not whether the baseline rates are or even should be the same in all syntactic contexts, but whether they are both similarly modulated by lexical frequency and conditional probability.

4.2.2 Liaison and production planning

The PPH predicts a parallel pattern of variability for liaison as was found for flapping in English (Kilbourn-Ceron, Wagner, and Clayards, 2016), since both alternations are dependent on a *following* context: Higher frequency words are easier to retrieve/plan, so higher frequency should correlate with higher likelihood of both words being planned within the same window, and consequently the realization of the contextual variant (i.e. the flap or liaison consonant). There is evidence from a previous study that probabilities do indeed affect liaison in this way: Côté (2013) found a correlation between liaison rate and the predictability of W2's syntactic category. For a given W1, the rate of liaison was highest for those words for which the category of the following word was most predictable. For example, *très* "very" appeared exclusively before adjectives/adverbs, and its liaison rate was the highest of all the adverbs they examined, at 84%. In contrast, *moins* "less" only appeared before adjectives/adverbs 53% of the time, and its liaison rate was only 14%. These and the other results of Côté (2013) could be compatible with PPH predictions: If syntactic predictability facilitates retrieval of W2, the PPH predicts higher rates of liaison for higher predictability. A study suggestive of this is Gahl and Garnsey (2004), who found higher

rates of t/d deletion when the complement of the verb was of a more predictable category (for that particular verb).

A difference to note between liaison and flapping is that the articulation of the liaison consonant itself does not necessarily have to be tightly coordinated with the gestures of the following vowel. Although the liaison consonant is normally resyllabified into the onset of W2 (Côté, 2011), it is possible to pronounce the liaison variant with a pause afterwards. This is not possible for a flap, which requires detailed coordination with the following vowel it is released into. This could mean that liaison is less sensitive to production planning effects than flapping, since less detailed coordination is necessary. For the purpose of finding evidence of production planning effects, the crucial difference between liaison and flapping is that liaison is non-reductive. Probabilistic reduction theories make overlapping predictions with the PPH for reductive processes like flapping, but make no or opposite predictions for non-reductive processes. If it is the case that lexical frequency of the liaison-triggering word modulates variability in the same way as it did for flapping, this would support the idea that online speech production planning effects play a role in shaping phonological variability.

4.3 Data set

The source of data for this study was the *Phonologie du Français Contemporain* corpus (PFC; Durand, Laks, and Lyche, 2002; Durand, Laks, and Lyche, 2009), a geographically diverse corpus of read speech and spontaneous conversations. The PFC authors annotated a subset of each speaker's data for liaison: the read text (same passage for all speakers), and five minutes of each speaker's spontaneous conversation data. This subcorpus was retrieved using the online search tool on the PFC website.[2] It contains contains data from 417 speakers recorded in 39 different regions, and a total of 53467 tokens.

[2]http://public.projet-pfc.net/liaison/

For the purposes of the PFC annotation, liaison was defined as "the pronunciation of any graphic consonant when the word (W2) following the linking word (W1) is vowel initial," and potential liaison sites to be coded were defined as the environments that Delattre (1966) defines as possible liaison contexts. The full protocol is described in Durand and Lyche (2003).

The annotation records 0 for no liaison, 1 for *liaison enchainée* (forward-syllabified liaison, the typical case). We restrict our data to observations with either of those annotations, removing cases coded as *liaison non-enchainée*, uncertain, and "epenthetic" liaison. This resulted in loss of less than 1% of the data, leaving 52953 tokens of potential liaison sites.

For the present study, this data was restricted to two syntactic contexts, as discussed above: PlNoun-Adj and Adj-Noun sequences. This was done by retrieving part-of-speech information from the Lexique database (New et al., 2001, Version 3.81), and matching it orthographically with the PFC data. Many words were ambiguous between different parts of speech, so the subsets were determined by selecting any pair of words that could potentially match the criteria of PlNoun-Adj or Adj-Noun. These subsets were subsequently manually verified, and any word-pairs that were determined to be mislabeled were removed, leaving 1451 tokens in the PlNoun-Adj dataset and 2865 in the Adj-Noun dataset.

Lexique is also the source for the lexical frequency information. Lexique provides several frequency measures. The frequency calculated from movie subtitles was used for this study, as it more closely approximates spoken French (see Brysbaert and New, 2009, for discussion of subtitle-based frequencies).

Conditional probability for liaison pairs was estimated by fitting a trigram language model to the French Gigaword corpus (First Edition, Graff, 2006), a large archive of French newswire text compiled by the Linguistics Data Consortium. Using a larger corpus like Gigaword instead of the PFC corpus allows more accurate estimates, especially for bigrams (two-word sequences) which are by definition more rare than individual words. The language model was fitted using the `lmplz` function from the KenLM language

model toolkit (Heafield et al., 2013), which uses modified Kneser-Ney smoothing without pruning. The calculated values were used to determine the conditional probability of W2 given W1 for each bigram in the data set. In each of the PlNoun-Adj and Adj-Noun subsets, there were liaison pairs that were not observed in the Gigaword corpus. It is possible to estimate bigram frequencies from separate unigram frequencies, but the estimate may be less reliable, so unobserved bigrams were removed from the data set. This resulted in a loss of 66 tokens (4.55%) for the PlNoun-Adj data set (n = 1385), and 360 tokens (12.57%) for the Adj-Noun data set (n = 2505).

The overall rate of liaison for the PlNoun-Adj context was 31.99% (n = 1385) by token, and the liaison rate by W1 type was 9.89% (n = 161). For the Adj-Noun context, the liaison rate was 89.3% (n = 2505) by token and 57.69% (n = 104) by W1 type.

In terms of regional variation, the realization of liaison was lower and varied more for the PlNoun-Adj context, ranging from about 65% in Lacaune, France (n = 34) to 0% in Chlef, Algeria (n = 48), with a mean of 32%. For the Adj-Noun context, the prevalence of liaison by dialect ranged from 98% in Brécy, France (n = 54), to 74% in Burkina Faso (n = 50), with Chlef being an outlier in this context at 40% realization of liaison (n = 20).

4.3.1 Predictors

To address the research questions outlined in Section 4.1, the pattern of liaison realization will be modeled as a function of a number of predictors. The two main variables of interest are the lexical frequency of W2 in the liaison pair, and the conditional probability of W2 given W1.

Speech rate was calculated in words per second for each utterance, using the start and endpoints and orthographic transcription provided in the PFC. Selkirk (1986) and Kaisse (1985) have stated that liaison may not be sensitive to speech rate, and Pak and Friesner (2006) found no effect of (self-selected) speech rate in a production experiment. On the other hand, speech rate may be correlated with speech style, and liaison application tends

to increase in more formal styles (Kaisse, 1985), which are associated with slower speech rates. Furthermore, there is some evidence that suggests a correlation between higher speech rate and larger planning scope (Wagner, Jescheniak, and Schriefers, 2010). This would predict the opposite: an increased liaison rate for faster speech. Hence, speech rate is included as a control, though investigating speech rate effects is not the focus of this study.

The number of syllables of W1 was determined from the Lexique database, and included as a control. Previous research has observed that monosyllabic words are associated with higher probability of liaison, though this is based on a limited set of closed-class words (Côté, 2011). There is also a correlation between frequency and number of syllables (Zipf, 1929), so these may have been confounded in previous studies. However, the PPH might predict more liaison in monosyllables beyond frequency effects. Griffin (2003) found that when instructed to say two-word sequences without a pause between the words, speakers took longer to initiate speech when the first word was *shorter*. This suggests in order to avoid disfluent pauses, speakers *extend* their planning window when the first word is shorter, since a short word would not give enough time for parallel planning and retrieval of the second word without an intervening pause. If the same strategy is used in spontaneous speech to avoid disfluent pauses, monosyllabic words should be more likely to be planned together with the words that follow them, leading to higher rates of liaison.

Figure 4.1 shows the correlation between the probability of liaison realization and the frequency of W2 (log-transformed) for the PlNoun-Adj context, and Figure 4.2 shows the same for the Adj-Noun context. Both plots suggest a positive correlation, i.e. the likelihood of a liaison consonant being realized increases as the frequency of the following word increases.

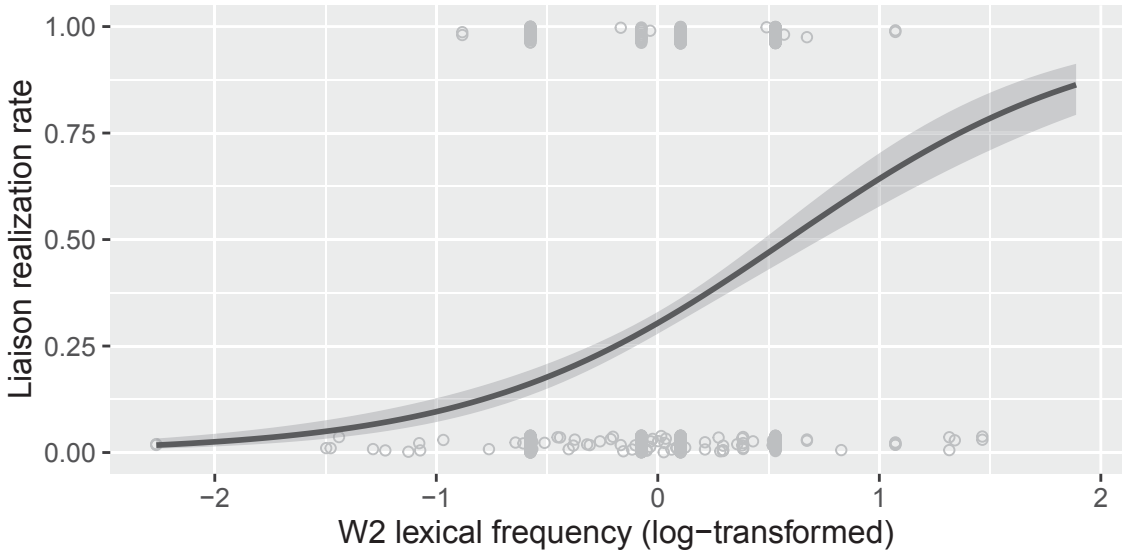

FIGURE 4.1: Empirical plot of liaison realization in **PlNoun-Adj** context as a function of W2 lexical frequency. Points are jittered for visibility, one point per token. Solid line is a GAM smoother with binomial logit-link, shading indicates 95% CI.

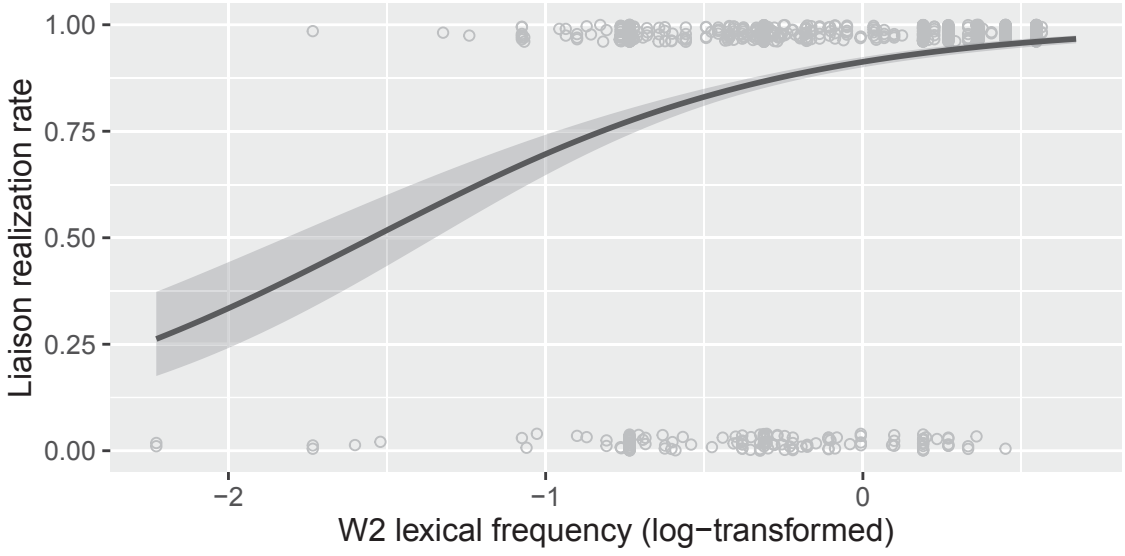

FIGURE 4.2: Empirical plot of liaison realization in **Adj-Noun** context as a function of W2 lexical frequency. Points are jittered for visibility, one point per token. Solid line is a GAM smoother with binomial logit-link, shading indicates 95% CI.

