Laboratory Phonology
Journal of the Association for Laboratory Phonology

Kilbourn-Ceron, O., et al. 2020 Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realizations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1): 5, pp.1–28. DOI: https://doi.org/10.5334/labphon.168

JOURNAL ARTICLE

Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realizations

Oriana Kilbourn-Ceron[1], Meghan Clayards[2,3] and Michael Wagner[2]

[1] Department of Linguistics, Northwestern University, Evanston, IL, US

[2] Department of Linguistics, McGill University, Montréal, QC, CA

[3] School of Communication Sciences and Disorders, McGill University, Montréal, QC, CA

Corresponding author: Oriana Kilbourn-Ceron (oriana.kilbourn@gmail.com)

Predictability has been shown to be associated with many dimensions of variation in speech, including durational variation and variable omission of segments. However, the mechanism or mechanisms that underlie these effects are still unclear. This paper presents data on a new aspect of predictability in speech, namely how it affects allophonic variation. We examine two coronal stop allophones in English, flap and glottal stop, and find that their relationship with predictability is quite different from what is expected under current theories of probabilistic reduction in speech. Flapping is more likely when the word that follows is more predictable, but is not influenced by the frequency of the word itself, while glottal stops are more likely in words that are *less* predictable. We propose that the crucial distinction between these two allophones is how they are conditioned by phonological context. This, we argue, interacts with online speech planning processes and gives rise to variability for context-dependent allophones. This hypothesis offers a specific, testable mechanism for certain predictability effects, and has the potential to extend to other factors that contribute to variability in speech.

Keywords: Phonological variation; predictability; speech production planning; corpus phonology

1. Introduction

The pronunciation of words in a sentence context can differ greatly from their citation forms. The sound sequences created by adjacent words may require phonological or phonetic adjustments, as is the case in many types of connected speech processes across a number of languages (Kaisse, 1985). The predictability of a word in its sentential context has also been found to influence its phonetic realization, including its duration, amplitude, vowel quality, consonant voicing and closure duration, and even omission of segments (Aylett & Turk, 2004; Bell et al., 2003; Ernestus, Lahey, Verhees, & Baayen, 2006; Fosler-Lussier & Morgan, 1999; Lieberman, 1963; Torreira & Ernestus, 2009). How do phonological context and contextual predictability come together to influence the distribution of pronunciation variants in running speech? This question remains open, though the intersection of phonology and predictability is an active area of research (Shaw & Kawahara, 2018). This paper contributes to our understanding of this question by presenting an empirical study of word-final coronal stop realizations in English, and elaborating our hypothesis about the relationship between predictability and the selection of phonologically conditioned pronunciation variants.

The idea we pursue is that speech production planning constrains cross-word interactions. A pronunciation variant which relies on phonological information in an upcoming word can only be chosen if that upcoming information is available at the time when the varying word is planned. Predictability can be understood as one of the factors which modulate the time course of speech production planning. This proposal, the Production Planning Hypothesis, makes predictions that are different from those of other mechanisms that have been proposed to explain predictability effects. It relates phonological variability to the interaction of phonological computation with other cognitive processes during real-time language processing, an idea that has recently garnered attention from several scholars working in different research traditions (Bürki, 2018; Kilbourn-Ceron, 2017b; MacKenzie, 2012, 2016; Tamminga, 2018; Tamminga, MacKenzie, & Embick, 2016; Tanner, Sonderegger, & Wagner, 2017; M. Wagner, 2011, 2012).

To explore the nature of these effects, a corpus study is presented which analyzes the pattern of two allophones of coronal stops in one variety of North American English: flaps and glottal stops. We examine the effect of different measures of predictability, and show they affect allophony in a more intricate way than simply causing phonetic reduction of predictable material.

Section 2 begins with a discussion of two pronunciation variants which are sensitive to phonological context: flapping and glottalization. This is followed by a review of the speech production planning literature, and how predictability affects speech planning in Section 2.2. Section 2.3 details our proposal, the Production Planning Hypothesis. The corpus analysis is presented in Section 3, showing that flapping and glottalization pattern differently with respect to predictability, a distinction unexplained by previous theories. The implications for these findings are discussed in Section 4, and Section 5 concludes.

2. Background

2.1. Phonologically-conditioned variation

The realization of coronal stops in American English, which we focus on in this paper, can be quite accurately predicted from syllabic position and identity of adjacent segments (Kahn, 1976; Randolph, 1989). The distribution of allophones can be described as the outcome of an input-output mapping under particular phonological conditions. For example, flapping has sometimes been described by the rule in (1)[1]:

(1) t,d → ɾ /V__V

This rule may oversimplify things in that there is some evidence that there are degrees of flapping, but it accurately captures a restriction on flaps in running speech—they almost never occur outside of this phonological environment.[2] Randolph (1989) found in a corpus analysis that out of 953 flaps, only 16 (1.7%) occurred outside of this environment (Randolph, 1989, p. 119–20). In a study of spontaneous speech, Patterson and Connine

[1] Although the mapping is presented here as a SPE-style transformational rule, this is not crucial for the present proposal. Any phonological input-output mapping system is compatible with the point, whether rule or constraint-based, categorical or probabilistic. We also acknowlede the assumption that the variation under discussion involves categorical changes; this will be discussed in Section 2.4.

[2] Flapping is also sensitive to stress. The canonical flapping environment is actually following a *stressed* vowel, and is rare within words when the following vowel is stressed. However, across word boundaries flapping is possible even when the following vowel is stressed, as in *at Olive's*, while aspirating the /t/ in this type of example is often said to be impossible. 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 Gussenhoven (1986); Kahn (1976) for discussion.

(2001) found a 93.9% flapping rate for word-internal, intervocalic /t/ following a stressed vowel. Zue and Laferriere (1979) found that 99% of t/d were flapped in the same environment during experimental trials involving read speech.

However, this rule also over-predicts the presence of flapping. Word boundaries after the potential flap significantly impact the likelihood of flapping. Randolph (1989) found that of 1389 alveolar stops preceded by a vowel or glide and followed by a vowel, only 937 (67.5%) were flapped. Under experimental conditions, Fukaya and Byrd (2005) found flapping rates in word-final, phrase-internal contexts as low as 30% for one of their participants (n = 60), though another participant flapped on all 60 trials. In the same experiment, phrase boundaries almost categorically blocked flapping (only 2 out of 179 trials flapped). So although the rule in (1) seems to capture an important generalization about flaps, it doesn't fully explain their distribution in spontaneous speech.

Voiceless coronal stops have an additional possible realization as a glottal stop. This variant is well-documented in many varieties of English, including American Englishes (Byrd, 1994; Eddington & Channer, 2010; Roberts, 2006). This allophone is restricted to coda position (though glottal stops may independently appear elsewhere), and is sensitive to the identity of the following segment (Huffman, 2005; Seyfarth & Garellek, 2015; Sumner & Samuel, 2005). Glottalization occurs mostly before consonants, and is much more common before sonorants than obstruents. However, it is not excluded from any context: Eddington and Channer (2010) found a 24.8% glottalization rate for /t/ followed by vowels in the Santa Barbara Corpus.

Glottalization is also sensitive to phrasal position: Both Huffman (2005) and Seyfarth and Garellek (2015) report reliably higher rates of glottalization in phrase-final position (i.e., when followed by an intonational phrase boundary). Additionally, Seyfarth and Garellek (2015) report that the effect of the following segment is significantly reduced in phrase final position, and that the rate of glottalization in the absence of any following segment is just over 50%.

Both flap and glottal stop allophones are sensitive to phonological context, but not to the same extent. Flapping is totally blocked in the absence of the right segmental context, or if the context is disrupted by large prosodic breaks, while glottalization can occur in any context but seems to be sensitive to adjacent segments under certain conditions.

How can these two ways of being sensitive to phonological context be reconciled? An early response was to attribute variability to non-phonological 'performance' factors, and proceed with categorical descriptions of phonological patterns, in the vein of (1) (Chomsky & Halle, 1968). Another analytical option pursed early in the sociolinguistic tradition (Labov, 1994), and more recently gaining traction in constraint-based frameworks (Coetzee & Pater, 2011), is to incorporate the probabilistic aspect of phonological patterns into the phonological description itself.

Our approach combines two key insights of these approaches, that (1) variability may be introduced into phonological patterns by 'performance' factors external to the phonological grammar itself, and (2) modeling variability is crucial to understanding phonological patterns. We propose specifically that systematic variability is introduced during the online processing of speech. Although we do not rule out the possibility that phonological knowledge itself may have a probabilistic component, we suggest that understanding different sources of variability is crucial to the development of parsimonious phonological models. We turn now to a discussion of the speech production system, and how it is influenced by word predictability.

2.2. Speech production

Planning to speak involves several stages. Speaking aloud requires the speaker to first formulate a message at the conceptual level, which then provides the starting point for linguistic processing, and ultimately becomes an articulatory plan ready to be externalized. Current models of spoken word production identify at least two distinct stages of linguistic processing: lexical selection and form encoding (using the terminology of Levelt, 2001; see Dell & O'Seaghdha, 1992; Levelt, Roelofs, & Meyer, 1999) for detailed articulations of two influential models, and Wheeldon & Konopka, 2018 for a recent review).

Lexical selection is the process whereby appropriate linguistic representations are selected to express the concepts in the speakers' intended message. In the case of single-word production, the result of this process is the selection of a unique *lemma*, which specifies the syntactic and semantic properties of the word to be spoken. This stage temporally precedes access to any phonological information, which only occurs later, during form encoding.

Form encoding begins with the retrieval of the phonological code associated with the selected lemma. Then, these phonemes are anchored to a metrical structure including at least syllabic and prosodic word levels. This representation guides the assembly of a more detailed phonetic code which can be passed forward for articulatory execution.

