UNIVERSITY OF CALIFORNIA

Los Angeles

The Role of Markedness in Phonological Processing Above the Word Level

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Arts in
Linguistics

by

Canaan Michael Breiss

2020

ABSTRACT OF THE THESIS

The role of markedness in phonological processing above the word level

by

Canaan Michael Breiss

Master of Arts in Linguistics

University of California, Los Angeles, 2020

Professor Jesse Harris, Co-Chair

Professor Bruce Hayes, Co-Chair

A rich body of psycholinguistic research demonstrates that the phonotactic well-formedness of a monomorphemic (non)word has an effect on the way that word is processed when perceived, and also impacts the cognitive effort involved its production in speech. This study examines the effect of phonological markedness across a word boundary on processing, looking at prosodically close English Adjective-Noun sequences in a well-formedness judgment task, a speeded production paradigm, and an accompanying speech-error analysis. I found that different types of phonotactic markedness are distinct in their strength of impact on processing in speech production, although only a subset of the constraints tested show a significant impact of markedness compared to unmarked items, and only one of them shows this effect in the expected inhibitory direction. These results raise interesting questions about the relationship between the neural-level language-production system and the symbolic-level phonological grammar and how the two interact in speech production.

The thesis of Canaan Michael Breiss is approved.

Kie Zuraw

Jesse Harris, Committee Co-Chair

Bruce Hayes, Committee Co-Chair

University of California, Los Angeles

2020

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and most importantly, many thanks are due to my committee, Jesse Harris, Bruce Hayes, and Kie Zuraw – their patient and insightful guidance, advice, support, and critique was instrumental in bringing this project to fruition, and in helping me grow as a researcher. I am also grateful for the tireless help of Liz Martin, Xi (Lily) Xu, and Erica Brain for invaluable assistance running participants and coding data, and to Eleanor Glewwe and Adam Royer for recording stimuli for the experiment. I'd also like to thank Megha Sundara for helpful initial discussions concerning methodology, as well as the audience at the 2018 California Meeting on Psycholinguistics and Amanda Rysling for helpful feedback.

Two years after this MA thesis was complete, I'd also like to thank Beth Sturman for encouraging me to file officially, and Christine Prechtel and Jeremy Steffman for much-needed pomodoro sessions and good company which helped this officially get done.

0. Introduction

A rich body of psycholinguistic research demonstrates that the phonotactic probability of a monomorphemic (non)word has an effect on the way that word is processed when perceived, and also impacts the cognitive effort involved its production in speech. Phonotactic probability, which is often equated with the more theoretically-nuanced concept of phonological markedness, is often calculated in terms of the sum of a word's phonemic positional probability and/or biphone probability. It has also been found to impact speech production in polymorphemic real words, mirroring the finding in monomorphemes. This study examines the effect of phonological markedness across a word boundary on processing, looking at prosodically close English Adjective-Noun sequences in a well-formedness judgment task, a speeded production paradigm, and an accompanying speech-error analysis.

Motivating the current study's focus on between-word phonological markedness, corpus-based evidence has shown that speakers avoid creating phonologically marked word-boundaries in speech, and researchers have argued that the types of markedness which drive this phrasal avoidance are related to the language's word-internal phonotactics. For example, English bans adjacent sibilants within roots (ex., *[ˈpɪʃ.sɚ] is not a possible word of English) and similarly speakers of English gradiently avoid creating sequences of adjacent sibilants across a word boundary (strings of words such as *push someone* are less likely than expected based on non-phonological factors). Based on these findings, I test the hypotheses that such marked sequences across a word-boundary give rise to increased processing cost in speech production, as measured by response latency in a production task, relative to unmarked word-transitions. I also conduct a post-hoc analysis of the production task to determine whether speech-errors are more likely in marked items than in unmarked items, and an online ratings experiment to test whether subjects judge marked two-word sequences as less well-formed than their unmarked counterparts.

In pursuit of this aim, I outline the current understanding of the intersection between phonological grammar and the speech production system, describe the methods and findings of each of the three experiments named above, and discuss their implications for phonological and psycholinguistic theories of speech production. I find that participants marginally prefer phonologically unmarked phrases to marked ones, and that contrary to predictions, different types of phonologically marked word-boundaries affected processing difficulty in different ways, with some types of markedness being associated with a distinct delay in response latency, and others being reliably linked to increased processing ease and lower response-latencies. I also find that speech errors are no more likely to occur in marked phrases than in unmarked ones, running counter to the predicted trend. Taken together, these findings indicate that the interface between the phonological grammar and the post-lexical stage of processing is modulated by a greater range of factors than previously assumed, and that cross-word phonological markedness has a nuanced effect on speech production. I close by outlining questions raised by these findings.

1. Background

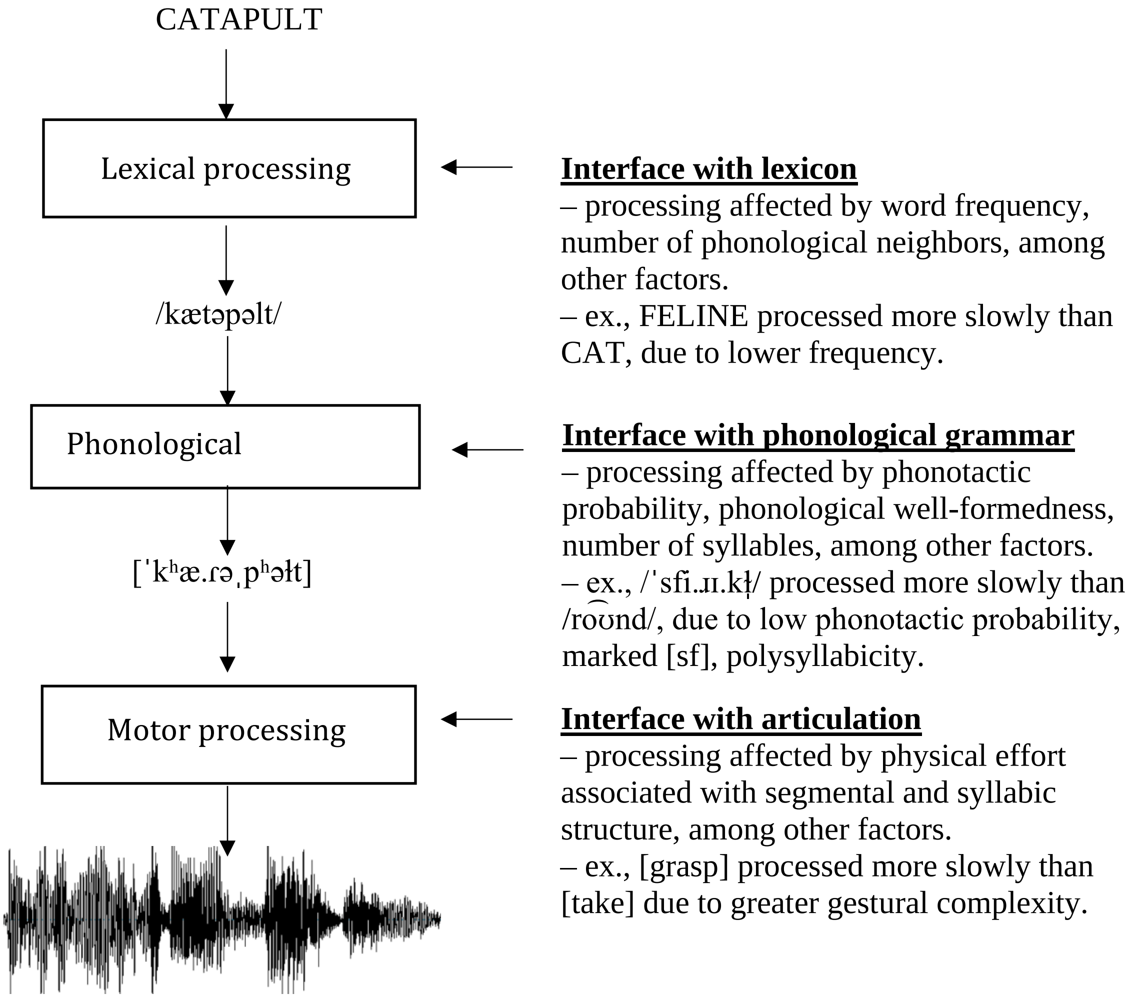
The investigation into the nature of the speech production mechanism, including the interfaces between the lexicon, phonological grammar, and speech production, has long been a central topic of study for linguists, psychologists, and cognitive scientists. Below, I outline the main points of scholarly agreement about the speech-production system and highlight how lexical and grammatical factors interface with these processes. I also discuss non-psycholinguistic findings which motivate the present study.

1.1 Speech production, phonotactics, and the lexicon: an overview

I couch the present studies in the model of speech production outlined by Goldrick & Rapp (2007) and Kirov (2014), which divide the speech production process into three broad stages: *lexical* processing, *post-lexical* processing (also called *phonological* processing), and *motor* processing. In the course of speaking an utterance, the speaker proceeds through each of the stages in turn. In the description that follows, I focus on the aspects and implications of this system for phonological aspects of speech production, although syntactic, semantic, and pragmatic factors make use of the system as well.

Given the intention to speak a certain word, the speaker first carries out lexical processing in which they retrieve the word's rough phonological form, an abstract set of symbols underspecified for syllabic, prosodic, and featural structure, from long-term memory in the lexicon. Accessing this representation is facilitated by the increased frequency of the target word in question, as well as by the similarity of the target word to other words in the lexicon (having a greater *phonological neighborhood size*). Having retrieved the phonological form of a word, roughly equivalent in information-content to the conception of an *underlying representation* in phonological theory (cf. Chomsky & Halle 1968), post-lexical processing is responsible for linking the string of phonemic symbols to a syllabic and metrical grid, filling in underspecified featural and prosodic structure, and applying phonological processes. This phonetically-specified form of a word, roughly equivalent to a *surface representation* in phonological theory, is then translated to a series of articulatory representations, which in turn drive motor action.

The diagram below (inspired by Kirov 2014) illustrates the system, along with its interfaces with the lexicon and phonological grammar, discussed below.



Interface with lexicon
– processing affected by word frequency, number of phonological neighbors, among other factors.
– ex., FELINE processed more slowly than CAT, due to lower frequency.

Interface with phonological grammar
– processing affected by phonotactic probability, phonological well-formedness, number of syllables, among other factors.
– ex., /ˈsfi.ɹɪ.kɫ̩/ processed more slowly than /ro͡ʊnd/, due to low phonotactic probability, marked [sf], polysyllabicity.

Interface with articulation
– processing affected by physical effort associated with segmental and syllabic structure, among other factors.
– ex., [grasp] processed more slowly than [take] due to greater gestural complexity.

Figure 1: A schematic overview of the speech-production system.

1.2 Interfaces in processing

The model structure and location of interfaces outlined above accounts for findings that the behavior at different levels of speech production and processing is impacted in discrete ways by lexical structure and phonological grammar. Specifically, Goldrick & Larson (2008) and Goldrick (2011) review compelling evidence that the fact that frequent words have lower response latency in production tasks, and that those with more phonological neighbors are produced more quickly and accurately (Vitevitch 2002), arises as a consequence of the structure of the lexicon interacting with lexical phonological processing. They also provide evidence that properties governed by a language's phonological grammar, and in particular the phonotactic markedness of an utterance's segments or syllables (as assessed by the similar, though not identical, metric of *phonotactic probability*), are significant predictors of response latency in production tasks, after controlling for lexical-level factors (Goldrick & Larson 2008). Goldrick (2011) reviews a range of evidence which supports this finding, and discusses the specific mechanism by which such facilitatory or inhibitory effects emerge from the interaction of the phonological grammar and post-lexical processing. He argues that phonotactic constraints shape the course of post-lexical phonological encoding processes which link the symbolic lexical representation to a fully-specified surface-form, and that the connections between lexical and phonetic representations vary with both absolute and relative phonotactic well-formedness of the sequence being encoded, based on information from the phonological grammar. This leads to well-formed structures being encoded more accurately and quickly, and ill-formed structure being more time-consuming and error-prone in their encoding. This is the mechanism which the current study seeks to investigate.

1.3 The domain of interaction between phonological markedness and processing

Among the studies which have found systematically differing response latencies and error rates in spoken production tasks for high- and low-phonotactic-probability stimuli, most have used single real or nonce words as stimuli (cf., e.g., Vitevitch & Luce 2005). Among the few studies which have looked at the effect of phonotactic well-formedness in polymorphemic stimuli, Goldberg (2010) shows that both mono- and polymorphemic stimuli which contained a tautosyllabic onset and coda sequence with similar or identical consonants (e.g., *kick, cog, dimmed*) have longer response latencies in a production task than unmarked forms, a finding which he situates in terms of the Obligatory Contour Principle (OCP), a preference to avoid identical repeated segments (cf., e.g., Yip 1988). Goldberg also shows that compounds which created a geminate at their join (e.g., *pen name*) have longer spoken response latencies than compounds which did not contain a such a heteromorphemic sequence. In another departure from much of the literature on phonotactic markedness in speech-production, Goldberg extends this finding beyond geminate-forming polymorphemes: he shows that other heteromorphemic phonotactically marked sequences, including iambic stress clash and obstruent clusters, were frequently repaired by speech errors in the production of a brain-damaged individual, giving a window into the way the phonology interacts with heteromorphemic markedness.

No study that I am aware of has looked at the effect of phonological well-formedness on processing in speech production *across* a word-boundary, however – the current study takes up this task, with an explicit assumption that the same processes and interfaces which interact to shape the time course of production of single marked words will extend to sequences across a word boundary. This assumption is justified in part by the work of Goldberg reviewed above, as well as an emerging body of evidence from corpus-phonology and subjective ratings tasks, reviewed below.

1.4 Corpus and behavioral evidence for phonotactic markedness across the word boundary

Outside of psycholinguistics, there is a large body of behavioral evidence showing that speakers possess fine-grained well-formedness judgments covering a range of phonological sequences, both those attested in their native language as well unattested sequences for which they have no direct lexical evidence to support their judgments. Investigations of speakers' subjective well-formedness judgments have been carried out on nonce-words of varying markedness (Albright 2009a, and references therein) as well as phonological markedness resulting from morphosyntactic derivation (Shih 2014, and references therein). Across such tasks, speakers exhibit robust, reliable preferences for attested, unmarked phonological sequences over attested, marked ones; and unattested, unmarked sequences over unattested, marked ones.

In recent years, evidence from large-scale lexical and corpus-based phonological analysis has demonstrated that the same phonological well-formedness judgments that speakers access in subjective item-rating tasks manifest themselves passively as well in shaping their lexicon and on-line lexical and constructional choice. Martin (2011) found that the lexicons of Navajo and English systematically under-represent compounds which violate language-specific phonological markedness-constraints at their join. For example, Martin finds that English underrepresents compounds which create geminate consonants at the join between the two compound members, such as *bookkeeper*, which contains a geminate [kː]. Martin explains this systematic under-representation of phonologically-marked lexical items as the cumulative result of generations of speaker choice about what potential compounds to create, and further, of those attested in their passive linguistic exposure, which items successive generations of speakers choose to use in their own speech. Thus, a synchronic dispreference for creating or using attested phonologically marked lexical items leads to a skewing of the lexicon over time to over-represent phonologically well-formed words, and underrepresent phonologically marked ones.

Extending Martin's line of inquiry, corpus-based work by Breiss & Hayes (2018) have shown that, after controlling for syntactic obligatoriness, fixed expressions, and relative word frequency, English speakers systematically underrepresent two-word sequences which create a phonologically marked sequence of phonemes across the join between the two words. Examining several distinct corpora of written and spoken English, they find that these patterns extend to multiple types of markedness, including the geminate-avoidance found in compounds by Martin. The table below displays the average over- or under-representation of a number of markedness constraints violated across word-joins in their six corpora, measured in standard deviations from the expected number of such marked word-sequences under the hypothesis that phonological markedness plays no role in word-concatenation (effect sizes of magnitude 2.4 or greater are significant at the $p < 0.01$ level).