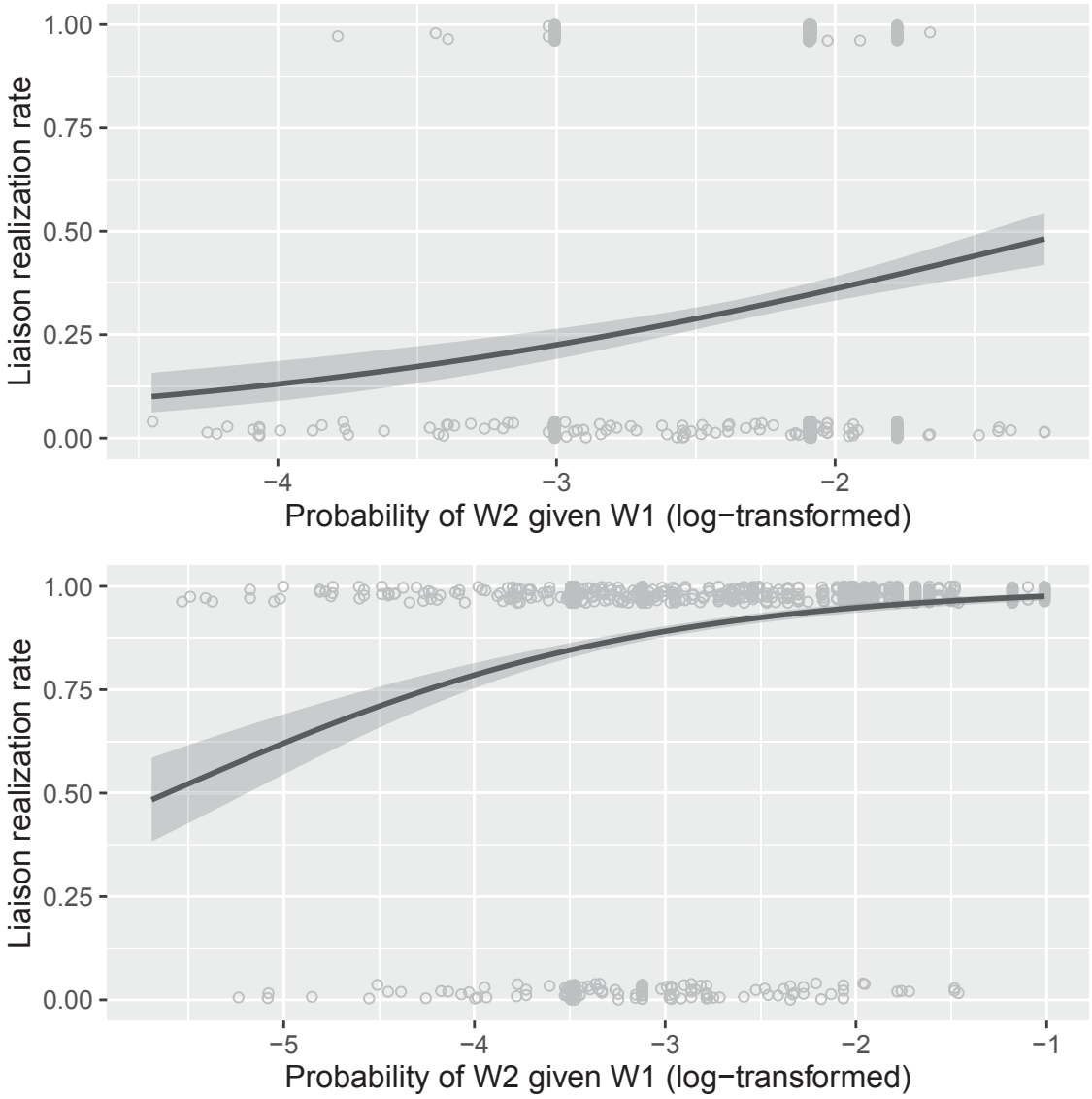

FIGURE 4.3: Empirical plot of liaison realization in **PlNoun-Adj** context (top) and **Adj-Noun** context (bottom) as a function of the conditional probability of W2 given W1 in the potential liaison pair. Points are jittered for visibility, one point per token. Solid line is a GAM smoother with binomial logit-link, shading indicates 95% CI.

4.4 Analysis

To verify that the empirical trends were statistically significant after controlling for duration and individual word, speaker, and regional differences, a mixed-effects logistic regression was fit to the data, with a separate model for each syntactic context. This type of regression predicts the log-odds of a binary outcome, in this case whether or not the liaison consonant is realized. Using a mixed-effects model allows the inclusion of both fixed effects, which estimate the influence of experimental and control variables, as well as random effects, which account for variability within groupings of observations (Gelman and Hill, 2007; Baayen, 2008).

4.4.1 Model structure

The dependent variable is coded as 0 (no liaison) or 1 (liaison). The model estimates the log-odds of liaison being realized, so positive effect estimates for the independent variables represent an increase in the predicted likelihood of liaison applying.

The predictors of interest for this study were included as fixed effects: W1 FREQUENCY and W2 FREQUENCY were log-transformed to bring the distribution closer to normality, and standardized (centered and divided by two standard deviations, Gelman and Hill, 2007) within each data set. The CONDITIONAL PROBABILITY of W2 given W1 was already given in log-transformed value from the calculation described in Section 4.3.1, and was also standardized.

The control predictor SPEECH RATE, in units of words-per-second, was standardized, and the predictor SYLLABLES was sum-coded with two levels, monosyllabic (1) and polysyllabic (-1).

Random effect structure for both models included random intercepts by location of data collection, and by speaker (nested within collection location). For the PlNoun-Adj

TABLE 4.1: Model results: Fixed effects coefficients, standard errors, z-scores, and p-values (Wald test) for all model predictors applied to the Plural Noun-Adjective data set.

Fixed effects	$\hat{\beta}$	se($\hat{\beta}$)	z	Pr(z)
Intercept	-1.994	0.406	-4.916	<0.001
W1 Frequency	5.195	1.116	4.655	<0.001
W2 Frequency	3.862	0.958	4.029	<0.001
Conditional Probability	1.106	0.393	2.813	0.005
Speech Rate	0.473	0.296	1.597	0.110
Syllables	0.143	0.111	1.284	0.199

model, it was not possible to fit a stable model with any further random effects structure[3]. This limits the generalizability of the results, and will be discussed below. The random effect structure for the Adj-Noun model included random intercepts for W1 and W2, and also by-speaker random slopes for W1 FREQUENCY, W2 FREQUENCY and CONDITIONAL PROBABILITY, which increases the accuracy of p-values and coefficient estimates for the corresponding fixed effects terms and helps ensure their generalizability across speakers (Barr et al., 2013).

4.4.2 Results

This section reports the results of the statistical analysis of liaison likelihood for the PlNoun-Adj and Adj-Noun contexts. Fixed-effects coefficients for the fitted models are presented in Tables 4.1 and 4.2.

Plural Noun-Adjective context

The coefficient estimates for the model fitted to the PlNoun-Adj data are presented in Table 4.1. The predicted probability of realizing liaison in this context, with all other predictors held at their mean values, is 11.98%, somewhat higher than the empirical by-W1 rate of 9.89%. Before discussing the fixed effects estimates, it is important to reiterate that in order to fit a stable model, it was necessary to exclude word-level random intercepts. Therefore, effects that are statistically significant in the model presented in Table 4.1 may not be generalizable across words or bigrams; generalizability should be tested in future research with a more balanced dataset that includes more bigram types.

The estimates for the effects of all the probability-based measures were positive, and significantly different from 0. The effect of W1 FREQUENCY was the largest ($\hat{\beta} = 5.195$, $p =$<0.001), suggesting that the frequency of W1 plays a role in shaping the variability of liaison realization. This is a correlation that has not been very much discussed in the literature, outside of noting that W1s in the obligatory liaison contexts are usually closed-class, high frequency words. Finding a W1 frequency effect in the PlNoun-Adj context is interesting, as it shows that frequency effects are not reducible to a functional vs content word distinction.

The frequency of W2 also has a positive effect on liaison rate ($\hat{\beta} = 3.862$, $p =$0). The effect is quite large, with the odds of liaison predicted to increase by 47.56 times with every increase in one standard deviation (SD) of W2 FREQUENCY. This is suggestive of a role for W2 lexical frequency in explaining liaison variability, but again due to the unbalanced nature of the data it is not possible to generalize across words, and further investigation will be needed to confirm the effect.

[3] Attempts to fit models with word-level (W1, W2, or bigram) random intercepts resulted in unstable models with estimates for those random effects terms being extremely high and non-normally distributed. Examination of the data suggests that sparsity may be the issue, as a significant amount of data comes from a small number of bigrams, which are mainly from the PFC read text. This issue should be overcome in future research with a more balanced dataset.

TABLE 4.2: Model results: Fixed effects coefficients, standard errors, z-scores, and p-values (Wald test) for all model predictors applied to the Adjective-Noun data set.

Fixed effects	$\hat{\beta}$	se($\hat{\beta}$)	z	Pr(z)
Intercept	3.244	0.778	4.171	<0.001
W1 Frequency	1.164	0.395	2.945	0.003
W2 Frequency	1.010	0.710	1.422	0.155
Conditional Probability	1.399	0.695	2.012	0.044
Speech Rate	0.577	0.298	1.937	0.053
Syllables	1.140	0.554	2.057	0.040

CONDITIONAL PROBABILITY also was estimated to have a positive effect ($\hat{\beta} = 1.106$, p =0.005). This estimate represents a predicted increase in odds of liaison by 3.02 times for every 1 SD increase in CONDITIONAL PROBABILITY.

Neither SPEECH RATE nor SYLLABLES had a statistically significant effect. SYLLABLES did have a positive coefficient estimate ($\hat{\beta} = 1.14$, p =0.04), which agrees with findings in the liaison literature that monosyllabic words are associated with a higher likelihood of liaison than polysyllabic ones.

Adjective-Noun context

In the Adj-Noun context, the overall probability of realizing liaison was much higher (96.25%), as expected from both examination of the empirical data and previous reports in the literature that this is an obligatory or at least highly prevalent liaison context. The fixed effects estimates for the model fitted to this data is presented in Table 4.2.

As in the PlNoun-Adj context, the effect of W1 FREQUENCY was statistically significant ($\hat{\beta} = 1.164$, p =0.003). This represents a predicted increase in odds of about 3.2 times for every 1 SD change in the value of W1 FREQUENCY. Since the model for Adj-Noun includes a random intercept for W1, this effect is predicted above and beyond any idiosyncratic tendencies for particular lexical items to have higher or lower liaison rates

overall, and the effect also seems to be robust across speakers, given that a by-speaker random slope was also included for this effect. Hence, this model shows with higher certainty that there is indeed a general correlation between increased lexical frequency and increased likelihood of liaison within the Adj-Noun context.