In running speech, these processes of selection and encoding must occur multiple times, along with additional processes to integrate these lexical items into the larger syntactic and prosodic context. The relative timing of these processes is not yet well understood. It is broadly agreed that speech is planned incrementally from the beginning of the utterance (Bürki, 2018; Levelt et al., 1999; Meyer, 1991; Shattuck-Hufnagel, 1979). The speaker can initiate articulation as soon as the motor plan for the first word, perhaps even the first syllable, is complete (Kawamoto, Liu, & Kello, 2015). Hence, linguistic planning and articulation occur in parallel, with planning racing just ahead of what is coming out of the speaker's mouth.

Some global details of the utterance must be computed before articulation begins, like prosodic phrasing and intonation contours (Keating & Shattuck-Hufnagel, 2002). However, these details can be fixed before form encoding is complete, since they don't rely on information about the phonemic make-up of words, which could be filled in later. In fact, F. Ferreira (1993) showed that the final slot in a prosodic phrase is assigned a fixed duration by speakers regardless of the length of the word that will be in that slot. As early as 1978, Sternberg, Monsell, Knoll, and Wright proposed that multi-word utterances are encoded as programs in which a number of sub-programs are embedded. This allows utterance-level variables to be set early, while phonetic details of the sub-programs are retrieved only as necessary as the utterance is unfolding. Speech could begin as soon as the first sub-program is ready to be articulated.

There is also variability with respect to the planning scope. Wheeldon and Lahiri (1997, 2002) provide evidence that utterance initiation time can depend on the number of prosodic words, and speakers take longer to initiate speaking for longer sentences when they had time to silently prepare the target sentence. However, when speakers were not given time to prepare in advance, their speaking latencies instead reflected the number of syllables in the initial prosodic word of the sentence. We argue that this variability in planning scope can explain some of the variability observed in sandhi processes.

However, it is common to see segmental interactions across prosodic words, suggesting that it is possible for the forms of multiple words to be encoded in tandem. Anticipatory speech errors are clear evidence for this (Fromkin, 1971): For example, errors like "the mirst of May" instead of "the first of May," show that the phonological code of /m/ from "May" can be active at least in time to affect the encoding of the intended "first." The

same logic can be applied to context-sensitive allophones such as flaps. Since a flap can only be planned in intervocalic contexts, flapping across the boundary of two prosodic words (e.g., "We upse[ɾ] Andy") shows that these words must have been encoded within the same window.

Though it must be possible for the form encoding process to span more than one word, it's not clear how far in advance this planning window extends, or under what conditions it does extend beyond a prosodic word (Bürki, 2018). In fact, it seems that the size of the window may vary, and can depend on several different factors. Wheeldon and Lahiri (1997, 2002) found that depending on the task, utterance initiation time can be driven more by the number of upcoming prosodic words (with delays between presentation of the stimulus and the cue to start speaking), or the internal complexity of the first upcoming prosodic word (in speeded tasks). In other words, how far a speaker plans ahead is task-dependent. The size of the planning window has also been found to depend on syntactic constituency and semantic coherence (Wheeldon, 2013), 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, & Watson, 1996), and been argued to decrease planning scope (F. Ferreira & Swets, 2002; V. Wagner, Jescheniak, & Schriefers, 2010). Individual differences in working memory also correlate with planning scope (Swets, Jacovina, & Gerrig, 2014). Michel Lange and Laganaro (2014) found evidence that speakers who initiate speech more quickly show less sensitivity to phonological details of upcoming words (Experiment 2). Finally, we note that the 'planning window' metaphor carries an implicit assumption of discrete units of planning, when in fact it may be more appropriate to think of planning a continuous process that involves gradient levels of activation (Pluymaekers, Ernestus, & Baayen, 2005a). This view of planning is also compatible with the logic of our study: Instead of asking about 'extent of planning window,' one would ask instead about what upcoming material has been activated to the degree that it could affect planning of the current word.

One factor which is known to have significant effects on linguistic processing is lexical frequency. Landmark studies from Oldfield and Wingfield (1964, 1965) showed that the time it takes to initiate speech in the task of picture-naming depends on the lexical frequency of the picture's name. Subsequent work has strongly supported the finding that lexical frequency influences the time it takes to name a given word (Griffin & Bock, 1998; Jescheniak & Levelt, 1994; Schilling, Rayner, & Chumbley, 1998). In multi-word utterances, Konopka (2012) found that sentences beginning with high-frequency words were initiated faster than those beginning with low-frequency words.

Levelt and colleagues' influential model of spoken word production locates lexical frequency effects at the level of form encoding, when the phonological code of a lemma is being retrieved (Levelt et al., 1999; though see Gahl, 2008 for evidence that frequency effects also arise at prior stages). So, all else being equal, higher frequency words will be retrieved and encoded sooner than lower frequency words. Extrapolating to the planning of multi-word sequences, we hypothesize that the lexical frequency of each word after the first may affect whether or not they are all phonologically encoded within the same window by either speeding or slowing retrieval. This is illustrated in **Figure 1**. Each box represents the duration of processing for a given word at a given stage. The green leftwards arrow and box represent conditions which are favorable for rapid initiation of form encoding for *word 2*, allowing it to begin before *word 1* form encoding has finished, and allowing for potential interaction between their phonological forms. On the other hand, the red rightwards arrow represents conditions under which form encoding is delayed, and *word 2* is prevented from exerting any influence on the form encoding of *word 1*.

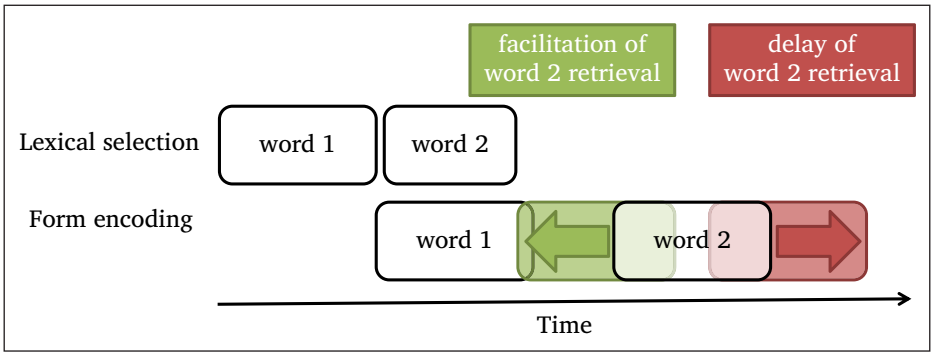

Figure 1: Schema of the time course of select stages of planning a two-word sequence. The vertical dimension represents the distinct stages of linguistic planning, lexical selection and form encoding. The boxes represent the duration of processing at each stage for each word. If conditions facilitate retrieval of the phonological code *word 2*, form encoding can begin sooner, and may overlap with form encoding of *word 1* (green scenario), while a delay could prevent interaction between *word 1* and *word 2* (red scenario).

The processing of *word 1* may also itself be facilitated or delayed by frequency effects, but how this might affect cross-word interactions is less clear. Miozzo and Caramazza (2003) found that in single word utterances, high-frequency distractors affect production latency less than low-frequency distractors. They conclude that frequent words are planned earlier relative to subsequent words, and therefore interfere less with following words. This would suggest that a high-frequency first word might make it less likely that it is planned together with the following word. Given the diagram in **Figure 1**, this would make sense if frequency effects apply at the level of the form encoding (as argued in Levelt et al., 1999), but leave the relative timing of lexical selection intact. In this case, two words should be less likely to be encoded at the same time when the first word is frequent compared to when it is less frequent, and *word 1* frequency should have a negative effect on flapping rate. Konopka (2012), on the other hand, found that a high-frequency first word leads to greater semantic interference with the following word, suggesting a high frequency of the first word makes it more likely for two words to be planned together, at least with respect to their semantics. This would make sense if frequency effects apply at the lexical selection stage and thus retrieval of the second word happens earlier relative to the phonological encoding of the first word, as argued in Alario, Costa, and Caramazza (2002). We would then expect that phonological planning of the second word would be more likely to happen while the first word is being planned, and flapping rate should *increase* with the frequency of the first word. Kittredge, Dell, Verkuilen, and Schwartz (2008) tried to arbitrate between these views of the level at which frequency effects apply. They found frequency effects at both stages, though age of acquisition effects were only found at the phonological level (see also the discussion Tanner et al., 2017). To conclude, it's less clear what to expect with respect to the effect of the frequency of the first word, but a higher frequency of the second word should make it more likely for two words to be planned together.

While lexical frequency reflects the general likelihood of encountering a word in any context, it's also clear that language users are sensitive to the contextual predictability of words. Griffin and Bock (1998) showed that speakers are much quicker to name objects when they had just heard a semantically congruent sentence. Beattie and Butterworth (1979) found that in spontaneous speech, words that are less predictable from context are more likely to be preceded by a hesitation. This effect remained even when the lexical frequency of the words was controlled, but only for low-frequency words. Konopka (2012) found that the scope of planning, indexed by a phonological priming effect, was expanded

when the first word in a sentence was one that subjects had recently produced (in an earlier, unrelated experimental task).

Measures of predictability have also been shown to have an effect on words' phonetic realization. Gregory, Raymond, Bell, Fosler-Lussier, and Jurafsky (1999), analyzing a subset of monosyllabic t/d final words from a corpus of telephone conversations, found that the highest frequency words were 22% shorter than the lowest frequency words. They also found that word duration was correlated with semantic relatedness and discourse repetition, with word duration decreasing for words that had been previously mentioned and which were semantically related to the words in the preceding conversation. Jurafsky, Bell, Gregory, and Raymond (2001) found that several measures of predictability are correlated with rates of final t/d deletion. Gregory et al. (1999) found similar results, and also that flapping of t/d is more likely between pairs of words with high mutual information. Gahl and Garnsey (2004) found that speakers were more likely to delete a verb-final t/d when the verb was in a syntactic frame that matched its usual syntactic complement. Torreira and Ernestus (2009) found an effect of bigram frequency with the following word on the acoustic realization of /t/ in French. Pluymaekers, Ernestus, and Baayen (2005b) 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. Raymond, Brown, and Healy (2016) found that the predictability of the following phonological environment (i.e., whether a given word is typically followed by a consonant or vowel) had a significant effect on rates of t/d deletion. In the flapping environment specifically, they found that the likelihood of deletion was positively correlated with the likelihood that the target would be followed by a vowel. These empirical results clearly establish that predictability affects the phonetic realization of specific segments within words.