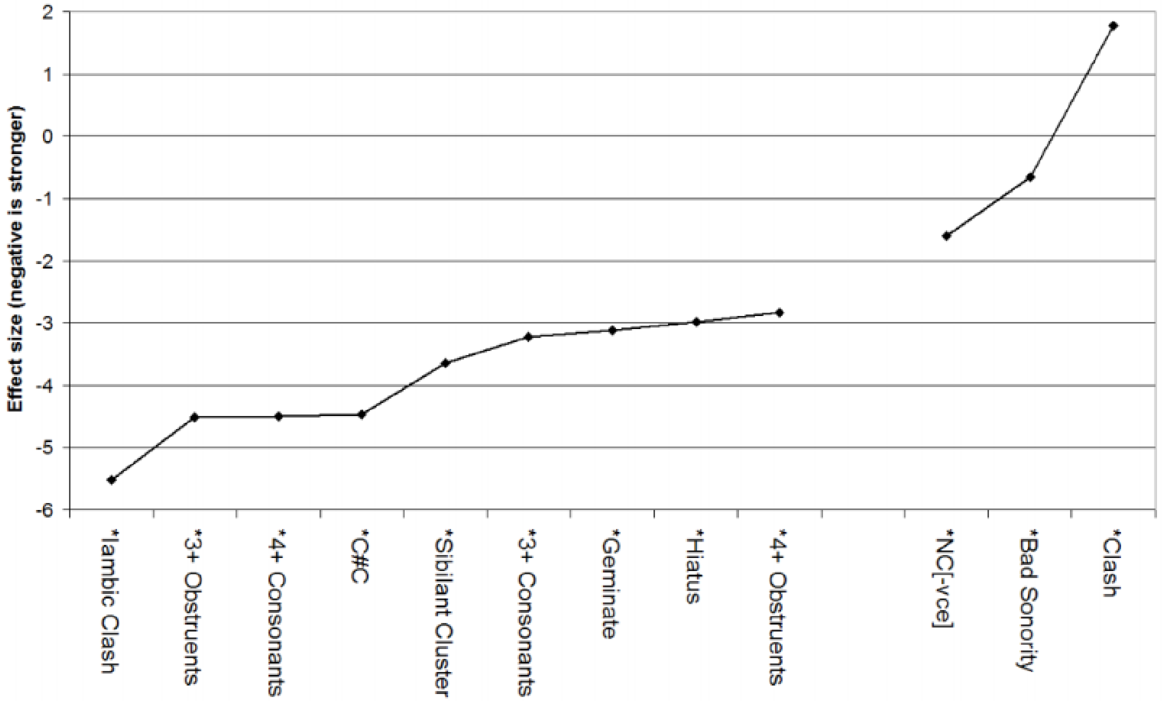


Figure 2: *Relative under- and over-representation of phonological context between words. The constraints *IAMBIC CLASH, *SIBILANT CLUSTER, *3+ CONSONANTS, and *GEMINATE are described below in 3.1, *3+ OBSTRUENTS, *4+ OBSTRUENTS, *C#C, and *4+ CONSONANTS are restrictions on consonant clusters of varying sizes across word boundaries, *HIATUS is a ban on any V#V sequence, *NC[-vce] is a restriction on voiceless consonants after nasals (see Pater 1999), *BAD SONORITY is a restriction on sequences violating the Syllable Contact Law (Vennemann 1988 [2011]), and *CLASH is a restriction on adjacent stressed syllables. The gap in the horizontal axis separates constraints which were significantly under-represented (on the left) from those which were not (on the right).*

2. Hypotheses

2.1 Hypotheses about the relationship between phonological grammar and processing

Prosodic domain	*Processing cost of markedness?*	*Marked status in the grammar?*
Monomorpheme / root	✓ (Vitevitch & Luce 2005, Goldrick & Rapp 2008)	✓ (Albright, 2009b on cumulative complexity effects in syllables)
Polymorpheme / inflected word	✓ (Goldberg 2010)	✓ (Chong, 2016 on derived environment effects)
Phrase	? *(the current study)*	✓ (Breiss & Hayes 2018, Shih 2017)

Table 1: *Schematic review of literature discussed to this point, and the place of the current study.*

Table 1 summarizes the relationship between the current study and the existing research on phonological markedness in speech production, as discussed in the preceding section. In this light, the current investigation of the processing penalty of phonological markedness in multi-word utterances builds on two strands of previous work: findings that markedness in mono- and poly-morphemes causes a response latency increase in spoken production, and studies showing that phonological markedness is active at the phrasal level. Given these findings, we can make two structured hypotheses about the way a multi-word utterance would be processed by and interact with the speech-production system, grammar, and lexicon, and from these situate each of the experiments described below with respect to particular predictions about their findings.

2.1.1 Hypothesis 1 (“the equality of domains”):

Between-word phonologically marked sequences should be processed in a similar way to within-word phonologically marked sequences.

Since inhibitory effects of phonological markedness have been shown to be active at the root- and inflected-word level, and it is well-established that phonological processing is also constrained by a “window”, extending from a few syllables to a word or two ahead of the one currently being produced (cf. Levelt 2001), we can hypothesize that multi-word utterances are manipulated by the same processes which operate on one-word utterances. That is, if two words have been retrieved and are being processed in post-lexical speech production at the same time, the algorithms by which abstract phonemic representations are related to their phonetically-specific “proto-surface-forms” should be impacted by the phonotactic markedness of the sequences of phonemes across the word boundary in the same way that they are impacted by the same markedness of the sequences across a word-internal morpheme-boundary, or within a root itself. The degree to which findings in the present studies deviate from replicating those effects on production which are well-known from the word-internal domain can be informative about the role that the word-boundary as a structure (prosodic or syntactic) is playing in phonological processing.

2.1.2 Hypothesis 2 (“the equality of markedness”):

All marked sequences will exert a similar impact on processing. The degree to which types of markedness impact processing differently can diagnose mismatches between (formal) phonological markedness and (psycholinguistic) phonotactic probability.

While the picture laid out by Goldrick (2011) about the interaction of the phonological grammar and post-lexical processing is quite clear, the question of what counts as a “phonotactically well-formed sequence” is still very much open for debate. I am unaware of other studies in the phonological processing literature which examine the possibility of markedness-type as a contributing factor to markedness-induced increases in response latency that this and other studies have found. In the majority of the speech-production literature reviewed here, phonological markedness is discussed primarily in terms of phonotactic probability – that is, the likelihood that a given phoneme or syllable will occur in the environment of another particular phoneme or syllable. This probability scale is often divided up into a binary “high probability” and “low probability” items (cf. e.g., Vitevitch 2002, Vitevitch et al. 2004, Vitevitch & Luce 2005), without reference to the phonemic or phonetic configurations of items in such categories.

From the phonological point of view, however, the designation of a phonemic sequence as “marked” or “unmarked,” while certainly meaningful, does not capture the whole story about the kinds of knowledge speakers have about phonological distributions and acceptability in their language. Specifically, it is implicitly acknowledged in generative phonology that marked structures are not all alike: different specific low-probability phoneme-sequences can give rise to marked structures which differ phonetically as well as in the way they are treated by the phonological system of the language in question. Thus from a phonological point of view, a description of a sequence of phonemes as “marked” without reference to the phonetic substance of the segments involved or the phonological system in which it is embedded, is somewhat incomplete.

For example, Tagalog (Austronesian, the Philippines) has phonological process which repairs a sequence of a nasal followed by an obstruent via nasal place assimilation to the obstruent and deletion of the obstruent itself – often known as “nasal substitution” (Pater 1999, Zuraw 2000, 2010, et seq.). For example, /maŋ + tulis/ → [manulis] “to write”. Thus, in Tagalog, such a sequence, abbreviated *NC, is considered marked. Tagalog also has a morphological process which infixes VC structure containing a bilabial nasal: for example, /RED + pili / → [pumili] “to choose”. However, for roots with initial bilabial sonorants, this process is blocked: the infixed form of an /m/-initial root is homophonous with the base. These structures are both marked in that they are repaired, but each is subject to a different repair - *NC-violations are repaired phonologically, while consecutive bilabial onsets are “repaired” via morphological blocking. Both sequences are marked, but they are not repaired in the same way in the language’s (morpho)phonological grammar – marked sequences are not all created (or repaired) equally; most often the repair is specific to the type of markedness in question.

Reconciling the phonological and production-oriented notions of markedness raises a number of questions, many which are larger in scope than the present study allows discussion for. It suffices, however, to note that the predictions that psycholinguistic findings make about phonotactic probability, interpreted here (and by Goldrick, 2011) as being roughly equivalent to phonological markedness, and the predictions about phonological markedness stemming from generative phonology are not the same with respect to the influences of phonological markedness-types on processing. In the psycholinguistic literature, the finding that “phonotactic probability” influences processing time in one-word items is fairly well established, as discussed above – however, the measures of phonotactic probability do not take into account the phonetic substance of the low-frequency, “marked” items, and implicitly predicts that any sufficiently low-probability sequence should impact processing in the same way. Phonotactic markedness as envisioned phonologically, however, is difficult to divorce from its particular phonemic substance – under this schema, it is not a given that any marked sequence will impact processing in the same way. The current study pits these notions of markedness against one another – it explores the effect of discrete types of phonological markedness in sequences which all are either unattested or rare in the English lexicon, and so have low phonotactic probability. Hypothesis 2 (the “equality of markedness” hypothesis) predicts, then, that formal phonological markedness will be equivalent to phonotactic probability in terms of its effect on response latency. I adopt it here as the null hypothesis, in line with much of the research reviewed above, but am not committed to its validity beyond this.

2.2 Predictions for experimental findings

Turning now from broader hypotheses about the architecture of the phonology-processing interface and its implications, the current study tests specific predictions about the effect of markedness and markedness-type on the subjective judgment of marked cross-word sequences, as well as the impact of such markedness on spoken response latency, and the incidence of speech errors in such tokens.

I will use experimental investigation to evaluate three predictions:

1. Phonologically marked items are rated as less well-formed than unmarked items in a subjective acceptability-judgment task. I term this prediction the *preference prediction*.

2. Phonologically marked items have longer response latencies than their unmarked counterparts in a production task, indicative of a greater degree of inhibition in post-lexical processing. Further, while I do not make a particular prediction about the manner in which markedness-type will contribute to explaining the potential impact of markedness on processing, I note that this question may be clarified in the process of addressing this prediction. I term this prediction the *response latency prediction*.

3. Phonologically marked items are more likely to contain speech-errors than their unmarked counterparts in the same production task. Similar open questions about the role of markedness-type may be addressed here as well. I term this prediction the *speech error prediction.*

3. Experiments

To evaluate the hypotheses laid out above, I conducted three experiments – two separate tasks, and one post-hoc analysis of a subset of the data from one of the tasks. To evaluate the *preference prediction*, I conducted an online forced-choice ratings task; to evaluate the *response latency prediction*, I carried out a speeded production experiment to measure the time to speech for marked vs. unmarked items, and to address the *speech error prediction* I carried out a post-hoc analysis on speech error-having tokens excluded from the production experiment.

Because the materials and participants are consistent throughout all three experiments, I describe them here.

3.1 Materials

Stimuli consisted of 100 two-word sequences (hereafter *items*), divided into 50 pairs (hereafter *groups*). Each of the items in a group had the same first word and differing second words, such that one item's *word-one* + *word-two* combination created a cross-word sequence which violated a specific phonological markedness constraint, and the other created an unmarked phonological sequence. Recalling the above discussion, the item *squeamish sergeant* creates a sequence of clashing sibilants between the two words – a sequence banned within English words and underrepresented in phrasal construction (Breiss & Hayes, 2018) – while its counterpart in its group, *squeamish agent*, contains an unmarked word-transition. I term an individual item's status as to whether it contains a marked word-transition its *condition*.

Five different types of markedness (hereafter *constraints*, in the sense of the Optimality-Theoretic phonological markedness constraint violated by marked items in a given group (Prince & Smolensky 2008 [1993])) were represented in the stimuli, each comprising 10 groups, making 10 marked and 10 unmarked items for each constraint. The constraints were chosen on the basis of the corpus work discussed above such that each markedness-type was avoided phrasally. The markedness-types are:

- *BADHIATUS – adjacent members of the set [ə, ʌ, a, æ, ɔ] across a word boundary. These particular vowels create hiatus which cannot be fixed by glide-formation, and so is hypothesized to be more marked. Sample stimuli violating this constraint are *magenta alpaca, raw agreement,* and *sepia arrangement*.

- *GEMINATE – a sequence of two identical consonants. English bans morpheme-internal geminate consonants, and avoids them in productive morphology (Martin 2011). Sample stimuli violating this constraint are *horrid deluge*, *solemn mentor*, and *gentle lagoon*.
- *IAMBICCLASH – a *word-one* with iambic stress (a final primary stress preceded by an unstressed syllable) followed by a *word-two* with initial primary stress), following Breiss & Hayes 2018). Sample stimuli violating this constraint are *aloóf heíress*, *opáque páttern*, and *supréme áctor*.
- *SIBILANTCLUSTER – a sequence of two sibilants (English /s, z, ʃ, ʒ/), either as independent phonemes (in the case of *word-two* onset) or as the fricative release from English phonemes /tʃ, dʒ/ (for a *word-one* coda). English bans word-internal sibilant clusters, and disprefers them as a result of morphosyntactic construction (cf. Shih et al. 2015). Sample stimuli violating this constraint are *dangerous zebra*, *rich salami*, and *hodgepodge sherry*.
- *THREECONSONANTS – a sequence of three or more consonants (Breiss & Hayes 2018). Sample stimuli violating this constraint are *frantic scratching, landmark statement,* and *hardened camouflage*.

In addition to being selected for their particular phonological properties, items within a group were also carefully selected to match *word-two*'s for a number of other lexical properties which have been shown to influence response latency in similar tasks (e.g. Fiez et al. 1999, Vitevitch & Sommers 2003). Specifically, with-group *word-two*'s were selected to match as closely as possible in phonological neighborhood size (roughly, the number of other lexemes created through the deletion, addition, local exchange, or substitution of a single phoneme of a given form – this metric is also known as having Levenshtein distance, or edit distance, of one), number of syllables, stress pattern (except for the *IambicClash constraint, where marked status crucially depends on the stress pattern of *word-two*), two different measures of lexical frequency (retrieved from the English Lexicon Project database, Balota et al. 2007). *Word-one*'s were identical within groups, but across groups within markedness types they were selected to contain similar stress patterns and number of syllables. Paired t-tests indicated no significant difference in phonological neighborhood size and frequency between marked and unmarked *word-two's*, both as a whole and within constraint. Since *word-one*'s were identical across items within groups, only syllable count and stress pattern was controlled for. Table 2 below shows the means and standard deviations for log-frequency and phonological neighborhood size broken down by condition and constraint.

	Log-frequency		Phonological neighbors	
	Marked	*Unmarked*	*Marked*	*Unmarked*
*BadHiatus	2.11 (0.20)	2.12 (0.28)	0.20 (0.42)	0.80 (0.20)
*Geminate	1.72 (0.22)	1.77 (0.20)	1.10 (0.62)	0.50 (0.27)
*IambicClash	2.39 (0.15)	2.35 (0.16)	3.70 (1.40)	3.00 (1.08)
*SibilantCluster	2.02 (0.23)	2.12 (0.28)	5.40 (2.88)	1.70 (0.07)
*ThreeConsonants	2.29 (0.22)	2.42 (0.22)	0.90 (0.41)	1.30 (0.45)
Overall	2.10 (0.09)	2.17 (0.09)	2.26 (0.69)	1.46 (0.30)

Table 2: *Means and standard errors (in parentheses) for word-twos broken down by constraint and condition.*

3.1.1 A note on the status of the *BadHiatus constraint

In the course of further analysis after the stimuli for this experiment were finalized, it was discovered that the phonological configuration denoted by *BadHiatus was in fact *not* underrepresented phrasally in Breiss & Hayes' corpus project– only hiatus-forming bigrams when examined as a whole are avoided, while the particular [*vowel # vowel*] sequence specified in *BADHIATUS does not drive phrasal avoidance. I proceed with the analysis as planned, but note this finding here for the sake of thoroughness.

3.2 Participants

Participants for all experiments were UCLA undergraduates recruited through the SONA Psychology Subject Pool, and were compensated with course credit for their participation. Only potential participants who self-identified as native speakers of English were eligible to sign up, and all participants completed a language-background questionnaire before or after each experiment to determine their status as either monolingual (i.e., speaking only English), or whether they spoke other non-English languages natively. Crucially, participants in the judgment task were ineligible to participate in the production task, and vice versa.

3.3 Experiment 1: Judgment task

In the vein of other acceptability-judgments on mono- and polymorphemic stimuli, I carried out an acceptability-judgment task to see whether speakers exhibited a preference for unmarked over marked items.

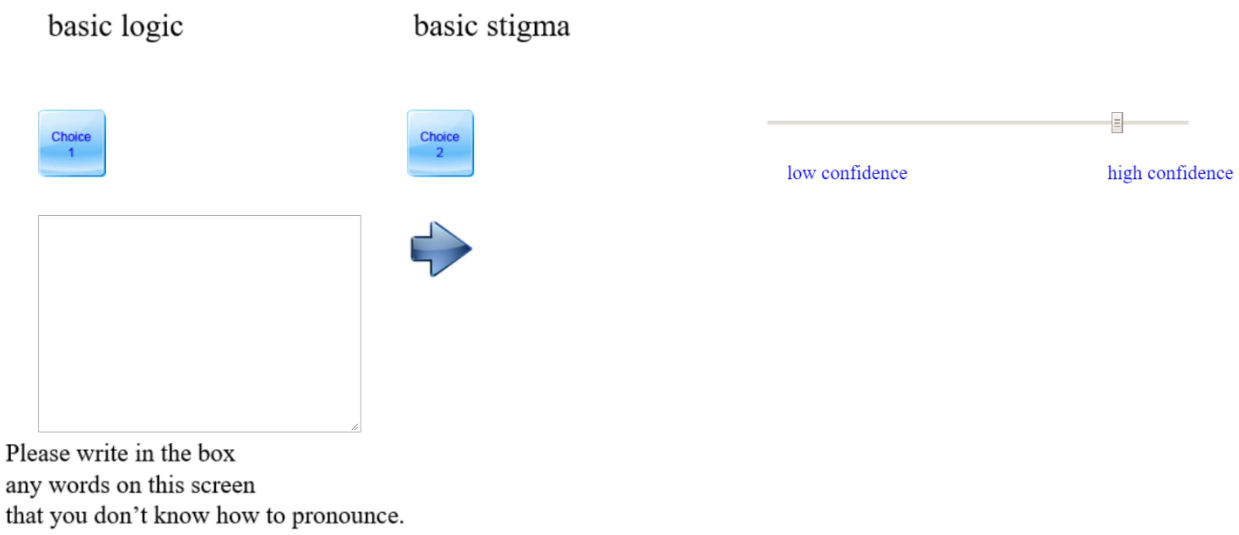
3.4 Procedure

To minimize between-experiment differences and to facilitate easy direct comparison of results, I used the same items, described above, which are used in the production study. The study was conducted online through the Appsobabble platform (Tehrani 2015), and Author Recognition Task (ART) scores and linguistic background information was collected via an IbexFarm (www.spellout.net/ibexfarm/) questionnaire prior to the experiment itself. ART scores are a common variable collected in many psycholinguistic tasks: a participant’s ART score is a rough measurement of their passive print exposure, and has been shown to be correlated with many performance-related factors in psycholinguistic research; I used an updated version of the test created by Moore & Gordon (2015), who built on the original by Stanovich & West (1989). I collected it here so as to be able to determine whether a participant’s passive exposure to print media is correlated with increased sensitivity to phonological well-formedness.