As for W2 FREQUENCY, the effect was not statistically reliable in the Adj-Noun model, though the coefficient estimate was positive, as in the PlNoun-Adj model ($\hat{\beta} = 1.01$, p =0.155). Since the Adj-Noun model included a random intercept for W2, as well as by-speaker random slope for W2 FREQUENCY, it's possible that that there was not enough data to reliably distinguish between random variance among W2 liaison likelihood and a specific frequency effect. Future work should investigate this effect with a larger and more balanced data set.

As in the PlNoun-Adj model, the CONDITIONAL PROBABILITY estimate for Adj-Noun was positive and statistically significant, and of a similar magnitude ($\hat{\beta} = 1.399$, p =0.044), with the odds of liaison realization increasing by 4.05 times for every 1 SD increase in the value of CONDITIONAL PROBABILITY.

The estimated SPEECH RATE coefficient was positive, but only marginally significant ($\hat{\beta} = 0.577$, p =0.053), in line with PPH predictions. SYLLABLES did have a positive and statistically significant coefficient estimate ($\hat{\beta} = 0.143$, p =0.199), which once again agrees with claims in the literature that monosyllabic W1 are associated with higher rates of liaison (Côté, 2011).

4.5 Discussion

This study investigated the relationship between word probability and liaison in the PFC corpus in order to address the main research question of this paper: Does the ease of retrieval/encoding of W2 in the liaison context affect the likelihood of liaison? The results of the statistical analysis provided some evidence that both local (conditional probability)

and global (lexical frequency) measures of word probability play a role in shaping the pattern of variability in liaison. This is consistent with the PPH prediction that liaison, and external sandhi in general, should be more likely to apply when the following, triggering word is easier to plan[4].

W2 Frequency It has been shown that lexical frequency has a facilitatory effect on word form retrieval (Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994), so the PPH predicted that higher W2 frequency should lead to higher liaison rates. There was a significant positive effect of W2 FREQUENCY, but only in the model fitted to PlNoun-Adj data. This model had only a limited random effect structure, so while these results are suggestive, they should be confirmed with further analysis in future research.

Conditional Probability The results of this study also provided suggestive evidence that the conditional probability of W2 given W1 was positively correlated with rate of liaison. That is, when W2 in a liaison context is *more predictable*, the liaison consonant is more likely to be realized. This effect could be understood under the PPH as predictability *facilitating* the retrieval/encoding of W2, which leads it to be available sooner and more likely to be planned within the same window as W1, enabling the possibility of liaison applying.

W1 Frequency The results showed a positive effect of W1 frequency in both the PlNoun-Adj and Adj-Noun contexts. The theories of probability effects presented so far do not make clear predictions as to how W1 frequency should influence liaison, so it is not clear how to interpret this effect. Previous studies that investigated lexical frequency effects

[4] Bürki, Frauenfelder, and Alario (2015) found that realizations of the singular indefinite determiner *un*, which participates in obligatory liaison, are influenced by the initial phonological segments of both the adjective and the noun in a determine-adjective-noun sequence. This suggests that at least for determiner liaisons, realization can be affected by properties of later words than the one immediately following. The PPH predicts that the ease of retrieval of these words should also have an effect. This could be an interesting direction to pursue in future work, although determiner liaison is almost categorical so it would be difficult to study its pattern of variability.

have found that more frequent words are durationally and segmentally reduced (Gregory et al., 1999; Jurafsky et al., 2001; Bell et al., 2003; Schuppler et al., 2012), supporting usage-based accounts, but liaison is not reductive. In an exemplar-based account, it might be possible to think of liaison realization as an exceptional variant of W1 which is more likely to be preserved in higher frequency words, analogous to irregular conjugations of verbs being more resistant to paradigm leveling (Bybee, 1985).

The prediction of the PPH depends on how W1 frequency would affect the relative timing of phonological encoding for W1 and W2. If W1 is encoded more quickly when it is more frequent, this might allow less time to retrieve W2 in time for W2 to influence W1's encoding, leading to lower liaison rates. On the other hand, Konopka (2012) found evidence that, under certain circumstances, higher frequency words were associated with an *extended* planning scope. Accordingly, the PPH would predict higher rates of liaison for higher frequency, more easily retrievable words, since they would be more likely to be planned together with the following word. See Tanner, Sonderegger, and Wagner (2017), Section 2.3 for related discussion of word probability effects and PPH predictions.

Other variables There was also evidence that liaison is more likely when W1 in the Adj-Noun context is monosyllabic, an effect observed across liaison contexts in previous studies (Mallet, 2008; Laks, 2009). The PPH could offer an interpretation for this effect: If monosyllabic words are more likely to be planned together with the word that follows them, then the PPH predicts higher rates of liaison for monosyllabic words. Findings from Griffin (2003) suggest that under experimental conditions, speakers *are* more likely to plan two words at once if the first is monosyllabic, possibly as a strategy to avoid disfluent pauses between the two words.

The PPH predicts correlations between planning window size and external sandhi application, and by extension with the variables that affect planning window size, for all

external sandhi processes in which the triggering environment is found across a following word boundary. There is supporting evidence for this pattern from previous studies on English coronal stop realizations (Tanner, Sonderegger, and Wagner, 2017; Kilbourn-Ceron, Wagner, and Clayards, 2016). The finding of a parallel effect of lexical frequency for both general reductive processes like coronal stop deletion and flapping in English, and for a non-reductive process like liaison in French is uniquely predicted by the PPH. Below, we turn to discussion of usage-based and functional accounts of liaison variability, and how they relate to our results and the PPH, followed by short remarks on the issue of the liaison consonant's lexical affiliation.

Probability-based accounts The effect of probability and predictability on pronunciation has been present in the literature for some time, with Zipf (1929) proposing a relationship between lexical frequency and word length. Gregory et al. (1999) and Jurafsky et al. (2001) unified previously studied usage-based variables under the umbrella of 'probability', and tested the relationship between word shortening/reduction and several probability-based measures. They propose that the more predictable or probable a word is, the more likely it is to be reduced. The results of both studies show that various measures of predictability are correlated with temporal and segmental reduction (duration, vowel reduction, and final t/d deletion, and t/d flapping). While all the dependent variables are correlated with some probability-based measure, it is not the same measure in all cases, leaving open the question of the exact mechanism by which these factors influence lexical production.

Kilbourn-Ceron, Wagner, and Clayards (2016) discusses how probability-based accounts of reduction like Jurafsky et al. (2001) make overlapping predictions with the PPH. Kilbourn-Ceron, Wagner, and Clayards (2016)'s finding that W2 frequency has a positive correlation with the likelihood of flapping could alternatively be interpreted as a reduction due to higher predictability of the following word and consequently a reduced flap

realization of the word-final coronal stop. However, given that liaison involves *extra* articulation, the positive frequency effect cannot be straightforwardly derived from this type of account. While probability-based reduction effects may well play a role in explaining certain kinds of variability, the finding of a W2 frequency effect in liaison points to a need to acknowledge the role that general speech production planning constraints play in shaping phonological variability. Indeed, the PPH proposes a specific cognitive mechanism that could be at the source of some probability-based reduction effects.

Côté (2013) investigates the role of transition probability in predicting the likelihood of liaison between two words. Côté proposes that for a given word, the 'productivity' of liaison is related to the frequency with which that word is followed by words of a particular syntactic category, a related but different transitional probability measure than the one used in this study. For example, she examines the adverb *très* 'very,' finding that it is followed by adjectives or adverbs in almost 100% of occurrences across the corpus, and that liaison is realized in 84% of potential liaison contexts. On the other hand, the adverb *trop* 'much' is only followed by adjectives/adverbs in 53% of cases, and liaison is only realized in 14% of liaison contexts. This analysis of summary statistics opens many interesting questions: how is (syntactic) probability stored or represented, and how/when does it enter into the computation of whether or not a liaison consonant is realized? Côté (2013) suggests a functional explanation for lack of liaison in low-predictability contexts: the tendency to syllabify the liaison consonant into the onset of W2 may interfere with lexical access, making it less favourable to realize liaison in these contexts, an idea reiterated in the Message Oriented Phonology framework proposed in Hall et al. (2016) (discussed below).

The PPH could offer a cognitive mechanism for transition probability effects. If predictability of W2's category given W1 makes it easier to retrieve and encode that word within the same planning window as W1, then the PPH predicts the pattern demonstrated

in Côté (2013). The results of Gahl and Garnsey (2004) suggest that syntactic predictability is indeed relevant for phonological realization. An interesting question to investigate would be whether for the same category of W2, it would be possible to find differences in liaison likelihood depending on the syntactic structure. The PPH predicts that a more complex syntactic structure should lead to lower rates of liaison.

Bybee (2001) proposes that the frequency of liaison is dependent on usage, with an analysis couched in Exemplar Theory (Bybee and Scheibman, 1999; Pierrehumbert, 2001; Bybee, 2007; Pierrehumbert, 2006). In this framework, lexical entries are made up of exemplars, stored representations of previously heard instances of that lexical item. Bybee (2001) argues that words that co-occur frequently tend to be stored as a unit, and this is what preserves idiosyncratic, lexically-conditioned alternations like liaison. It is suggested that this type of usage effect can take place at the level of *constructions*, which could be as specific as a sequence of lexical items, or consist of grammatically-defined 'slots' such as NOUN + PLURAL + ADJECTIVE. Bybee (2001) proposes that the liaison consonant is part of a multi-word construction that behaves parallel to a suppletive forms of single words. Under this account, the liaison 'variant' will tend to be preserved only when the construction is highly frequent, behaving more like a single unit, otherwise it will 'regularize' to the non-liaison construction.

This type of account is challenged by the prevalence and productivity of liaison between prenominal adjectives and nouns. Bybee notes that bigram frequency for individual adjective-noun pairs is not as high as other 'fully grammatical morphemes' (i.e. functional morphemes) that exhibit liaison. She points to the fact that prenominal adjectives are of a relatively limited class, and even so predicts that liaison should be less stable in this context than for other highly frequent constructions. But as suggested by the original classification of this context as 'obligatory' liaison, and as we have seen above in the results of this study, the likelihood of liaison in this context is quite high. Future work should further compare measures of predictability such as bigram frequency, although an

exemplar-based account of pronunciation is not incompatible with the PPH: the retrieval of W1 and W2 as separate units or a single unit would still interact with planning scope.

Information-theoretic approaches A related stream of research has tied phonetic and phonological variation to effective message transmission (Aylett and Turk, 2004; Hall et al., 2016; Cohen Priva, 2017). Under this view, phonetic and phonological patterns reflect the pressure to encode a uniform amount of information throughout the utterance. For example, more frequent words are temporally compressed or articulatorily reduced because they contribute less information to the listener.

Hall et al. (2016) discuss the predictions of their approach for boundaries that separate 'units of meaning'. If the message is predictable from context, then accurately identifying boundaries is less crucial than when uncertainty is high. They posit that the pattern of liaison reported by Côté (2013), can be explained by assuming that syllabification of the liaison consonant into the onset of W2 reduces its identifiability. Hence, liaison is avoided when W2 is less predictable in order to improve boundary identification.

However, previous work has investigated the perception of resyllabification in liaison contexts, and found that it does not interfere with recognition of the vowel-initial word (Gaskell, Spinelli, and Meunier, 2002; Spinelli, McQueen, and Cutler, 2003). In fact, Gaskell, Spinelli, and Meunier (2002) suggest that syllabification across word boundaries can work in conjunction with other acoustic and lexical cues to facilitate identification of the vowel initial word. They found that liaison consonants, which are syllabified as onsets of the vowel initial word, are of consistently shorter duration (10–18%) than in a sequence of identical segments where the conosonant underlyingly belongs to the second word. For example, in the liaison context *dernier oignon* "last onion", the [ʁ] pronounced between the two words is consistently shorter than in *dernier rognon* "last kidney", where the [ʁ] that is pronounced comes from the second word. They hypothesize that listeners can make use

of these subtle durational differences to distinguish ambiguities that arise due to resyllabification. If this were indeed the case, a message-oriented framework would predict a *higher* likelihood of liaison in lower predictability contexts, the opposite prediction of the one suggested by Hall et al. (2016), and contrary to the results presented in this study.