This evidence suggests that predictability, in both the sense of prior probability and contextual probability, have an important effect on the time course of linguistic processing in general and form encoding in particular. We propose that the variability of the planning window, and its interaction with form encoding, is crucial to understanding how predictability affects allophonic variation.

2.3. The Production Planning Hypothesis

How does predictability modulate the selection of context sensitive allophones? We draw on a recent line of investigation which relates intra-speaker variability to the dynamics of speech planning (Bürki, 2018; Kilbourn-Ceron, 2017b; MacKenzie, 2012, 2016; Tamminga, 2018; Tamminga et al., 2016; Tanner et al., 2017; M. Wagner, 2011, 2012). Our proposal, in brief, is that predictability affects the size of the form encoding window. The form encoding window, in turn, restricts the size of the input to the phonological input-output mapping. Information that falls outside this window cannot affect allophone selection—even if that information is found in the very next word. This is the case illustrated in the lower portion of **Figure 1** in red. If the trigger of a phonological process is not planned soon enough, the process cannot apply, a situation that Tamminga (2018) aptly names a *co-presence failure*. In the case of t/d, a co-presence failure with a following vowel would remove the opportunity for the flapping rule to apply, therefore the rate of flapping should be modulated by the likelihood that the conditioning environment (i.e., the following vowel) has been planned early enough.

For example, consider a two-word sequence like *cat attack*. Of the several possible phonological encodings of the word *cat*, the flapping rule in (1) predicts a flap when the following word is *attack*. However, this assumes that the segmental information of *attack* has been (at least partially) retrieved and is available at the time that the encoding of *cat*

is taking place. If *cat* must be encoded in the absence of information about the following word, then the flapping rule could not come into play at all, and some other variant should be selected. We propose that both of these scenarios are possible, as illustrated in the lower half of **Figure 1**, and that this is the source of some of the variability of context-sensitive cross-word interactions. Thus the Production Planning Hypothesis predicts that factors that affect speech planning also affect phonological interactions between words.

Furthermore, the Production Planning Hypothesis makes the 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 retrieve the phonological detail of more than one word ahead of time.[3]

Tanner et al. (2017) present evidence that the rate of coronal stop deletion in British English is affected by the following phonological context, and that production planning modulates this relationship. They found that longer pauses between words and higher word frequency reduced the effect of the following context on deletion. They hypothesized that longer pauses between words and higher word frequency both reduce the chances that a word is planned at the same time as the following word. They did not however find strong evidence that the effect of following context was modulated by the conditional probability of the two words, so the relationship between predictability and phonological context effects is still not entirely clear. The present study seeks to find clearer evidence by examining a process that is more closely dependent on the segmental context, flapping, and comparing to another which is much less dependent, glottalization.

2.4. Alternative accounts of predictability effects

Most previous work on the relationship between predictability and pronunciation variation has focused on phonetic reduction in particular. Several types of proposals have emerged from this research, many of which are mutually compatible with each other and with the proposal of this paper. However, few address the question of how the distribution of phonologically-sensitive pronunciation variants is affected by predictability. We briefly review some existing proposals and outline their predictions in the context of our study.

Words and segments that are more predictable have been consistently found to be phonetically reduced, especially when considering duration (see end of Section 2.2). Production ease accounts consider phonetic reduction a reflex of easier planning conditions, though accounts differ on whether the planning difficulty of previous, current, or future material is the focal point (see for example V. S. Ferreira & Dell, 2000; Pluymaekers et al., 2005a; Watson, Buxó-Lugo, & Simmons, 2015). Under the assumption that lexical frequency eases planning, higher lexical frequency should be associated with a realization more dissimilar from the citation form. Communicative accounts propose that phonetic reduction is driven by the speaker's desire to efficiently and accurately transmit their intended message (Aylett & Turk, 2004; Hall, Hume, Jaeger, & Wedel, 2018; Jaeger, 2010; Turk, 2010, among others). Use of a pronunciation variant like a flap, which neutralizes the phonemic t/d distinction, or a glottal stop, which removes place of articulation cues, presumably decreases intelligibility. On this assumption, a communicative account would

[3] A reviewer brings up the question of whether these types of effects may also come into play word-internally. This depends on how incremental word-internal planning turns out to be, for example whether the morphemes in a complex word "ration-al-iz-ation" are retrieved one-by-one or in parallel. This question could be particularly interesting to investigate in polysynthetic languages where single 'words' can comprise a large number of lexical morphemes. We leave this interesting question to future work.

predict that these variants should be used when message predictability is relatively high, as would be the case for higher frequency words.

Finally, representational accounts attribute predictability effects to accumulation of stored experiences of specific words and phrases (Bybee, 2001, 2007; Pierrehumbert, 2001). Frequent repetition leads to lenition, so under this account, both the frequency of the variable word and its frequency of co-occurrence with the following word would have a positive correlation with use of the context-sensitive variant.

What is clear from this work is that there is a strong negative correlation between predictability and duration. This presents the possibility that many of the apparently qualitative changes associated with high predictability, including the change from coronal stop to a flap or glottal stop, could in reality be a gradient process that arises from temporal compression of gestures, rather than mappings like the rule in (1). There is evidence that flapping is (or at least can be) a gradient process that involves degrees of flapping, (e.g., Fox & Terbeek, 1977), even if the acoustic consequences of flapping often appear to be rather categorical (De Jong, 1998). A gradient account is also made plausible by the fact that consonants other than t/d are subject to similar temporal reductions in flapping environments (Browman & Goldstein, 1992; Turk, 1992), and 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 (Braver, 2011; Herd, Jongman, & Sereno, 2010; Malécot & Lloyd, 1968). This pattern is unexpected if flapping involves a categorical phonological change (though see Bermudez-Otero, 2011).

Our proposal is compatible with many aspects of this account. In particular, we emphasize that although our discussion of flapping and glottalization implies a categorical alternation, our discussion and conclusions are compatible with a gradient analysis of these variants. The transcription in the Buckeye corpus on which this study is based is just a coarse proxy measure based on perception. Though the articulatory reality is undoubtedly more complex (Fukaya & Byrd, 2005; Purse, 2019), perceptual annotation is a reasonable starting point for investigation given the finding of De Jong (1998) that even gradient articulatory overlap can lead to categorical perceptual results.

A major point of difference is the assumption that phonological representations are invariant and that qualitative variation is attributable solely to temporal compression. The PPH relies on the assumption that contextual changes in phonetic form are encoded during the speech planning process. Studies on coarticulation have found evidence that anticipatory coarticulation is planned, rather than being an automatic articulatory process. Whalen (1990) investigated anticipatory coarticulation between a word-initial /a/ and a consonant or vowel in the next syllable. The F1 of /a/ was lower if a /b/ followed compared to a /p/, and F2 was lower if an /u/ followed, compared to an /i/. However, when participants were asked to initiate speech with part of the word missing (e.g., A_I or AB_), coarticulatory effects disappeared for the missing segment, even though the missing segment was immediately revealed and integrated into the utterance. Liu, Kawamoto, Payne, and Dorsey (2018) used a similar paradigm, and tested a greater number of participants. They found that several participants showed the same pattern of anticipatory coarticulation as the three participants in Whalen (1990), though others showed no differences between conditions. This provides evidence that articulatory plans are actively adjusted during the planning process as a function of what information is available about upcoming segments.

We can make a similar point based on the data used in this study by looking for evidence that speakers make choices about the articulatory plan, rather than simply compressing the existing articulatory plan differently depending on temporal factors. We will look at

two qualitatively different outcomes of reduction, glottalization and flapping, which are not degrees of the same gradient reduction process, but involve different planning choices by the speaker.

2.5. Summary and predictions

The relationship between predictability and phonetic reduction is clear: Many measures of predictability are positively correlated with reduction. However, there has not yet been much research on how predictability affects allophones, which may also be considered reductions, but differ in important ways. In particular, some allophones like flaps require information about the phonological context to be available during their planning.

Our first research aim is to establish a clearer empirical picture of how predictability and allophone distribution relate to each other. Some previous work has found that higher predictability is associated with a higher probability of phonological interactions like flapping (Gregory et al., 1999) and voicing assimilation (Ernestus et al., 2006).

Given this previous work, we expect to find a significant positive correlation between flapping and our measures of predictability. As for glottal stops, we are not aware of any previous studies that have reported rate of glottalization in relation to measures of predictability. If glottalization of /t/ is considered a general reductive process, being a reduction of the tongue tip gesture, it would be expected that lexical frequency of the target word should be positively correlated with glottalization.

Secondly, we put our account of predictability effects to the test. The mechanism proposed by the Production Planning Hypothesis makes different predictions about how different allophonic processes will be affected by predictability based on whether they are sensitive to phonological context. Increased predictability facilitates planning, potentially widening the advance planning window, and therefore increases the rate at which a context-sensitive process like flapping applies in North American English.

In contrast, glottalization of /t/ in North American English does not strictly require a particular phonological context to be realized. It is not excluded from any context: Eddington and Channer (2010) found a 24.8% glottalization rate for /t/ followed by vowels in the Santa Barbara Corpus, and Seyfarth and Garellek (2015) found rates between around 25% and 90% for different types of following consonants in the Buckeye corpus. Before a pause, i.e., in the absence of a following segment, the likelihood of a glottal stop was just over 50%. Since the present study is restricted to intervocalic context, we expect that contexts which promote the inclusion of a following vowel in the same planning window should slightly *decrease* the likelihood of glottalization, since the 'default' rate of glottalization when no segment follows is relatively high (i.e., around 50%), while pre-vocalic glottalization is relatively low.