The experiment itself was a two-alternative forced-choice task, with instructions as follows:

For each set of phrases shown in lower-case letters, you will be asked to choose between, the two phrases will appear on the screen, each paired with a button. It is important that you say each phrase out loud at least three times, to help you get an idea of how it sounds and feels. Click on the button corresponding to the phrase you feel sounds best to you. Please make the determination of which phrase to choose based solely on your subjective feeling of how good the combination of the sounds of the two words sound together. Please do not consider the meaning of each of the phrases (be they funny, implausible, etc.) as a factor in your decision.

Once a participant indicated a choice, they were asked to rate how confident they were in the choice, and were allowed to indicate which, if any, of the words in either of the items they did not know how to pronounce.



Please write in the box
any words on this screen
that you don’t know how to pronounce.

Figure 3: *An example trial screen from Experiment 1.*

Items within group were randomized for which item appeared to the left or the right of the screen, and groups were presented in a random order for each participant. Catch-groups consisting of items of the form “CHOOSE THIS PHRASE” and “DON’T CHOOSE THIS PHRASE” were included so as to weed out non-compliant participants. Time-on-screen information was collected, as well as presentation order of each of the items, and whether the marked item corresponded to the right or left button for any given trial.

3.5 Data processing

60 participants were recruited to participate in the task, of which 53 successfully completed the experiment. 10 participants had to be excluded due to failure to correctly enter their randomly-generated SubjectID when moving between online platforms between pre-experimental language-background survey and the experiment itself, 7 were excluded for not beginning the experimental portion of the task after completing the language-background questionnaire and 4

had to be excluded for not producing the full number of tokens in the experimental task (likely due to closing the browser tab in which the experiment was running before they had completed the experiment). One further participant had to be excluded because they systematically chose the wrong phrase on the catch items (all other participants exceptionlessly chose the correct catch item). This left 35 participants, who generated in total 1925 tokens. Excluding tokens produced on catch items (175 tokens) and tokens which contained one or more words which the participant didn't know (76 tokens), the final dataset contained 1674 tokens.

In addition to group, constraint, and condition information discussed above, each token was further annotated for the amount of time the participant spent on screen deciding which item to choose (hereafter *decision time*), the previous trial's decision time, the previous trial's choice (marked or unmarked), the button (left or right) which was associated with the current trial's choice, the button associated with the previous trial's choice, whether the participant was monolingual or non-monolingual, whether the participant identified their native language as English or not, the participant's ART score, and the confidence with which each trial's choice was marked by the participant. The ratios and counts of unmarked vs. marked choices is shown in Table 3 below, broken down by constraint:

	Unmarked	*Marked*	*Difference (preference for unmarked)*
*BADHIATUS	54% (177)	46% (152)	~ 8% (25)
*GEMINATE	52% (168)	48% (154)	~ 4% (14)
*IAMBICCLASH	53% (185)	47% (162)	~ 6% (23)
*SIBILANTCLUSTER	53% (178)	47% (156)	~6% (22)
*THREECONSONANTS	51% (176)	49% (166)	~ 2% (10)
Overall	53% (884)	47% (790)	~ 5% (94)

Table 3: *Counts and percentages of marked and unmarked choices from Experiment 1, showing a slight preference for unmarked items.*

A by-subjects examination of the response data confirmed that rates of choice at or close to chance were not the artefactual result of a bimodal distribution between participants whose choices were uninformed by phonological markedness and those whose choices were highly sensitive to it: the distribution of percent choice “unmarked” was roughly normal with a slight skew towards a greater number of unmarked choices (mean percent choice *unmarked*: 53, first quartile 49%, third quartile 56%). Thus, an exploratory analysis of the raw data reveals a slight preference for the unmarked item in a group.

3.6 Modeling

Although the trend in the raw data in favor of participants choosing an unmarked item more often than chance appears promising, a statistical model does not support this possibility. I fit a generalized linear mixed-effects regression model using the *lme4* package (Bates et al. 2015) in R (R Core Team 2013), with the sum-coded markedness of the item chosen on each trial (marked or unmarked) as the dependent variable. I initially ran a model with all of the fixed effects discussed above, plus random slopes for subject and group. Using multiple likelihood-ratio tests I removed non-contributing predictors, retaining only those which significantly contributed ($p < 0.05$) to the fit of the model, plus those which we hypothesized might affect participants’ performance in the task: ART, and trial order. The final model contained fixed effects of confidence, ART, and trail order, and random intercepts for subject and group.

The model described above finds a trend ($p = 0.07$) towards unmarked items being chosen more often than marked ones, as well as an effect of confidence which indicates that choices of marked items are more likely to have higher confidence ratings than choices of unmarked items. This is an unexpected finding, because it appears to mean that as participants feel more sure that they have settled on the right choice, or as their phonological preference increases in strength, they are more and more drawn to the marked member of the group.

Model structure:

```
glmer(Condition ~ ART + TrialOrder + Confidence (normalized) +
(1|Subject) + (1|Group), family = "binomial")
```

Model output:

```
Fixed effects:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.266413   0.147774  -1.803   0.0714 .
ART                0.004696   0.004316   1.088   0.2766
TrialOrder         0.002120   0.003233   0.656   0.5120
Confidence         0.115734   0.052361   2.210   0.0271 *
```

Table 4: *Best-fitting generalized logistic mixed effects regression model of response data for all constraints, from Experiment 1.* Significance codes: *'***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1.*

Sub-analyses were run with each constraint's data individually to test whether specific constraints might be individually more likely to motivate unmarked item choices. These models were fit in the same way described above, except that in these models the intercept is the likelihood that unmarked items of a given constraint are chosen more frequently than marked items for that constraint. These analyses found that *BADHIATUS items exhibited the expected effect, with marked items significantly less likely to be chosen than unmarked ones, and that no other constraints had a significant effect on the choice of unmarked items over marked ones. The final model of the *BADHIATUS data alone is shown below.

Model structure:

```
glmer(Condition ~ ART + PriorButton + Confidence (normalized) x
TrialOrder + (1 | Subject) + (1 | Group), family = "binomial")
```

Model output:

```
Fixed effects:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -0.773063   0.349246  -2.213   0.0269 *
ART                         0.003244   0.009827   0.330   0.7413
TrialOrder                  0.010757   0.007299   1.474   0.1406
Confidence                  0.408483   0.232296   1.758   0.0787 .
PriorButton = Button 2      0.475135   0.235195   2.020   0.0434 *
TrialOrder x Confidence    -0.013726   0.007228  -1.899   0.0576 .
```

Table 5: *Best-fitting generalized logistic mixed effects regression model of response data for* *BADHIATUS *items from Experiment 1.* Significance codes: ‘***’ $p < 0.001$; ‘**’ $p < 0.01$; ‘*’ $p < 0.05$; ‘.’ $p < 0.1$.

The significant negative intercept in the model indicates that marked items are chosen significantly less often than chance. The model contains a trend towards significance for confidence with a positive intercept, which mimics the model containing all constraints. This means that overall, choices of marked items were associated with higher confidence ratings. The model also contains a trending interaction between confidence and trial order with a negative intercept, indicating that the effect of confidence as dependent on choice of item changes throughout the course of the experiment. Finally, the model contains a significant effect of PriorButton, indicating that choosing the item associated with Button 2 on the immediately preceding trial increases the likelihood of a marked item being chosen on the current trial. This effect is likely spurious, since the condition of the item shown on each button was randomized for each participant in each trial, and I discard it as such.

Additional analysis found no significant effect of markedness of any other constraint.

3.7 Discussion

This experiment did not support the *preference prediction*, which held that participants would prefer unmarked items over marked items in a forced-choice task. However, a sub-analysis revealed that *BADHIATUS tokens exhibited the expected dispreference for marked items, albeit weakly. This mixed result could have several possible sources: first, and most directly, it could be the case the participants in fact do not possess any strong or reliable intuitions about relative phonological well-formedness across a word boundary. An alternative hypothesis is that the setup of the experiment, being conducted online rather than in a lab setting, discouraged participants from attending to the task as intently as they might otherwise in a more formal setting with no possible distractors. A third factor could be that the online experiment involved two steps, first filling out a language-background and demographic questionnaire on the IbexFarm platform, and then navigating to a second web-page, hosted on the Appsobabble platform, to complete the forced-choice task. As discussed in section 3.5, there was a large amount of participant attrition between the two parts of the experiment, as well as a good deal of participant error generally, perhaps due to its being conducted online. Finally, the null result could be from a lack of power – in the descriptive statistics in Table 3, all unmarked items were chosen with greater frequency than marked ones. It could be the case that with such a subtle effect, more data is needed to establish the finding firmly. One other confound to consider is the confounding fact that I used attested lexical items of English as stimuli; while I attempted to make each item a low-probability bigram in the speech of participants, it is possible that semantic, pragmatic, or other lexically-oriented factors overpowered or confounded participant's phonological judgments about the stimuli.

3.8 Experiment 2: Speeded production task

This experiment assesses the *response latency prediction*, namely that phonotactically marked sequences across a word-boundary impact post-lexical phonological processing in speech production. I used a speeded reading paradigm in which participants were visually presented with stimuli and asked to read them out loud as quickly and as accurately as possible, in two different experimental manipulations. The *response latency prediction*, in combination with the specifics of the experimental design outlined below, allows us to infer that if we find a reliable increase in response latency for items containing a phonologically marked sequence, we can attribute this change to an increase in phonological processing cost.

3.9 Procedure

Each participant took part in the experiment in a sound-attenuated room, with the stimuli presented visually on a computer screen and responses recorded using a headset-mounted USB microphone. Each participant took part in one of two slightly different experimental manipulations (hereafter *paradigms*), and saw one of two lists of stimuli (hereafter *List A* and *List B*, described below). Participants were divided into paradigms such that one sequential pair of participants were run in the same paradigm, and then the next two were run in the alternate paradigm, modulo some slight experimenter error which resulted in occasional deviation from this system. Aside from some occasional experimenter error, participants were alternatingly shown the different lists of stimuli.

3.10 Lists

The 100 items described above were divided into two counterbalanced lists, such that neither list contained both the marked and unmarked item in a given bigram, and that both lists contained the same number of marked and unmarked items for each constraint.

3.11 Paradigms

Both paradigms had in common the following setup. The participant was asked to fixate on a point on the computer screen, and after 1000ms a random item selected from the relevant lists appeared on screen. In one paradigm (hereafter called the FAST paradigm) the participant read the phrase out loud as quickly and accurately as possible as soon as they saw it. In the other setup (called the DELAY paradigm) the stimulus was shown to participants for 750ms before they needed to speak it aloud. This was implemented via a color-change in the display: participants were instructed that when the stimulus was shown in red letters they should silently read the item and prepare to speak it, and when the stimulus turned green they should speak as quickly as possible. In both paradigms, participants were instructed to press the spacebar as soon as they had finished speaking to indicate their completion of the task. If they did not do so within 4000ms of the stimulus appearing (in the FAST paradigm) or the stimulus turning green (in the DELAY paradigm), a warning screen was displayed instructing participants to speak faster. After the completion of a given trial, a screen appeared instructing participants to press the spacebar when they were ready to proceed to the next item. Each token was recorded separately, beginning from the appearance of an item on the screen (in the FAST paradigm) or from the stimulus turning green (in the DELAY paradigm), and terminating when the participant pressed the spacebar to proceed to the next item. After completing one list of 50 items in a given paradigm, plus five additional practice items before the start of the list to allow them to allow them to become accustomed to the task, participants completed a survey asking them to indicate which, if any, of the words used during the production task they did not know how to pronounce or did not feel confident in their ability to define.

3.11.1 Why paradigms?

The choice to compare response latency between subjects and paradigms was motivated by the literature discussed above which documents the combined impact of lexical-level factors (frequency, phonological-neighborhood size), post-lexical phonological factors (phonotactic probability), and motor-level factors (gross articulatory complexity) on response latency in speeded repetition tasks. Response latency in the FAST paradigm encompasses the time it takes for a participant to complete lexical, post-lexical, and motor-level processing, while the DELAY response latency for the same item only captures the effect of motor planning. By subtracting a given item's DELAY response latency from its FAST response latency, we can factor out interference from motor-processing in the response latency we submit to analysis. This is crucial to the experiment because several of the constraints investigated here such as *THREECONSONANTS, necessarily involve greater articulatory complexity than others, such as *GEMINATE. A failure to account for the confound of motor processing could lead to a false conclusion that a specific type of markedness caused difficulty for phonological processing, while in reality the effect was due to associated articulatory complexity. By using the DELAY paradigm to separate motor-encoding from lexical and post-lexical processing, I hypothesize that the response latency in the DELAY condition will vary between items only to the extent that motor-level processing impacts the production of that item. Having factored out the effect of motor processing through the use of dual paradigms, to isolate the contribution of post-lexical processing I controlled for lexical-level effects within groups, having items within a group share the same *word-one*, and matching the lexical qualities of *word-two*'s. This piecewise approach to controlling for the two primary confounding factors influencing response latency in this context allows us to directly compare marked and unmarked phonological contexts in their impact on phonological processing.

3.12 Interpretation of experimental results

The interpretation of the response-latencies of individual item productions across paradigms and speakers is as follows. If, for the marked item in a group, the difference between response latencies between the DELAY and FAST paradigms is significantly greater than the difference between response latencies between the DELAY and FAST paradigms for the unmarked item in that same group, we can hypothesize that it is the phonological markedness present in the marked item which is the source of the difference. More broadly, if marked items, regardless of constraint, exhibit a greater difference in response latency between-paradigms than unmarked items in their group do, then we can attribute this difference to the marked status of those items, reflecting an increased post-lexical processing difficulty. Lastly, if, for a given pair of constraints, if one constraint's marked items show systematically greater differences between productions in the DELAY and FAST paradigms compared to unmarked items in their group, and this difference is greater than the same within-items-across-paradigms difference for the other constraint's marked and unmarked items, we can say that the constraint with the greater such difference is exerting a greater effect on post-lexical phonological. I return to this piecewise mode of analysis in section 3.14.2 when examining the raw data.

3.13 Data processing

The production experiment yielded 3150 tokens, each of which was manually annotated for response latency by the author or one of three trained research assistants, measured from the onset of stimulus presentation to the onset of speech. Each token was also annotated for speech error information, non-error disfluencies, and the presence of unavoidable but potentially confounding behavioral variants. Tokens were marked for containing any of the following characteristics:

Speech errors

- The participant misspoke one or both of the item words, substituting real words
- The participant misspoke one or both of the item words, not creating real words
- The participant restarted the item from the beginning, or from the beginning of the second word

Behavioral miscues

- The participant left off one or more syllables of the item
- The participant made no noticeable effort to speak in an accelerated or urgent manner
- The participant didn’t speak the item at all

Tolerable behavioral variation

- The token contained an audible smack of the lips before speaking
- The token contained an audible breath before speaking

These annotations were used in part to determine which tokens were contained irredeemable errors (those listed under *behavioral miscues*), which tokens contained speech errors of the type described above (listed under *speech errors*), and which tokens were error-free or contained tolerable behavioral variation (discussed section 3.16 as part of modeling).

65 participants were recruited to take part in the task. Two subjects were excluded due to experimenter error, and one subject was excluded due to excessively slow speech (assessed qualitatively). 330 tokens were excluded because participants indicated that they contained words with which they were not familiar, and 219 tokens were excluded due to non-speech error disfluencies in production (coughs, sneezes, lack of speech on the recording, adjusting the headset while speaking, etc.). For items in the *IAMBICCLASH constraint, I attempted to choose *word-one*'s which did not readily lend themselves to repairing the stress clash with the following word via stress-shift to their initial syllable (a process commonly termed the "rhythm rule," cf. Liberman & Prince 1977) – however, 19 tokens contained production of marked *IAMBICCLASH items in which the stress was placed non-canonically. Because the effect of this sort of external sandhi on speech processing is not known, I excluded these tokens from both the clean and error-containing datasets. Of the remaining 2599 tokens, 306 included speech errors and 2293 were error-free. I also excluded from the response latency analysis 68 tokens which had a log-transformed response latency more than two standard deviations from the mean of their paradigm. The final number of tokens in each paradigm were: FAST – 1448 (including 224 errors), and DELAY – 1150 (including 83 errors).

The error-free tokens were submitted to an analysis of response latency to evaluate the *response latency prediction*, discussed here, and the error-free and speech-error-containing tokens were recombined for post-hoc analysis for the interaction of speech-error likelihood with phonological markedness to evaluate the *speech error prediction*, discussed in section 2.2.

3.14 Data Exploration

In the following section I present descriptive statistics for the data which I submit to regression analysis in section 3.17.

3.14.1 Response latencies by condition, paradigm, and constraint

Table 6 displays mean and standard errors for response latencies, in deciseconds, broken down by paradigm and constraint.