Under the PPH, the prediction of more liaison in lower predictability contexts could not be derived, as it would imply that a process that makes reference to information in the upcoming word applies *more* when the word is *less* likely to be planned within the same window.

Lexical affiliation of liaison consonant In this study we have made the assumption that the liaison consonant is final in W1—now, we briefly address an alternative view.

It has been proposed in previous work that the liaison consonant is in fact affiliated with both W1 and W2, both lexically and in terms of surface syllabification (L'Esperance, 2015; Smolensky and Goldrick, 2016). Smolensky and Goldrick (2016) propose that both W1 and W2 are associated with a final and initial consonant respectively, and both consonants carry some 'activation weight'. Only when the weights are combined and reach above a certain threshold is the liaison consonant realized. In their formal Harmonic Grammar analysis, this is implemented by having the liaison consonant surface as the single exponent of both underlying, partially-activated consonants in W1 and W2. This account of the phonological conditions on liaison is compatible with the PPH explanation of locality of production planning effects: crucially, detailed segmental information about W2 is necessary for the realization of the liaison consonant. From the point of view of assessing PPH predictions, the only difference in the Smolensky and Goldrick (2016) account is that the information that must be available is a partially-activated consonant segment rather than that a vowel-initial word is upcoming, as in more traditional accounts.

An observation that has been brought up in experimental literature is that liaison consonants are shorter in duration than their counterparts where the consonant starts out as

an onset of the next word.

4.6 Conclusion

This study has presented evidence that the variability in the realization of a liaison consonant in French is dependent on the predictability of the second word in the liaison pair, both in terms of its local predictability given W1, and its global probability as measured by lexical frequency. This finding lends support to the idea that locality of production planning plays a role in phonological variability above and beyond temporal and gestural reduction effects: liaison, a non-reductive sandhi process, shows the same pattern of variability as reductive processes like that for flapping, shown in Chapter 3 (Kilbourn-Ceron, Wagner, and Clayards, 2016). This prediction is crucially not derivable from hypotheses that based on probabilistic reduction, signal redundancy, or information modulation. The PPH, on the other hand, explains these parallel patterns by reference to the size of the planning window for phonological encoding, predicting that variants that depend on following context, like liaison and flapping, can only be realized if the following context is sufficiently planned. Factors like lexical frequency and predictability (among many others), which delay retrieval and encoding, modulate the availability of the following context and therefore reduce the probability of planning the contextually-triggered variant. Future work could test the effect on external sandhi of a range of other factors that have been shown to modulate difficulty of speech production planning, such as syntactic complexity of an upcoming constituent (Ferreira, 1991), number of words in the utterance (Wheeldon and Lahiri, 2002; Wheeldon, 2012), and even codability[5] of the noun in an upcoming constituent (Griffin, 2001; Lee, Brown-Schmidt, and Watson, 2013).

[5]Codability refers to potential ambiguity between names for a noun, with more ambigious nouns being less codable. For example, an image of an apple is highly codable since there is only one clear name, *apple*, while a boat is less codable since it could be called by *ship, boat* or *sailboat* (Griffin, 2001, and references therein).

Accounting for production planning effects is important to understanding which parts of pronunciation variability are part of a speaker's knowledge of their language, and which parts are a consequence of general cognitive processes. The PPH provides a mechanism that can account for many probabilistic effects while maintaining a framework in which phonological processes simply apply when their structural description is met.

Chapter 5

Conclusion

This dissertation has investigated the variability observed in external sandhi phenomena, and tested the idea that it is structured by constraints on speech production planning. Case studies of three different external sandhi processes showed that prosodic, syntactic, functional, and lexical factors play a role in explaining the variation of cross-word processes. The effect of these factors was hypothesized to be mediated by online speech production planning constraints, according to the **Production Planning Hypothesis** (PPH) presented in Chapter 1. The PPH proposes a speech planning-based mechanism underlying phonological variability effects in external sandhi, offering a unified account of the gradient boundary effects and predictability effects presented in the results of Chapters 2, 3 and 4.

Section 5.1 presents a general discussion of the results from this thesis, including a brief summary of the results, with Section 5.1.1 considering prior accounts of locality effects, and Section 5.1.2 discussing previous research on variation. General implications of the PPH account and the findings are this thesis are presented in Section 5.2, along with directions for future research. Section 5.3 concludes the thesis.

5.1 General Discussion

The first case study, presented in Chapter 2, investigated the effects of boundary phenomena on high vowel devoicing (HVD) in spontaneous Japanese. The empirical results established that between voiceless consonants, the prototypical segmental environment for HVD (Fujimoto, 2015), there is more variability across word boundaries than within words. Results also showed that among HVD environments that spanned a word boundary, there was a consistently lower likelihood of devoicing as the strength of the boundary increased, and this effect was above and beyond the effect of pauses or boundary-related lengthening.

Pauses were also found to have a negative correlation with the likelihood of devoicing, but the magnitude of the effect was dependent on the strength of the prosodic boundary, with pauses decreasing HVD likelihood more drastically at weaker boundaries. It was also found that the effects of speech rate and lexical frequency on likelihood of devoicing were modulated by the type of prosodic boundary that followed the vowel.

Our analysis of this pattern was that there are two separate devoicing processes: one that is dependent on the voiceless consonant in the upcoming word, and one that is associated with strong prosodic boundaries. We suggested the overall inhibitory effect of prosodic boundaries on HVD can be explained by the PPH: stronger boundaries delay planning of the following word, making it less likely for the crucial voiceless consonant in the following word to be available in time to trigger planning of a voiceless vowel. However, this effect appears to be attenuated at strong boundaries with pauses because there is a separate devoicing process that applies exactly in that set of environments.

Chapter 3 examined the effects of two planning-related variables on the realization of flapping in North American English. Ferreira (1991) showed that syntactic complexity of the immediately upcoming constituent affected planning time, establishing the link between production planning and syntax. This was tested in the production experiment

in Chapter 3, which showed that flaps are more likely when the upcoming constituent is the object of the /t/-final verb rather than the subject of a new clause. Crucially, this effect was gradient, in the sense that a clause boundary did not totally block flapping, but only decreased its likelihood. The second planning-related factor tested in Chapter 3 was lexical frequency, an estimate of the global probability of a word's occurrence. Based on well-established facilitatory effects of lexical frequency on word form retrieval (Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994), it was predicted that the frequency of the trigger-containing word should have a positive correlation with flapping likelihood, a result which was borne out by analysis of flapping in the Buckeye Corpus of Conversational Speech.

The aim of Chapter 4 was to test whether probability effects are the same in a non-reductive external sandhi process, which the PPH predicts that they should be. The process examined was liaison in French, an alternation in which a latent consonant is realized before a vowel-initial following word. The effects of two probability-based measures were tested: lexical frequency of the following word, as in Chapter 3, and conditional probability of the following word given the target word. This was tested for two syntactic contexts: a plural noun followed by an adjective, and an adjective followed by a noun. Results from the plural noun-adjective context showed tentative support for positive correlations between the predictability measures and liaison, with the caveat that the effect was not confirmed to be generalizable across word types. The adjective-noun results showed a more robust positive effect of conditional probability on liaison. Consistent with PPH predictions, ease of retrieval/encoding of the trigger-containing word, as measured by predictability, has a positive correlation with sandhi application in liaison as well as in flapping.

5.1.1 Locality: variability at boundaries

Most frameworks that address locality effects do not model gradient differences between degrees of phono-syntactic locality. In Direct Reference theories, syntactic contexts either allow or block external sandhi application. Similarly, in Prosodic Phonology, a phonological process is bounded to a particular prosodic domain outside of which it cannot apply. These dichotomous accounts of locality alone are insufficient to describe the patterns shown in either Chapter 2 for HVD or in Chapter 3 for flapping. The results of the HVD study showed that there is a gradient difference in devoicing rate between at least four prosodically-based boundaries annotated in the corpus (which are likely highly correlated with syntactic constituency). These differences were found to be reliable even when phonetic factors like speech rate, pauses, and normalized durations were controlled. A gradient difference between two degrees of syntactic distance was also found in the flapping case study, where results showed a small but consistent difference in flapping rate depending on whether a clause boundary followed the potential flap. Under the PPH, these gradient effects can be understood as arising from the effect of syntax on the phonological encoding window.

Explaining gradient locality effects by reference to production planning does not preclude the existence of other direct morpho-syntactic effects. For example, French liaison might *also* be constrained by syntactic categories or structure, as proposed by Kaisse (1985) and Pak (2008). However, this type of account can be paired with a theory of online speech production planning constraints as a part of a complementary account of 'performance' factors that shape the data. This can yield a better understanding of which patterns of variability *grammatical* theories are responsible for, and which are due to grammar-external factors.

5.1.2 Variability: usage and function

The application of external sandhi processes is variable, and the results from Chapters 3 and 4 suggest that the variability is structured by factors related to language use: lexical frequency and conditional probability. How should these patterns bear on theories of phonological knowledge? This section discusses some influential views on this question, and how the PPH account of usage patterns fits with these views.

In this thesis, probability measures were tested as proxies for difficulty of advance planning. Chapter 3 tested lexical frequency effects in the realization of flaps in English, and Chapter 4 tested the effects of conditional probability on realization of liaison consonants in French. In both of these studies, it was found that higher predictability of the trigger-containing word makes the sandhi more likely to apply. Under the PPH, these effects can be explained by the same underlying mechanism: words that are more probable/predictable are planned sooner, more likely to be within the same planning window with the previous word (all else being equal), and therefore more likely to trigger the external sandhi process.

In the case of flapping, we explored an explanation in terms of probabilistic reduction, which associates phonetic reduction with increased word probability (Jurafsky et al., 2001). The effect of probability on flapping is argued to show that word probabilities must be represented and used in the calculation of pronunciations (Gregory et al., 1999). Under the PPH interpretation of the correlation of probability with likelihood of flapping, it would be possible to argue that at least part of the probability effect is indirect, via its effect on the scope of planning. The PPH provides a mechanism for probabilistic effects that makes explicit, testable predictions about the directions of the effects.

Probabilistic reduction accounts specifically address only reductive processes. The case study on liaison showed that similar predictability patterns exist for non-reductive

processes. Accounts of probability effects for these types of processes must draw on different mechanisms. The exemplar-based account of liaison of Bybee (2001) is based on the premise that bigrams that occur more frequently *resist* regularization to more general paradigms. In the case of liaison, the general paradigm is *no* alternation, and liaison is the 'irregular' pattern that is preserved by frequent usage.

The strong prediction of the PPH is that any process which depends on details in a later planning unit should be variable, with the likelihood of the process applying decreasing as the likelihood of the following information being available decreases. The scope of this prediction is very general: the external sandhi could be a reductive process or a non-reductive process, an assimilation or a dissimilation. The factors that could influence the availability of the following information are numerous, including syntactic-semantic facilitation or inhibition effects, since they are prior to phonological encoding, individual differences in planning scope, or task-specific effects on planning scope.

5.2 Implications and future directions

This thesis has shown that part of the variability observed in external sandhi processes can be understood as consequences of speech production planning constraints. Specifically, gradient inhibitory effects of syntactic-prosodic boundaries as well as usage-based predictability effects on variability show patterns predicted by the PPH. These two are factors that figure prominently in the literature on phonological variation, but PPH predictions extend to a number of other variables that could be tested in future work. It would be particularly interesting to test factors that have previously been associated with differences in planning scope, but that have not yet been associated with external sandhi prevalence. For example, cognitive load (Ferreira and Swets, 2002; Wagner, Jescheniak, and Schriefers, 2010), syntactic priming (Konopka, 2012), and task demands (Wheeldon

and Lahiri, 1997) have been argued to affect the scope of advance planning. These factors could all be tested in relation to sandhi application.