3. Corpus study

How does predictability affect allophone distribution? We address this question by analyzing the pattern of t/d realization in the Buckeye Corpus of conversational speech (Pitt et al., 2007). Predictability is operationalized using two distinct but mathematically related variables, lexical frequency of the trigger word and the conditional probability of the trigger word given the target word. Although this presents some complications for statistical analysis, both variables were included since they track conceptually independent sources of planning facilitation. The studies reviewed in Section 2.2 found that frequency is a good predictor of planning times, but this has mostly been investigated in single-word naming contexts. To the extent that this index of planning time accurately reflects processing during multi-word utterances, finding that *trigger* word frequency affects the realization of the preceding word would be in line with the predictions of the Production Planning Hypothesis. However, we also expect that conditional probability is an important

measure of planning ease in spontaneous discourse, since semantic and syntactic context affect speech latencies (Griffin & Bock, 1998; Konopka, 2012). Disentangling the relative contributions of each of these variables is not one of the goals of this paper; we merely aim to provide an empirical picture of how both measures relate to allophonic variability, and suggest that the findings are compatible with our proposal.

First we present the dataset, then present the statistical model used for analysis. We show the results of fitting the model to our dataset, followed by a discussion of the implications.

3.1. Dataset

The set of observations used for our analysis were pairs of words in which the first ended in /t/ or /d/ immediately preceded by a vowel (hereafter *target words*), and followed by a vowel-initial next word (hereafter *trigger words*), e.g., "ended up," "quite easy." The were collected from the Buckeye corpus of conversational speech (Pitt et al., 2007), a corpus of sociolinguistic interviews with 40 speakers native to central Ohio, totaling about 300000 words. The speakers were balanced by age (over/under 40), gender of speaker, and gender of interviewer (Kiesling, Dilley, & Raymond, 2006).

The corpus contained 11863 qualifying word pairs, which were extracted along with existing time-aligned phonetic transcription using the Montreal Corpus Tools software (McAuliffe, Stengel-Eskin, Socolof, & Sonderegger, 2017).[4] The Pitt et al. (2007) transcriptions were prepared automatically and subsequently hand-corrected by phonetically trained research assistants. For flaps [dx], annotators were instructed to only include segments with sustained voicing throughout the phone. For glottal stops [tq], transcribers were instructed to "label all /t/ or /d/ phones which show glottalization the phoneme label /tq/" (Kiesling et al., 2006). A test of labeling consistency using four transcribers and four one-minute samples from the corpus yielded an inter-transcriber agreement of 92.9% for stop consonants (Pitt, Johnson, Hume, Kiesling, & Raymond, 2005, 80.3% overall). For our analysis, we grouped the observations into four categories based on the realization of the underlying coronal stop in the target word: "full" if the surface transcription matched the underlying (21.46% of tokens), "flapped" if transcribed with [dx] (54.83%), "glottalized" if transcribed with [tq] (14.24%), and "other" for any other transcribed segment.

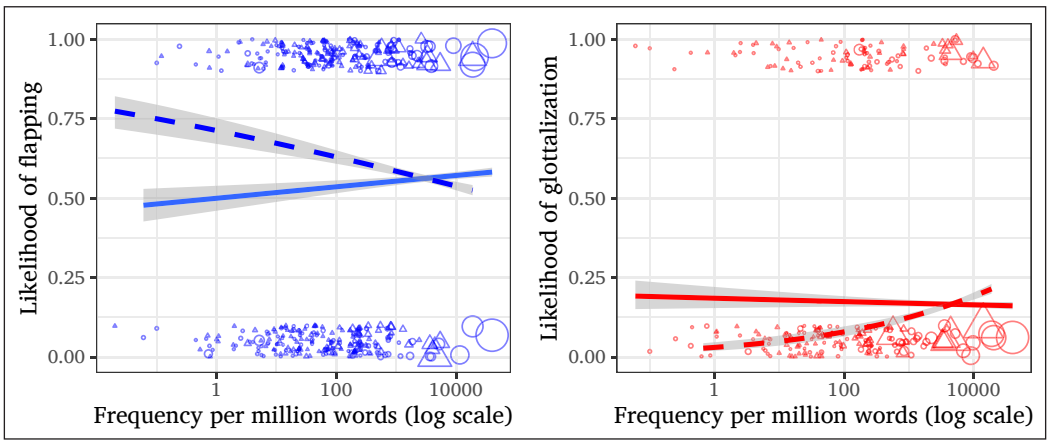

Figure 2: Relationship between SUBTLEX-US word frequency (per million words) and proportion of tokens transcribed as flaps [dx] (left panel, blue), glottal stops [tq] (right panel, red) in the Buckeye corpus. Solid lines show trigger word frequency, dashed lines show target word frequency, with shading showing 95% confidence intervals of a linear smooth (GLM, logit-link). Bubbles (trigger word frequency) and triangles (target word frequency) show density of observations, with size reflecting the number of observations at a given value on the x-axis.

[4] We gratefully acknowledge the assistance of Michael McAuliffe in extracting and preparing these data.

We removed observations in which the trigger word was a disfluency marker like "um"[5] (20% of tokens), and those in which the trigger word was reduced to a syllabic sonorant on the surface (0.09% of tokens). This left 8428 tokens for analysis.

We enriched the dataset with information about the probabilities of the observed words. The prior probability for each word was estimated by retrieving its lexical frequency from SUBTLEX-US, a database of word frequencies based on a 51 million-word corpus of film and television subtitles (Brysbaert & New, 2009). Frequencies ranged from 39971 per million to 0.2 per million for words that occurred only once in the corpus. The range of values for target and trigger words was comparable, with median values of 15.44 and 16.41 words per million respectively. The empirical correlations between each of these measures and flapping and glottalization are shown in **Figure 2**. Distributions are plotted in Appendix A.

We also calculated a measure of probability which takes into account the likelihood of each word pair as a collocation. There are many ways pairwise likelihood could be calculated. The bigram frequency is simply the likelihood of two words occurring together. This value is highly dependent on the frequencies of the individual words in the bigram, since the bigram cannot be more frequent than either of the words individually. It also cannot distinguish between pairs where the first word is infrequent and the second is highly frequent, or vice versa.

We chose to focus on the conditional bigram probability of the trigger word given the target word (hereafter *conditional probability*), which controls for the base frequency of the target word. For example, "out of" and "instead of" are sequences with similar relative bigram frequency, occurring equally often, but very different conditional probabilities: "instead" is highly likely to be followed by "of" (about 90% likely), while "out" is only followed by "of" about 20% of the time.

In order to estimate the conditional probabilities for the words in our dataset, we fitted a bigram language model to the SWITCHBOARD corpus (Godfrey, Holliman, & McDaniel, 1992), a corpus of spontaneous telephone conversations comprising about 3 million words. Using this larger corpus as the basis for the language model allows for more accurate estimates, especially for two-word sequences. The language model was fitted using the `lmplz` function from the KenLM language model toolkit (Heafield, 2011), which uses modified Kneser-Ney smoothing without pruning. The conditional probabilities calculated by the language model were matched orthographically to the two-word sequences from the Buckeye dataset. The empirical relationships of conditional probability to flapping and to glottalization are shown in **Figure 3**, and the empirical distribution is presented in Appendix A.

Several variables were also included in the dataset to act as controls in the statistical analysis. The underlying voicing of the target words' final segment was recorded; /t/ was flapped in 56.1% of cases, while /d/ was flapped slightly less at 52%. Number of syllables is highly correlated with word frequency, as the most frequent words are monosyllabic. Target and trigger words were labeled as either monosyllabic or polysyllabic, with each syllabic segment in the Buckeye surface transcription counting as one syllable. Most of the observations consisted of two monosyllabic words (77.8%), followed by monosyllabic-polysyllabic pairs (13.5%), polysyllabic-monosyllabic pairs (7.6%), and 86 polysyllabic pairs (1%). Flapping rates were comparable within these groups (53%–57%), and glottalization rates showed some variation (7%–17%).

We also included duration as a control measure for several reasons. The first relates to an articulatory account of flapping that views it as an automatic result of durational

[5] These words were *uh, um, okay, oh, and, um-huh, uh-uh, uh-hum, um-hum.*

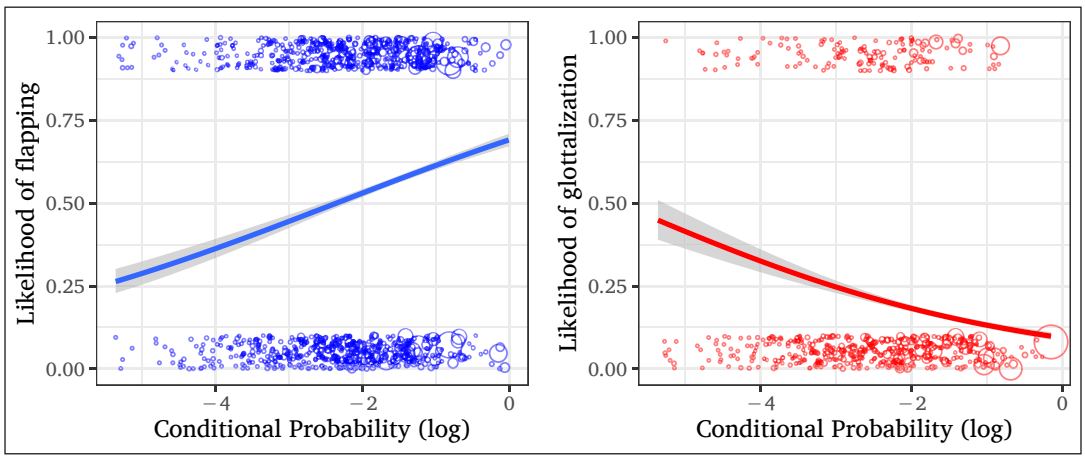

Figure 3: Relationship between Conditional Probability (of trigger word given target word) and proportion of tokens transcribed as flaps [dx] (blue, left panel) or glottal stops [tq] (red, right panel) in the Buckeye corpus. Solid lines and shading are linear smooths (GLM, logit-link) with 95% confidence intervals. Bubbles show density of observations, with size reflecting the number of observations at a given value on the x-axis.

compression, as it is entertained for example in the literature on articulatory phonology (Browman & Goldstein, 1992; Byrd & Saltzman, 2003).