DELAY:

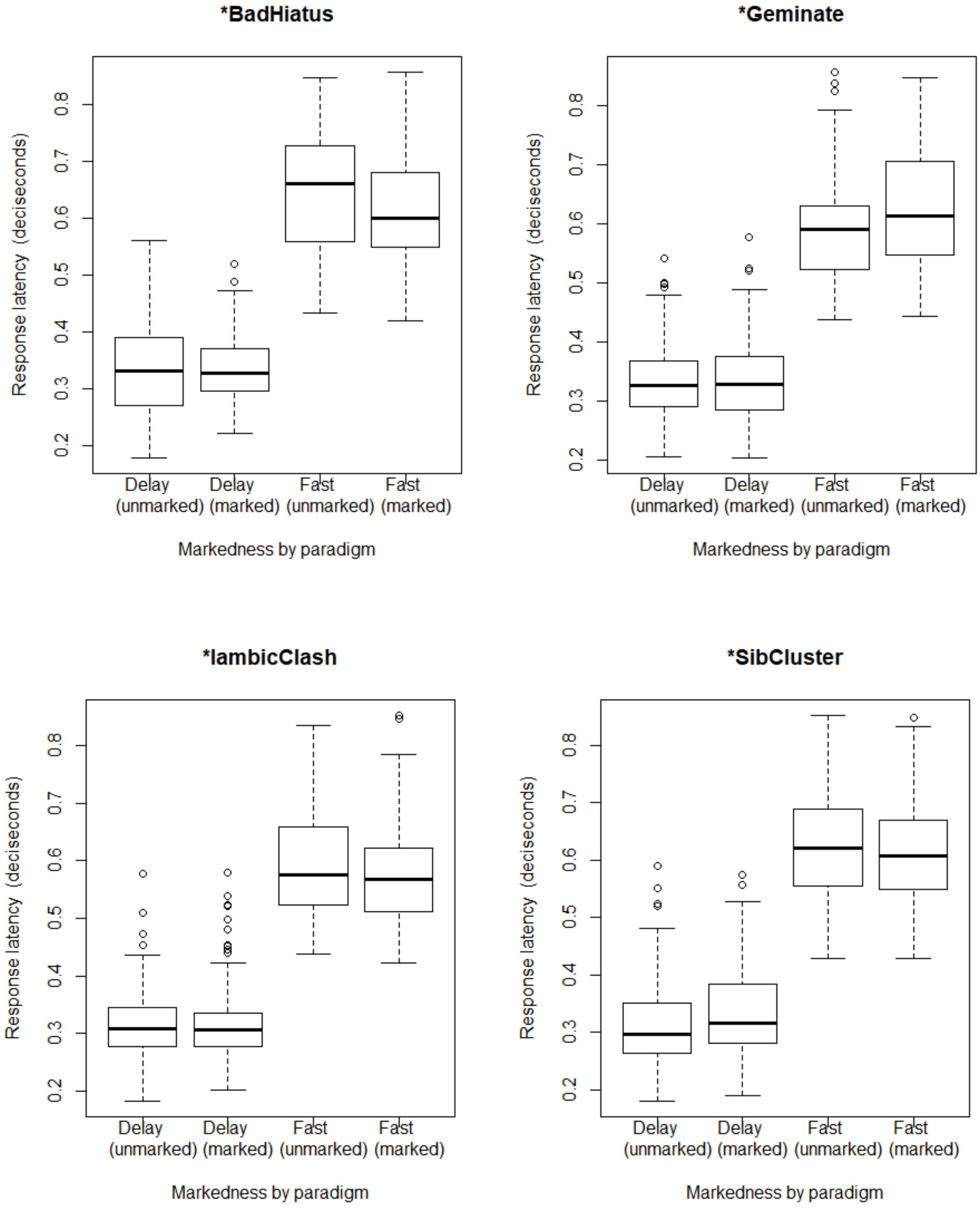
Constraint	*Unmarked*	*Marked*
*BadHiatus	3.342 (0.088)	3.349 (0.064)
*Geminate	3.366 (0.075)	3.343 (0.071)
*IambicClash	3.216 (0.060)	3.207 (0.069)
*SibilantCluster	3.175 (0.072)	3.386 (0.080)
*ThreeConsonants	3.187 (0.055)	3.305 (0.066)
Overall	3.2572 (0.070)	3.318 (0.070)

FAST:

Constraint	*Unmarked*	*Marked*
*BadHiatus	6.429 (0.102)	6.220 (0.095)
*Geminate	5.919 (0.092)	6.252 (0.096)
*IambicClash	5.968 (0.083)	5.774 (0.078)
*SibilantCluster	6.108 (0.077)	6.105 (0.091)
*ThreeConsonants	5.902 (0.078)	6.140 (0.082)
Overall	6.0652 (0.086)	6.0982 (0.088)

Table 6: *Means and standard-errors (in parentheses) for response latencies broken down by paradigm, condition, and constraint.*

The boxplots in Figure 4 display the mean, first, and third quartiles for the response latency for each constraint broken down by paradigm and condition.



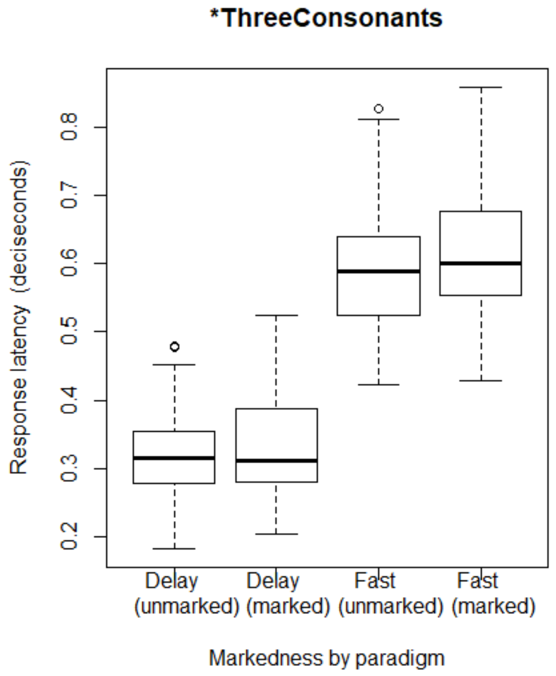


Figure 4: *Box-plots showing the mean, first, and third quartiles of responses for each constraint, broken down by paradigm and condition.*

3.14.2 Isolating the contribution of phonological processing

Returning to the interpretive roadmap from section 3.12, we can isolate the contribution of markedness to phonological processing time (across all constraints) by subtracting the response latency for a given item in the DELAY condition from its response latency in the FAST condition – the difference encapsulates the total time taken up by lexical and phonological processing. Because we controlled for lexical characteristics of items within groups, we can subtract the cost of lexical and phonological processing of unmarked items (2.803 deciseconds) from the same cost for marked items (2.778 deciseconds), finding that the contribution of markedness to response latency – a difference of 0.025 deciseconds – favors marked items as being processed faster. This test reveals that marked and unmarked items, agnostic of constraint, are virtually identical in phonological processing time. To explore whether this effect is due to a uniform lack of effect of markedness, whether the effect of markedness for one constraint in particular might be getting lost in the rest of the data, or whether the lack of effect is the result of a number of overlapping facilitatory and inhibitory effects for different constraints, the same metrics are shown in below, broken down by constraint.

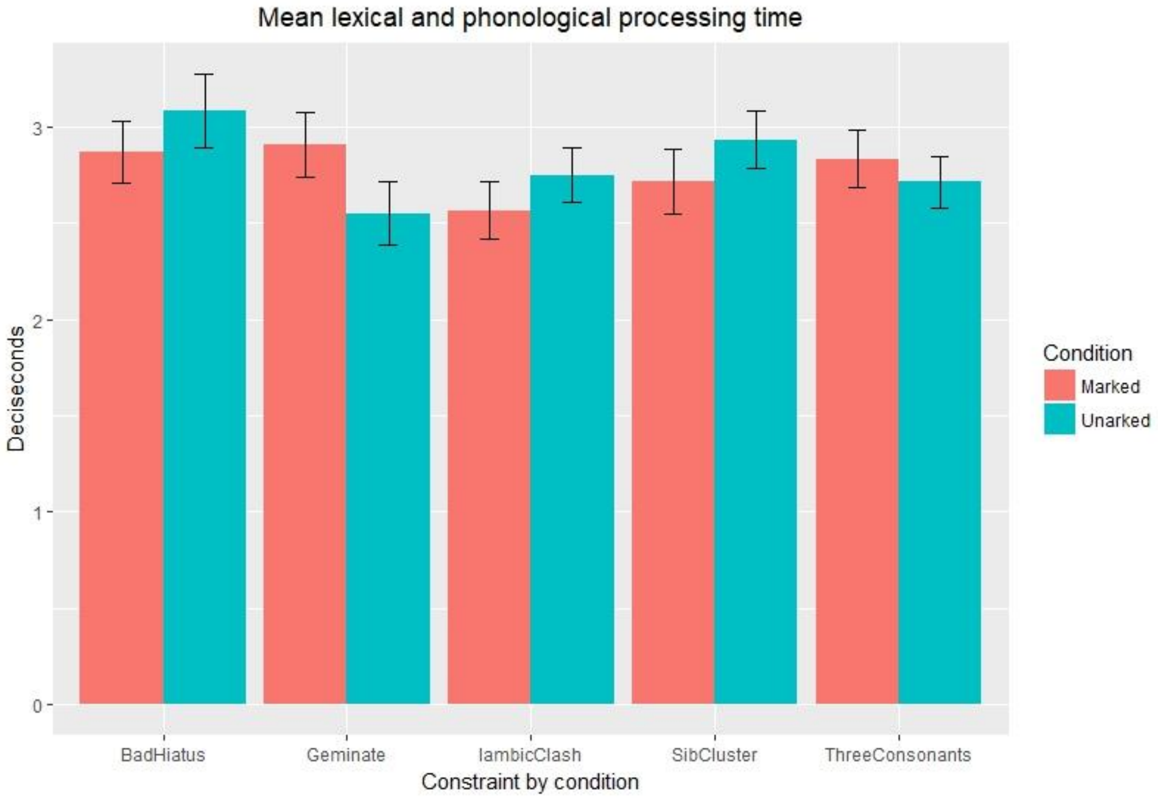


Figure 5: *Visualization of the effect of phonological markedness on combined lexical and phonological processing time, in deciseconds.*

	Marked	*Unmarked*	*Difference*
*BadHiatus	2.871 (0.159)	3.087 (0.190)	-0.216 (0.349)
*Geminate	2.909 (0.167)	2.553 (0.167)	0.356 (0.334)
*IambicClash	2.567 (0.147)	2.752 (0.143)	-0.185 (0.290)
*SibilantCluster	2.719 (0.171)	2.933 (0.149)	-0.214 (0.320)
*ThreeConsonants	2.835 (0.148)	2.715 (0.133)	0.120 (0.281)

Table 7: *Means and standard error (in parentheses) for differences between FAST and DELAY paradigms by condition, broken down by constraint. Table results are the combined effect of lexical and phonological processing, in deciseconds.*

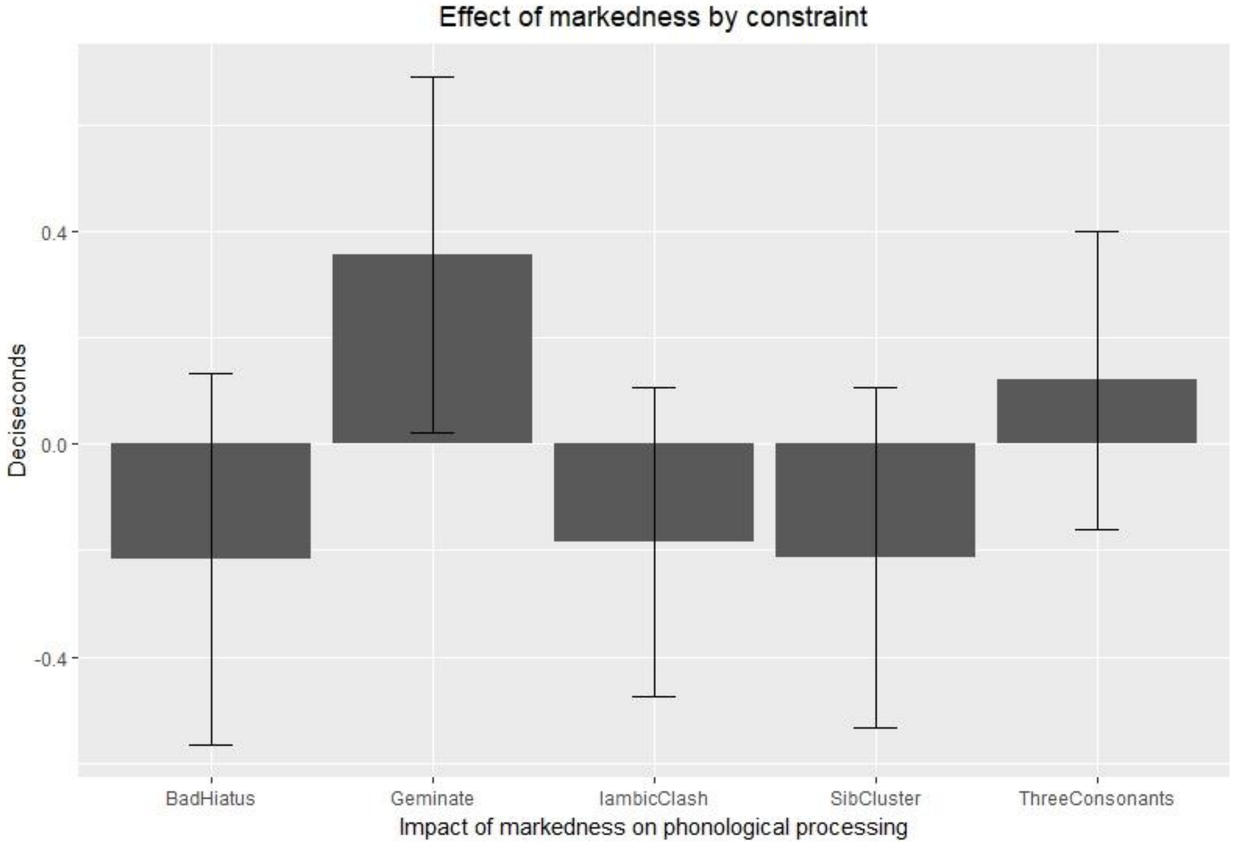


Figure 6: *The isolated effect of markedness on phonological processing (values from Figure 13): positive values indicate an increase in processing time for marked items, negative values indicate a decrease in processing time for marked items.*

In contrast to the near-identical combined lexical and phonological processing times for marked and unmarked items discussed above, breaking down these same measures by constraint reveals a range of facilitatory and inhibitory trends of markedness in the raw data. Subtracting the lexical and phonological processing item for unmarked items from that of marked items for each constraint, we can isolate the effect of each type of markedness on phonological processing, shown in Figure 6 above. This preliminary data exploration suggests that phonological markedness has no unidirectional across-the-board effect on phonological processing; for

*BADHIATUS, *IAMBICCLASH, and *SIBILANTCLUSTER marked items are, on average, processed more quickly than their unmarked counterparts, while for *GEMINATE and *THREECONSONANTS items the reverse is true, with marked items having an inhibitory effect on processing compared to unmarked ones. In the following section I verify these findings using logistic regression.

3.15 Analysis

Several rounds of analysis were performed on the cleaned dataset. First, I tested for a main effect of item markedness on response latency, independent of constraint, and then tested for the effect of markedness broken down by constraint, using the full set of cleaned data.

3.16 Model terms

The dependent variable in the model was *response latency*; this value, originally recorded by annotators in seconds, was multiplied by 10 to aid in convergence by putting it on roughly the same scale of magnitude as the other numerical predictors included in the model. The three predictors of primary interest are *constraint*, *condition*, and *paradigm,* as well as their interaction.

The models presented below include fixed effects for several of the measures controlled for within groups during stimulus selection: *phonological neighborhood* for each item's *word-one* and *word-two,* the natural log of the *frequency* in the SUBTITLE-US database (Brysbaert & New 2009) for *word-one* and *word-two,* the number of syllables in *word-one* and *word-two*. The model included a fixed effect of *trial order,* as well as two behavioral factors: *breath* (coded true for tokens which included an audible intake of breath before the onset of speech, false otherwise), and *lipsmack* (coded true for tokens which included an audible pre-speech smacking of the lips). These last two effects were included on the hypothesis that the presence of these two behaviors could meaningfully impact the response time, as well as the possibility that speakers might use them as a sort of “delaying tactic” to give themselves more time while processing marked sequences. The model also included an effect of *monolingualism*, valued at true if the participant self-identified as a native speaker of only English, and false otherwise.

Following Baayen & Milin (2010)’s guidance for analyzing reaction time data, I also included the *previous trial’s response latency* as a fixed effect, multiplied by 10 to match the scale of the dependent variable. One of the standard assumptions made when using logistic regression is that each individual data point submitted to the model is independent from each other data point so included. In this case, the data points are individual tokens of one participant speaking one item in one paradigm. However, Baayen & Milin demonstrate that reaction-time data of the sort which this study models often contains non-trivial between-trial dependencies, known as autocorrelation, which violate this assumption of independence which logistic regression relies on. To mitigate the effect of violating the assumption of independence in the data, I added for each token the *previous trial’s response latency* if that previous token was error-free. I also included p*rior error* as a fixed effect, coded true if the previous token contained an error, and false otherwise.

The model included random intercept for s*ubject*, on the hypothesis that there is some level of individual variation in performance on this sort of task. I also included a random slope by *condition* nested within *subject* to allow for individual differences in the speed of post-lexical phonological processing of marked sequences. I also included a random intercept by *item*, to control for residual between-item differences arising from other factors than those discussed in section 3.13.

The binary-valued predictors were coded using dummy coding so that each level of the model could be compared to its reference level. The reference levels for the predictors all the models described here are as follows: DELAY is the reference *paradigm*, *unmarked* is the default condition, *BADHIATUS is the default constraint, and *breath*, *lipsmack*, and *prior error* are at their false values. The reference value for continuous factors are their average values.