The PPH is built up from research on the scope of advance planning, but the literature on speech production planning has not looked at external sandhi phenomena in these experimental paradigms. Phonological alternations are highly relevant for diagnosing the scope of planning, as they by definition involve interactions between information at a distance. Hence, the application of a context-sensitive sandhi process could be used as a diagnostic for scope of phonological encoding.

For example, many studies of speech planning scope use onset latency as a measure of advance planning, with longer latencies indicating more advance planning. Griffin (2003) argues that production of two-word sequences have longer onset latencies when the first word is monosyllabic than when it is disyllabic. This is interpreted as there being more advance planning in the monosyllable condition: in order to avoid pauses between words, both words must be planned in advance. This interpretation could be supported by running a similar experiment that included conditions where external sandhi could apply between the two words to be uttered. If it is the case that longer onset latencies signal more advance planning, then longer onset latencies should also be correlated with more consistent application of external sandhi.

Directional asymmetry The PPH also makes interesting predictions about directional asymmetries. In the studies presented in this thesis, we have only investigated processes where there is a trigger *following* the target of the alternation, and the asynchrony between the planning of the target and following trigger leads to variability. However, if the trigger precedes the target, then triggering context will always have been planned before the target, and none of the production planning effects should hold. That is, the PPH predicts that only external sandhi with a *regressive* component should show sensitivity to production planning window effects, whereas processes that are purely *progressive*

will not demonstrate this type of variability, a prediction which should be tested in future work.[1]

There is some evidence of this regressive/progressive asymmetry in coarticulation from Whalen (1990). He tested whether coarticulation on vowels would still be realized if the trigger of the coarticulation was only presented after the onset of speech. Results showed that anticipatory coarticulations were mitigated when the triggering segment was unknown until after speech onset, while perserverative effects were not affected. Whalen ultimately argues that both types of coarticulation are planned, but does not develop an explanation for the asymmetry between perserverative and anticipatory processes. The PPH naturally explains this aspect of Whalen's results: even if both processes are planned, only in anticipatory coarticulation is the target of coarticulation planned in a window where phonological information about the 'trigger' is not yet known.

5.3 Conclusion

The three studies presented in this thesis investigated the patterns of variability in external sandhi, examining boundary-related and probability-related factors in particular. It was found that the effects of these variables can be understood as affecting the scope of online speech production planning, which in turn constrains the application of external sandhi. This approach offers a unified account of gradient syntactic-prosodic boundary effects as well as probability-related effects. Understanding the interaction between phonological computation and the process of online speech production planning can offer a clearer picture of other sources of variability in phonological patterns, including what kinds of variability should be directly modeled by phonological theory.

[1] There may in fact be a different class of production planning effects that apply in the case of progressive processes. Although a trigger that precedes the target is necessarily planned first, it may be subject to subject to memory decay, for example. Future work should test how long phonological material of an already-encoded word remains active, and subsequently PPH predictions could be tested for progressive phonological processes like assimilation or vowel harmony.

Bibliography

Anttila, Arto (1997). "Deriving variation from grammar". In: *Variation, change, and phonological theory*. Ed. by Frans Hinskens, Roeland van Hout, and W. Leo Wetzels. Current Issues in Linguistic Theory 146. Amsterdam/Philadelphia: John Benjamins, pp. 35–68.

Anttila, Arto (2007). "Variation and optionality". In: *The Cambridge handbook of phonology*. Ed. by Paul V. De Lacy. Cambridge: Cambridge University Press, pp. 519–536.

Aylett, Matthew and Alice Turk (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, pp. 31–56.

Baayen, R.H. (2008). *Analyzing linguistic data*. Cambridge: Cambridge University Press.

Banel, Marie-Hélène and Nicole Bacri (1994). "On metrical patterns and lexical parsing in French". *Speech Communication* 15.1, pp. 115–126.

Barnes, Jonathan (2006). *Strength and Weakness at the Interface: Positional Neutralization in Phonetics and Phonology*. Berlin: de Gruyter.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal". *Journal of Memory and Language* 68.3, pp. 255–278.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-5. URL: `http://CRAN.R-project.org/package=lme4`.

Beckman, Mary and Janet Pierrehumbert (1988). *Japanese tone structure*. Linguistic Inquiry Monographs 15. Cambridge: MIT Press.

Beckman, Mary E. (1996). "When is a syllable not a syllable?" In: *Phonological structure and language processing: Cross-linguistic studies*. Ed. by Takashi Otake and Anne Cutler. Berlin: Walter de Gruyter, pp. 95–123.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea (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, pp. 1001–1024.

Bermudez-Otero, Ricardo (2011). "Cyclicity". In: *The Blackwell companion to phonology*. Ed. by Marc van Ostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice. Oxford: Wiley-Blackwell.

Boersma, Paul (1997). "How we learn variation, optionality, and probability". In: *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*. Vol. 21, pp. 43–58.

Braver, Aaron (2011). "Incomplete neutralization in American English flapping: A production study". *University of Pennsylvania Working Papers in Linguistics* 17.1, pp. 31–40.

Breen, Mara, Duane G. Watson, and Edward Gibson (2011). "Intonational phrasing is constrained by meaning, not balance". *Language and Cognitive Processes* 26.10, pp. 1532–1562.

Browman, Catherine P. and Louis Goldstein (1992). "Articulatory phonology: An overview". *Phonetica* 49.3-4, pp. 155–180.

Brysbaert, Marc and Boris New (2009). "Moving beyond Kučera 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, pp. 977–990.

Bürki, Audrey, Ulrich H Frauenfelder, and F-Xavier Alario (2015). "On the resolution of phonological constraints in spoken production: Acoustic and response time evidence". *The Journal of the Acoustical Society of America* 138.4, pp. 429–434.

Bybee, Joan (1985). *Morphology: A study of the relation between meaning and form*. Typological Studies in Language 9. Amsterdam/Philadelphia: John Benjamins.

Bybee, Joan (2001). "Frequency effects on French liaison". In: *Frequency and the emergence of linguistic structure*. Ed. by Joan Bybee and Paul Hopper. Typological Studies in Language 45. Amsterdam: John Benjamins, pp. 337–360.

Bybee, Joan (2007). *Frequency of use and the organization of language*. New York: Oxford University Press.

Bybee, Joan and Joanne Scheibman (1999). "The effect of usage on degrees of constituency: the reduction of *don't* in English". *Linguistics* 37.4, pp. 575–596.

Byrd, Dani and Elliot Saltzman (2003). "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening". *Journal of Phonetics* 31, pp. 149–180.

Cedergren, Henrietta J. and David Sankoff (1974). "Variable rules: Performance as a statistical reflection of competence". *Language* 50.2, pp. 333–355.

Cedergren, Henrietta J. and Louise Simoneau (1985). "La chute des voyelles hautes en français de Montréal: As-tu entendu la belle syncope?" In: *Les tendances dynamiques du français parlé à Montréal*. Ed. by Monique Lemieux and Henrietta Cedergren. Québec: Bibliothèque nationale du Québec, pp. 57–144.

Chen, Matthew Y. (1987). "The syntax of Xiamen tone sandhi". *Phonology Yearbook* 4, pp. 109–49.

Chen, Matthew Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge Univeristy Press.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Coetzee, Andries W. and Shigeto Kawahara (2013). "Frequency biases in phonological variation". *Natural Language & Linguistic Theory* 31.1, pp. 47–89.

Coetzee, Andries W. and Joe Pater (2011). "The place of variation in phonological theory". In: *The handbook of phonological theory*. Ed. by John A. Goldsmith, Jason Riggle, and Alan C. L. Yu. Oxford: Wiley-Blackwell, pp. 401–434.

Cohen Priva, Uriel (2015). "Informativity affects consonant duration and deletion rates". *Laboratory Phonology* 6.2, pp. 243–278.

Cohen Priva, Uriel (2017). "Not so fast: Fast speech correlates with lower lexical and structural information". *Cognition* 160, pp. 27–34.

Cooper, William E. and Jeanne Paccia-Cooper (1980). *Syntax and speech*. Cognitive science series 3. Cambridge: Harvard University Press.

Côté, Marie-Hélène (2005). "Le statut lexical des consonnes de liaison". *Langages* 2, pp. 66–78.

Côté, Marie-Hélène (2011). "French liaison". In: *The Blackwell companion to phonology*. Ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Karen Rice. Oxford: Wiley-Blackwell.

Côté, Marie-Hélène (2013). "Understanding cohesion in French liaison". *Language Sciences* 39, pp. 156–166.

Côté, Marie-Hélène (2017). "La liaison en diatopie: esquisse d'une typologie". *Journal of French Language Studies* 27.1, pp. 13–25.

Crothers, John H., James Lorentz, Donald Sherman, and Marilyn Vihman (1979). *Handbook of phonological data from a sample of the world's languages: a report of the Stanford Phonology Archive*. Department of Linguistics, Stanford University.

Cutler, Anne and Sally Butterfield (1992). "Rhythmic cues to speech segmentation: Evidence from juncture misperception". *Journal of Memory and Language* 31.2, pp. 218–236.

Cutler, Anne, Delphine Dahan, and Wilma Van Donselaar (1997). "Prosody in the comprehension of spoken language: A literature review". *Language and Speech* 40.2, pp. 141–201.

Cutler, Anne and Dennis Norris (1988). "The role of strong syllables in segmentation for lexical access". *Journal of Experimental Psychology: Human Perception and Performance* 14.1, pp. 113–121.

Dauer, Rebecca M. (1980). "The reduction of unstressed high vowels in Modern Greek". *Journal of the International Phonetic Association* 10.1, pp. 17–27.

Davis, Matthew H., William D. Marslen-Wilson, and M. Gareth Gaskell (2002). "Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition." *Journal of Experimental Psychology: Human Perception and Performance* 28.1, pp. 218–244.

De Jong, Kenneth (1998). "Stress-related variation in the articulation of coda alveolar stops: Flapping revisited". *Journal of Phonetics* 26.3, pp. 283–310.

Delattre, Pierre (1955). "Les facteurs de la liaison facultative en français". *The French Review* 29.1, pp. 42–49.

Delattre, Pierre (1966). *Studies in French and comparative phonetics: selected papers in French and English*. The Hague: Mouton.

Dell, Gary S. and Padraig G. O'Seaghdha (1992). "Stages of lexical access in language production". *Cognition* 42.1-3, pp. 287–314.

Den, Yasuharu (2015). "Some phonological, syntactic, and cognitive factors behind phrase-final lengthening in spontaneous Japanese: A corpus-based study". *Laboratory Phonology* 6.3-4, pp. 337–379.

Den, Yasuharu and Hanae Koiso (2015). "Factors affecting utterance-final vowel devoicing in spontaneous Japanese". In: *Proceedings of the 15th International Congress of Phonetic Sciences*. Ed. by The Scottish Consortium for ICPhS 2015. Paper 582. Glasgow: University of Glasgow.

Durand, Jacques, Bernard Laks, and Chantal Lyche (2002). "La phonologie du français contemporain: usages, variétés et structures". In: *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language.* Ed. by Claud Pusch and Wolfgang Raible. Tübingen: Gunter Narr, pp. 93–106.

Durand, Jacques, Bernard Laks, and Chantal Lyche (2009). "Le projet PFC (phonologie du français contemporain): une source de données primaires structurées". In: *Phonologie, variation et accents du français*. Paris: Hermès, pp. 19–61.

Durand, Jacques and Chantal Lyche (2003). "Le projet 'Phonologie du Français Contemporain'(PFC) et sa méthodologie". In: *Corpus et variation en phonologie du français*. Ed.

by Élisabeth Delais-Roussaire and Jacques Durand. Toulouse: Presses Universitaires du Mirail, pp. 213–276.