The PPH is very compatible with a gradient account of flapping, and with proposals about gradient gestural overlap. What distinguishes the PPH from prior articulatory overlap accounts is that it does not view flapping as an automatic consequence of temporal compression. By including duration as a control measure, we want to ensure that temporal compression alone is not sufficient to explain the observed patterns of variability.

A second reason to include a durational control is that the phonological literature on flapping has related the process to prosodic phrasing, and holds that the effect of other factors such as syntax is mediated by the presence or absence of certain prosodic junctures (Nespor & Vogel, 1986). The Production Planning Hypothesis is very compatible with the idea that prosodic phrasing will modulate flapping rate, but it also predicts that it should not be the *only* factor affecting the likelihood of flapping. A durational measure can serve as a proxy measure for prosodic boundary strength (Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992), and including it in the model will help with the argument that the observed variability is not purely a result of variability in the prosodic phrasing of utterances.

The durational measure we chose for this study was the ratio of observed/expected duration of the target word. The expected duration was calculated by adding together the mean durations of each phone in the surface transcription. The mean phone durations were calculated over the entire Buckeye Corpus. A value below 1 indicates that the target word is shorter than expected based on the average durations of its component phones, and a value greater than 1 means it is longer than expected. In addition to any predictability-induced durational effects, this variable also captures compression due to faster speech rate, or expansion due to boundary-induced final lengthening (Wightman et al., 1992). Its distribution is illustrated by the plots in Appendix A.

Another control measure we included is the presence or absence of pauses, clearly a factor in coronal stop realization. The rate of glottalization before a pause is much higher than when no pause follows, 47.4% versus 8.4% (n = 1143 with pause, 7285 without pause), and flapping before a pause was rare (1.3%). Therefore we included a variable tracking whether or not a pause was annotated in the Buckeye transcription between the target and trigger words.

3.2. Analysis

The occurrences of flapping and glottalization were analyzed in separate logistic regression models with elastic net regularization. This technique penalizes large coefficient estimates, which allows (1) the shrinkage and/or removal of the least predictive variables, and (2) mitigation of collinearity-induced estimate inflation. This is of particular concern since the probability-based predictors are correlated by definition. **Figure 4** shows that Trigger Word Frequency and Conditional Probability have a strong positive correlation, as expected since the latter is calculated using the former. Using the penalized regression technique may lead to dropping whichever one of these variables is less predictive. However, this does not necessarily constitute evidence against an independent effect of the less predictive variable – we return to this issue in the discussion.

Following the procedure outlined in Tomaschek, Hendrix, and Baayen (2018), the models were fit using the `glmnet` (Friedman, Hastie, & Tibshirani, 2010) package for `R` (R Core Team, 2013). We used the `cv.glmnet` function which performs k-fold cross-validation and returns possible values for lambda, the penalty imposed on non-zero coefficients. The value for alpha was 1, equivalent to the lasso model, which yields the model stringent penalty on non-zero coefficients (but may sacrifice accuracy). We selected a value of lambda within one standard error of the minimum mean squared error (MSE) of the cross-validated models, as per the recommendation of Tomaschek et al. (2018, and references cited therein). This resulted in a model with the smallest number of non-zero coefficients

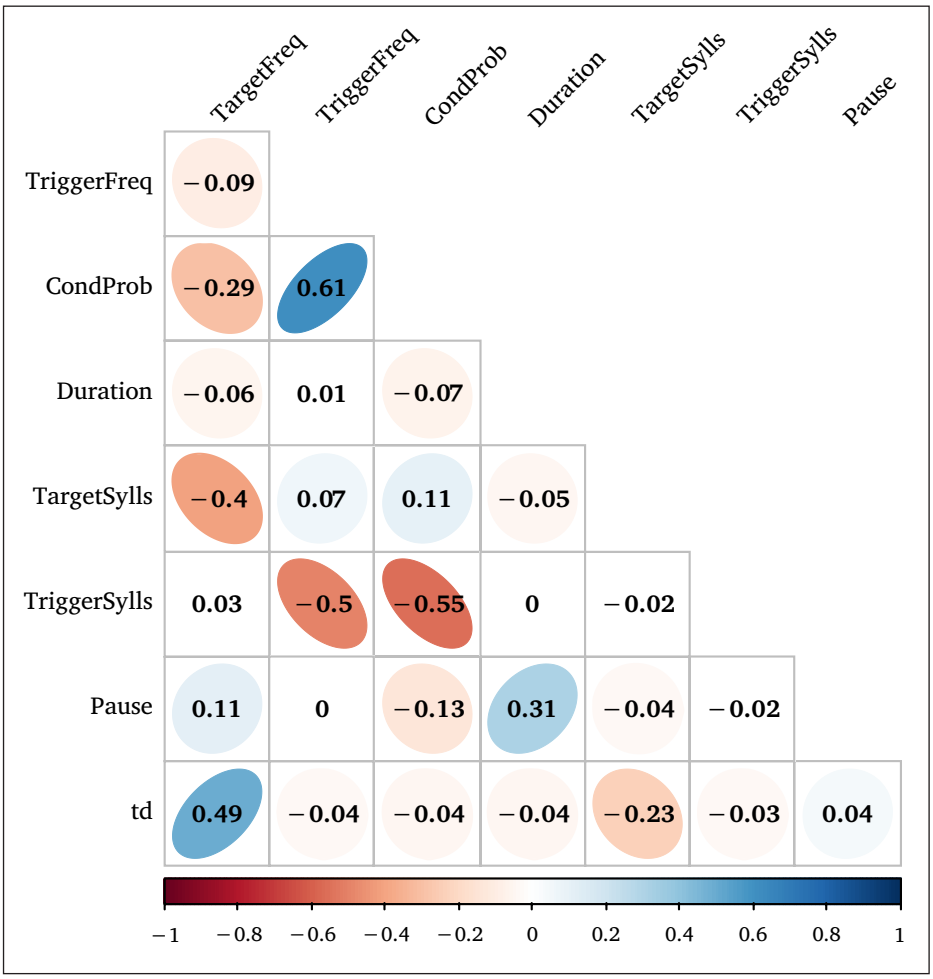

Figure 4: Correlations between variables to be included in model. Opacity and ellipse size indicate magnitude, and no ellipse indicates non-significant correlations.

while maintaining a reasonable MSE. If a coefficient remains in the model, it can be taken as evidence that it is an important part of explaining the variance in the dataset. Reliable standard errors are not available for regularized regression models, so we have refrained from reporting *p*-values in the tables below. Results of non-penalized logistic regressions, including standard errors and *p*-values, are reported in Appendix B, and were qualitatively similar.

Each model predicted the log-likelihood of a particular variant (either flap or glottal stop) as a function of the variables described in the previous section. Both models included the following fixed effects: `Target Word Frequency`, which was standardized by subtracting the mean and dividing by two standard deviations; `Duration`, log-transformed to approach normality and also standardized; `Pause`, a categorical variable with "no pause" as the reference level (0) and "pause" set to 1; `Target # of Syllables`, `Trigger # of Syllables` (monosyllabic or polysyllabic) were binary variables, which were centered around 0 by subtracting the mean value. Additionally, the model for flapping included `Underlying t/d`, tracking the underlying voicing of the target word's final segment (also a centered, binary variable). For glottalization, the model excluded all data with /d/-final words, since these segments are very rarely realized as glottal stops (11 of 1561 /d/-final tokens in the current dataset), and therefore also excluded the `Underlying t/d` variable.

The `glmnet` package does not support inclusion of random effects in the model. However, we report non-penalized regressions in Appendix B with random intercepts by-speaker and by-target word, with qualitatively similar results. Additional models which included all variables and maximum identifiable random-effects structure were also fitted, again with qualitatively similar results.

Tables 1 and **2** show the model estimates for the fixed effects coefficients in the fitted models. Each coefficient represents the estimated change in log-odds of the outcome when other predictors are held at their mean observed values, except `Pause` which is held at 0 (no pause).

3.3. Results

3.3.1. Target Word Frequency

Our analysis did not retain `Target Word Frequency` as an important predictor of the likelihood of flapping, once other variables were controlled. This is in line with the finding of Gregory et al. (1999) that only mutual information, but not target word frequency, is

Table 1: Coefficients for elastic net regression model of **flapping** data, alpha = 1 and lambda = 0.01473092, Binomial deviance = 1.13155.

	Coefficient estimates
(Intercept)	0.5979912
Target Word Frequency	.
Trigger Word Frequency	.
Conditional Probability	0.3339739
Duration	.
Target # of Syllables	.
Trigger # of Syllables	.
Pause	–3.4928831
Underlying t/d	0.0609548

Table 2: Coefficients for elastic net regression model of **glottalization** data, alpha = 1 and lambda = 0.02664108, Binomial deviance = 0.7604593.

	Coefficient estimates
(Intercept)	–2.0495764
Target Word Frequency	.
Trigger Word Frequency	.
Conditional Probability	–0.1668624
Duration	0.1239920
Target # of Syllables	.
Trigger # of Syllables	.
Pause	1.9088963

predictive of flapping. The analysis of glottalization also revealed no significant effect of `Target Word Frequency`. While much previous work has investigated the effect of frequency and predictability on *deletion* of word-final coronal stops (Guy, 2007; Jurafsky et al., 2001; Raymond et al., 2016; Tanner et al., 2017), as far as we are aware this is the first time such results have been reported for glottalization, and only the second for flapping.