3.17 Testing for an effect of markedness on processing: an initial minimal model

The primary hypothesis being addressed in this experiment is whether there is an effect of phonological markedness on phonological processing. To test this hypothesis, I fit a simple model with the primary predictors of interest *paradigm* and *condition*, and their interaction, with random intercepts for *subject* and *item*, and a random slope for *condition* nested within *subject*. A significant interaction of these fixed effects would indicate that marked items have a greater impact on phonological processing than unmarked items when spoken unprepared, above and beyond marked items taken as a group and unprepared items taken as a group. This would indicate that phonological markedness contributes significantly to the response latency of such items, confirming the *response latency prediction.* I begin with a very sparse model to ensure that the behavioral and lexical characteristics do not swamp any potentially small effect of markedness. The model specifications and output are outlined below.

Model structure:

```
lmer(ResponseLatency ~ Paradigm*Condition + (1 +
Condition|Subject) + (1|Item))
```

Model output:

```
Fixed effects:
                                   Estimate Std. Error       df t value
Pr(>|t|)
(Intercept)             3.29155    0.10914  78.4000  30.159   <2e-16 ***
Condition = marked      0.06965    0.07208 130.9000   0.966    0.336
Paradigm = FAST         2.79611    0.13414  58.6000  20.845   <2e-16 ***
Condition = marked x Paradigm = FAST  -0.02851 0.05525 512.5  -0.516   0.606
```

Table 8: *Minimal linear mixed effects regression model for response latency data from Experiment 2, examining only the interaction of paradigm and markedness. Significance codes: '***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1.*

This model finds no significant or trending main effect of markedness, not any effect of markedness by condition. This finding could be compatible with a result in which there was no significant effect of markedness on processing, or one where the noise in the data due to lexical and behavioral data is masking an effect of markedness.

3.18 Incorporating behavioral and lexical information into the model

I turn now to a model which *includes* behavioral and lexical information, on the possibility that a small, significant effect of markedness, if present, could be revealed by accounting for sources of variance in the experimental data stemming from non-phonological factors. I began with a maximal model containing *all* predictors described in 3.16, and systematically removed each factor (other than those of primary interest, *constraint, condition,* and *paradigm*), from the model if they did not significantly contribute to the fit of the model to the data, as assessed by the likelihood-ratio test. The final best-fitting model contains fixed effects of *paradigm*, *condition, breath, lipsmack, prior response latency, log-frequency word-one, log-frequency word-two*, and random intercepts of *subject* and *item*, with a random slope of *constraint* nested within *subject*. Comparing this model to the minimal model in in 3.17 via an ANOVA test, I find that the current model which includes the behavioral and lexical information described above captures significantly more of the variance in the dataset. The best-fitting model improves over the minimal model by 781.79 points on an Akaike Information Criterion (AIC) score; for reference, AIC differences of more than 10 points are generally accepted to signal a significant improvement in model fit (Burnham & Anderson 2004). As before, the model also was fit with an interaction of condition and paradigm.

Model structure:

```
lmer(ResponseLatency ~ Paradigm*Condition + Breath + Lipsmack +
PriorResponseLatency + Log-FreqWordOne + Log-FreqWordTwo +
(1|Item) + (1+Condition|Subject))
```

Model output:

```
Fixed effects:
                             Estimate Std. Error         df t value Pr(>|t|)
(Intercept)                 3.74449    0.18041  179.70000  20.755  < 2e-16 ***
Condition = marked          0.07168    0.07129  101.00000   1.006 0.317045
Paradigm = FAST             2.64669    0.13861   73.50000  19.094  < 2e-16 ***
Breath = True               0.38516    0.07959 1754.10000   4.839 1.42e-06 ***
Lipsmack = True             0.16467    0.07123 1738.10000   2.312 0.020910 *
PriorResponseLatency        0.04563    0.01826 1767.10000   2.499 0.012553 *
Log-Freq. Word-one         -0.14679    0.04198   93.40000  -3.497 0.000723 ***
Log-Freq. Word-two         -0.14926    0.04641   91.40000  -3.216 0.001796 **
Cond. = marked x Par. = FAST -0.05618 0.06759   52.00000  -0.831 0.409686
```

Table 9: *Best-fitting linear mixed effects regression model of response latency data from Experiment 2, collapsing across constraints.* Significance codes: *'***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1.*

Supporting the raw means discussed in section 3.14.2, the model finds that there is no significant effect of markedness on response latency, shown in Figure 7. Other significant effects in the model are explored below – note that these same effects are generally significant in all of the analyses run on this dataset; I discuss them here and omit discussion in future models unless warranted.

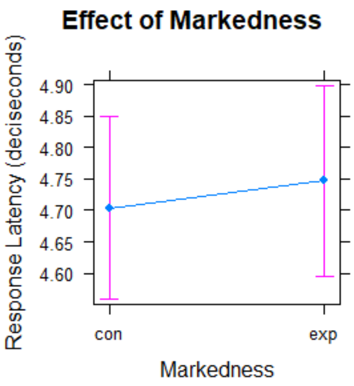


Figure 7: *Partial effect plot of the non-significant effect of markedness on response latency; based on regression model in Table 9.*

3.18.1 Lexical and behavioral predictors

Unsurprisingly, the experimental manipulation of *paradigm* plays a significant role in determining the response latency of any token. Behavioral measures such as Breath and Lipsmack also significant increase the response latency for tokens where they occur. Lexical-level factors such as the log-frequency of both words in an item also significantly impact the response latency for a given token, with increased response latency for lower-frequency items, and vice versa. Additionally, *prior response latency* is positively correlated with response latency on the current trial, as predicted by Baayen & Milin (2010) – the effect contributes only moderately to the fixed-effects-only R^2 (an increase of approx. 0.01, from 0.727 to 0.737), with a smaller increase in the R^2 when random effects are considered (an increase of approx. 0.005, from 0.850 to 0.855). A Wald type-II chi-squared test for the independence of *prior response latency* from the dependent variable indicates that the relation between these variables is significant, $\chi^2(1, N = 1827) = 6.24$, $p = 0.01$. Comparing the model in Table 9 to an identical model that lacks *prior response latency* reveals that the direction and significance levels of the other fixed effects are qualitatively unchanged, indicating that while *prior response latency* does contribute to predicting response subjects' response latencies, the general impact of the predictor is likely in reducing variance and thus aiding in more accurate assessment of the fixed-effect coefficients, but *prior response latency* does not itself significantly impact the theoretical interpretation of the fixed effects of interest. The partial effects of these predictors are visualized below.

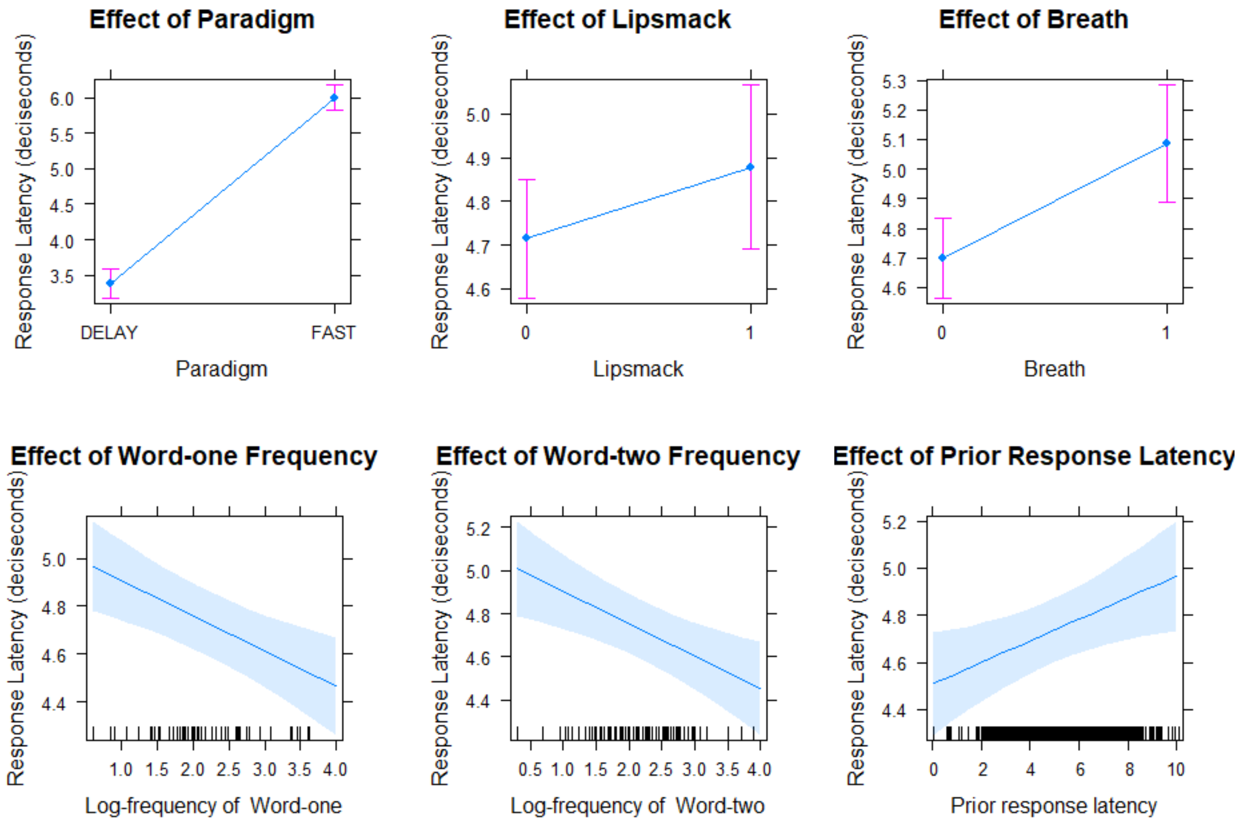


Figure 8: *Partial effect plots of behavioral and lexical factors which significantly modulate response latency; based on regression model in table 9.*

3.19 Sub-analysis of the effect of markedness by constraint on response latency

3.19.1 By-subject and by-item t-tests

To begin to explore the individual effects of each constraint's markedness-type on processing, Table 10 shows by-subject and by-item paired t-tests for differences between marked and unmarked items in both paradigms. Their interpretation is discussed below.

	DELAY		FAST	
Constraint	Subject (*df* = 26)	Item (*df* = 9)	Subject (*df* = 35)	Item (*df* = 9)
	t-score (p-val)		*t-score (p-val)*	
*BadHiatus	-0.11 ($p = 0.91$)	-0.06 ($p = 0.95$)	1.08 ($p = 0.29$)	0.026 ($p = 0.80$)
*Geminate	0.09 ($p = 0.92$)	0.00 ($p = 0.99$)	-2.17 ($p < 0.04$)	-2.66 ($p > 0.03$)
*IambicClash	-0.08 ($p = 0.94$)	-0.10 ($p = 0.92$)	1.96 ($p = 0.06$)	1.60 ($p = 0.14$)
*SibilantCluster	-2.44 ($p < 0.03$)	-1.54 ($p = 0.16$)	-1.03 ($p = 0.31$)	0.66 ($p = 0.52$)
*ThreeConsonants	-1.39 ($p = 0.18$)	-0.74 ($p = 0.48$)	-3.44 ($p < 0.00$)	-0.10 ($p = 0.92$)

Table 10: *By-subject and by-item t-tests for differences between marked and unmarked items by paradigm and constraint.*

By-subjects t-tests consider the differences within subjects for marked and unmarked items within a paradigm, ignoring differences between items. These tests show significant differences between marked and unmarked conditions in the FAST paradigm for the *Geminate, *IambicClash, and *ThreeConsonants constraints. Two of these, *Geminate and *ThreeConsonants, have negative coefficients, indicating that marked items are significantly slower to produce than unmarked items in the FAST paradigm, with *IambicClash has a positive coefficient, indicating that it has a facilitatory effect on processing. In the DELAY paradigm, the only significant by-subject difference between marked and unmarked items is for *SibilantCluster, which, with a negative coefficient, indicates that marked items are produced significant more slowly in the DELAY condition than unmarked items. Interestingly, this difference is not present in the FAST paradigm – I interpret this to mean that when taking only motor-encoding into account, the articulatory difficulty involved in the subtle changes of place of articulation for two non-identical sibilants lead to significant motor processing differences between marked and unmarked items in the *SibilantCluster constraint. This same effect is not seen for FAST items, perhaps because other sources of variation between marked and unmarked items in lexical and phonological processing swamp the effect. This difference, while relatively minor, justifies my choice of a between-paradigm comparison to factor out motor-encoding.

By-item t-tests collapse across differences between subjects and focus on by-item differences between paradigms for marked and unmarked conditions. These tests are less revealing, and do not match well with the results of the same tests collapsing across items. When comparing performance on each individual item within paradigms without regard for the subject producing the token, only *GEMINATE exerts an effect on processing in the raw data. I interpret this divergence between types of t-test to signal that the dataset contains a great deal of by-item variation, but that a given item is treated in the same way by many different participants. This finding is borne out in the modeling below.

3.19.2 Regression analysis of by-constraint effect of markedness on processing

The regression in section 3.18 confirms the analysis of the raw data in 3.14.2, finding no significant difference in response latency between marked and unmarked items after controlling for lexical and articulatory factors. I now turn to the subsequent finding that different constraints have fairly large differences in effect on processing (shown in Figures 5 and 6). To verify whether marked items for each constraint had significantly longer response latencies than unmarked ones, I fit five individually models with the same specifications as the model in section 3.18, one for each constraint's tokens. A significant effect of markedness and paradigm in these models would indicate whether, taken separately, each constraint has an impact on phonological processing. One downside to these sub-analyses, however, is that there is a rather severe loss of power in the analysis because each model is fit now to only roughly one fifth of the number of tokens as the full model. The means and standard errors for the combined lexical and phonological processing times of marked and unmarked items for each constraint, and their significance, is reported below.

	Marked	*Unmarked*	*Difference*	*Signif.*
*BADHIATUS	2.871 (0.159)	3.087 (0.19)	-0.216 (0.349)	*
*GEMINATE	2.909 (0.167)	2.553 (0.167)	0.356 (0.334)	.
*IAMBICCLASH	2.567 (0.147)	2.752 (0.143)	-0.185 (0.29)	
*SIBILANTCLUSTER	2.719 (0.171)	2.933 (0.149)	-0.214 (0.32)	
*THREECONSONANTS	2.835 (0.148)	2.715 (0.133)	0.120 (0.281)	

Table 11: *Mean lexical and phonological processing time in deciseconds (std. error), plus differences in response latency attributable solely to phonological markedness, by constraint. Significance codes: '***' $p < 0.001$; '**' $p < 0.01$; '*' $p < 0.05$; '.' $p < 0.1$.*

We can see that *BADHIATUS-violating items have a significant facilitatory effect on phonological processing, and that *GEMINATE items have a marginal inhibitory effect on phonological processing. Other types of markedness do not significantly differ from their unmarked counterparts.

To determine whether each type of markedness differs in its impact on phonological processing from other types of markedness, I reran the full dataset in a model identical to the one in section 3.18 above, but with a three-way interaction of Condition*Paradigm*Constraint. This model captures significantly more variation than the model of section 3.18 ($\chi^2 = 29.892$, $p < 0.018$), indicating that the additional degrees of freedom which come with allowing each constraint to have an independent effect on phonological processing are justified by the increased accuracy of the model. This indicates that, abstracting away from particular pairwise differences between constraints, allowing constraints to differ in their effect on processing significantly improves the model – intuitively, this can be understood as supporting the finding that not all types of markedness have the same effect on phonological processing, as discussed above. I return to this point in relation to the notion of phonotactic probability later in this paper. This parallels the finding shown in Figure 6, that different markedness constraints have widely divergent effects on phonological processing.

By rerunning the model with each markedness-type held as the reference-level in turn, we can see the pairwise comparisons between the different constraints, shown below. The significance of the difference between marked and unmarked items is indicated by the vertical indicators to the right of each constraint's bar, and the significance of the difference between the *effects* of each constraint is shown by the horizontal indicators spanning the two constraints whose aggregate effect is significant.

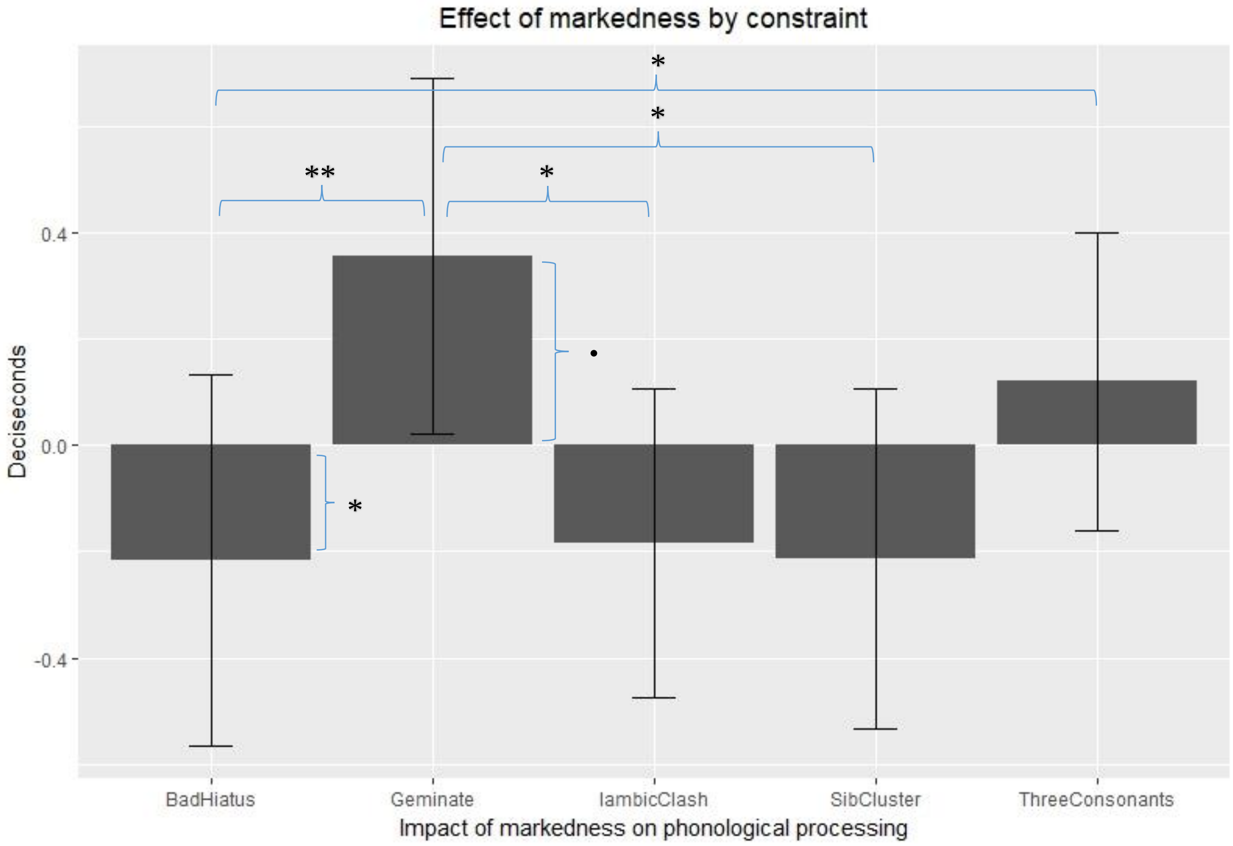


Figure 9: *Significant differences between the effect of different types of markedness on phonological processing. Significance codes: '***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1.*

3.20 Effect of markedness on response latency excluding sparse data

In an attempt to mitigate the effects of the data-exclusion measures described in section 3.13 on the total count of items in each paradigm, and to increase data integrity by excluding data from participants who lost many tokens to error-removal procedures, I created a sub-dataset, which I will call the *reduced* dataset, which excluded the bottom 10% of participants and the bottom 10% of groups by token.