Durand, Jacques and Chantal Lyche (2008). "French liaison in the light of corpus data". *Journal of French Language Studies* 18.01, pp. 33–66.

Eddington, David and Caitlin Channer (2010). "American English has goʔ a loʔ of glottal stops: Social diffusion and linguistic motivation". *American Speech* 85.3, pp. 338–351.

Eddington, David and Dirk Elzinga (2008). "The phonetic context of American English flapping: Quantitative evidence". *Language and Speech* 51.3, pp. 245–266.

Ernestus, Mirjam, Mybeth Lahey, Femke Verhees, and R. Harald Baayen (2006). "Lexical frequency and voice assimilation". *The Journal of the Acoustical Society of America* 120.2, pp. 1040–1051.

Ferreira, Fernanda (1988). "Planning and timing in sentence production: The syntax-to-phonology conversion". PhD thesis. University of Massachusetts at Amherst.

Ferreira, Fernanda (1991). "Effects of length and syntactic complexity on initiation times for prepared utterances". *Journal of Memory and Language* 30.2, pp. 210–233.

Ferreira, Fernanda (1993). "Creation of prosody during sentence production." *Psychological Review* 100.2, pp. 233–253.

Ferreira, Fernanda and Benjamin Swets (2002). "How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums". *Journal of Memory and Language* 46, pp. 57–84.

Féry, Caroline (2004). "Gradient prosodic correlates of phrasing in French". In: *Nouveaux départs en phonologie*. Ed. by Trudel Meisenburg and Maria Selig. Tübingen: Gunter Narr, pp. 161–182.

Fosler-Lussier, Eric and Nelson Morgan (1999). "Effects of speaking rate and word frequency on pronunciations in convertional speech". *Speech Communication* 29.2, pp. 137–158.

Fougeron, Cécile and Patricia A. Keating (1997). "Articulatory strengthening at edges of prosodic domains". *The Journal of the Acoustical Society of America* 101.6, pp. 3728–3740.

Fox, Robert A. and Dale Terbeek (1977). "Dental flaps, vowel duration and rule ordering in American English". *Journal of Phonetics* 5, pp. 27–34.

Fromkin, Victoria A. (1971). "The non-anomalous nature of anomalous utterances". *Language* 47.1, pp. 27–52.

Fujimoto, Masako (2015). "Vowel devoicing". In: *The Handbook of Japanese Phonetics and Phonology*. Ed. by Haruo Kubozono. Berlin: Mouton de Gruyter, pp. 167–214.

Fukaya, Teruhiko and Dani Byrd (2005). "An articulatory examination of word-final flapping at phrase edges and interiors". *Journal of the International Phonetic Association* 35.1, pp. 45–58.

Gahl, Susanne and Susan M. Garnsey (2004). "Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation". *Language*, pp. 748–775.

Garrett, Merrill F. (1988). "Processes in language production". In: *Linguistics: The Cambridge survey: Volume 3. Language: Psychological and biological aspects*. Ed. by Frederick J. Newmeyer. Cambridge University Press, pp. 69–96.

Gaskell, M. Gareth and William D. Marslen-Wilson (2002). "Representation and competition in the perception of spoken words". *Cognitive Psychology* 45.2, pp. 220–266.

Gaskell, M Gareth, Elsa Spinelli, and Fanny Meunier (2002). "Perception of resyllabification in French". *Memory & Cognition* 30.5, pp. 798–810.

Gelman, Andrew and Jennifer Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Goldwater, Sharon and Mark Johnson (2003). "Learning OT constraint rankings using a maximum entropy model". In: *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl. Stockholm: Department of Linguistics, Stockholm University, pp. 111–120.

Gordon, Matthew (1998). "The phonetics and phonology of non-modal vowels: a cross-linguistic perspective". In: *Proceedings of the Twenty-Fourth Annual Meeting of the Berkeley Linguistics Society*. Ed. by Benjamin Bergin, Madeleine Plauché, and Ashlee Bailey. Berkley: Berkley Linguistics Society, pp. 93–105.

Gorman, Kyle, Jonathan Howell, and Michael Wagner (2011). "Prosodylab-aligner: A tool for forced alignment of laboratory speech". *Canadian Acoustics* 39.3, pp. 192–193.

Graff, David (2006). *Gigaword First Edition LDC2006T17*. URL: `https://catalog.ldc.upenn.edu/LDC2006T17`.

Gregory, Michelle, William D. Raymond, Alan Bell, Eric Fosler-Lussier, and Daniel Jurafsky (1999). "The effects of collocational strength and contextual predictability in lexical production". In: *Proceedings of the 35th Meeting of the Chicago Linguistic Society*. Ed. by Sabrina Billings, John Boyle, and Aaron Griffith, pp. 151–166.

Griffin, Zenzi M (2001). "Gaze durations during speech reflect word selection and phonological encoding". *Cognition* 82.1, B1–B14.

Griffin, Zenzi M (2003). "A reversed word length effect in coordinating the preparation and articulation of words in speaking". *Psychonomic Bulletin & Review* 10.3, pp. 603–609.

Gussenhoven, Carlos (1986). "English plosive allophones and ambisyllabicity". *Gramma* 10, pp. 119–141.

Guy, Gregory R. (1991). "Explanation in variable phonology: An exponential model of morphological constraints". *Language Variation and Change* 3.1, pp. 1–22.

Guy, Gregory R. (2011). "Variability". In: *The Blackwell companion to phonology*. Ed. by Marc van Oostendorp, Colin J Ewen, Elizabeth V Hume, and Keren Rice. Vol. IV. Oxford: Wiley-Blackwell, pp. 2109–2213.

Hall, Kathleen Currie, Elizabeth Hume, T. Florian Jaeger, and Andrew Wedel (2016). "The message shapes phonology". Ms. Univ. of British Columbia, Univ. of Canterbury, Univ. of Rochester, Univ. of Arizona.

Han, Mieko Shimizu (1962). "Unvoicing of vowels in Japanese". *Onsei no kenkyuu* 10, pp. 81–100.

Harrell Jr, Frank E. (2014). *rms: Regression Modeling Strategies*. R package version 4.2-0. URL: http://CRAN.R-project.org/package=rms.

Harrell Jr, Frank E. et al. (2015). *Hmisc: Harrell Miscellaneous*. R package version 3.17-1. URL: https://CRAN.R-project.org/package=Hmisc.

Hasegawa, Nobuko (1979). "Fast speech vs. casual speech". In: *Papers from the Fifteenth Regional Meeting of the Chicago Linguistics Society*, pp. 126–137.

Hayes, Bruce and Zsuzsa Cziráky Londe (2006). "Stochastic phonological knowledge: The case of Hungarian vowel harmony". *Phonology* 23.1, pp. 59–104.

Hayes, Bruce and Colin Wilson (2008). "A maximum entropy model of phonotactics and phonotactic learning". *Linguistic Inquiry* 39.3, pp. 379–440.

Hayes, Bruce, Colin Wilson, and Anne Shisko (2012). "Maxent grammars for the metrics of Shakespeare and Milton". *Language* 88.4, pp. 691–731.

Heafield, Kenneth, Ivan Pouzyrevsky, Jonathan H. Clark, and Philipp Koehn (2013). "Scalable Modified Kneser-Ney Language Model Estimation". In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria, pp. 690–696. URL: {http://kheafield.com/professional/edinburgh/estimate_paper.pdf}.

Herd, Wendy, Allard Jongman, and Joan Sereno (2010). "An acoustic and perceptual analysis of /t/ and /d/ flaps in American English". *Journal of Phonetics* 38.4, pp. 504–516.

Hirayama, Manami (2009). "Postlexical prosodic structure and vowel devoicing in Japanese". PhD thesis. University of Toronto.

Hirayama, Teruo (1985). "Zennippon no hatsuon to akusento [Pronunciation and accent in all Japan]". In: *Nihongo Hatsuon Akusento Jiten [Japanese Pronunciation Accent Dictionary]*. Ed. by Nihon Hoosoo Kyookai. Tokyo: Nihon Hoosoo Shuppan Kyookai, pp. 37–69.

Holst, Tara and Francis Nolan (1995). "The influence of syntactic structure on [s] to [ʃ] assimilation". In: *Phonology and phonetic evidence: Papers in Laboratory Phonology IV*. Ed. by Bruce Connell and Amelia Arvaniti. Cambridge: Cambridge University Press, pp. 315–333.

Hualde, José Ignacio (2013). "Intervocalic lenition and word-boundary effects: evidence from Judeo-Spanish". *Diachronica* 30.2, pp. 232–266.

Igarashi, Yosuke, Hideaki Kikuchi, and Kikuo Maekawa (2006). "Inritsu joohoo [Prosodic information]". In: *Nihongo hanashi kotoba koopasu no koochikuhoo [Construction of The Corpus of Spontaneous Japanese], Kokuritsu Kokugo Kenkyuujo [National Institute for Japanese Language (NIJAL)] Report 124*, pp. 347–453.

Imai, Terumi (2004). "Vowel devoicing in Tokyo Japanese: A variationist approach". PhD thesis. Michigan State University.

Inkelas, Sharon and Draga Zec (1990). *The phonology-syntax connection*. Chicago: University of Chicago Press.

Inkelas, Sharon and Draga Zec (1995). "Syntax-phonology interface". In: *The handbook of phonological theory*. Ed. by John A. Goldsmith. Oxford: Basil Blackwell, pp. 535–549.

Ito, Junko and Armin Mester (2012). "Recursive prosodic phrasing in Japanese". In: *Prosody matters: Essays in honor of Elisabeth Selkirk*. Ed. by Tony Borowsky, Shigeto Kawahara, Mariko Sugahara, and Takahito Shinya. Sheffield & Bristol, Conn.: Equinox, pp. 280–303.

Jaeger, T. Florian (2010). "Redundancy and reduction: Speakers manage syntactic information density". *Cognitive Psychology* 61.1, pp. 23–62.

Jäger, Gerhard (2007). "Maximum entropy models and stochastic Optimality Theory". In: *Architectures, rules, and preferences: variations on themes by Joan W. Bresnan*. Ed. by Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Chris Manning. Stanford: CSLI Publications, pp. 467–479.

Jannedy, Stefanie (1995). "Gestural phasing as an explanation for vowel devoicing in Turkish". *Ohio State University Working Papers in Linguistics* 45, pp. 56–84.

Jescheniak, Jörg D. and Willem J. M. Levelt (1994). "Word frequency effects in speech production: Retrieval of syntactic information and of phonological form". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.4, p. 824.

Jun, Sun-Ah and Mary Beckman (1993). *A gestural-overlap analysis of vowel devoicing in Japanese and Korean*. Paper presented at the 1993 Annual Meeting of the LSA, Los Angeles, January 7–10.

Jurafsky, Dan, Alan Bell, Michelle Gregory, and William D. Raymond (2001). "Probabilistic Relations Between Words: Evidence from reduction in lexical production". In: *Frequency and the emergence of linguistic structure*. Ed. by Joan Bybee and Paul Hopper. Typological Studies in Language 45. Amsterdam: John Benjamins, pp. 229–254.

Kahn, David (1976). *Syllable-based generalizations in English phonology*. New York: Garland.

Kaimaki, Marianna (2015). "Voiceless Greek vowels". In: *Proceedings of the 15th International Congress of Phonetic Sciences*. Ed. by The Scottish Consortium for ICPhS 2015. Paper 791. Glasgow: University of Glasgow.

Kaisse, Ellen M. (1985). *Connected speech: the interaction of syntax and phonology*. Orlando: Academic Press.

Kaisse, Ellen M. and Patricia A. Shaw (1985). "On the theory of Lexical Phonology". *Phonology* 2.01, pp. 1–30.