3.3.2. Trigger Word Frequency

The model for flapping in **Table 1** does not retain `Trigger Word Frequency` as an important predictor. This is somewhat unexpected, based on the empirical trend observed in **Figure 2**, which suggested a positive correlation. This empirical trend may simply be due to the correlation of `Trigger Word Frequency` with `Conditional Probability`, which is calculated in part from the trigger word frequency. The model for glottalization in **Table 2** also does not provide a non-zero estimate for `Trigger Word Frequency`, suggesting that it does not have much predictive power above and beyond the other variables included in the model.

In light of the strong correlation between `Trigger Word Frequency`, we carried out additional analyses using model comparison to assess whether this variable merits further investigation. A non-penalized logistic regression was fitted which excluded `Conditional Probability`. In this model, `Trigger Word Frequency` did have a statistically significant positive estimate ($\hat{\beta} = 0.24, p = 0.019$), in line with the empirical trend. Based on a likelihood ratio test, Trigger Word Frequency significantly improves the predictions of the model compared to a model which includes the control variables plus `Target Word Frequency` ($\chi^2(1) = 9.14, p = 0.0025$). However, dropping `Trigger Word Frequency` from the full model shows that it does not significantly improve the model over and above `Conditional Probability` ($\chi^2(2) = 1.44, p = 0.49$).

Another non-penalized regression was fitted for glottalization in which `Conditional Probability` was dropped. In this model, the estimate for `Trigger Word Frequency` was quite different, with a negative sign, and no longer statistically significant ($\hat{\beta} = -0.15, p = 0.39$). A likelihood ratio test showed that including the `Trigger Word Frequency` fixed effect and associated random slope terms did not significantly improve the model compared to a baseline with only `Target Word Frequency` and control variables ($\chi^2(3) = 7.73, p = 0.05$). Further comparison of the full model with a model in which `Trigger Word Frequency` terms were dropped showed that those variables did significantly contribute to explaining the variance in the data over and above `Conditional Probability` ($\chi^2(3) = 13.95, p = 0.003$).

These exploratory analyses suggest that `Trigger Word Frequency` may indeed play a role in explaining variability of flapping and glottalization, but more data is necessary to ascertain the sign and magnitude of its effect. Recent results from a randomized-control experiment investigating word frequency effects on flapping suggest that in production of short phrases, flapping is indeed sensitive to trigger word frequency when conditional probability is controlled (Kilbourn-Ceron & Goldrick, 2019).

3.3.3. Conditional Probability

Flapping is estimated to be more likely as `Conditional Probability` increases, that is, the easier the trigger word is to predict from the target word ($\hat{\beta} = 0.334$). This is somewhat in agreement with the finding of Gregory et al. (1999) that mutual information is predictive of flapping, though they did not find an effect of conditional probability alone. However, their analysis included consonant-initial trigger words and excluded tokens whose final t/d was deleted, which likely resulted in a significant difference in the observed proportion of flapped tokens. For glottalization, the opposite correlation was found. Words ending in /t/ are much *less* likely to be pronounced as glottal stops when `Conditional Probability` is high ($\hat{\beta} = -0.331$).

3.3.4. Control Variables

The estimated effect of `Duration` on flapping was shrunken to zero by the Lasso penalty. In contrast, `Duration` was a significant predictor of glottalization: Words with unexpectedly long durations were more likely to be pronounced with a final glottal stop ($\hat{\beta} = 0.273$). The number of syllables in target and trigger words did not receive non-zero estimates in the model, for either flapping or glottalization.

The `Pause` estimate was large for both flapping and glottalization, and of relatively large magnitude compared to other effects. For flapping, the estimate was negative ($\hat{\beta} = -3.493$), confirming that flapping in the presence of a pause is very rare. The opposite is true for the glottalization model, where pause is predicted to have a significant positive effect ($\hat{\beta} = 1.982$).

4. Discussion

Our results show that predictability has an influence on the realization of word-final coronal stops that goes beyond straightforward reduction of predictable material.

Addressing our first research question, we have shown that the distribution of flaps and glottal stops is significantly related to the predictability of the trigger word, but in different ways for each allophone. To address our second question, we discuss the pattern for each allophone, and evaluate how well our results match the predictions of the theories presented in Section 2.

4.1. Flapping

Word-final coronal stops are more likely to be flapped when the word that follows is highly predictable, according to at least one of the variables that we investigated. `Conditional Probability`, the probability of the trigger word given the target word, had a significant positive effect on flapping. The frequency of the trigger word itself also appeared to have a positive effect, but the estimate was not significant in the statistical. This may have been due to issues with suppression because of a high correlation between `Conditional Probability` and `Trigger Word Frequency` (Tomaschek et al., 2018) – further work in more controlled paradigms is necessary to ascertain whether these are two independent effects (see Kilbourn-Ceron & Goldrick, 2019, for recent findings on this question).

These results are compatible with the predictions of the Production Planning Hypothesis. The predictability of the trigger word, whether global or local, affects how quickly the word is accessed and encoded during speech planning. The more predictable the trigger word is, the more likely it is to be planned within the same window as the previous (target) word. Since simultaneous availability of the following vowel is a necessary condition for flapping, increased availability of the trigger word is predicted to have a positive effect on the likelihood of flapping. The fact that the `Conditional Probability` effect is larger than, and possibly masks, the `Trigger Word Frequency` effect may make sense based on previous speech production findings. Beattie and Butterworth (1979) found that hesitations were consistently correlated with contextual probability, but lexical frequency effects were no longer significant when contextual probability was held constant. Konopka (2012) found that the extension of planning scope based on the frequency of the first word in a sentence only took place if the structure of the sentence had been primed, making it easier to plan. It may be that lexical frequency effects in running speech are too subtle to be detected or only come into play under certain planning conditions. Controlled studies are needed to discover the contribution of lexical frequency to planning of allophonic variants.

In terms of probabilistic reduction accounts, the effect of `Conditional Probability` supports the idea that ease of planning of upcoming material leads to reduction of words being currently planned, if flapping is considered a reduction. On the other hand, we failed to detect any effect of `Target Word Frequency` on flapping, similar to the results of Gregory et al. (1999), suggesting that the effects of predictability on durational reduction and segmental deletion may be qualitatively different from how predictability interacts with flapping. Under a representational account of probabilistic reduction, it might be possible to argue that the underlying driver of reduction is the frequency of the word pair, which is correlated with the `Conditional Probability` measure. Under a communicative account, it could be argued that a more complex relationship is at play. For example, Turnbull, Seyfarth, Hume, and Jaeger (2018) proposed that there is a trade-off of inferrability between the target and trigger words in nasal place assimilation. Their results supported this trade-off idea, with target words showing a higher degree of coarticulatory effects on F2 when target word predictability was high, but also when *trigger* word predictability was *low*. This is the opposite of what we found for flapping, which also encodes information about the upcoming word, namely that it begins with a vowel. The PPH would predict that the increased predictability of the trigger word should also facilitate nasal assimilation. These conflicting results suggest that gradient coarticulatory effects could be an interesting future testing ground for these two types of accounts. It could be interesting to compare nasal place assimilation and coronal stops realizations directly, since nasal assimilation neutralizes to another phoneme, while flapping and glottalization do not, and so may have different informational consequences.

We also found that the presence of a pause was a significant predictor of flapping. Pauses are associated with larger prosodic boundaries, which have been found in earlier work to block flapping (Patterson & Connine, 2001; Scott & Cutler, 1984). In addition to acting as proxies for prosodic boundaries, pauses may have also been indications of hesitations, disfluent speech, and/or planning difficulties. It would be interesting in future work to try to disentangle the effect of prosodic boundaries from those of planning-induced pauses and lengthening (F. Ferreira, 1993). A preliminary look at the observations in our corpus that were followed by filled pause words shows that flapping is much lower than average for these words, with 20.3% for “um” ($n = 133$) and 15.2% for “uh” ($n = 303$), even though they technically fit the segmental description to trigger flapping.

For `Duration`, we found no significant effect. This suggests that for a given sequence of phones, its duration compared to the mean of that same sequence does not predict whether or not the target word contains a flap. Flapping was more likely when the underlying phone was /d/ rather than /t/, which is unsurprising since /t/ has the additional possible realization of glottal stop competing with flapping. There was no detectable effect of number of syllables for either the target or trigger word.

4.2. Glottalization

Our analysis of glottal stops revealed that `Conditional Probability` had a negative correlation with the likelihood of glottalization. This is in the direction predicted by the Production Planning Hypothesis: An extension of the planning scope to include the following vowel initial word would make it more likely that glottalization should be suppressed, since the flap variant will be chosen instead. We note that since our analysis only includes intervocalic contexts, the canonical flapping environment, we are cautious in interpreting this result as evidence that glottalization is highly sensitive to segmental context. It could be that the lack of glottalization is mirroring the increase in flapping in intervocalic contexts. However, previous work such as Seyfarth and Garellek (2015) has shown that glottal stops are still sensitive to segmental context in pre-consonantal environments, so production planning effects should still be in force. If the analysis were to be repeated with only pre-consonantal contexts, the opposite effect would be predicted, with the highest correlation between glottalization and `Conditional Probability` in pre-sonorant contexts.

This type of pattern is unexpected under accounts of probabilistic reduction, assuming that glottalization is a reduction relative to the full /t/ gesture. It could be possible to argue, as an anonymous reviewer suggested, that glottalization involves reinforcement through addition of a glottal gesture. If this is so, then a production ease account could explain the negative correlation as the addition of a gesture in difficult-to-plan contexts, though it is unclear why the tongue tip gesture would simultaneously be dropped. Under a communicative account, glottal reinforcement could be considered a way to strengthen a cue to word boundaries in environments where the upcoming word may be difficult for the listener to retrieve. However, there is evidence that realization of a /t/ as a glottal stop hinders recognition of the target word. Garellek (2013) found that subjects are significantly less accurate in recognizing minimal pairs like “dent-den” when the final /t/ is realized as a glottal stop (though see Chong & Garellek, 2018 for further results on recognition of glottalized vowel-consonant sequences). Hence, there might have to be a trade off between intelligibility of the target and trigger words under this account, as suggested by Turnbull et al. (2018). Furthermore, an explanation based on glottal reinforcement does not predict an opposite effect of conditional probability before sonorants.