Rerunning the same analyses described above in this reduced dataset, the effects of markedness on phonological process is qualitatively similar, but quantitatively more extreme. An updated version of Table 11, showing the significance of the difference in lexical and phonological processing time for marked and unmarked items for each constraint, is shown below.

	Marked	*Unmarked*	*Difference*	*Signif.*
*BadHiatus	2.746 (0.171)	3.101 (0.207)	-0.355 (0.378)	**
*Geminate	2.877 (0.185)	2.502 (0.176)	0.375 (0.361)	*
*IambicClash	2.552 (0.152)	2.766 (0.143)	-0.214 (0.295)	
*SibilantCluster	2.737 (0.171)	2.975 (0.165)	-0.238 (0.336)	.
*ThreeConsonants	2.86 (0.156)	2.695 (0.139)	0.165 (0.295)	

Table 12: *Mean lexical and phonological processing time in deciseconds (std. error), plus differences in response latency attributable solely to phonological markedness, by constraint, in the reduced dataset.*

Figure 10 is structured in the same way as Figure 9, but demonstrates the stronger effect of markedness seen in the reduced dataset.

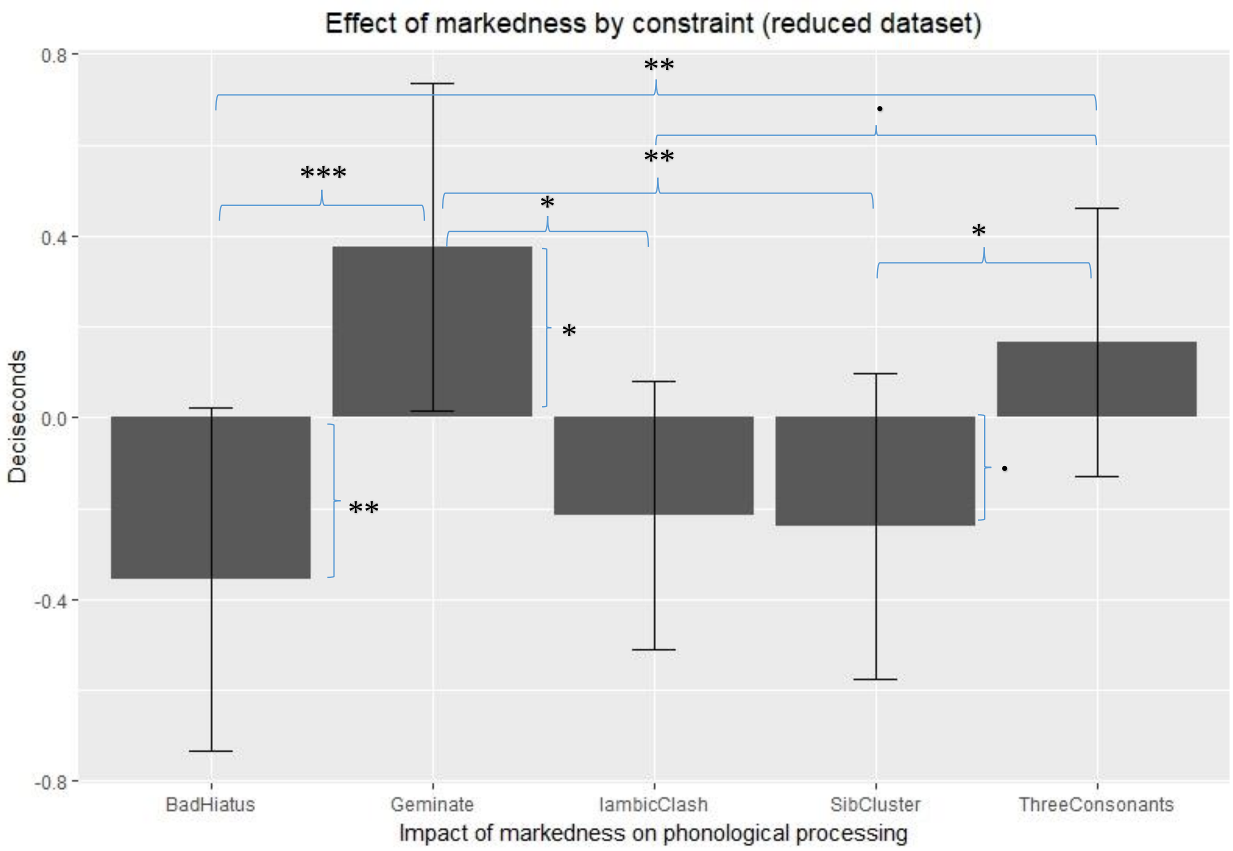


Figure 10: *Effect of markedness on phonological processing within and between constraints, using the reduced dataset.*

We can see that when controlling for sparse data, the patterns present in the data at sections 3.18 and 3.19 are enhanced – this lends support to the idea that phonological markedness exerts varying impacts on phonological processing, although the findings lack the across-the-board consistency which would be expected for a result supporting the *response latency hypothesis*.

3.21 Conclusion

While the production task did not find an across-the-board effect of markedness on response latency which would be required to support the *response latency prediction*, certain subsets of the data do show this pattern. Specifically, marked *BADHIATUS items ease production relative to unmarked *BADHIATUS items, decreasing response latency, while marked *GEMINATE items exert the greatest amount of delay on post-lexical processing speech production compared to their unmarked counterparts, significantly impacting processing time. All other constraints' marked items do not have significantly longer response latencies than their unmarked items. However there are significant differences between constraints' effect of markedness in the FAST condition, the specific environment where Hypothesis 2 ("equality of markedness") had predicted that variation in markedness might occur.

Broadly speaking, these findings are unexpected: the equality of markedness hypothesis had suggested that if phonological processing was sensitive to the phonotactic markedness of the content which is being processed, as argued for by, among others, Goldrick & Larson (2008) and Goldrick (2011), we would expect to see a processing cost for markedness in all constraints, not just one. Even in a world where phonotactic probability *alone* impacted phonological processing time, the constraints examined here should all show an inhibitory effect on phonological processing. The current finding runs counter to this, not supporting the *response latency prediction,* nor the equality of markedness hypothesis from which it depended. I tentatively interpret this to mean that the relationship between a phonemic configuration which is marked in the phonological grammar and one which exerts a cost in phonological processing is not a one-to-one correspondence, and is perhaps mediated by the word-boundary which intervenes in the middle of the marked sequence. I return to these findings in section 4 below.

3.22 Experiment 3: Speech error analysis

To determine whether a marked item is more likely than an unmarked item to be produced with a speech error, I ran a generalized linear mixed-effects model on a superset of the data analyzed above in the production experiment, combining error-containing tokens with the same tokens used for the response-latency analysis used in Experiment 2.

3.23 Data

This analysis combined the 2293 tokens in the speech production analysis with the 306 tokens which contained speech-errors, yielding 2599 tokens. Note that the 20 *IambicClash tokens which were produced with non-canonical stress placement via the “rhythm rule” were excluded from this set.

Table 13 below displays the percentages and raw counts of errors by paradigm and constraint.

DELAY

	Unmarked	*Marked*
*BadHiatus	11.6% (13/112)	8.8% (10/113)
*Geminate	7.9% (8/101)	3.6% (4/110)
*IambicClash	7.3% (9/124)	6.0% (7/116)
*SibilantCluster	5.9% (7/118)	8.7 % (10/114)
*ThreeConsonants	5.8% (7/121)	5.8% (7/121)
Overall	7.6% (44/576)	6.6% (38/574)

FAST

Constraint	*Unmarked*	*Marked*
*BadHiatus	23.6% (32/136)	17.9% (26/145)
*Geminate	21.1% (28/133)	16.4% (23/140)
*IambicClash	14.1% (23/163)	15.5% (23/148)
*SibilantCluster	8.6% (13/150)	22.7% (30/132)
*ThreeConsonants	9.9% (15/151)	7.3% (11/151)
Overall	15.1% (111/733)	15.8% (113/716)

Table 13: *Raw totals of error-containing tokens broken down by paradigm, condition, and constraint.*

The plots below display the relative ratio of speech errors to clean tokens by paradigm, condition, and constraint.

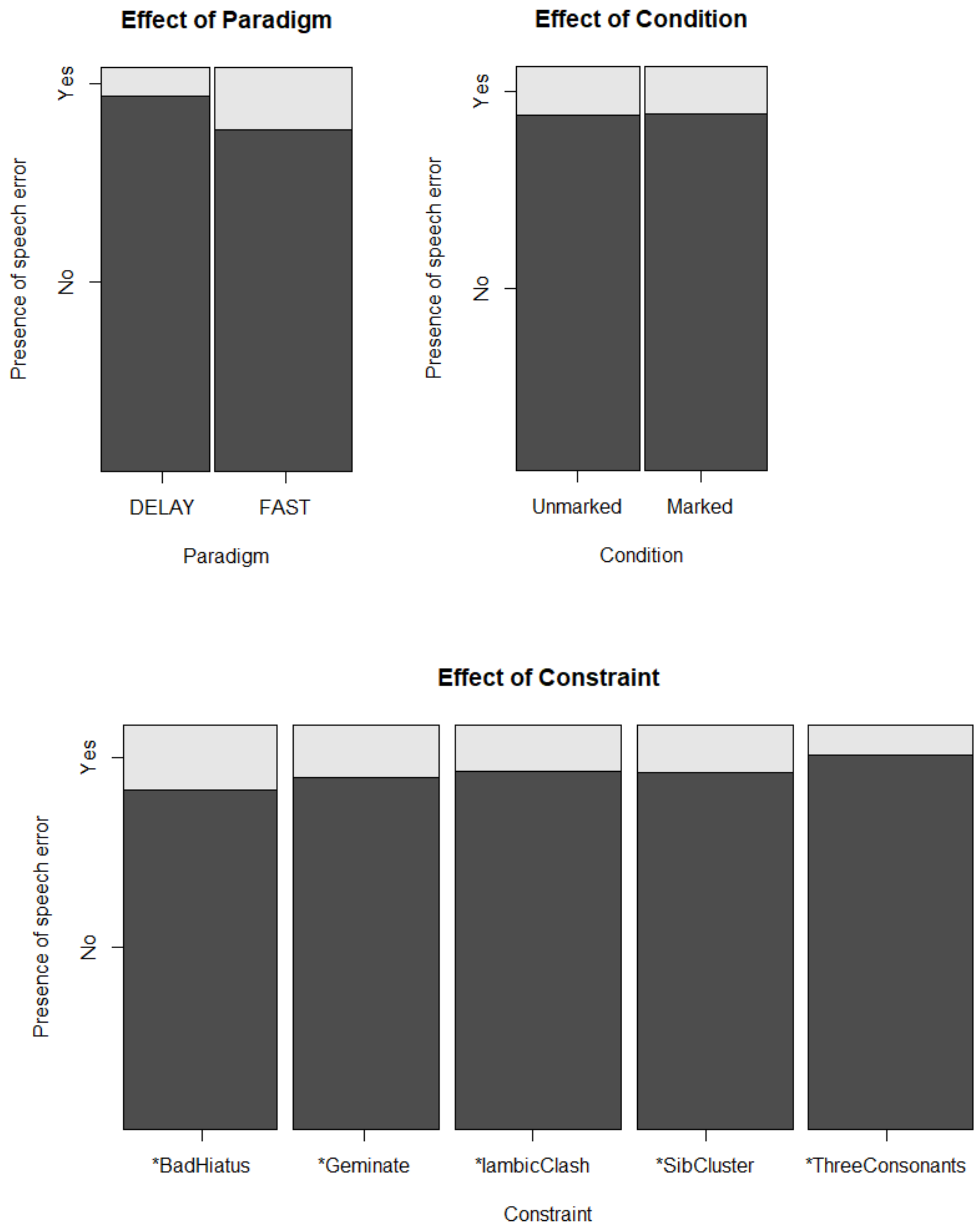


Figure 11: *Plots of the ratio of error-containing tokens to error-free tokens across paradigms, conditions, and constraints.*

3.24 Analysis

The model's dependent variable was binary: whether or not the token in question contained a speech error. In building the model, I took the same top-down approach as in I did in analyzing the previous experimental data, beginning with a model containing all available predictors and removing those which did not contribute significantly to the fit of the model to the data. I retained the same random effect structure as for previous analyses, as well as the predictors of primary interest, *constraint*, *paradigm*, and *condition*, across models. As above, subsequent analyses examine their interaction. The final model includes fixed effects of *condition, constraint, paradigm, lipsmack, number of syllables* - *word-one, phonological neighborhood size* - *word-one, log-frequency – word-one* and *log-frequency – word-two*. The model also included a random intercept for subject and item, and a random slope for condition nested within subject.

3.25 Results

The model described above found that the degree to which marked items are more likely than unmarked items to contain a speech error in the FAST paradigm was not significantly different than the degree to which marked items are more likely than unmarked items to contain a speech error in the DELAY paradigm.

Model structure:

```
glmer(SpeechError ~ Condition*Paradigm + Lipsmack + Num. Syll.
Word-One + Phon. Neighbors Word-one + Log-FreqWordOne + Log-
FreqWordTwo + (1|Group) + (1+Condition|Subject), family =
"binomial")
```

Model output:

```
Fixed effects:
                                      Estimate  Std. Error z value Pr(>|z|)
(Intercept)                            0.83271   0.80697    1.032 0.302122
Condition = marked                    -0.26277   0.29850   -0.880 0.378709
Paradigm = FAST                        0.88250   0.25887    3.409 0.000652 ***
Lipsmack = True                       -1.18860   0.42079   -2.825 0.004732 **
Num. Syll. Word-one                   -0.66551   0.29290   -2.272 0.023078 *
Phon. Neighbors Word-one              -0.04937   0.02325   -2.123 0.033729 *
Log-freq. Word-one                    -0.36271   0.14897   -2.435 0.014901 *
Log-freq. Word-two                    -0.60317   0.15129   -3.987 6.69e-05 ***
Condition = marked x Paradigm = FAST   0.27247   0.29429    0.926 0.354521
```

Table 14: *Best-fitting generalized linear mixed effects regression model of speech-error likelihood data from Experiment 3.* Significance codes: *'***' p < 0.001; '**' p < 0.01; '*' p < 0.05; '.' p < 0.1.*

The following plots show the main effect of condition (not significant) and paradigm.

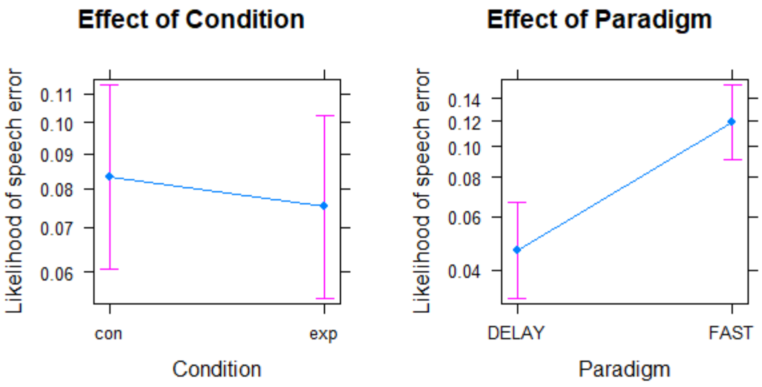


Figure 12: *Partial effect plots of the effects of condition (con = unmarked, exp = marked) and paradigm on likelihood of a token containing a speech error; based on regression model in Table 14.*

The main effect of condition (marked or unmarked) is not a significant predictor of speech error rate. Further significant predictors show that that a token being produced in the FAST paradigm increases the log-odds that it contains a speech error – this is consistent with the hypothesis that the DELAY paradigm allows participants to plan their upcoming speech, and thus to an extent optimize their production, while the FAST paradigm permits participants no such advantage. This finding also is in accordance with the theory that many experimentally-induced speech errors arise via mistakes in lexical processing (Goldrick, Folk, & Rapp 2010) or post-lexical processing (Onishi, Chambers, & Fisher 2002), rather than purely in motoric actuation. Non-phonological factors of significance were the lexical-level factors of log-frequency of *word-one* and *word-two*, number of phonological neighbors for *word-one*, and the number of syllables in *word-one*, all of which decrease the log-odds of a token containing a speech error. Intriguingly, while Breath did not meaningfully contribute to the prediction of whether a given token would contain a speech error, the presence of an overt smacking of the lips before speaking did decrease the log-odds of the spoken token containing a speech error. This finding seems to bear out the speculation, proposed above in 2, idea that participants may on some level deploy a pre-speech lipsmack as a means of “stalling for time” while they plan an utterance, since the presence of a lip-smack is also associated with greater response latency. Interestingly, however, response latency itself, though highly correlated with lip-smack, did not significantly predict a decrease in likelihood of a speech error.

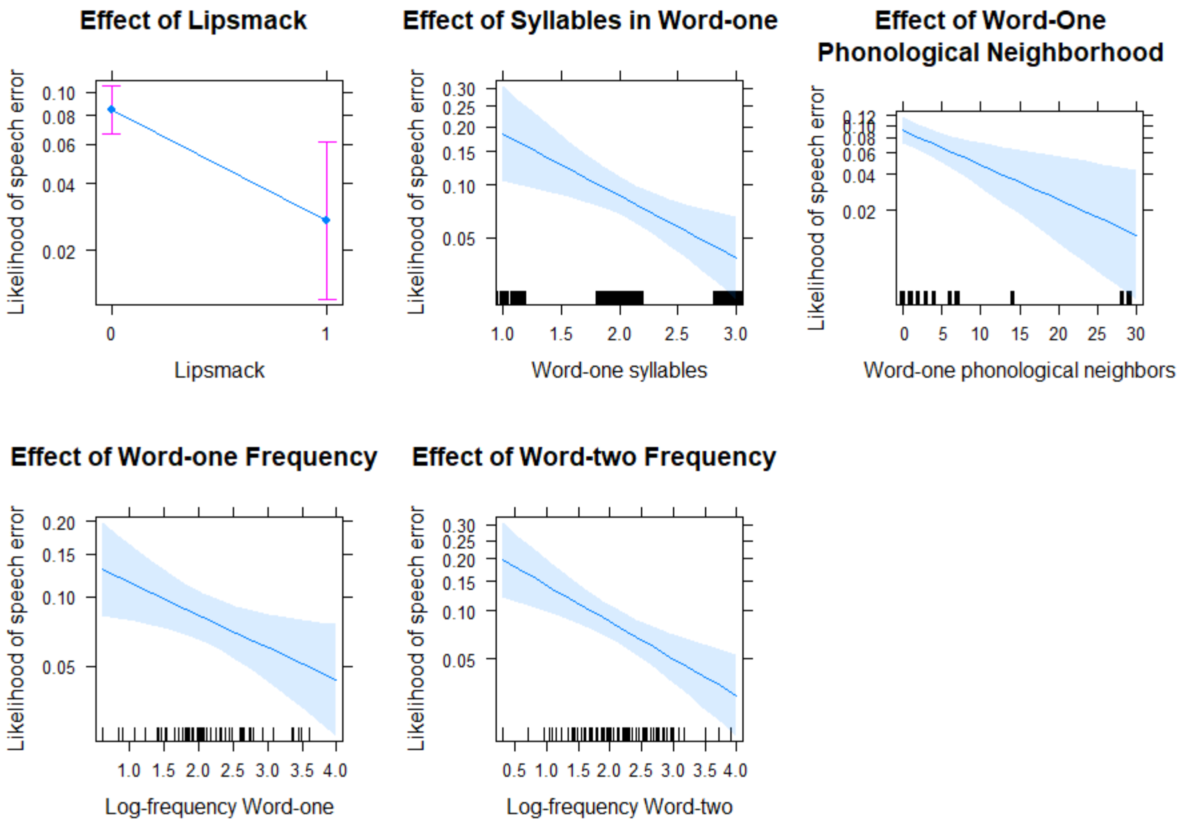


Figure 13: *Partial effect plots of behavioral and lexical factors which significantly modulate the likelihood of a token containing a speech error; based on regression model in Table 14.*

3.26 Further analyses

Rerunning the same model with instead a three-way interaction of Constraint*Paradigm*Condition yielded no new significant finding for any other predictors, and the same two analyses performed on the dataset without *BadHiatus tokens also yielded no further significant results.

3.27 Discussion

The models above found no significant impact of markedness, or of the interaction of markedness and paradigm, on the likelihood of a token to contain a speech error. I interpret these results to mean that this experiment found that speech errors are no more likely to occur in marked tokens than in unmarked tokens. This finding does not bear out the *speech error prediction* made above in section 2.2, which suggested we might find a greater incidence of speech errors in marked tokens, particularly in the FAST paradigm. This result mirrors the findings of Vitevitch et al. (2004) which showed that in unprepared speech elicited in a speeded picture-naming paradigm, words with low phonotactic probability were not significantly more likely to contain a speech error than those with high phonotactic probability.

In contrast to Vitevitch et al., however, I compared the unprepared speech of the FAST paradigm to speech after the off-line completion of lexical and phonological processing, and find that speaking unprepared leads to significantly more errors. This intuitively makes sense: if participants have time to perform lexical and phonological processing without the added pressure to begin speaking as fast as possible, their processing at these stages is less error prone (or, alternatively, they have the opportunity to correct mistakes in processing before they become evident in speech). If the imperative is to begin speaking quickly, on the other hand, participants may be more rushed during processing and lack an opportunity to correct any mistakes they notice before beginning speech, leading to a more error-prone production process. However, we cannot say whether the errors are caused predominantly by phonological markedness, because marked items are no more likely to contain an error than unmarked items, nor can we confidently state what stage of processing any given error occurs in (except perhaps that the majority of errors in the DELAY paradigm occur during motor processing). In summary, then, findings from Experiment 3 lead to similar conclusions as those in Vitevitch et al. (2004): in a simple speech-production task, speech errors are not notably correlated with phonological markedness, suggesting that an experimental paradigm designed specifically to elicit speech errors might be more successful in probing the role of phonological markedness and phonotactic probability in causing speech errors.

4. General discussion

4.1 Review of findings

The initial goal of this study was to probe the ways in which phonotactic markedness impacts post-lexical processing of phonological material in the production of multi-word sequences. I conducted three studies which sought to replicate findings for effects of markedness on processing and well-formedness judgment tasks which have been demonstrated on single-word stimuli using off-line acceptability judgments, response latencies in speech production, and speech-error analysis. The results, null findings for speech-error and subjective rating tasks, and mixed, somewhat unexpected results for the investigation of the processing cost of between-word markedness, are summarized directly below, and then discussed in light of their ability to contribute to our knowledge about the phonology-processing interface in speech production.

Returning to my original hypotheses and predictions laid out in section 2, I find that the *preference prediction*, that speakers will disprefer marked items to unmarked ones in a subjective ratings task, trends towards significance, but does not rise to the level of rejecting the null hypothesis that phonological markedness does not inform participant preference for one item over another. I hypothesize that the small size of the effect is due to the fact that participants are asked to make acceptability judgments based on a relatively minor phonological context: markedness created between two words, a context where such markedness is often attested in speech and prose. The task also involved real lexical items of English which very possibly carried lexical or semantic information that overcame participants’ phonological intuitions.

The *response latency prediction*, which stated that marked items would have significantly longer response times than unmarked ones, is partially supported: I found no main effect of markedness when collapsing constraints together, but found that two constraints, *BADHIATUS and *GEMINATE, had opposite, significant effects on response latency. Marked *BADHIATUS items had significantly shorter response latencies than unmarked *BADHIATUS items, while marked *GEMINATE items had significantly longer response latencies than unmarked *GEMINATE items. Further, when comparing the magnitude and direction of the effect of markedness for each constraint (whether significant on their own or not) to those of other constraints, I find that there is a robust internal differentiation in how types of markedness impact post-lexical processing of phonological material, borne out through multiple regression. These effects emerge more strongly when considering a subset of the full production task data excluding tokens from participants who have high rates of error-containing tokens and from groups which have very sparse data in the four cells of the Condition*Paradigm design.

The *speech error prediction*, that marked items would have a greater chance of containing a speech error, was not borne out. There was no significant interaction between *condition* and *paradigm* in the model, the expected location to find markedness-induced speech errors, nor were other theoretically interpretable possible interactions and fixed effects significant. This mimics similar findings by Vitevitch & Luce (2004), who found that low phonotactic probability nonce words were no more likely to contain a speech error than high-probability words.

4.2 Interpreting null and mixed findings

The clearest null result of this study, that speech errors were no more likely to arise in marked than in unmarked items, has two possible causes. The first and most interpretable one is that the results of the experiments are veridical: markedness between words does not impact speech-errors. The second possible cause is that factors specific to my experimental investigation and / or analysis did not accurately assess the way phonological markedness impacts the likelihood of speech errors – that is, markedness between words may (or may not) give rise to an increased incidence of speech errors, but I failed to accurately measure that here. Although I did not use an experimental design specifically targeted to elicit speech errors (cf. those of Baars 1992), the number of speech errors relative to the number of clean tokens (306 to 2293, or roughly 13.4%) is in fact greater than the rate of speech errors in a paradigm designed to elicit them overtly (cf. Rose & King 2007 with 6.9% errors in a speech-error-elicitation paradigm). One possible experimental confound is the use of attested English words, which (I assume) speakers have spoken before and thus have practice producing alongside other English words.

The results of the other two studies presented here are more ambiguous and difficult to interpret. I found that in a judgment task participants identify unmarked items as more well-formed than marked items at barely significant rates, an effect which cut across constraint with relative uniformity, but in the production task there was no uniform effect of phonological markedness on processing. Contrary to expectations, when looking at marked items as a whole, they are not processed significantly slower than unmarked items. Sub-analyses reveal that this lack of an across-the-board effect is deceiving: constraints vary in the effect of markedness on phonological processing, whether facilitative (*BADHIATUS) or inhibitory (*GEMINATE), and, even when not exerting a significant impact on processing, they differ significantly with respect to one another. For example, even though marked *SIBILANTCLUSTER items are not processed significantly slower or faster than unmarked *SIBILANTCLUSTER items, the effect of having *SIBILANTCLUSTER-violating markedness is significantly different from having *GEMINATE-violating markedness, with respect to impact on phonological processing. Taken together the fact that the suggestive results of the acceptability judgment task do not line up with the findings in the production task is puzzling – it suggests that the factors that influence intuitions about phonological acceptability and those which influence processing are not necessarily the same. I return to this question below, in connection to a discussion of the sources of phonological markedness.

4.3 Addressing the "equality of markedness" hypothesis

The contradictory findings in Experiments 1 and 2, as well as the unexpectedly non-uniform results of Experiment 2, are difficult to reconcile under a theory such as the one supporting the equality of markedness hypothesis which states that all phonologically marked material should be processed in a similar way. To understand why this expectation is not borne out, I return to the initial basis of comparing markedness in phonological grammar with psycholinguistic markedness in speech production.

As discussed in the introduction, the standard metric for phonological markedness in speech production literature is phonotactic probability, which is usually calculated using the summed positional probabilities or biphone probabilities of each stimulus. Phonotactic probability has been shown to significantly impact speech production, and is the basis of further discussion (notably in Goldrick 2011) of the larger role of phonotactic restrictions in speech production, which draws on data from speech-error analysis as well as speech production studies.

Phonotactic probability, however, does not have the same specificity as most notions of phonological markedness: as discussed in section 2.1.2, phonological markedness is most often discussed in the context of its phonemic substance, or what Optimality-Theoretic markedness constraint it violates in the language's grammar. Phonotactic probability, on the other hand, simply indexes the probability of a sequence given the lexicon of a language, without regard to phonemic substance or the phonological situation which assigns markedness to that particular sequence.

Given these contrasting notions of markedness, the equality of markedness hypothesis was set up with the assumption that phonological markedness was, in this context, approximately equivalent to phonotactic markedness – it is possible that this assumption was false, leading to the mixed results of Experiment 2. However, *THREECONSONANTS-violating sequences are not unattested in English words but are merely uncommon in the lexicon, *BADHIATUS-violating words are vanishingly rare in English, as are *GEMINATE-violating ones, while words violating *IAMBICCLASH and *SIBILANTCLASH are completely absent from the lexicon. Even under theories in which phonological processing responded *only* to phonotactic probability, and not to the interaction of phonotactic well-formedness with phonological grammar, as hypothesized here, such sequences should be ill-formed and result in a delay in processing. The fact that the only constraint to have an inhibitory effect on processing was *GEMINATE is therefore puzzling, and suggests that either the current experiment did not adequately control for confounding factors in response latency, or, failing that, that the interface between processing and the phonological grammar is more complex and nuanced than previously thought. On the other hand, one encouraging note on *GEMINATE is that my findings mirror one by Goldberg (2010) that polymorphemic and compound lexical items which contained a geminate across the smaller (prosodic) boundary increased response latency significantly, though he did not investigate other phonological markedness constraints in this context.

Returning to the comparison between Experiment 1 and Experiment 2, the fact that participants may be sensitive to phonological well-formedness of phrases in judgment tasks raises the question of the source of these judgments, if indeed they are confirmed by further study. Phonological markedness has long been theorized to have multiple functional sources, including perceptual, articulatory, and processing-related biases against words which have characteristics which are not handled well by these different sub-parts of the language faculty (cf. Hayes et al. 2004, Flack 2007, among many others). It is likely, then, that when participants draw on their phonological grammar in determining a well-formedness judgment, this judgment is informed by a conglomeration of these different sources, thus explaining the difference between Experiments 1 and 2. The primary processing-related functional motivation for phonological markedness to date has been motivation for the OCP, which has been argued to arise from a higher activation threshold required to re-activate recently activated phonemes (Frisch 2004). Turning to my Experiment 2, this motivation could explain why *GEMINATE-violating marked items alone exhibited a processing slow-down, as well as Goldberg's similar finding. However, this explanation alone does not explain the lack of phonological processing delay for items with low phonotactic probability: *SIBILANTCLUSTER and *IAMBICCLASH sequences are arguably *lower*-probability than geminate sequences, and they exhibit no such inhibition in production. I am not able to satisfactorily resolve the question of why the results of Experiment 2 pattern as they do; the question of what factors specifically contribute to response latency increases – phonotactic probability alone, phonological markedness alone, or some combination of the two – is not yet understood, and requires further study.

4.4 Addressing the "equality of domains" hypothesis

In the foregoing discussion the focus has primarily been on attempting to explain the perplexing lack of agreement between phonological probability as indexed by phonological markedness and the results of Experiment 2. One possibility which has gone undiscussed, however, is that the prosodic boundary between the two words, across which the marked sequence is created, played a role in modulating the effect of markedness on response latency. Kilborn-Ceron et al. (2018) have shown that the presence of a word-boundary does impede production planning in spontaneous speech for the purposes of the application of phrasal phonological processes (see also Wagner 2012 for an introduction to this research program), Although the nature of the current task is different from the data which Kilborn-Ceron et al. base their conclusion on, it may be the case that the intervening word boundary impedes retrieval of phonological material such that the marked sequence created by the [-*coda # onset*-] sequence in the current experiment is not actually being processed at the same time, and thus is not marked from the point of view of phonological processing. This explanation could also tentatively explain why *GEMINATE was the only type of markedness to show an effect – functional motivation for difficulty processing geminate sequences is argued to lie in difficulty achieving repeated sequential activation of a given phonemic representation in speech production, rather than in post-lexical phonological processing, an effect which could arise even if the geminate sequence was not being processed as a single string of phonemes. This could explain why we did not observe a significant effect of phonological markedness, even when that markedness aligned with low phonotactic probability. Under this interpretation of the findings, the equality of domains hypothesis is not borne out because the word-boundary significantly inhibits between-word processing of marked sequences with the result that phonological markedness created between words did not impact processing directly. This reasoning is also consistent the finding that geminate consonants were the only marked sequences which led to significant processing delays. This explanation, while promising, is still tentative – more investigation would be needed to support the assumptions behind it, including a verification of the *lack* of impact of phonological markedness between words on processing times, and a confirmation that the same marked sequences do indeed induce a processing penalty in single words.

4.5 The relationship between corpus and experimental findings

In section 1.4 I reviewed findings which served as part of the impetus for the current study, which showed that speakers gradiently underrepresent phonotactically-marked word-transitions in speech and writing. The current findings that between-word markedness exerts an impact on processing in speech-production and that this effect is modulated by the type (that is, the phonemic substance) of the markedness in question, when taken in conjunction with the corpus-based findings, raise interesting questions about the causal relationship between the phonological grammar and speech-production mechanisms. I see two possible ways to interpret these findings together: either that there is a direct causal relationship between phonological processing difficulty and avoidance of phonologically-marked word-transitions, or that the two phenomena are both the result of an earlier stage of speech-planning in the grammar.

The first option, that the processing difficulty incurred by marked word-transitions *causes* the avoidance seen in corpora, is not a viable interpretation of the facts, I argue, because for this to be the case post-lexical phonological processing difficulty would have to influence constructional and lexical choice and planning at a stage far earlier than that which actually incurs the penalty. Even allowing for cascading activation or simultaneous planning at several levels (as work by, e.g., Bishop & Kim (2018) suggests), the fact that speech planning at the level of constructional choice and lexical selection of arguments does not necessarily proceed in the same order observed in the surface string makes the prospect that the avoidance of future processing difficulty would drive the effect seen in corpora implausible.

The possibility that both the production penalty for marked word-transitions and their gradient avoidance in the output of the grammar whole-sale both emerge as a by-product of a third, grammar-internal factor, I think, is the more likely explanation that unites these findings. This analysis views processing-difficulty as reflective of phonological markedness, but rather than having markedness-avoidance in speech be driven by the processing cost it incurs, it maintains that phonological markedness is avoided across word-boundaries due to interaction between phonological and syntactic well-formedness constraints on constructional and lexical choices which occurs before the moment of speech production, which results in the gradient avoidance of phonological sequences which, when produced, incur processing penalties due to their grammar-internal markedness. This tentative explanation allows for the finding that when *forced* to speak phonotactically marked phrases, as was the case for participants in Experiment 2 above, some of those phrases are processed with greater difficulty than unmarked sequences. This explanation also allows that such sequences are disfavored during speech-planning, through lexical and constructional selection, and this pre-production stage which disfavors phonological markedness gives rise to the illusion that it is the processing penalty driving the avoidance. Experiments 2 and 3 tested the predictions of this theory with regard to increased difficulty in processing marked phrases in speech production; to test the prediction that such phrases are avoided at the utterance-planning level, another experimental paradigm would be required. One possible experimental manipulation might involve a free-choice picture-naming task involving a choice between two nearly-synonymous syntactic constructions, such as Rosenbach (2003) on the English genitive alternation or Bresnan (2007) on the English dative alternation, or the choice between two nearly-synonymous lexical items in the same phrasal-phonological context.

5. Conclusion

The three experiments laid out here have endeavored to probe the relationship between phonological processing within single words and multi-word utterances, on the one hand, and the relationship between phonological markedness and phonotactic probability on the other. The first point, that phonological markedness affects multi-word processing, is partially supported in that what impact of processing I found occurred across word-boundaries, as predicted by the model laid out in Goldrick & Rapp (2008). I found that different types of phonotactic markedness are distinct in their strength of impact on processing in speech production, although only a subset of the constraints tested show a significant impact of markedness compared to unmarked items, and only one of them shows this effect in the expected inhibitory direction. Further, markedness in this task did not significantly impact the likelihood of speech-errors, which were not more likely in phonologically-marked two-word sequences than in unmarked ones. These findings do not align well with the results of a subjective well-formedness ratings task, however, which suggest that participants have an across-the-board preference for unmarked items, albeit a small one if the effect is confirmed.

If the results of the production study are to be taken as an accurate assessment of the facts, then it seems that there is a rather large mismatch between phonological markedness and phonotactic probability, which confounds any strong claim about the domain of phonological markedness impacting processing beyond what has been stated above. However, an alternative explanation is that the word-boundary which intervenes in the marked sequences is leading to a disruption of the speech-planning process with the effect that the entire marked sequence is not subject to phonological processing at once, and so does not exert an inhibitory effect on this process. Unfortunately on the basis of this study alone it is not possible to distinguish which of these several explanations (mismatch between phonological markedness and phonotactic probability, insensitivity to phonological markedness in phonological processing, or an inhibitory effect of a word-boundary) is responsible for the current findings. More study is necessary to explore the precise role of the word boundary in processing as a part of speech production, as well as any potential mismatch between phonological markedness and phonotactic probability. These results raise interesting questions about the relationship between the neural-level language-production system and the symbolic-level phonological grammar and how the two interact in speech production, and serve as a jumping-off point for potential further inquiry into the relationship between processing and phonological grammar with respect to markedness, frequency, and prosodic structure.

Appendix: Stimuli

The table below contains the stimuli used in all three experiments, as well as their lexical statistics.

Item	*Group*	*Condition*	*Log-freq SUBTLWF - Word 1*	*Phonological Neighbors - Word 1*	*Num. Syllables - Word 1*	*Log-freq SUBTLWF - Word 2*	*Phonological Neighbors - Word 2*	*Num. Syllables - Word 2*	*List*
<u>Constraint: *THREECONSONANTS</u>									
medium sculpture	1	marked	2.449	3	3	2.204	1	2	A
medium icon	1	unmarked	2.449	3	3	1.949	0	2	B
stubborn anger	2	unmarked	2.74	0	2	2.997	2	2	A
stubborn courage	2	marked	2.74	0	2	3.082	1	2	B
modest profits	3	marked	2.48	0	2	2.452	0	2	A
modest assets	3	unmarked	2.48	0	2	2.417	0	2	B
frantic itching	4	unmarked	2.07	0	2	2.065	2	2	A
frantic scratching	4	marked	2.07	0	2	2.279	4	2	B
landmark statement	5	marked	1.813	0	2	3.174	0	2	A
landmark order	5	unmarked	1.813	0	2	3.902	3	2	B
hardened overalls	6	unmarked	1.771	6	2	1.724	0	3	A
hardened camouflage	6	marked	1.771	6	2	1.978	0	3	B
difficult struggle	7	marked	3.487	0	3	2.834	1	2	A
difficult autumn	7	unmarked	3.487	0	3	2.146	4	2	B
basic logic	8	unmarked	2.789	2	2	2.535	1	2	A
basic stigma	8	marked	2.789	2	2	1.447	0	2	B
jovial staccato	9	marked	1.079	1	2	0.954	0	3	A
jovial concerto	9	unmarked	1.079	1	2	1.556	1	3	B
earnest authority	10	unmarked	1.863	0	2	2.899	0	4	A
earnest capacity	10	marked	1.863	0	2	2.522	2	4	B
<u>Constraint: *GEMINATE</u>									
horrid deluge	11	marked	1.982	4	2	1.23	0	2	A
horrid outback	11	unmarked	1.982	4	2	1.681	1	2	B
gentle mirage	12	unmarked	2.927	7	2	1.903	0	2	A

gentle lagoon	12	marked	2.927	7	2	1.914	0	2	B
illicit toddler	13	marked	1.532	1	3	1.613	6	2	A
illicit youngster	13	unmarked	1.532	1	3	1.875	0	2	B
spastic gyration	14	unmarked	1.51	0	2	0.301	0	3	A
spastic contortion	14	marked	1.51	0	2	0.699	0	3	B
madcap pirate	15	marked	1.23	0	2	2.575	3	2	A
madcap rascal	15	unmarked	1.23	0	2	2.305	2	2	B
solemn helper	16	unmarked	2.053	4	2	2.017	2	2	A
solemn mentor	16	marked	2.053	4	2	2.253	1	2	B
secret telescope	17	marked	3.455	0	2	1.945	0	3	A
secret cantaloupe	17	unmarked	3.455	0	2	1.591	0	3	B
nonstop avocado	18	unmarked	1.929	0	2	1.799	0	4	A
nonstop pomegranate	18	marked	1.929	0	2	1.146	0	4	B
massive verdict	19	marked	2.614	3	2	2.747	1	2	A
massive lecture	19	unmarked	2.614	3	2	2.728	0	2	B
bulletproof hogwash	20	unmarked	2.029	0	3	1.477	0	2	A
bulletproof falsehood	20	marked	2.029	0	3	1.041	0	2	B

Constraint: *BADHIATUS

raw agreement	21	marked	2.643	29	1	2.951	0	3	A
raw protection	21	unmarked	2.643	29	1	3.079	1	3	B
magenta flamingo	22	unmarked	1.66	0	2	1.785	0	3	A
magenta alpaca	22	marked	1.66	0	2	1.041	0	3	B
gamma addiction	23	marked	1.82	3	2	2.301	0	3	A
gamma prescription	23	unmarked	1.82	3	2	2.576	1	3	B
legato soprano	24	unmarked	0.903	0	2	2.134	0	3	A
legato arrangement	24	marked	0.903	0	2	2.684	0	3	B
guerilla unrest	25	marked	1.544	0	3	1.724	0	2	A
guerilla revolt	25	unmarked	1.544	0	3	1.903	1	2	B
scuba contraption	26	unmarked	1.898	0	2	1.881	1	3	A
scuba apparel	26	marked	1.898	0	2	1.556	0	3	B
beta avoidance	27	marked	2.004	2	2	1.398	0	3	A
beta dependence	27	unmarked	2.004	2	2	1.342	2	3	B
shallow distraction	28	unmarked	2.393	14	2	2.207	1	3	A
shallow objection	28	marked	2.393	14	2	2.559	1	3	B
jumbo allowance	29	marked	2.061	2	2	2.283	1	2	A
jumbo commission	29	unmarked	2.061	2	2	2.64	1	3	B
sepia reflection	30	unmarked	0.602	0	2	2.354	0	3	A
sepia arrangement	30	marked	0.602	0	2	2.588	0	3	B

Constraint: *SibilantCluster

gorgeous shaman	31	marked	3.09	1	2	2.11	11	2	A
gorgeous warlock	31	unmarked	3.09	1	2	2.27	2	2	B
freakish agent	32	unmarked	1.415	1	2	3.719	1	2	A
freakish sergeant	32	marked	1.415	1	2	3.51	0	2	B
rich salami	33	marked	3.613	28	1	2.06	2	3	A
rich papaya	33	unmarked	3.613	28	1	1.431	0	2	B
hodgepodge potion	34	unmarked	0.903	0	2	2.58	4	2	A
hodgepodge sherry	34	marked	0.903	0	2	2.286	29	2	B
oversize salmon	35	marked		5	3	2.52	8	2	A
oversize mustard	35	unmarked		5	3	2.52	4	2	B
famous guardian	36	unmarked	3.361	0	2	2.555	0	3	A
famous chaperone	36	marked	3.361	0	2	2.004	0	3	B
generous showmanship	37	marked	2.931	0	3	1.146	0	3	A
generous thoughtfulness	37	unmarked	2.931	0	3	1.079	0	3	B
dangerous camel	38	unmarked	3.378	0	3	2.41	6	2	A
dangerous zebra	38	marked	3.378	0	3	2.111	1	2	B
obnoxious sardine	39	marked	2.255	0	3	1.491	0	2	A
obnoxious giraffe	39	unmarked	2.255	0	3	1.887	0	2	B
flawless diehard	40	unmarked	1.845	0	2	0.699	0	2	A
flawless zealot	40	marked	1.845	0	2	0.954	3	2	B

Constraint: *IambicClash

extreme contest	41	marked	2.747	0	2	2.98	1	2	A
extreme affair	41	unmarked	2.747	0	2	3.18	5	2	B
absurd finesse	42	unmarked	2.6	0	2	1.778	1	2	A
absurd gimmick	42	marked	2.6	0	2	1.699	1	2	B
discreet laughter	43	marked	2.32	0	2	2.848	2	2	A
discreet applause	43	unmarked	2.32	0	2	2.76	2	2	B
serene perfume	44	unmarked	1.716	2	2	2.77	0	2	A
serene lotion	44	marked	1.716	2	2	2.22	5	2	B
superb odor	45	marked	2.17	0	2	2	15	2	A
superb array	45	unmarked	2.17	0	2	1.69	10	2	B
supreme cadet	46	unmarked	2.651	0	2	2.248	1	2	A
supreme actor	46	marked	2.651	0	2	2.76	5	2	B
aloof heiress	47	marked	1.462	0	2	1.613	5	2	A
aloof neglect	47	unmarked	1.462	0	2	1.903	0	2	B
obese elite	48	unmarked	1.398	2	2	2.121	7	2	A
obese athlete	48	marked	1.398	2	2	2.279	0	2	B
acute insult	49	marked	2.121	1	2	2.73	0	2	A

acute reply	49	unmarked	2.121	1	2	2.265	0	2	B
opaque design	50	unmarked	0.845	1	2	2.747	4	2	A
opaque pattern	50	marked	0.845	1	2	2.719	3	2	B

References

Albright, Adam. "Feature-based generalisation as a source of gradient acceptability." *Phonology* 26.1 (2009a): 9-41.

Albright, Adam. "Cumulative violations and complexity thresholds: Evidence from Lakhota." *17th Manchester Phonology Meeting*. 2009b.

Baars, Bernard J. "A dozen competing-plans techniques for inducing predictable slips in speech and action." *Experimental slips and human error*. Springer, Boston, MA, 1992. 129-150.

Balota, David A., et al. "The English lexicon project." *Behavior research methods* 39.3 (2007): 445-459.

Bates, Douglas, Martin Maechler, Ben Bolker, Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. (2015).

Bishop, Jason and Boram Kim. "Anticipatory shortening: Articulation rate, phrase length, and lookahead in speech production". Talk given at Speech Prosody 8 (2018)

Breiss, Canaan & Bruce Hayes. "More on leaking grammars: Sentence construction respects phonological markedness constraints." Talk given at LSA 2018, Salt Lake, UT.

Bresnan, Joan. "Is syntactic knowledge probabilistic? Experiments with the English dative alternation." *Roots: Linguistics in search of its evidential base* 96 (2007): 77-96.

Brysbaert, M. & New, B. (2009) *Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English.* Behavior Research Methods, 41 (4), 977-990.

Burnham, Kenneth P., and David R. Anderson. "Multimodel inference: understanding AIC and BIC in model selection." *Sociological methods & research* 33.2 (2004): 261-304.

Chomsky, Noam, and Morris Halle. "The sound pattern of English." (1968).

Chong, Adam J. "Learning consequences of derived-environment effects." *Proceedings of the Linguistic Society of America* 1 (2016): 11-1.

Fiez, Julie A., et al. "Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading." *Neuron* 24.1 (1999): 205-218.

Flack, Kathryn Gilbert. *The sources of phonological markedness*. UMass Amherst, 2007.

Frisch, Stefan A. "11 Language processing and segmental OCP effects." *Phonetically based phonology* (2004): 346.

Goldberg, Ariel M. *Gradient well-formedness across the morpheme boundary*. The Johns Hopkins University, 2010.

Goldrick, Matthew, and Brenda Rapp. "Lexical and post-lexical phonological representations in spoken production." *Cognition*102.2 (2007): 219-260.

Goldrick, Matthew, Jocelyn R. Folk, and Brenda Rapp. "Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure." *JML* 62.2 (2010): 113-134.

Goldrick, Matthew, and Meredith Larson. "Phonotactic probability influences speech production." *Cognition* 107.3 (2008): 1155-1164.

Goldrick, Matthew. "Using psychological realism to advance phonological theory." *The Handbook of Phonological Theory, Second Edition* (2011): 631-660.

Harald Baayen, R., and Petar Milin. "Analyzing reaction times." *International Journal of Psychological Research* 3.2 (2010).

Hayes, Bruce, Robert Kirchner, and Donca Steriade, eds. *Phonetically based phonology*. Cambridge University Press, 2004.

Kilbourn-Ceron, Oriana, and Morgan Sonderegger. "Boundary phenomena and variability in Japanese high vowel devoicing." *NLLT* 36.1 (2018): 175-217.

Kirov, Christo Nikolaev. *A Bayesian approach to speech production*. Diss. 2014.

Levelt, Willem JM. "Spoken word production: A theory of lexical access." *Proceedings of the National Academy of Sciences* 98.23 (2001): 13464-13471.

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

Moore, M., & Gordon, P.C. (2015). *Reading ability and print exposure: Item Response Theory analysis of the Author Recognition Test*. Behavior Research Methods, 47, 1095-1109.

Onishi, Kristine H., Kyle E. Chambers, and Cynthia Fisher. "Learning phonotactic constraints from brief auditory experience." *Cognition* 83.1 (2002): B13-B23.

Pater, Joe. "Austronesian nasal substitution and other NC effects." *The prosody-morphology interface* (1999): 310-343.

Prince, Alan, and Joseph Liberman. "On stress and linguistic rhythm." *LI*. 1977.

Prince, Alan, and Paul Smolensky. *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons, 2008.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. 2013.

Rose, Sharon, and Lisa King. "Speech error elicitation and co-occurrence restrictions in two Ethiopian Semitic languages." *Language and Speech* 50.4 (2007): 451-504.

Rosenbach, Anette. "Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English." *Topics in English Linguistics* 43 (2003): 379-412.

Shih, Stephanie Sin-yun. *Towards optimal rhythm*. Diss. Stanford University, 2014.

Shih, Stephanie, et al. "Rhythm's role in genitive construction choice in spoken English." *Rhythm in cognition and grammar*(2015): 207-234.

Stanovich, K. E., & West, R. F. (1989). *Exposure to print and orthographic processing.* Reading Research Quarterly, 24, 402–433.

Tehrani, Henry. "Appsobabble [online app platform]", www.appsobabble.com. 2015.

Vennemann, Theo. *Preference laws for syllable structure: and the explanation of sound change with special reference to German, Germanic, Italian, and Latin*. Walter de Gruyter, 2011.

Vitevitch, Michael S. "The influence of phonological similarity neighborhoods on speech production." *J. Experimental Psychology: Learning, Memory, & Cognition* 28.4 (2002): 735.

Vitevitch, Michael S., and Mitchell S. Sommers. "The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults." *Memory & Cognition* 31.4 (2003): 491-504.

Vitevitch, Michael S., Jonna Armbrüster, and Shinying Chu. "Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density." *Journal of Experimental Psychology: Learning, Memory, and Cognition*30.2 (2004): 514.

Vitevitch, Michael S., and Paul A. Luce. "Increases in phonotactic probability facilitate spoken nonword repetition." *Journal of memory and language* 52.2 (2005): 193-204.

Wagner, Michael. "Locality in phonology and production planning." *McGill Working Papers in Linguistics* 22.1 (2012): 1-18.

Yip, Moira. "The Obligatory Contour Principle and phonological rules: A loss of identity." *Linguistic inquiry* 19.1 (1988): 65-100.

Zuraw, Kie Ross. *Patterned exceptions in phonology*. Diss. 2000.

Zuraw, Kie. "A model of lexical variation and the grammar with application to Tagalog nasal substitution." *Natural Language & Linguistic Theory* 28.2 (2010): 417-472.

Zuraw, Kie, and Yu-An Lu. "Diverse repairs for multiple labial consonants." *Natural Language & Linguistic Theory* 27.1 (2009): 197-224.