Katz, Jonah (2016). "Lenition, perception and neutralization". *Phonology* 33.1, pp. 43–85.

Kawahara, Shigeto (2011). "Japanese loanword devoicing revisited: A rating study". *Natural Language & Linguistic Theory* 29.3, pp. 705–723.

Keating, Patricia A. (2006). "Phonetic encoding of prosodic structure". In: *Speech production: Models, phonetic processes, and techniques*. Ed. by Jonathan Harrington and Marija Tabain. New York: Psychology Press, pp. 167–186.

Keating, Patricia A. and Stefanie Shattuck-Hufnagel (2002). "A prosodic view of word form encoding for speech production". *UCLA Working Papers in Phonetics* 101, pp. 112–156.

Kiesling, Scott, Laura Dilley, and William D. Raymond (2006). "The variation in conversation (ViC) project: Creation of the Buckeye Corpus of Conversational Speech". Ms., Department of Psychology, Ohio State University, Columbus, OH. URL: http://buckeyecorpus.osu.edu/BuckeyeCorpusmanual.pdf.

Kilbourn-Ceron, Oriana (2015). "The influence of prosodic context on high vowel devoicing in spontaneous Japanese". URL: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0932.pdf.

Kilbourn-Ceron, Oriana (2017). "Speech production planning affects phonological variability: a case study in French liaison". In: *Proceedings of the 2016 Annual Meeting on*

Phonology. URL: http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4004.

Kilbourn-Ceron, Oriana and Morgan Sonderegger (2017). "Boundary phenomena and variability in Japanese high vowel devoicing". *Natural Language & Linguistic Theory*, pp. 1–43. URL: https://link.springer.com/article/10.1007/s11049-017-9368-x.

Kilbourn-Ceron, Oriana, Michael Wagner, and Meghan Clayards (2016). "The effect of production planning locality on external sandhi: a study in /t/". In: *The proceedings of the 52nd Meeting of the Chicago Linguistics Society*. lingbuzz/003119.

Kingston, John (2008). "Lenition". In: *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Ed. by L. Colantoni and J. Steele. Somerville, MA: Cascadilla Press, pp. 1–31.

Kiparsky, Paul (1979). "Metrical structure assignment is cyclic". *Linguistic Inquiry* 10, pp. 421–441.

Kiparsky, Paul (1982). "Lexical phonology and morphology". In: *Linguistics in the Morning Calm*. Ed. by Linguistic Society of Korea. Seoul: Hansin, pp. 3–35.

Kiparsky, Paul (1985). "Some consequences of lexical phonology". *Phonology* 2.1, pp. 85–138.

Kondo, Mariko (1997). "Mechanisms of vowel devoicing in Japanese". PhD thesis. University of Edinburgh.

Kondo, Mariko (2005). "Syllable structure and its acoustic effects on vowels in devoicing environments". In: *Voicing in Japanese*. Ed. by Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara. Studies in Generative Grammar 84. Berlin: Walter de Gruyter, pp. 229–246.

Konopka, Agnieszka E. (2012). "Planning ahead: How recent experience with structures and words changes the scope of linguistic planning". *Journal of Memory and Language* 66.1, pp. 143–162.

Kowal, Sabine, Mary R. Bassett, and Daniel C. O'Connell (1985). "The spontaneity of media interviews". *Journal of Psycholinguistic Research* 14.1, pp. 1–18.

Krivokapić, Jelena (2007). "Prosodic planning: Effects of phrasal length and complexity on pause duration". *Journal of Phonetics* 35.2, pp. 162–179.

Krivokapić, Jelena and Dani Byrd (2012). "Prosodic boundary strength: An articulatory and perceptual study". *Journal of Phonetics* 40.3, pp. 430–442.

Kuriyagawa, Fukuko and Masayuki Sawashima (1989). "Word accent, devoicing and duration of vowels in Japanese". *Annual Bulletin of the Research Institute of Language Processing* 23, pp. 85–108.

Kuwabara, Hisao and Kazuya Takeda (1988). "Analysis and prediction of vowel devocalization in isolated Japanese words". *The Journal of the Acoustical Society of America* 83.S1, S29–S29.

Labov, William (1969). "Contraction, deletion, and inherent variability of the English copula". *Language*, pp. 715–762.

Labrune, Laurence (2012). *The phonology of Japanese*. Oxford: Oxford University Press.

Laks, Bernard (2009). "Dynamiques de la liaison en français". In: *Le français d'un continent à l'autre*. Ed. by Luc Baronian and France Martineau. Presses de l'Université de Laval, pp. 237–267.

Lee, Eun-Kyung, Sarah Brown-Schmidt, and Duane G Watson (2013). "Ways of looking ahead: Hierarchical planning in language production". *Cognition* 129.3, pp. 544–562.

Lehiste, Ilse (1960). "An acoustic–phonetic study of internal open juncture". *Phonetica* 5.Suppl. 1, pp. 5–54.

L'Esperance, Marie-Josee (2015). "The Phonetics and Phonology of Liaison Consonants in Montreal French". MA thesis. Cornell University.

Levelt, Willem J. M. (1989). *Speaking. From Intention to Articulation*. Cambridge: MIT Press.

Levelt, Willem J.M., Ardi Roelofs, and Antje S. Meyer (1999). "A theory of lexical access in speech production". *Behavioral and Brain Sciences* 22.1, pp. 1–38.

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

Lindblom, Björn (1995). "A view of the future of phonetics". In: *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm: University of Stockholm*, pp. 462–9.

Lindsley, James R. (1975). "Producing simple utterances: How far ahead do we plan?" *Cognitive Psychology* 7.1, pp. 1–19.

Lombardi, Linda (1991). "Laryngeal features and laryngeal neutralization". PhD thesis. University of Massachusetts.

Lovins, Julie B (1976). "Pitch accent and vowel devoicing in Japanese A preliminary study". *Annual Bulletin of Research Institute of Logopedics and Phoniatrics, University of Tokyo* 10, pp. 113–125.

MacKenzie, Laurel (2013). "Variation in English auxiliary realization: A new take on contraction". *Language Variation and Change* 25.01, pp. 17–41.

Maekawa, Kikuo and Hideaki Kikuchi (2005). "Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report". In: *Voicing in Japanese*. Ed. by Jeroen van de Weijer, Kensuke Nanjo, and Tetsuo Nishihara. Studies in Generative Grammar 84. Berlin: Mouton de Gruyter, pp. 205–228.

Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara (2000). "Spontaneous Speech Corpus of Japanese." In: *Proceedings of the Second International Conference of Language Resources and Evaluation (LREC)*. Vol. 2, pp. 947–952.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi, and Jennifer J Venditti (2002). "X-JToBI: an extended J-ToBI for spontaneous speech." In: *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP2002)*. Denver, pp. 1545–1548.

Malécot, André and Paul Lloyd (1968). "The/t/ : /d/ distinction in American alveolar flaps". *Lingua* 19.3-4, pp. 264–272.

Mallet, Géraldine-Mary (2008). "La liaison en français: descriptions et analyses dans le corpus PFC". PhD thesis. Paris 10.

Martin, Andrew (2011). "Grammars leak: Modeling how phonotactic generalizations interact within the grammar". *Language* 87.4, pp. 751–770.

Martin, Andrew, Akira Utsugi, and Reiko Mazuka (2014). "The multidimensional nature of hyperspeech: Evidence from Japanese vowel devoicing". *Cognition* 132.2, pp. 216–228.

Mattys, Sven L. and James F. Melhorn (2007). "Sentential, lexical, and acoustic effects on the perception of word boundaries". *The Journal of the Acoustical Society of America* 122.1, pp. 554–567.

Mattys, Sven L., James F. Melhorn, and Laurence White (2007). "Effects of syntactic expectations on speech segmentation." *Journal of Experimental Psychology: Human Perception and Performance* 33.4, p. 960.

Mattys, Sven L., Laurence White, and James F. Melhorn (2005). "Integration of multiple speech segmentation cues: a hierarchical framework". *Journal of Experimental Psychology: General* 134.4, p. 477.

McCawley, James D. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.

McClelland, James L. and Jeffrey L. Elman (1986). "The TRACE model of speech perception". *Cognitive Psychology* 18.1, pp. 1–86.

McQueen, James M., Anne Cutler, Ted Briscoe, and Dennis Norris (1995). "Models of continuous speech recognition and the contents of the vocabulary". *Language and Cognitive Processes* 10.3-4, pp. 309–331.

Michelson, Karin (1999). "Utterance-final phenomena in Oneida". In: *Proceedings of LP'98: Item and order in language and speech*. Ed. by Osamu Fujimura, Brian D. Joseph, and Bohumil Palek. Prague: Charles University Press, pp. 31–45.

Mitchell, Heather L., Jeannette D. Hoit, and Peter J. Watson (1996). "Cognitive-linguistic demands and speech breathing". *Journal of Speech, Language, and Hearing Research* 39.1, pp. 93–104.

Mohanan, Karuvannur Puthanveettil (1982). "Lexical phonology". PhD thesis. Massachusetts Institute of Technology.

Mohanan, Karuvannur Puthanveettil (1986). *The theory of lexical phonology*. Dordrecht: D. Reidel.

Morin, Yves-Charles (2003). "Remarks on prenominal liaison consonants in French". In: *Living on the Edge: 28 Papers in Honour of Jonathan Kaye*. Ed. by Stefan Ploch. Studies in Generative Grammar 64. Berlin: Mouton de Gruyter, pp. 385–400.

Morin, Yves-Charles and Jonathan D. Kaye (1982). "The syntactic bases for French liaison". *Journal of Linguistics* 18.2, pp. 291–330.

Nagy, Naomi and Bill Reynolds (1997). "Optimality Theory and variable word-final deletion in Faetar". *Language Variation and Change* 9.1, pp. 37–55.

Nakatani, Lloyd H. and Kathleen D. Dukes (1977). "Locus of segmental cues for word juncture". *The Journal of the Acoustical Society of America* 62.3, pp. 714–719.

Nespor, Marina and Irene Vogel (1986). *Prosodic phonology*. Berlin: Walter de Gruyter.

New, Boris, Christophe Pallier, Ludovic Ferrand, and Rafael Matos (2001). "Une base de données lexicales du français contemporain sur internet: LEXIQUE". *L'Année Psychologique* 101, pp. 447–462.

Newell, Heather (2008). "Aspects of the morphology and phonology of phases". PhD thesis. McGill University.

Newell, Heather and Glyne Piggott (2014). "Interactions at the syntax–phonology interface: Evidence from Ojibwe". *Lingua* 150, pp. 332–362.

NHK (1991). *Nihongo Hatsuon Akusento Jiten (Japanese Pronunciation Accent Dictionary)*. Tokyo: Nihon Hoosoo.

Nielsen, Kuniko Y. (2015). "Continuous versus categorical aspects of Japanese consecutive devoicing". *Journal of Phonetics* 52, pp. 70–88.

Nolan, Francis, Tara Holst, and Barbara Kühnert (1996). "Modelling [s] to [ʃ] accommodation in English". *Journal of Phonetics* 24.1, pp. 113–137.

Norris, Dennis (1994). "Shortlist: A connectionist model of continuous speech recognition". *Cognition* 52.3, pp. 189–234.

Odden, David (1987). "Kimatuumbi phrasal phonology". *Phonology* 4.1, pp. 13–36.

Ogasawara, Naomi (2013). "Lexical representation of Japanese vowel devoicing". *Language and Speech* 56.1, pp. 5–22.

Oi, Mutsumi (2013). "The interaction between accent contrast and vowel devoicing in Tokyo Japanese". MA thesis. University of Ottawa.

Oldfield, Richard C. and Arthur Wingfield (1964). "The time it takes to name an object". *Nature* 202, pp. 1031–1032.

Oldfield, Richard C. and Arthur Wingfield (1965). "Response latencies in naming objects". *Quarterly Journal of Experimental Psychology* 17.4, pp. 273–281.

Pak, Marjorie (2008). "The postsyntactic derivation and its phonological reflexes". PhD thesis. University of Pennsylvania.

Pak, Marjorie and Michael Friesner (2006). "French phrasal phonology in a derivational model of PF". In: *Proceedings of NELS 36*. Ed. by Christopher Davis, Amy Rose Deal, and Youri Zabbal. Amherst: GLSA University of Massachusetts Amherst, pp. 480–491.

Patterson, David and Cynthia M Connine (2001). "Variant frequency in flap production". *Phonetica* 58.4, pp. 254–275.

Pierrehumbert, Janet B. (2001). "Exemplar dynamics: Word frequency, lenition and contrast". In: *Frequency and the emergence of linguistic structure*. Ed. by Joan Bybee and Paul Hopper. Typological Studies in Language 45. Amsterdam: John Benjamins, pp. 137–158.

Pierrehumbert, Janet B. (2006). "The next toolkit". *Journal of Phonetics* 34.4, pp. 516–530.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume, and Eric Fosler-Lussier (2007). *Buckeye corpus of conversational speech (2nd release)*. Columbus, OH: Department of Psychology, Ohio State University (Distributor). URL: `www.buckeyecorpus.osu.edu`.

Pluymaekers, Mark, Mirjam Ernestus, and R. Harald Baayen (2005). "Lexical frequency and acoustic reduction in spoken Dutch". *The Journal of the Acoustical Society of America* 118.4, pp. 2561–2569.

Post, Brechtje Maria Bowine (2000). *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, and Michael Becker (2010). "Harmonic Grammar with linear programming: From linear systems to linguistic typology". *Phonology* 27.01, pp. 77–117.

Price, Patti J., Mari. Ostendorf, Stefanie Shattuck-Hufnagel, and Cynthia Fong (1991). "The use of prosody in syntactic disambiguation". *Journal of the Acoustical Society of America* 90.6, pp. 2956–2970.

Purse, Ruaridh and Alice Turk (2016). "'/t, d/Deletion': Articulatory gradience in variable phonology". In: *The 15th Annual Meeting of the Society for Laboratory Phonology*. Cornell Universtiy, Ithaca, NY.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Randolph, Mark A. (1989). "Syllable-based constraints on properties of speech sounds". PhD thesis. MIT.

Reynolds, William Thomas (1994). "Variation and phonological theory". PhD thesis. University of Pennsylvania.

Salverda, Anne Pier, Delphine Dahan, and James M McQueen (2003). "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension". *Cognition* 90.1, pp. 51–89.

Scheer, Tobias (2012). *Direct Interface and One-Channel Translation*. Studies in Generative Grammar 68. Boston: Mouton de Gruyter.

Schuppler, Barbara, Wim A van Dommelen, Jacques Koreman, and Mirjam Ernestus (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, pp. 595–607.

Scott, Donia R. and Anne Cutler (1984). "Segmental phonology and the perception of syntactic structure". *Journal of Verbal Learning and Verbal Behavior* 23.4, pp. 450–466.

Seidl, Amanda (2001). *Minimal indirect reference: A theory of the syntax-phonology interface*. Outstanding dissertations in Linguistics. New York: Routledge.

Selkirk, Elisabeth (1974). "French liaison and the X' notation". *Linguistic Inquiry* 5.4, pp. 573–590.

Selkirk, Elisabeth (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.

Selkirk, Elisabeth (1986). "On derived domains in sentence phonology". *Phonology Yearbook* 3, pp. 371–405.

Selkirk, Elisabeth (2011). "The syntax-phonology interface". In: *The Handbook of Phonological Theory*. Ed. by John Goldsmith, Jason Riggle, and Alan Yu. Vol. 2. Wiley-Blackwell, pp. 435–483.

Shattuck-Hufnagel, Stefanie (1979). "Speech errors as evidence for a serial-ordering mechanism in sentence production". In: *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Ed. by William E. Cooper and Edward Walker. Hillsdale, NJ: L. Erlbaum Associates, pp. 295–342.

Shattuck-Hufnagel, Stefanie (2000). "Phrase-level phonology in speech production planning: Evidence for the role of prosodic structure". In: *Prosody: Theory and experiment: Studies Presented to Gösta Bruce*. Ed. by Merle Horne. Dordrecht: Springer, pp. 201–230.

Shattuck-Hufnagel, Stefanie (2015). "Prosodic Frames in Speech Production". In: *The handbook of speech production*. Ed. by Melissa A. Redford. Wiley Online Library.

Shih, Stephanie S. (2016). "Super additive similarity in Dioula tone harmony". In: *Proceedings of the 33rd West Coast Conference on Formal Linguistics*. Ed. by Kyeong min Kim et al. Somerville, Mass.: Cascadilla Proceedings Project, pp. 361–370.

Smith, Caroline L. (2003). "Vowel devoicing in contemporary French". *Journal of French Language Studies* 13.2, pp. 177–194.

Smolensky, Paul and Matthew Goldrick (2016). "Gradient Symbolic Representations in Grammar: The case of French Liaison". Ms. Johns Hopkins University and Northwestern University.

Smolensky, Paul and Géraldine Legendre (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge: MIT Press.

Snijders, Tom A.B. and Roel J. Bosker (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications. ISBN: 9780761958901.

Sohn, Ho-min (1975). *Woleaian reference grammar*. Honolulu: Univ. of Hawaii Press.

Spinelli, Elsa, James M. McQueen, and Anne Cutler (2003). "Processing resyllabified words in French". *Journal of Memory and Language* 48.2, pp. 233–254.

Steriade, Donca (2008). "The phonology of perceptibility effects: The P-map and its consequences for constraint organization". In: *The nature of the word: studies in honor of Paul Kiparsky*. Ed. by Kristin Hanson and Sharon Inkelas. Cambridge: MIT Press, pp. 151–179.

Sternberg, Saul, Stephen Monsell, Ronald L. Knoll, and Charles E. Wright (1978). "The latency and duration of rapid movement sequences: Comparisons of speech and typewriting". In: *Information processing in motor control and learning*. Ed. by George Stelmach. New York: Academic Press, pp. 117–152.

Sternberg, Saul, Ronald L. Knoll, Stephen Monsell, and Charles E. Wright (1988). "Motor programs and hierarchical organization in the control of rapid speech". *Phonetica* 45.2-4, pp. 175–197.

Stevens, Mary (2012). "A phonetic investigation into"raddoppiamento sintattico" in Sienese Italian". PhD thesis. University of Melbourne.

Šurkalović, Dragana (2016). "The No-Reference Hypothesis: A modular approach to the syntax-phonology interface". PhD thesis. University of Tromsø.

Swets, Benjamin, Matthew E. Jacovina, and Richard J. Gerrig (2014). "Individual differences in the scope of speech planning: Evidence from eye-movements". *Language and Cognition* 6.1, pp. 12–44.

Takeda, Kazuya, Yoshinori Sagisaka, and Hisao Kuwabara (1989). "On sentence-level factors governing segmental duration in Japanese". *The Journal of the Acoustical Society of America* 86.6, pp. 2081–2087.

Tamminga, Meredith, Laurel MacKenzie, and David Embick (2016). "The dynamics of variation in individuals". *Linguistic Variation* 16.2, pp. 300–336.

Tanner, James, Morgan Sonderegger, and Michael Wagner (2017). *Production planning and coronal stop deletion in spontaneous speech*. Ms. to appear in *Laboratory Phonology*.

Torreira, Francisco and Mirjam Ernestus (2009). "Probabilistic effects on French [t] duration". In: *10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*. Causal Productions Pty Ltd., pp. 448–451.

Torreira, Francisco and Mirjam Ernestus (2011). "Vowel elision in casual French: The case of vowel /e/ in the word *c'était*". *Journal of Phonetics* 39.1, pp. 50–58.

Tsuchida, Ayako (1997). "Phonetics and phonology of Japanese vowel devoicing". PhD thesis. Cornell University.

Tsuchida, Ayako (2001). "Japanese vowel devoicing: Cases of consecutive devoicing environments". *Journal of East Asian Linguistics* 10.3, pp. 225–245.

Turk, Alice (1992). "The American English flapping rule and the effect of stress on stop consonant durations". *Working papers of the Cornell phonetics laboratory* 7, pp. 103–133.

Turk, Alice (2010). "Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis". *Laboratory Phonology* 1.2, pp. 227–262.

Vance, Timothy J. (1992). "Lexical phonology and Japanese vowel devoicing". In: *The joy of grammar: a festschrift in honor of James D. McCawley*. Ed. by Gary N. Larson, Lynn A. MacLeod, James D. McCawley, and Diane Brentari. Amsterdam: John Benjamins, pp. 337–350.

Vance, Timothy J. (2008). *The Sounds of Japanese*. Cambridge: Cambridge University Press.

Varden, John Kevin (1998). "On high vowel devoicing in standard modern Japanese: Implications for current phonological theory". PhD thesis. University of Washington.

Varden, John Kevin (2010). "On Vowel Devoicing in Japanese". *The MGU Journal of Liberal Arts Studies* 4.1, pp. 223–235. URL: http://hdl.handle.net/10723/83.

Venditti, Jennifer J. (2005). "The J_ToBI model of Japanese intonation". In: *Prosodic typology: The phonology of intonation and phrasing*. Ed. by Sun-Ah Jun. Oxford: Oxford University Press, pp. 172–200.

Venditti, Jennifer J., Kazuaki Maeda, and Jan P.H. van Santen (1998). "Modeling Japanese boundary pitch movements for speech synthesis". In: *The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*, pp. 317–322.

Venditti, Jennifer J., Kikuo Maekawa, and Mary E. Beckman (2008). "Prominence marking in the Japanese intonation system". In: *Handbook of Japanese linguistics*. Ed. by Natsuko Tsujimura. Oxford: Blackwell Publishing Limited, pp. 456–512.

Wagner, Michael (2011). "Production-planning constraints on allomorpy". *Proceedings of the Acoustics Week in Canada. Canadian Acoustics* 39.3, pp. 160–161.

Wagner, Michael (2012). "Locality in Phonology and Production Planning". *McGill Working Papers in Linguistics* 22.1, pp. 1–18.

Wagner, Valentin, Jörg D. Jescheniak, and Herbert Schriefers (2010). "On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36.2, p. 423.

Whalen, Douglas H. (1990). "Coarticulation is largely planned". *Journal of Phonetics* 18, pp. 3–35.

Wheeldon, Linda (2012). "Producing spoken sentences: The scope of incremental planning". In: *Speech planning and dynamics*. Ed. by Susanne Fuchs, Melanie Weirich, Daniel Pape, and Pascal Perrier. Vol. 1. Speech Production and Perception. Peter Lang, pp. 97–118.

Wheeldon, Linda and Aditi Lahiri (1997). "Prosodic units in speech production". *Journal of Memory and Language* 37, pp. 356–381.

Wheeldon, Linda and Aditi Lahiri (2002). "The minimal unit of phonological encoding: prosodic or lexical word". *Cognition* 85.2, B31–B41.

Wheeldon, Linda R., Mark C. Smith, and Ian A. Apperly (2011). "Repeating words in sentences: Effects of sentence structure." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37.5, pp. 1051–1064.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price (1992). "Segmental durations in the vicinity of prosodic phrase boundaries". *Journal of the Acoustical Society of America* 92, pp. 1707–1717.

Yoshida, N. and Y. Sagisaka (1990). "Factor analysis of vowel devoicing in Japanese [in Japanese]". In: *ATR Technical Report TR-I-0159*. ATR Interpreting Telephony Research Laboratories, Kyoto.

Zipf, George Kingsley (1929). "Relative frequency as a determinant of phonetic change". *Harvard Studies in Classical Philology* 40, pp. 1–95.