There were no significant effects of lexical frequency on glottalization. However, exploratory analyses based on model comparison suggest that frequency may play a role in explaning some of the variation in glottlization; further research is needed to clarify this issue.

The effects of `Duration` and `Pause` were significant and in the expected positive directions. Glottal stops are common before pauses (Seyfarth & Garellek, 2015), and glottal voice quality in general is highly associated with intonational phrase boundaries (Redi & Shattuck-Hufnagel, 2001), which are typically lengthened. This may indicate that in addition to being sensitive to upcoming segments, the planning of the glottal stop itself may be related to details of its position in larger prosodic structure, and glottalization may serve as a cue for a boundary, or increase the perceived strength of a boundary. The

negative effect of frequency could then be a reflection that boundary strength correlates negatively with the predictability of a following constituent (Turk, 2010). There was no detectable effect of number of syllables for either the target or trigger word.

4.3. Theoretical interpretation

Our analysis of flapping and glottalization reveal that the distribution of these variants is not straightforwardly explained by assuming that they are predictability-motivated reductions. Target word frequency, previously found to be a reliable predictor of durational compression and segmental deletion (Aylett & Turk, 2006; Jurafsky et al., 2001), was not predictive of either allophone. The main finding of our analysis is that conditional probability of the trigger word given the target word had significant and *opposite* effects on flapping and glottalization. This is consistent with the predictions of the Production Planning Hypothesis, and adds to recent studies in support of speech production effects on phonological variability (Kilbourn-Ceron, 2017a; Lamontagne & Torreira, 2017; Tamminga, 2018; Tanner et al., 2017).

Obviously 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 Production Planning Hypothesis makes very concrete predictions in this regard, which differ from the predictions of alternative hypotheses, and that our data support these predictions.

A relevant question that remains open is at what stage allophonic variants are selected. For example, syllable-initial aspiration of voiceless stops in English could be implemented during phonological encoding, as long as syllabification is done first. Or, it could arise during phonetic encoding if an aspirated stop motor program is selected directly from a phonological representation which is unspecified for aspiration. It may well be that both are possible mechanisms for contextual variation. Although the answer to this question does not change the logic of our study, we find this to be an interesting avenue for future research that might allow us to make even more detailed predictions about variable sound patterns.

5. Conclusion

This paper presents a novel empirical investigation of the relationship between allophone distribution and predictability. Flapping and glottal stops are affected by predictability in a way that is different from the pattern previously found for reductions like durational compression and articulatory lenition. They are also different from each other, with flapping increasing with the predictability of the trigger word, while glottalization became *less* likely in predictable contexts. To explain these patterns, we have invoked the Production Planning Hypothesis, a proposal that relates predictability to allophonic variability through its effect on speech production planning.

The results of our corpus analysis showed that allophonic variation patterns in different ways with respect to predictability depending on whether the allophone is sensitive to segmental properties of adjacent words, a distinction not drawn in other theories of predictability effects. Flapping, and to a lesser extent glottalization, is a process that depends on the phonological content of a trigger word. Some aspects of the variability of these context-sensitive allophones, we argued, are explained by the fact that the phonological representation of the trigger word is not always available at the time when the current word is phonologically encoded. The Production Planning Hypothesis makes the 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 that a context-sensitive process relies on.

An area of inquiry that may further distinguish the predictions of the Production Planning Hypothesis is the study of non-reductive alternations. Processes 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. For such non-reductive processes, theories that refer directly to predictability 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 communicative account like Hall et al. (2018) could also predict a negative relationship: Liaison encodes information about the upcoming word, namely that it begins with a vowel, and might therefore in principle help with its retrieval. The Production Planning Hypothesis, on the other, is incompatible with an effect in this direction. A few pieces of evidence so far support the idea that liaison increases in predictable contexts. Côté (2013) argues that liaison is more likely when the transitional probability between a word and the syntactic category of the next word is high. Kilbourn-Ceron (2017a) found in an analysis of liaison patterns in adjective-noun and noun-adjective sequences that in both cases, the frequency of the second word increases the likelihood of liaison. This suggests to us that further work on production planning effects on liaison and other types of cross-word processes is a fruitful avenue for future research.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Plots showing the distributions of the variables used for statistical modeling. DOI: https://doi.org/10.5334/labphon.168.s1
- **Appendix B.** Tables reporting the results of non-penalized mixed-effects logistic regressions. DOI: https://doi.org/10.5334/labphon.168.s2

Acknowledgements

We extend our thanks to Morgan Sonderegger, Heather Goad, Timo Roettger, and five reviewers for their feedback on earlier versions of this manuscript. We also thank James Tanner, Jeffrey Lamontagne, Kie Zuraw, and attendees of AMP 2016 and Atelier de phonologie MOT 2017 for helpful discussion.

Funding Information

This work was supported in part by funding from the Social Sciences and Humanities Research Council-Conseil de recherches en sciences humaines. SSHRC-CRSH awarded Insight Grant 435-2014-1504 to MC and MW, and Joseph-Armand Bombardier CGS 767-2012-1089 to OKC.

Competing Interests

The authors have no competing interests to declare.

References

Alario, F.-X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*(3), 299–319. DOI: https://doi.org/10.1080/01690960143000236

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. DOI: https://doi.org/10.1177/00238309040470010201

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*(5), 3048–3058. DOI: https://doi.org/10.1121/1.2188331

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*(3), 201–211. DOI: https://doi.org/10.1177/002383097902200301

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, *113*(2), 1001–1024. DOI: https://doi.org/10.1121/1.1534836

Bermudez-Otero, R. (2011). Cyclicity. In M. van Ostendorp, C. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell companion to phonology*. Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444335262.wbctp0085

Braver, A. (2011). Incomplete neutralization in American English flapping: A production study. *University of Pennsylvania Working Papers in Linguistics*, *17*(1), 5.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3–4), 155–180. DOI: https://doi.org/10.1159/000261913

Brysbaert, M., & New, B. (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), 977–990. DOI: https://doi.org/10.3758/BRM.41.4.977

Bürki, A. (2018). Variation in the speech signal as a window into the cognitive architecture of language production. *Psychonomic Bulletin & Review*, 1–32. DOI: https://doi.org/10.3758/s13423-017-1423-4

Bybee, J. (2001). Frequency effects on French liaison. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 337–360). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tsl.45.17byb

Bybee, J. (2007). *Frequency of use and the organization of language*. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*(1–2), 39–54. DOI: https://doi.org/10.1016/0167-6393(94)90039-6

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 149–180. DOI: https://doi.org/10.1016/S0095-4470(02)00085-2

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

Chong, A. J., & Garellek, M. (2018). Online perception of glottalized coda stops in American English. *Laboratory Phonology*, *9*(1), 4. DOI: https://doi.org/10.5334/labphon.70

Coetzee, A. W., & Pater, J. (2011). The place of variation in phonological theory. In J. A. Goldsmith, J. Riggle & A. C. L. Yu (Eds.), *The handbook of phonological theory, second edition* (pp. 401–434). Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444343069.ch13

Côté, M.-H. (2013). Understanding cohesion in French liaison. *Language Sciences*, *39*, 156–166. DOI: https://doi.org/10.1016/j.langsci.2013.02.013

De Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, *26*(3), 283–310. DOI: https://doi.org/10.1006/jpho.1998.0077

Dell, G., & O'Seaghdha, P. (1992). Stages of lexical access in language production. *Cognition*, *42*(1–3), 287–314. DOI: https://doi.org/10.1016/0010-0277(92)90046-K

Eddington, D., & Channer, C. (2010). American English has go? a lo? of glottal stops: Social diffusion and linguistic motivation. *American Speech*, *85*(3), 338–351. DOI: https://doi.org/10.1215/00031283-2010-019

Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *The Journal of the Acoustical Society of America*, *120*(2), 1040–1051. DOI: https://doi.org/10.1121/1.2211548

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*(2), 233–253. DOI: https://doi.org/10.1037/0033-295X.100.2.233

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84. DOI: https://doi.org/10.1006/jmla.2001.2797

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. DOI: https://doi.org/10.1006/cogp.1999.0730

Fosler-Lussier, E., & Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in convertional speech. *Speech Communication*, *29*(2), 137–158. DOI: https://doi.org/10.1016/S0167-6393(99)00035-7

Fox, R., & Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics*, *5*, 27–34. DOI: https://doi.org/10.1016/S0095-4470(19)31111-8

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1. DOI: https://doi.org/10.18637/jss.v033.i01

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*(1), 27–52. DOI: https://doi.org/10.2307/412187

Fukaya, T., & Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association*, *35*(01), 45–58. DOI: https://doi.org/10.1017/S0025100305001891

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496. DOI: https://doi.org/10.1353/lan.0.0035

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 748–775. DOI: https://doi.org/10.1353/lan.2004.0185

Garellek, M. (2013). *Production and perception of glottal stops* (Unpublished doctoral dissertation). UCLA.

Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Proceedings of IEEE ICASSP-92 International Conference on Acoustics, Speech, and Signal Processing* (pp. 517–520). DOI: https://doi.org/10.1109/ICASSP.1992.225858

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Proceedings of the 35th annual meeting of the Chicago Linguistic Society* (Vol. 2, pp. 151–166).

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*(3), 313–338. DOI: https://doi.org/10.1006/jmla.1997.2547

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

Guy, G. R. (2007). Lexical exceptions in variable phonology. *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 9.

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard*. DOI: https://doi.org/10.1515/lingvan-2017-0027

Heafield, K. (2011). Kenlm: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation* (pp. 187–197). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2132960.2132986

Herd, W., Jongman, A., & Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics*, *38*(4), 504–516. DOI: https://doi.org/10.1016/j.wocn.2010.06.003

Huffman, M. K. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics*, *33*(3), 335–362. DOI: https://doi.org/10.1016/j.wocn.2005.02.004

Jaeger, T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. DOI: https://doi.org/10.1016/j.cogpsych.2010.02.002

Jescheniak, J. D., & Levelt, W. J. (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), 824. DOI: https://doi.org/10.1037/0278-7393.20.4.824

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language*, *45*, 229–254. DOI: https://doi.org/10.1075/tsl.45.13jur

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

Kaisse, E. M. (1985). *Connected speech: The interaction of syntax and phonology*. Academic Press.

Kawamoto, A. H., Liu, Q., & Kello, C. T. (2015). The segment as the minimal planning unit in speech production and reading aloud: Evidence and implications. *Frontiers in Psychology*, *6*, 1457. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2015.01457. DOI: https://doi.org/10.3389/fpsyg.2015.01457

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, *101*, 112–156.

Kiesling, S., Dilley, L., & Raymond, W. D. (2006). *The variation in conversation (ViC) project: Creation of the Buckeye corpus of conversational speech.* Retrieved from http://buckeyecorpus.osu.edu/BuckeyeCorpusmanual.pdf (Ms., Department of Psychology, Ohio State University, Columbus, OH)

Kilbourn-Ceron, O. (2017a). Speech production planning affects phonological variability: A case study in French liaison. In *Proceedings of the 2016 Annual Meeting on Phonology*. DOI: https://doi.org/10.3765/amp.v4i0.4004

Kilbourn-Ceron, O. (2017b). *Speech production planning affects variation in external sandhi* (Unpublished doctoral dissertation). McGill University, Montréal, Québec.

Kilbourn-Ceron, O., & Goldrick, M. (2019). The scope of phonological encoding in connected speech. Poster presented at the Psychonomic Society Annual Meeting, November 14–18, 2019. Palais des Congrès, Montréal, QC.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*(4), 463–492. DOI: https://doi.org/10.1080/02643290701674851

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162. DOI: https://doi.org/10.1016/j.jml.2011.08.003

Labov, W. (1994). *Principles of linguistic change. Volume I: Internal factors* (No. 20). Oxford: Blackwell.

Lamontagne, J., & Torreira, F. (2017). Directionality in production planning: Cross-word hiatus resolution in Spanish. In *Meeting of the Canadian Linguistics Association.*

Levelt, W. J. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*(23), 13464–13471. DOI: https://doi.org/10.1073/pnas.231459498

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. DOI: https://doi.org/10.1017/S0140525X99001776

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*(3), 172–187. DOI: https://doi.org/10.1177/002383096300600306

Liu, Q., Kawamoto, A. H., Payne, K. K., & Dorsey, G. N. (2018). Anticipatory coarticulation and the minimal planning unit of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 139. DOI: https://doi.org/10.1037/xhp0000443

MacKenzie, L. (2012). English auxiliary contraction as a two-stage process: Evidence from corpus data. In *Proceedings of the West Coat Conference on Formal Linguistics 29* (pp. 152–160).

MacKenzie, L. (2016). Production planning effects on variable contraction in English. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 121–130.

Malécot, A., & Lloyd, P. (1968). The/t:/d/distinction in American alveolar flaps. *Lingua*, *19*(3–4), 264–272. DOI: https://doi.org/10.1016/0024-3841(68)90084-3

McAuliffe, M., Stengel-Eskin, E., Socolof, M., & Sonderegger, M. (2017). Polyglot and Speech Corpus Tools: A system for representing, integrating, and querying speech corpora. In *Proceedings of Interspeech 2017*. DOI: https://doi.org/10.21437/Interspeech.2017-1390

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69–89. DOI: https://doi.org/10.1016/0749-596X(91)90011-8

Michel Lange, V., & Laganaro, M. (2014). Inter-subject variability modulates phonological advance planning in the production of adjective-noun phrases. *Frontiers in Psychology*, *5*. DOI: https://doi.org/10.3389/fpsyg.2014.00043

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, *132*(2), 228. DOI: https://doi.org/10.1037/0096-3445.132.2.228

Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. *Journal of Speech, Language, and Hearing Research*, *39*(1), 93–104. DOI: https://doi.org/10.1044/jshr.3901.93

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.

Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature*, *202*, 1031–1032. DOI: https://doi.org/10.1038/2021031a0

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. DOI: https://doi.org/10.1080/17470216508416445

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production. *Phonetica*, *58*(4), 254–275. DOI: https://doi.org/10.1159/000046178

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–158). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tsl.45.08pie

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech (2nd release) [www.buckeyecorpus.osu.edu]*. Columbus, OH: Department of Psychology, Ohio State University.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*(1), 89–95. DOI: https://doi.org/10.1016/j.specom.2004.09.001

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*(2–4), 146–159. DOI: https://doi.org/10.1159/000090095

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569. DOI: https://doi.org/10.1121/1.2011150

Purse, R. (2019). The articulatory reality of coronal stop 'deletion'. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 1595–1599).

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Randolph, M. (1989). *Syllable-based constraints on properties of speech sounds* (Unpublished doctoral dissertation). MIT.

Raymond, W. D., Brown, E. L., & Healy, A. F. (2016). Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. *Language Variation and Change*, *28*(2), 175–202. DOI: https://doi.org/10.1017/S0954394516000041

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407–429. DOI: https://doi.org/10.1006/jpho.2001.0145

Roberts, J. (2006). As old becomes new: Glottalization in Vermont. *American Speech*, *81*(3), 227–249. DOI: https://doi.org/10.1215/00031283-2006-016

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281. DOI: https://doi.org/10.3758/BF03201199

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 450–466. DOI: https://doi.org/10.1016/S0022-5371(84)90291-3

Seyfarth, S., & Garellek, M. (2015). Coda glottalization in American English. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS) in Glasgow*.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342).

Shaw, J., & Kawahara, S. (2018). Predictability and phonology: Past, present and future. *Linguistics Vanguard*, *4*(2). DOI: https://doi.org/10.1515/lingvan-2018-0042

Sternberg, S., Monsell, S., Knoll, R., & Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information Processing in Motor Control and Learning*, 117–152. DOI: https://doi.org/10.1016/B978-0-12-665960-3.50011-6

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*(3), 322–338. DOI: https://doi.org/10.1016/j.jml.2004.11.004

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: Evidence from eye-movements. *Language and Cognition*, *6*(1), 12–44. DOI: https://doi.org/10.1017/langcog.2013.5

Tamminga, M. (2018). Modulation of the following segment effect on English coronal stop deletion by syntactic boundaries. *Glossa: A Journal of General Linguistics*, *3*(1), 1–27. DOI: https://doi.org/10.5334/gjgl.489

Tamminga, M., MacKenzie, L., & Embick, D. (2016). The dynamics of variation in individuals. *Linguistic Variation*, *16*(2), 300–336. DOI: https://doi.org/10.1075/lv.16.2.06tam

Tanner, J., Sonderegger, M., & Wagner, M. (2017). Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology*, *8*(1). DOI: https://doi.org/10.5334/labphon.96

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, *71*, 249–267. DOI: https://doi.org/10.1016/j.wocn.2018.09.004

Torreira, F., & Ernestus, M. (2009). Probabilistic effects on French [t] duration. In *10th annual conference of the International Speech Communication Association (Interspeech 2009)* (pp. 448–451).

Turk, A. (1992). The American English flapping rule and the effect of stress on stop consonant durations. *Working Papers of the Cornell Phonetics Laboratory*, *7*, 103–133.

Turk, A. (2010). Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology*, *1*(2), 227–262. DOI: https://doi.org/10.1515/labphon.2010.012

Turnbull, R., Seyfarth, S., Hume, E., & Jaeger, T. (2018). Nasal place assimilation trades off inferrability of both target and trigger words. *Laboratory Phonology*, *9*(1), 15. DOI: https://doi.org/10.5334/labphon.119

Wagner, M. (2011). Production-planning constraints on allomorpy. *Proceedings of the Acoustics Week in Canada. Canadian Acoustics*, *39*(3), 160–161.

Wagner, M. (2012). Locality in phonology and production planning. In J. Loughran & A. McKillen (Eds.), *Proceedings of phonology in the 21 century: Papers in honour of Glyne Piggott* (Vol. 22, pp. 1–18). Montreal, QC.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (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), 423. DOI: https://doi.org/10.1037/a0018619

Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In *Explicit and implicit prosody in sentence processing* (pp. 85–98). Springer. DOI: https://doi.org/10.1007/978-3-319-12961-7_5

Whalen, D. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*, 3–35. DOI: https://doi.org/10.1016/S0095-4470(19)30356-0

Wheeldon, L. R. (2013). Producing spoken sentences: The scope of incremental planning. In P. Perrier & P. L. Verlag (Eds.), *Cognitive and physical models of speech production, speech perception, and production-perception integration* (pp. 97–156).

Wheeldon, L. R., & Konopka, A. E. (2018). Spoken word production. In S.-A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford handbook of psycholinguistics* (2nd ed., p. 335). Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198786825.013.15

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*(3), 356–381. DOI: https://doi.org/10.1006/jmla.1997.2517

Wheeldon, L. R., & Lahiri, A. (2002). The minimal unit of phonological encoding: Prosodic or lexical word. *Cognition*, *85*(2), 31–41. DOI: https://doi.org/10.1016/S0010-0277(02)00103-8

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *92*, 1707–1717. DOI: https://doi.org/10.1121/1.402450

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America*, *66*(4), 1039–1050. DOI: https://doi.org/10.1121/1.383323

How to cite this article: Kilbourn-Ceron, O., Clayards, M., & Wagner, M. 2020 Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realizations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1): 5, pp. 1–28. DOI: https://doi.org/10.5334/labphon.168

Submitted: 14 September 2018 **Accepted:** 20 April 2020 **Published:** 10 June 2020

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS