Contextual predictability influences word and morpheme duration in a morphologically complex language (Kaqchikel Mayan)

Kevin Tang[1] and Ryan Bennett[2]

[1]Department of Linguistics, Zhejiang University, No. 866 Yuhangtang Road, Hangzhou, Zhejiang, 310058, China

[2]Department of Linguistics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, California, 95064-1077, USA

Corresponding author: Kevin Tang (linguist@kevintang.org)

Probability is one of the many factors which influence phonetic variation. Contextual probability, which describes how predictable a linguistic unit is in some local environment, has been consistently shown to modulate the phonetic salience of words and other linguistic units in speech production (the PROBABILISTIC REDUCTION EFFECT). In this paper we ask whether the probabilistic reduction effect, as previously observed for majority languages like English, is also found in a language (Kaqchikel Mayan) which has relatively rich morphology. Specifically, we examine whether the contextual predictability of words and morphemes influences their phonetic duration in Kaqchikel. We find that the contextual predictability of a word has a significant effect on its duration. The effect is manifested differently for lexical words and function words. We also find that the contextual predictability of certain prefixes in Kaqchikel affects their duration, showing that contextual predictability may drive reduction effects at multiple levels of structure. While our findings are broadly consistent with many previous studies (primarily on English), some of the details of our results are different. These differences highlight the importance of examining the probabilistic reduction effect in languages beyond the majority, Indo-European languages most commonly investigated in experimental and corpus linguistics.

I. Introduction

The contextual probability of a linguistic unit—a segment, syllable, morpheme, word, or even phrase—refers to the likelihood of that unit occurring in a particular local linguistic environment. Contextual probability has been consistently shown to modulate the phonetic salience of words, segments, and other units in speech production (Arnon and Cohen Priva, 2013; Aylett and Turk, 2004, 2006; Bell et al., 2002, 2003, 2009; Bürki et al., 2011; Cohen, 2014; Cohen Priva, 2015; Gahl et al., 2012; Gregory et al., 1999; Hanique and Ernestus, 2011; Jurafsky et al., 2001; Kuperman and Bresnan, 2012; Lieberman, 1963; Pluymaekers et al., 2005b; Raymond et al., 2006; Tily and Kuperman, 2012; Torreira and Ernestus, 2009; Schuppler et al., 2012; Seyfarth, 2014; van Son and Pols, 2003; van Son et al., 2004; van Son and van Santen, 2005). This relationship between predictability and phonetic form can be termed the *probabilistic reduction effect*.

Most prior studies investigating the probabilistic reduction effect in speech production have drawn on data from majority languages like English and Dutch. This raises the fundamental question of whether the probabilistic reduction effect is cross-linguistically robust. In particular, the languages which have been studied in connection with the probabilistic reduction effect are largely Indo-European languages, with morphological systems that would be typically characterized as analytic (few morphemes per word) rather than synthetic or agglutinating (many morphemes per word). In an analytic language, complex semantic concepts such as causation ('Z makes X do Y') are often expressed using several independent words.

In agglutinative languages, those same concepts may be instead be encoded into a single, internally-complex word, with a high degree of morphological and phonological coherence (e.g. Kaqchikel *xiruwartisaj* /ʃ-i-ɾu-war-tis-aχ/ '(s)he made me go to sleep'). Presumably, these structural differences have consequences for the probabilistic reduction effect: statistical dependencies which hold between words in analytic languages, conditioning word-level predictability, may hold more strongly between *morphemes* in agglutinating languages, lessening or even eliminating the effect of contextual predictability on production at the word level.

In this paper we ask whether the probabilistic reduction effect, as observed for majority languages like English, may still be observed in a language (Kaqchikel Mayan) which has comparably rich morphology. Specifically, we first examine whether the contextual predictability of a word influences its phonetic duration in Kaqchikel; second, we examine whether the contextual predictability of a morpheme within a word influences its phonetic duration, above and beyond the duration of the word itself (focusing specifically on verbal aspect markers). Our study is motivated by the substantial morphological differences between Kaqchikel and Indo-European languages, and by the general lack of research on the probabilistic reduction effect in languages with relatively complex morphological systems.

A. The current studies

Kaqchikel is a K'ichean-branch Mayan language spoken by over half a million people in southern Guatemala. The morphological system of Kaqchikel is moderately agglutinating, especially in the areas of verbal derivation and inflection (see Chacach Cutzal 1990; Kaufman 1990; García Matzar et al. 1999; Brown et al. 2010; Coon 2016). Across lexical categories, the prefixal field is mostly reserved for inflectional affixes, while the suffixal field is composed of derivational affixes (see (1) and (2); the adjectival root *ch'u'j* /t͡ʃˀuʔχ/ 'crazy' is in bold).[1]

(1) x-i-b'e-ki-**ch'uj**-ir-is-aj

ASP-1SG.ABS-DIR-3PL.ERG-crazy-INCH-CAUSE-TRANS
‘they went somewhere to drive me crazy’

(2) qa-**ch’uj**-ir-is-ax-ik
1PL.ERG-crazy-INCH-CAUSE-PASS-NOM
‘our being driven crazy’

While the probabilistic reduction effect has, to our knowledge, never been examined in an agglutinative language, there is nonetheless reason to suspect that such an effect could be found in Kaqchikel. Pluymaekers et al. (2005b), focusing on the effects of lexical frequency (context-free predictability) on durational reduction, examined this effect within morphologically complex words in spoken Dutch. They focused on four Dutch affixes (three prefixes *ge-*, *ver-*, and *ont-*, and one suffix *-lijk*), and found that the token frequency of affixed words was inversely correlated with the duration of the entire affix and the durations of the individual segments in the affix. This suggests that lexical frequency affects not only the duration of whole words, as has been frequently reported, but also the duration of smaller units such as affixes and segments. In a different study, Caselli et al. (2016) examined the probabilistic reduction effect in morphologically complex words in spoken English, similarly finding that whole-word frequency and root frequency had independent effects on word duration.

Pluymaekers et al. (2005a) investigated the effect of contextual predictability of a word, given the previous or the following word, on the duration of the seven most frequent Dutch words ending in the adjectival suffix *-lijk* (considering the duration of the whole word, the stem, and the suffix separately). Contextual predictability given the previous word affected the duration of stems for just two out of the seven word types in this study. Contextual predictability given the following word affected stem duration for all seven word types, and the suffix duration of two word types. Despite the inconsistent effect of predictability across items, this study suggests that word-level contextual predictability, like context-free lexical frequency (Pluymaekers et al., 2005b), may condition the duration of whole words as well as sublexical units.

In another study, Arnon and Cohen Priva (2013) examined the probabilistic reduction effect in multi-word sequences (e.g. *I don't know*) using a combination of experimentally-induced lab speech and a corpus of spontaneous speech. This study found that high frequency word sequences have shorter durations overall. Crucially, this effect holds both within and across syntactic units, and is not reducible to the frequency of the individual words within each sequence. In connection with our study, we note that there is a potential parallel between such multi-word sequences and individual words in agglutinative languages like Kaqchikel: morphologically complex words in agglutinating languages, like multi-word sequences in more analytic languages, often subsume many meaning-bearing units which may be statistically interdependent. In sum, the studies mentioned above suggest that the predictability of an internally-complex structure (a word or multi-word sequence) can modulate phonetic duration at the level of the entire structure or its subparts (e.g. segments, morphemes, words), above and beyond what could be predicted from morphological and syntactic structure alone.

In our study, we considered whether the probabilistic reduction effect might manifest differently for function words and lexical words, and for morphologically simple vs. morphologically complex words. Previous studies of English have treated function words differently from content words, either by analyzing them separately (e.g. Bell et al. (2009)) or by excluding them from analysis completely (the majority of past studies). While there is reason to believe that function words are processed differently from content words (e.g. Levelt et al. (1999)), the lexical∼functional distinction is less clear-cut for agglutinative languages, in which words have a high likelihood of containing both lexical and functional material, and in which there may (perhaps as a result) be a smaller overall number of independent function words. For instance, tense/aspect distinctions are often expressed by independent auxiliaries in English (*will, have*, etc.), but by affixes in Kaqchikel (e.g. *y-, xt-*, etc.; see

Section IV). As a second example, Mayan languages typically have only a few independent prepositions, expressing most spatial relationships by means of inflected nouns known as *relational nouns* (e.g. Kaqchikel *w-ik'in* 1SG.ERG-with 'with me'; see Coon 2016; Henderson 2016 and references there). As a practical consequence, it becomes harder to see how one can exclude functional material from analysis in a language like Kaqchikel, as functional morphemes are so frequently contained within larger lexical words. That said, for practical reasons we follow past work in making a distinction between function words and lexical words in Kaqchikel, with the understanding that many lexical words, though built on a single core lexical category root, also contain one or more functional affixes.

Many studies which relate phonetic reduction to contextual predictability have focused on whole words as the unit of analysis. Indeed, there is a large body of evidence supporting the effect of inter-word contextual predictability on duration (e.g. Bell et al. (2003, 2009); Gregory et al. (1999); Jurafsky et al. (2001); Tily and Kuperman (2012)). Fewer studies have considered whether similar effects might hold at the level of the morpheme as well. Past studies exploring predictability and reduction at the morpheme level have focused on paradigmatic probability (Schuppler et al., 2012; Hanique and Ernestus, 2011; Hanique et al., 2010; Kuperman et al., 2007) rather than contextual probability. Unlike contextual predictability, which describes how likely a linguistic unit such as a word or morpheme is in a given context, paradigmatic probability describes how likely a linguistic unit is to be chosen from a set of related forms (e.g. a set of morphologically complex words belonging to the same inflectional or derivational paradigm). While both of these effects tap into morphological structure, Cohen (2014, 2015) shows that paradigmatic probability may affect phonetic salience in production, independent of contextual predictability. Indeed, the effect of paradigmatic probability on speech production is qualitatively distinct from the effect of contextual predictability, as forms with high paradigmatic probability seem to be phonetically enhanced rather than reduced. Cohen (2014) examined how contextual probability and paradigmatic probability jointly affect the duration of the subject-verb agreement suffix *-s* in

English. It was found that the higher the contextual probability, the shorter the suffix, and the higher the paradigmatic probability, the longer the suffix. Cohen (2015) extended this result by investigating how contextual probability and paradigmatic probability jointly affect the production of verbal inflectional suffixes in Russian (the neuter singular suffix *-o* and the plural suffix *-i*). Two types of paradigmatic probabilities were examined in this study. Cohen (2015) found that as the contextual probability of singular agreement increases, the first formant of *-o* decreases, reducing the acoustic distance between *-o* and *-i*. To the extent that this acoustic shift weakens the phonetic contrast between *-i* and *-o*, it can be viewed as a reduction effect (see also Lindblom 1990 and many others). While one type of paradigmatic probability was associated with a phonetic enhancement effect in this study (pairwise paradigmatic probability), the other measure (lexeme paradigmatic probability) had a less consistent effect on phonetic detail in production. Together, these two studies suggest that the contextual predictability of a morpheme may lead to morpheme-level reduction effects, while the paradigmatic predictability of a morpheme may lead to morpheme-level enhancement effects, at least in English and Russian. In our study, we also considered whether probabilistic reduction might manifest at the morpheme level in Kaqchikel, an agglutinative language.

This paper sets out to achieve three goals. The first is simply to establish whether word-level contextual probability influences word duration in Kaqchikel. The second goal is to determine if the effect of predictability on durational reduction might hold across different types of morphological structures. This second goal is motivated by two questions: (a) whether the probabilistic reduction effect interacts with the morphological complexity of words, and (b) whether the effect can be found in functional morphemes that are independent words, rather than affixes. The third goal is to determine if morpheme-level contextual probability can independently influence morpheme duration for affixes, apart from other factors known to affect morpheme duration in production.

In Study I, we analyze whether a reduction effect associated with word-level contextual probability holds for lexical words, and whether the number of morphemes contained in a word interacts with the hypothesized reduction effect. In Study II, we analyze whether such an effect might hold for function words as well. In Study III, we analyze whether there is an effect of morpheme-level contextual probability on morpheme duration, with a focus on verbal aspect markers.

II. Materials and methods

A. Word duration data

Word durations were extracted from a spoken corpus of Kaqchikel. The corpus in question is a collection of audio recordings of spontaneous spoken Kaqchikel, made in Sololá, Guatemala in 2013. Sixteen speakers of the Sololá variety of Kaqchikel contributed to this corpus and shared short, spontaneous narratives of their own choosing for the recording.

Fifteen (out of 16) of the speakers were born in the department of Sololá. The remaining speaker was born in the nearby department of Sacatepéquez. As of 2013, the speakers were all living in the department of Sololá, with six living in the city of Sololá, and ten in other towns. Six speakers were male, and 10 female; their ages ranged from 19-84 years old (mean = 33 years, median = 28 years, SD = 15.4). The speakers all had self-reported native-level fluency in Kaqchikel. Most speakers reported using Kaqchikel as the primary language of communication at home. Fluency was also assessed impressionistically by co-authors [ANON] during the recording sessions ([ANON] is a native speaker of Kaqchikel, and [ANON] an L2 learner).

In total, the corpus amounts to about 4 hours of recorded speech (≈ 40,000 word tokens). The entire corpus was transcribed orthographically by a native speaker of Kaqchikel. A subset of this corpus (≈ 80 minutes) was divided into utterances using PRAAT (Boersma and Weenink, 2014). For this purpose, an utterance was defined as a breath group, which is a stretch of speech set off by substantial silent pauses at its beginning and end, often flanked by audible inhalations which are visible on a spectrogram. Utterances in this sense often (but not always) coincide with a sentence or clause in the corpus. For this study, we took a subset of the corpus, consisting of approximately 3.5 minutes of audio per speaker (about 50 minutes in total), and annotated it phonetically on the word and segment levels using the Prosodylab-Aligner (`http://prosodylab.org/tools/aligner/`; Gorman et al. 2011; see [ANON] In press, 2018 for a more detailed description of the corpus and alignment process). Word durations were extracted from the resultant aligned corpus. Tokens were excluded from analysis if they a) were produced disfluently, b) were not attested in the written corpus of Kaqchikel (described in the next section) which we used to estimate predictability measures, or c) were found only once in the spoken corpus, as it is impossible to statistically model word-specific variation in duration from single tokens of a given word (see e.g. Pierrehumbert 2002; Coetzee and Pater 2011 for discussion of word-specific phonetic effects). In total, the durations of 8430 word tokens (694 word types) met these criteria and were included in the analysis.

In order to examine the effect of word class (functional vs. lexical) and morphological complexity as predictors of word-level duration, as well as their interaction with contextual predictability, we manually tagged each word type as being a function word or a lexical word. We also tagged word types for the number of morphemes they contain. Tagging was done by one of the authors ([ANON]), a second-language learner of Kaqchikel and a specialist in Mayan languages. Twenty-three word types were identified as typos and excluded from analysis.

The dataset is summarized in Table 1, which contains the number of distinct word tokens and word types divided by word class (functional vs. lexical) and morpheme count. Lexical words in our dataset have morpheme counts ranging from one to five. The distribution of morpheme counts is sparse for the function words, with no function word containing more than two morphemes. The majority of function words are monomorphemic (5223 word tokens and 141 word types). Bimorphemic function words (392 word tokens and 38 word types) amount to 7.5% of all function word tokens; many of these are relational nouns like *awoma* 2SG.ERG-reason 'because of you'. Given the sparsity of function words with higher morpheme counts, in Study II only the monomorphemic function words were analyzed. In sum, 2745 lexical word tokens and 492 word types were analyzed in Study I, and 5223 function word tokens and 141 word types were analyzed in Study II.

Morpheme count →	All		1		2		3		4		5	
↓ Word class	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types
Lexical	2,745	492	864	119	891	169	854	160	121	37	15	7
Functional	5,615	179	5,223	141	392	38	0	0	0	0	0	0

Table 1: Summary of word duration data. Token and type counts for this data are divided by word class in the first column and morpheme count across the table.

B. Probabilistic language model

In order to estimate measures of contextual predicability, we needed access to a reasonably large corpus of Kaqchikel. While it might be possible to estimate such variables using a spoken corpus, as Seyfarth (2014) did for English, our spoken corpus is likely too small to estimate the variables of interest (see Brysbaert and New 2009). This required the use of a written corpus: however, to the best of our knowledge there are no structured corpora of digitized, written Kaqchikel currently available for public use. It was therefore necessary to create a novel, digitized written corpus of Kaqchikel.

Our written corpus was constructed from existing religious texts, spoken transcripts, government documents, medical handbooks, and other educational books written in Kaqchikel—essentially all the materials we could find that were already digitized or in an easily digitizable format (see [ANON] In press, 2018 for more details on the construction of this written corpus). The written corpus contains approximately 0.7 million word tokens and 29,355 word types. Each word in the written corpus was phonemically transcribed using an automated grapheme-to-phoneme conversion script. All predictability variables were estimated using this written corpus.

Two bigram language models were constructed using the written corpus. One model describes the probability of each word given the word before it (the previous word), and the other model describes the probability of each word given the word after it (the following word). Bigram models were chosen over larger n-gram models because it has been found that using a larger window (e.g. a trigram model) often makes a negligible contribution to predicting word duration after bigram probabilities have been taken into account (Jurafsky et al., 2001). Model construction was carried out using the MIT Language Modeling (MITLM) toolkit (Hsu, 2009). The probabilities in the language models were smoothed using the modified Kneser-Ney method (Chen and Goodman, 1999) with the default smoothing parameters provided by the toolkit. These two models were used to estimate the contextual predictability of each word in the spoken corpus.

Phonotactic probability is also known to be a potential predictor of word duration (Gahl et al., 2012). In order to estimate the phonotactic probability of each word, an additional language model was constructed which estimated the probability of segmental transitions within words. Unlike the word-level models, a trigram model was chosen in favor of a bigram model for the calculation of phonotactic probability. This decision is motivated by the fact that the dominant shape of root morphemes in Kaqchikel is tri-segmental /CVC/, and /CVC/ roots are also domains for certain phonotactic restrictions ([ANON] 2016b; see also Hayes and Wilson 2008). The other modeling parameters were identical to the word-level language models.

C. Variables included in the statistical models

In both Study I and Study II we fit linear mixed-effects models to our data, attempting to predict word durations in our spoken corpus from a set of lexical, morphological, phonological, and contextual predictors. As noted above, two different word-level bigram probabilities were considered in investigating whether contextual predictability conditions word duration in Kaqchikel (i.e. the probabilistic reduction effect). These are the bigram probability of a word given the previous word (forward bigram probability), and the bigram probability of a word given its following word (backward bigram probability). Along with these predictors, additional control variables suggested by previous research were also included in our statistical model, in order to ensure that the effect of contextual predictability, if observed, is genuine and independent of any other potential predictors of word duration. These additional predictors are described below.

1. Baseline duration

Baseline duration is a crucial statistical control for investigating the probabilistic reduction effect. The aim of our study is to identify whether contextual predictability can modulate word duration, relative to the *expected* (or 'baseline') duration that each word should have, given other properties of that word which are independent of contextual predictability. Previous work on the probabilistic reduction effect has used a number of methods to estimate baseline durations for words. In most such studies, the number of segments and the number of syllables are used as predictors of baseline word duration. However, these are fairly crude measures of expected duration, as they draw no distinctions between different segment or syllable types (e.g. on average the consonant /t͡ʃ/ might be longer than the consonant /n/). To tackle this, another common method is to estimate the average duration of a segment type in the corpus, and sum the average segment durations for each segment contained in a given word type (e.g. Bell et al. 2009). Variations on this method could involve extending the sublexical units considered from single segments to bigrams or hierarchical structures like syllables, in order to capture the effects that syllable structure and phonotactic context might have on segmental duration (e.g. onset /l/ might not have the same average duration as coda /l/; Sproat and Fujimura 1993). Recently, Demberg et al. (2012) and Seyfarth (2014) used a fairly sophisticated technique which estimates word duration using a text-to-speech synthesis system trained on spoken speech.

In this study, our choice of a method for estimating duration baselines is restricted by the fact that Kaqchikel is an under-resourced language. There exist no text-to-speech synthesis systems for Kaqchikel, or any other Mayan language, which rules out the approach of Demberg et al. (2012) and Seyfarth (2014). Second, our spoken corpus is likely too small to estimate average bigram durations. The corpus contains merely 13,003 syllable tokens, which is too sparse to reliably estimate the durations of all segmental bigrams in the corpus. Kaqchikel has 22 consonant phonemes and 10 vowel phonemes; even assuming just two syllable types, CV and VC, 440 bigrams are possible given this phonemic inventory. Apart from the fact that Kaqchikel permits more complex syllable shapes than just CV and VC (e.g. *xtän* /ʃtən/ 'girl'), the complex morphology of the language produces additional consonant clusters, thus giving rise to even more bigram types (e.g. *nretamaj* /n-r-etam-aχ/ '(s)he learns it'). These data sparsity issues rule out using average bigram durations as a predictor of baseline word duration. Taken together, such limitations forced us to use a segment-level baseline method, as segmental durations can be more reliably estimated from our spoken corpus than the durations of larger units which are more sparsely attested.

Instead of summing the average durations of the segments contained in a given word to calculate its baseline duration, we employed an alternative method suggested to us by Uriel Cohen Priva (p.c.). This baseline method is similar to the method used by Bell et al. (2009), inasmuch as it involves predicting the duration of each word token from the counts of each phoneme type found in that word. It differs in that it uses a regression model to estimate the contribution of each segment, rather than computing the average durations of each segment type directly. To do this, we computed a regression model for the duration of each word token in our spoken corpus. There were 32 predictors in this model, one for each phoneme of Kaqchikel. For each word, the value for each of its predictors is the number of times the corresponding phoneme is found in the word. For example, the word *ninwatinisaj* /n-inw-atin-is-aχ/ 'I bathe him/her/it' contains one instance each of /w t s χ/, two instances of /a/, three instances each of /n i/, and zero instances of all other phonemes. A simple linear regression model was constructed to predict the duration of the 8430 word tokens in the spoken corpus based on their phoneme content. The fitted model was then used to re-predict word durations for each of the original word types. These predicted values then served as the baseline duration for each word type.

This method has the advantage of not relying on obtaining the segment durations directly from the spoken corpus, while allowing for each segment type to contribute differently to the overall duration of a word. Generally speaking, forced-alignment methods can obtain more accurate word-level alignments than segment-level alignments, because segment-level alignment is more dependent on the quality of the original phonetic transcriptions than word-level alignment. Therefore, this method is especially appropriate when segment-level phonetic transcriptions might not match the actual acoustic signal, due to e.g. unanticipated variation in production (such as lenition of segments) or simply human error. These factors are potentially relevant for segment-level alignments in our spoken corpus, as those alignments have not yet been manually corrected.

2. Syllable count

The number of syllables in each word type was included as a predictor of duration. This variable serves two purposes. First, it provides another statistical control for the expected baseline duration of each word, since the baseline estimate used here is dependent only on segments and not on syllables. Second, given Menzerath's law (Menzerath and de Oleza, 1928), and the specific sub-case of polysyllabic shortening (e.g. Turk and Shattuck-Hufnagel 2000 and references there), mean syllable duration may decrease with the number of syllables in the word. This means that syllable count could negatively correlate with overall word duration, once other factors (e.g. segment count) have been taken into consideration.

3. Speech rate

Speech rate was included as a control predictor, since speech rate negatively correlates with word duration essentially by definition. Speech rate was estimated as the number of syllables per second in each utterance, with ‘utterance’ defined as a breath group (see Section II A). This is a fairly standard measure of speech rate in phonetics (e.g. De Jong and Wempe 2009 and citations there).

4. Word position

It is well-known that word duration varies by phrasal position, with phrase-final and phrase-initial words showing some degree of lengthening relative to phrase-medial words (e.g. Klatt 1976; Wightman et al. 1992 and many others). We therefore included two categorical predictors related to phrasal context: one predictor for initial vs. non-initial position and another for final vs. non-final position.

5. Disfluency

Words that occur near disfluencies have been shown to lengthen relative to other words (Bell et al., 2003; Fox Tree, 1997). The relevant sense of ‘disfluency’ here includes both silent pauses in utterance-medial position, and so-called ‘filled pauses’ (such as English ‘uh’, ‘um’, and the like). We therefore included a categorical binary predictor in our analysis, coding if a word is adjacent to a silent pause or not. We did not analyze the potential effect of filled pauses because at present filled pauses are not consistently coded in our spoken corpus of Kaqchikel.

6. Word frequency

The number of occurrences of each word type in the written corpus was used as an estimate of overall word frequency. We expected that word duration would decrease as word frequency increases (Wright, 1979). That said, previous research which has assessed the effect of both word frequency and bigram probability jointly has shown more mixed results concerning the role of word frequency (significant, for instance, in Bell et al. 2002, Gahl et al. 2012 and Tily and Kuperman 2012, but not, for instance, in Seyfarth 2014).

7. Backward and forward bigram probability

Both backward and forward bigram probability were estimated using the word-level language models described above.[2] These two variables are the conditional probability of a word given the previous word (forward bigram probability) or the following word (backward bigram probability), as estimated from the smoothed language models. Previous work has shown that forward bigram probability (probability of word W given the previous word) may have a weaker, or even insignificant effect on phonetic reduction when compared to backward bigram probability (probability of a word W given the following word) (Jurafsky et al., 2001; Bell et al., 2009; Gahl et al., 2012; Seyfarth, 2014). However, these measures may not have an independent effect on word duration once raw, context-free word frequency is taken into account (Bell et al., 2002).

8. Neighborhood Density

The number of phonological neighbors for each word type was estimated using the written corpus. In this study, a phonological neighbor is defined as a word that is one phoneme different from the target word, by a single operation of insertion, deletion, or substitution (i.e. a Levenshtein distance of 1; Luce 1986).

Neighborhood density is known to affect accuracy in word production (Stemberger, 2004; Vitevitch, 1997) as well as naming latencies (Vitevitch, 2002; Vitevitch and Sommers, 2003). Most relevantly, Gahl et al. (2012) has shown that, all else being equal, higher neighborhood density is correlated with shorter word duration in speech production (see also Yao 2011; Vitevitch and Luce 2016). Hence, neighborhood density was included as another predictor in our model.

9. Phonotactic probability

The phonotactic probability of a word is based on the probabilities of the segmental sequences it contains, estimated using the segment-level language model described above. The phonotactic probability of a word is calculated as the sum of the log probabilities of the individual trigrams it contains, with the consequence that longer words will also tend to be less phonotactically probable.[3] Previous work has shown that phonotactic probability affects accuracy in word production (Goldrick and Larson, 2008) as well as naming latencies (Vitevitch et al., 2004). Gahl et al. (2012) found that, unlike neighborhood density, phonotactic probability has an inconsistent effect on word duration, varying with the choice of probability measure and other particulars of model construction. However, Gahl et al. (2012) only dealt with /CVC/ words, while our study examines words across a range of segmental lengths (from 2 to 11 segments, with a median word length of 3 segments). It is well known that neighborhood density is strongly correlated with phonotactic probability, but the strength of the correlation weakens as word length increases: this is because long words have fewer neighbors (Yao, 2011, Ch.2) but not necessarily a lower phonotactic probability (though see Daland 2015). Therefore, we might expect phonotactic probability to have a stronger effect than neighborhood density when words are relatively long.

10. Morpheme count

The number of morphemes a word contains was included in Study I to examine whether the probabilistic reduction effect interacts with morphological complexity. To do so, five interaction terms were included in the model, crossing morpheme count with word frequency, forward bigram probability, backward bigram probability, neighborhood density, and phonotactic probability. Note that a graded (multi-level) coding of morphological complexity was chosen over a binary one (on gradient structure in morphology, see Hay and Baayen 2005).

11. Initial model assessment

Our statistical models contain both continuous and categorical variables. Following standard practice in regression modelling, the continuous variables were first log-transformed (base 10) then z-score normalized (e.g. Baayen 2008, §2.2). Z-score normalization allows us to compare the relative strength of our continuous predictors directly. Categorical predictors were sum-coded to improve the interpretability of the regression coefficents and the collinearity of variables, and to avoid model convergence issues (Wissmann et al., 2007; Jaeger, 2009a,b).

Given that a large number of variables were included in our models, we needed to assess the possibility of collinearity between predictors. We computed the condition number (Belsley et al., 1980) for the model following guidelines in Baayen (2008, p.200), using the function *collin.fnc* in the library *languageR*. According to Baayen (2008, p.200), a model with a condition number $\leqslant 6$ has effectively no collinearity; a condition number ≈ 15 indicates a moderate level of collinearity, and a condition number $\geqslant 30$ indicates a high level of collinearity. For Study I the condition number was 6.17, which should present no danger of collinearity. For Study II the condition number was 9.90, a low level of collinearity.

D. Variables excluded from the statistical models

A number of variables that are known to affect word duration were not included in our statistical models. These decisions are individually justified below.

1. Segment count

Similar to syllable count, segment count can serve as a further statistical control, negatively correlating with word duration after other factors are taken into account (Arnon and Cohen Priva, 2014). The independent contribution of segment count may reflect the compression effects described by Katz (2012) and others: the amount of vowel compression (shortening) in a syllable increases with the number of consonants adjacent to that vowel; similar effects are observed for consonants in clusters (see also Browman and Goldstein 1988). However, segment count was not included in the analysis because it correlates strongly with our baseline duration measure (R^2 = 0.82 with lexical words, and 0.88 with function words). The inclusion of segment count as a predictor might therefore have led to troublesome collinearity with other fixed effect variables.

2. Orthographic length

Previous work (Warner et al., 2004; Gahl et al., 2012; Seyfarth, 2014) on English and Dutch has shown that the orthographic length of a word can affect word duration, even in regression models that include phonological variables like segment and syllable count. However, orthographic length was not included as a predictor in our models because it correlates strongly with segment count and baseline duration (the Kaqchikel orthography is relatively shallow, with a fairly close correspondence between graphemes and phonemes). Additionally, literacy rates are sufficiently low in Kaqchikel that we see little reason to believe that the orthography has a strong influence on Kaqchikel speakers' mental representation of their language (on literacy in Mayan languages, see Fischer and Brown 1996; Richards 2003; England 2003; Brody 2004; Holbrock 2016 and references there).

3. Part of speech

Previous work (e.g., Gahl et al. (2012); Seyfarth (2014)) suggests that certain parts of speech show greater reduction effects in the domain of word duration than other parts of speech. Part of speech was not included in our models because the spoken corpus is not yet annotated syntactically.

4. Repetition

Previous work (Fowler, 1988; Fowler and Housum, 1987) has shown that words which are repeated within some timeframe in a corpus are sometimes reduced in production compared to the first mention of those words in the corpus. However, word repetition does not seem to have a consistent effect on word duration when other factors have been taken into account, such as the intonational contour on new vs. repeated words (Hawkins and Warren, 1994; Aylett and Turk, 2004). Given the inconsistent effect of this predictor and the relatively small size of our dataset, this variable was not included.

5. Informativity

Informativity is defined as the average predictability of a word in context (Cohen Priva, 2008; Piantadosi et al., 2011; Seyfarth, 2014). While it is possible to compute this measure for Kaqchikel using our written corpus, informativity was not included in our analyses. The reason for this exclusion was that we would first like to establish whether more basic measures of contextual predictability have an effect on word duration in Kaqchikel. The examination of the average predictability of a word in context is therefore beyond the scope of this paper, and left for future research.

E. Model procedure

Linear mixed-effects models were used to predict the duration of each word token using the variables outlined above as predictors. The models were constructed in the statistical software platform R (R Core Team, 2017), using the `lmer` function in the `lme4` library (Bates et al., 2015).

We fit two separate mixed models for our analysis. In Study I, we fit a model for word duration over lexical words alone. In Study II, we fit a separate model for word duration over all monomorphemic function words. Polymorphemic function words were not analyzed, as there were not sufficient word tokens or types to analyze durational effects for words of this class (Table 1).

While Barr et al. (2013) recommend fitting the most complex random effects structure justified by the data, we chose not to follow this recommendation because it has been recently suggested that such models may not converge. Furthermore, even when models with maximal random effects structures do converge, they are not always readily interpretable (Baayen et al., 2017), and the inclusion of a large number of random effects can also lead to a reduction of statistical power (Matuschek et al., 2015). Instead, we specified our models' structures (fixed and random) by focusing on the variables of greatest theoretical interest, within the confines of a conservative model design.

In Study I (lexical words), the fixed effects included baseline duration, syllable count, speech rate, word position (initial vs. non-initial), word position (final vs. non-final), word frequency, backward and forward bigram probability, neighborhood density, phonotactic probability, and morpheme count. We also included interaction terms between morpheme count and each of word frequency, forward bigram probability, backward bigram probability, neighborhood density, and phonotactic probability. In Study II, which focused on monomorphemic function words, the fixed effects included all of the above, with the exception of morpheme count and the interaction terms between morpheme count and each of the five variables related to contextual predictability (word frequency, backward and forward bigram probability, neighborhood density, and phonotactic probability).

Table 2 and Table 3 summarize the distribution of the variables (both word duration and the predictors) in Study I and Study II respectively. The tables show the mean, standard deviation, interquartile range and range (max-min) for the continuous variables and count information for the categorical variables.[4]

	Mean	SD	IQR	Range
Word duration (log10, millisecond)	2.640	0.164	0.225	1.396
Baseline duration (log10, millisecond)	2.640	0.107	0.150	0.614
Syllable count (log10)	0.300	0.174	0.176	0.699
Speech rate (number of syllables per sec) (log10)	0.693	0.093	0.119	0.808
Word frequency (log10)	2.164	0.815	1.086	3.710
Neighborhood density (log10)	0.900	0.327	0.415	1.644
Phonotactic probability (log10)	-5.345	1.921	2.393	12.642
Forward bigram probability (log10)	-3.069	1.181	1.478	6.097
Backward bigram probability (log10)	-3.103	1.168	1.440	6.227
Morpheme count (log10)	0.277	0.206	0.477	0.699
Word position (Initial vs Non-initial)	Initial: 370; Non-initial: 2375			
Word position (Final vs Non-final)	Final: 717; Non-final: 2028			
Disfluency	True: 838; False: 1907			

Table 2: Descriptive statistics of variables in Study I

	Mean	SD	IQR	Range
Word duration (log10, millisecond)	2.299	1.069	0.316	1.412
Baseline duration (log10, millisecond)	2.299	0.133	0.163	0.895
Syllable count (log10)	0.051	0.127	0.000	0.477
Speech rate (number of syllables per sec) (log10)	0.697	0.093	0.115	0.935
Word frequency (log10)	3.754	0.927	1.219	4.943
Neighborhood density (log10)	1.507	0.306	0.171	1.839
Phonotactic probability (log10)	-2.462	1.414	1.385	11.723
Forward bigram probability (log10)	-1.774	1.017	1.288	6.209
Backward bigram probability(log10)	-1.767	1.043	1.384	6.034
Word position (Initial vs Non-initial)	Initial: 803; Non-initial: 4420			
Word position (Final vs Non-final)	Final: 560; Non-final: 4663			
Disfluency	True: 1541; False: 3682			

Table 3: Descriptive statistics of variables in Study II

In addition to these fixed effects, we included by-word random intercepts and slopes, and by-participant random intercepts and slopes, to take into account durational variability which might reflect idiosyncratic properties of individual word types or individual speakers. Given the size of our data set, we were not able to fit random effects for all the variables included in our fixed effects structure. Instead, we focused on the two bigram probability effects (forward word bigram probability and backward word bigram probability), which seem from past work to have a stronger effect on word duration than context-free predictors such as word frequency. The inclusion of by-word and by-participant random effects for backward and forward bigram probability ensure that our estimates of the effects of these factors will be relatively conservative (Barr et al., 2013). For the model structure of the initial models, see Appendix A: Study I and Study II.

To avoid overfitting our data, these initial models were then simplified following a step-down, data-driven model selection procedure which compared nested models using the backward best-path algorithm (e.g. Gorman and Johnson 2013; Barr et al. 2013), making use of the `anova()` function and likelihood ratio test provided by R. The two bigram probability fixed effects (the individual terms) and the two random slopes of bigram probabilities by participants and items were never considered for exclusion, since the key interest of this study is the effect of contextual predictability. In other words, only the control variables and the higher order variables (if any) were considered for exclusion. The random intercepts for both PARTICIPANT and WORD were never considered for exclusion, as it is standard practice to include these random effects in models of this type (e.g. Jaeger 2008). We chose a relatively liberal threshold of $\alpha = 0.1$ to be conservative in our model selection procedure, preferring to include potentially relevant predictors in the final model if they were reasonably well-justified. A set of models which are minimally simpler than the superset model (i.e. with one less predictor or interaction term) were generated and were then compared with the superset model. If the likelihood ratio test resulted in a p-value of 0.1 or higher, the simpler model was taken to be an improvement on the superset model. If there were multiple subset models which exceeded this α threshold, the subset model with the strongest evidence (the highest p-value) was selected. The step-down procedure began from the higher order fixed effects (the interaction terms) to the lower order fixed effects (the individual terms). The principle of marginality was adhered to, such that a lower order fixed effect were kept if there were a higher order fixed effect including it in the model. For the model structure of the best models, see Appendix A: Study I and Study II. The condition numbers for our final models in Study I and Study II were 4.56 and 4.38 respectively, again posing basically no danger of collinearity between predictors.

After each model was fitted, it underwent a process of model criticism. To ensure the normality of the residuals of the model, the dataset used to fit each model was trimmed by removing data points with an associated residual at least 2.5 standard deviations above or below the mean. Each of these trimmed datasets was then refitted using the original model structure. No more than 3% of the data points was trimmed in each dataset.

The statistical significance of the individual predictors in all the models was evaluated by bootstrapping. This is especially appropriate given the size of our dataset, which is potentially too small to reliably estimate p-values for predictors without bootstrap estimation. Bootstrapping was carried out using the `bootMer` function in the `lme4` library. 1000 bootstrap simulations were performed for each model. Bootstrapped p-values and confidence intervals at 95% were computed for each predictor in each model. We follow the conventional alpha-level of 0.05 for significance. Therefore, we will refer to any p-value below 0.05 as 'significant'. However, given the fact that we are dealing with small data, and some effects might reach significance with more data, we refer to effects that have a p-value greater than 0.05 but smaller than 0.1 as 'near-significant'.

III. Results

A. Study I

Table 4 summarizes the fixed effects in Model 1, which is fitted over lexical words.

	β	SE	t	$\text{CI}_{Lower95\%}$	$\text{CI}_{Upper95\%}$	$p_{Bootstrapped}$
Baseline duration	0.4786	0.0283	16.9176	0.4222	0.5340	$<.001^{***}$
Syllable count	0.1777	0.0281	6.3099	0.1210	0.2359	$<.001^{***}$
Speech rate	−0.3913	0.0118	−33.1016	−0.4138	−0.3683	$<.001^{***}$
Word position (Final vs. Non-final)	0.2916	0.0256	11.3795	0.2399	0.3427	$<.001^{***}$
Neighborhood density	−0.0486	0.0179	−2.7108	−0.0846	−0.0131	$.008^{**}$
Bigram prob. (previous)	−0.0383	0.0154	−2.4780	−0.0698	−0.0075	$.02^{*}$
Bigram prob. (following)	0.0062	0.0158	0.3915	−0.0244	0.0368	$.718^{n.s.}$

Level of significance: · ($p \leqslant 0.1$), * ($p \leqslant 0.05$), ** ($p \leqslant 0.01$), *** ($p \leqslant 0.001$).

Table 4: Fixed effects summary for Model 1 (Lexical words). β: coefficient; SE: standard error; t: t-value; $\text{CI}_{Lower95\%}$ and $\text{CI}_{Upper95\%}$: 95% confidence intervals of the coefficient from bootstrapping; $p_{Bootstrapped}$: p-value from bootstrapping simulations; all continuous variables were first log-transformed (base 10) then z-score normalized, and all categorical predictors were sum-coded.

We first examine the non-predictability control variables. Three of the control variables for word duration were highly significant in the expected directions: these are baseline duration (β: 0.4786, SE = 0.0283, $p < .001$), syllable count (β: 0.1777, SE = 0.0281, $p < .001$) and speech rate (β: −0.3913, SE = 0.0118, $p < .001$). Unsurprisingly, the longer the baseline (expected) duration, the longer the word duration; the higher the syllable count, the longer the word duration; and the faster the speech rate, the shorter the word duration. Given that the coefficient of syllable count was positive, we did not find an effect of polysyllabic shortening (Menzerath and de Oleza, 1928).

Words in utterance-initial position showed no significant differences relative to non-initial words, since this predictor was dropped from the final model. However, utterance-final words *were* lengthened relative to non-final words (β: 0.2916, SE = 0.0256, $p < .001$). That is, phrase-final lengthening was observed, but not phrase-initial lengthening. Of the remaining non-predictiability control variables, disfluency and morpheme count did not make a significant contribution to predicting word duration and were dropped from the model.

Having examined the control variables unrelated to contextual predictability, we move on to the three predictability-related control variables. Context-free word frequency and phonotactic probability did not make a significant contribution to predicting word duration, and were dropped from the model. As noted above, the effect of context-free word frequency on duration has been negligible in past work which also takes into account contextual measures of predictability (i.e. bigram probability; e.g. Seyfarth 2014). The effect of neighborhood density was significant (β: -0.0486, SE = 0.0179, p = .008), indicating that the more neighbors a word has, the shorter its word duration is. This facilitatory effect is in line with previous speech production studies (e.g. Goldrick and Larson 2008; Vitevitch et al. 2004). Phonotactic probability was not a significant predictor of word duration; unlike Gahl et al. (2012), we failed to find a facilitatory effect of phonotactic likelihood (the more phonotactically probable a word is, the shorter its duration).

Finally, we examined the two contextual bigram probability variables (probability given the previous/following word). While backward bigram probability (probability given the following word) did not reach significance (β = 0.0062, SE = 0.0158, p = .718), forward bigram probability (probability given the previous word) did ($\beta = -0.0383$, SE = 0.0154, p = .02). The coefficient for forward bigram probability suggests that the more predictable a word is given the previous word, the shorter its duration. To sum up, two out of five of the predictability variables reached significance, and did so in the direction predicted by the probabilistic reduction hypothesis.

Finally, we examined the five interaction terms. None of them make a significant contribution to predicting word duration and were dropped from the model. This suggests that none of the predictability variables have a significant interaction with morpheme count. In particular, neighborhood density and forward bigram probability, themselves significant predictors, did not change with the degree of morphological complexity.

B. Study II

Table 5 summarizes the fixed effects of Model 2, fitted over monomorphemic function words. Like Study I, baseline duration, syllable count and speech rate were all significant predictors, with effects in the expected direction (the longer the baseline duration, the longer the word duration ($\beta = 0.4034$, SE = 0.0366, $p < .001$); the higher the syllable count, the longer the word duration ($\beta = 0.1301$, SE = 0.0349, $p < .001$); and the faster the speech rate, the shorter the word duration ($\beta = -0.2806$, SE = 0.0098, $p < .001$)).

	β	SE	t	$\mathrm{CI}_{Lower95\%}$	$\mathrm{CI}_{Upper95\%}$	$p_{Bootstrapped}$
Baseline duration	0.4034	0.0366	11.0123	0.3332	0.4758	$<.001^{***}$
Syllable count	0.1301	0.0349	3.7255	0.0604	0.1969	$<.001^{***}$
Speech rate	−0.2806	0.0098	−28.6109	−0.3003	−0.2613	$<.001^{***}$
Word position (Initial vs. Non-initial)	0.1264	0.0266	4.7490	0.0760	0.1778	$<.001^{***}$
Word position (Final vs. Non-final)	0.5297	0.0310	17.1021	0.4661	0.5908	$<.001^{***}$
Disfluency	0.1337	0.0205	6.5046	0.0948	0.1744	$<.001^{***}$
Bigram prob. (previous)	−0.0067	0.0165	−0.4111	−0.0389	0.0258	$.66^{n.s.}$
Bigram prob. (following)	−0.0734	0.0230	−3.1801	−0.1204	−0.0260	$.002^{**}$

Level of significance: · ($p \leqslant 0.1$), * ($p \leqslant 0.05$), ** ($p \leqslant 0.01$), *** ($p \leqslant 0.001$).

Table 5: Fixed effects summary for Model 2 (Monomorphemic function words). β: coefficient; SE: standard error; t: t-value; $\mathrm{CI}_{Lower95\%}$ and $\mathrm{CI}_{Upper95\%}$: 95% confidence intervals of the coefficient from bootstrapping; $p_{Bootstrapped}$: p-value from bootstrapping simulations; all continuous variables were first log-transformed (base 10) then z-score normalized, and all categorical predictors were sum-coded.

Both positional effects were significant, indicating that monomorphemic function words were lengthened in both utterance-final ($\beta = 0.1264$, SE = 0.0266, $p < .001$) and utterance-initial position ($\beta = 0.5297$, SE = 0.0310, $p < .001$). The finding of utterance-initial lengthening differs from the results of Study I (lexical words). Disfluency was also a significant variable ($\beta = 0.1337$, SE $= 0.0205$, $p < .001$), indicating that if a word is adjacent to a silent pause, it is lengthened relative to other words (again, unlike our finding for lexical words in Study I).

Having examined our non-predictability control variables, we move onto the five predictability variables. Just as in Study I, word frequency and phonotactic probability were not significant predictors of word duration in Study II and were dropped from the model. Unlike Study I, neighborhood density did not reach significance, and was dropped from the model. Although forward bigram probability was not significant, it was in the expected direction ($\beta = -0.0067$, SE $= 0.165$, $p = .66$). Backward bigram probability, in contrast, was a significant predictor of word duration ($\beta = -0.0734$, SE $= 0.0230$, $p < .01$).

IV. Study III

Verbs in Kaqchikel are inflected for aspect, a grammatical category which indicates the relationship between some reference time and the time of the event described by the verb (e.g. *x-in-tz'ët* 'I see it (before some contextually-specified reference time)' (ASP-1SG.ERG-see); e.g. Reichenbach 1947; Robertson 1992). In Kaqchikel, there are three basic verbal aspect categories: *x-* /ʃ-/ COMPLETIVE, *y-/n-* /j-/∼/n-/ INCOMPLETIVE, *k-/t-* /k-/∼/t-/ POTENTIAL (the 2$^{\text{nd}}$ member of each /A/∼/B/ pair occurs before phonetically null 3SG.ABS agreement, e.g. *n-∅-in-tz'ët* 'I see it' (ASP-3SG.ABS-1SG.ERG-see)).

In this study, we asked whether the duration of aspect markers can be predicted from their contextual probability. As Kaqchikel is a morphologically rich language, and one with obligatory aspect, person, and number inflection on verbs, aspect markers provide a potentially fruitful testing ground for the hypothesis that contextual predictability affects phonetic duration at the level of the individual morpheme, and not just at the level of the word. This question is important to the extent that morphologically rich languages might be expected to show different patterns of contextual predictability than languages with relatively analytic morphological systems (Section I).

In Kaqchikel, aspect markers can be followed by a range of different morphemes. They are commonly followed by ergative or absolutive agreement markers (e.g. *xe'atin* [ʃ-eʔ-atin] 'they (3PL.ABS) bathed' or *xawatinisaj* [ʃ-aw-atin-is-aχ] 'you (2SG.ERG) bathed him/her/it'). They can also be followed by verb stems directly, if the verb is intransitive and has a 3SG.ABS subject (e.g. *xatin* [ʃ-atin] 'he/she/it bathed').

We focused on three aspect markers in this study: /ʃ-/ COMPLETIVE, and both realisations of /j-/∼/n-/ INCOMPLETIVE. The aspect markers /k-/∼/t-/ POTENTIAL are substantially less frequent in our corpus than /ʃ-/ or /j-/∼/n-/, which makes it difficult to reliably compute the effect of contextual predictability on the duration of these morphemes. For that reason, we do not analyze the duration of /k-/∼/t-/ here.

A. Materials and methods

1. Morpheme duration data

The phonetic durations of the aspect markers in our audio corpus were measured using the segment-level (i.e. 'phone-level') annotations described in Section II A. The dataset is summarized in Table 6, which contains the number of distinct verb tokens and types in the audio corpus divided by morpheme count. In total, the durations of aspect markers from 1016 verb tokens (199 verb types) were included in the analysis. Of these 1016 verb tokens, 375 were marked with /ʃ-/ COMP, 506 with /n-/ INCOMP.3SG.ABS, and 135 with /j-/ INCOMP.

Morpheme count →	All		1		2		3		4		5	
	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens	Types
	1016	199	0	0	241	46	667	115	93	31	15	7

Table 6: Summary of word duration data used in Study III. Type and token counts for verbs are shown divided by morpheme count across the table.

The phone-level segmentations produced by our forced-alignment are imperfect, and contain errors. These errors are not likely to be evenly distributed across segments. Segmentation of voiceless fricatives and nasals is a much easier task than the segmentation of glides (e.g. Turk et al. 2006; DiCanio et al. 2013), and so we expected (incorrectly, it turns out) that our automated segmentation for /j-/ INCL would be less accurate than our segmentation for /ʃ-/ COMP and /n-/ INC.3SG.ABS.

As a rough assessment of the accuracy of our forced alignment model across segment types, we hand-corrected a subset of the TextGrids produced by forced alignment, and compared them to the original, automatically aligned output.[5] The median alignment error for /ʃ/ (as it occurred in any morpheme) was 10ms; for /n/, 8.5ms; and for /j/, 10ms. For /n/, 50% of automated alignments were within one millisecond of our hand-corrected standard; this 1ms error criterion was met by 45% of alignments for /j/, and 39% of alignments for /ʃ/. If we set this error criterion to 20ms, it is met by 80% of alignments for /n/ and /ʃ/, and by 70% of alignments for /j/. Within each category, errors appear to be normally distributed, with a large peak below 10ms and a much thinner, long tail extending upward (particularly for /j/).

Given that glides are difficult for both human coders and forced-aligners to segment, it's possible that the relatively low error rate for /j/ reflects the fact that our hand-corrected alignments simply contain the same errors that were produced by the automatic alignment procedure. Our qualitative results, described below, remain the same whether or not we include *y-* /j-/ INCL in our analysis of duration and contextual predictability for aspect markers.

2. Probabilistic language model

In order to estimate morpheme-level measures of contextual predicability, we needed a morphologically parsed corpus of Kaqchikel. At the time of writing, a morphological parser has not yet been developed for Kaqchikel. Manually parsing our entire written corpus would be prohibitively time-consuming, and so we opted instead to use our smaller spoken corpus to estimate contextual probability measures at the morpheme level.

Given our focus on verbal aspect markers, we manually parsed all the verbal word types containing aspect markers which occurred in the spoken corpus (i.e. the same corpus used for computing word and morpheme durations). Morphological parsing was done by hand by one of the authors [ANON], a second-language learner of Kaqchikel and a specialist in Mayan languages. Decisions about how to segment verb forms were generally easy to make, as Kaqchikel is a fairly agglutinating language which normally has clear boundaries between morphemes, particularly among verbal, inflectional prefixes.

The token frequency of each word type was also computed from the same spoken corpus used to measure the duration of aspect markers, and their contextual probability.

A bigram language model was constructed using the parsed spoken corpus of verbal word types. The model describes the probability of each morpheme given the following morpheme in the same word (aspect markers are always word-initial in Kaqchikel). The model construction was carried out using the MIT Language Modelling (MITLM) toolkit with the same parameters as the word bigram language models described in Section II B. The resultant model was used to estimate the backward bigram probability of each aspect marker in the spoken corpus.

3. Variables included in the statistical models

In Study III we fit linear mixed-effects models to our data, attempting to predict the duration of the aspect markers in our spoken corpus from their contextual predictability and other control variables. As noted above, one measure of morpheme-level contextual probability—backward morpheme bigram probability, i.e. the likelihood of an aspect marker given the following morpheme—was included as a possible predictor of the duration of these aspect markers.

As found in Study I and Study II, variables related to word-level predictability, as well as a number of control variables, had an effect on word duration. To take these word-level effects into account, in our analysis of the duration of aspect markers, we included the actual word duration as a control variable. Furthermore, we included two segment-level control variables. The first segment-level control variable is the target segment type. This variable would allow the duration of each of the three segment types (/ʃ-/, /n-/, /j-/) to be different from each other: for instance, we might expect the fricative /ʃ-/ to be longer than the sonorants /n-/ and /j-/. The second segment-level control variable is whether the segment following the aspect marker is a consonant or a vowel, since the aspect marker could have different phonetic properties in different segmental context. This variable therefore serves to control for possible differences in the syllabification of the aspect markers across forms.

The two studies (Cohen, 2014, 2015) known to us which directly examined the effect of morpheme-level contextual predictability on reduction also included paradigmatic probability as a factor (Section I A). Both studies showed that paradigmatic probability has an enhancement effect on the phonetic realization of morphemes. However, paradigmatic probability was not included in the current study, as we would first like to establish whether the effect of contextual predictability might exist at all at the morpheme-level in Kaqchikel. Furthermore, computing paradigmatic probability reliably would most likely require a larger, more thoroughly parsed corpus of written or transcribed Kaqchikel. The joint examination of both paradigmatic probability and contextual predictability is therefore beyond the scope of this paper, and left for future research.

The same model assessment steps were followed as in Section II C 11. The continuous predictors were first log-transformed (base 10) then z-score normalized. The categorical predictors were sum-coded. To assess the possibility of collinearity between predictors, the condition number was computed. The condition number was 2.75, presenting no danger of collinearity.

4. Model procedure

The same model procedure was followed as in Section II E. Linear mixed-effects models were used to predict the duration of each aspect marker of each word token using the variables outlined above as predictors.

The fixed effects included word duration, target segment (/ʃ-/, /n-/, /j-/), following segment type (consonant vs. vowel) and backward morpheme bigram probability. In addition to these fixed effects, we included by-word random intercepts and slopes, and by-participant random intercepts and slopes. We focused on the backward morpheme bigram probability for the by-word and by-participant random slopes to ensure that our estimate of the effect would be relatively conservative.

Table 7 summarizes the distribution of variables (both aspect marker duration and the predictors) in Study III. The tables show the mean, standard deviation, interquartile range and range (max-min) for the continuous variables and count information for the categorical variables.

	Mean	SD	IQR	Range
Marker duration (log10, millisecond)	1.924	0.261	0.415	1.301
Word duration (log10, millisecond)	2.643	0.164	0.217	1.396
Backward morpheme bigram probability (log10)	-0.381	0.323	0.333	1.836
Target segment	/ʃ-/: 375; /n-/: 506; /j-/: 135			
Following segment type	Consonant: 351; Vowel: 665			

Table 7: Descriptive statistics of variables in Study III

For the model structure of the initial model, see Appendix A: Study III. This initial model was subjected to nested model comparisons. Given that morpheme bigram probability (following morpheme) is our key variable of interest, just as Model 1 and Model 2, only the control variables were considered for exclusion to avoid overfitting. For the model structure of the best model, see Appendix A: Study III. The condition number for our final model was 2.54, presenting essentially no danger of collinearity between predictors.

5. Results

Table 8 summarizes the fixed effects in Model 3, which is fitted over the duration of aspect markers.

	β	SE	t	$\mathrm{CI}_{Lower95\%}$	$\mathrm{CI}_{Upper95\%}$	$p_{Bootstrapped}$
Word duration	0.4491	0.0272	16.541	0.3959	0.5022	$<.001^{***}$
Target segment (/ʃ-/ vs. /n-/)	−0.4981	0.1365	−3.6500	−0.7709	−0.2357	$<.001^{***}$
Target segment (/ʃ-/ vs. /j-/)	−0.1077	0.1865	−0.5770	−0.4786	−0.2638	$.526^{n.s.}$
Morpheme bigram prob. (following)	−0.1451	0.0492	−2.9450	−0.2489	−0.0460	$.01^{*}$
Level of significance: · (p ⩽ 0.1), * (p ⩽ 0.05), ** (p ⩽ 0.01), *** (p ⩽ 0.001).						

Table 8: Fixed effects summary for Model 3 (Aspect markers). β: coefficient; SE: standard error; t: t-value; $\mathrm{CI}_{Lower95\%}$ and $\mathrm{CI}_{Upper95\%}$: 95% confidence intervals of the coefficient from bootstrapping; $p_{Bootstrapped}$: p-value from bootstrapping simulations; all continuous variables were first log-transformed (base 10) then z-score normalized, and all categorical predictors were sum-coded.

We first examine the control variables. The effect of word duration was highly significant in the expected direction with a positive estimate (β = 0.4491, SE = 0.0272 , p = .001). Unsurprisingly, the longer the word duration, the longer the duration of the aspect marker. The overall effect of target segment type was significant in the nested model comparison, suggesting that target segments have significantly different durations from each other. A further examination of the two contrasts (/ʃ-/ (base) vs. /n-/ and /ʃ-/ (base) vs. /j-/)

indicates that the aspect marker /n-/ was significantly shorter than the aspect marker /ʃ-/ ($\beta = -0.4981$, SE $= 0.1365$, $p < .001$) but the aspect marker /j-/ was not significantly different from the aspect marker /ʃ-/ ($\beta = -0.1077$, SE $= 0.1865$, $p = .526$). The following segment type (consonant vs. vowel) was dropped from the model, suggesting that potential differences in syllabification did not significantly affect the duration of the aspect marker.

Having examined the control variables, we examine the key variable of interest, backward morpheme bigram probability. Backward morpheme bigram probability was significant in the expected direction with a negative estimate ($\beta = -0.1451$, SE $= 0.0492$, $p = .01$). This suggests that the more probable the aspect marker is given the following morpheme, the shorter its duration.

V. Discussion

In this study, we set out to examine three questions: a) whether the probabilistic reduction effect can be found in Kaqchikel, b) whether the effect (if any) holds across different morphological structures, and c) whether the effect can also be found between morphemes in the same word.

In Study I and Study II, we examined a number of predictability variables. In Study I, we found neighborhood density and forward bigram probability to be significant variables in our model of word duration for lexical words. In Study II, we found backward bigram probability (but not other predictability variables) to have a significant effect on word duration for monomorphemic function words. In Study I, we specifically examined whether morphological complexity interacts with any of our predictability variables, but found no support for any such interactions. Comparing across Study I and Study II, it is clear that contextual predictability affects word duration in different ways for lexical vs. function words. Overall, there is no strong evidence that morphological complexity interacts with the probabilistic reduction effect in Kaqchikel, but the effect of word class (lexical vs. functional) seems clear to the extent that Study I and Study II uncovered some qualitatively different results.

In Study III, we shifted our focus from words to morphemes; specifically we examined whether the contextual predictability of aspect markers given the following morpheme conditions their durations. We found a reduction effect on the duration of morphemes conditioned by their morpheme-level contextual predictabilities. Together, these results support the existence of a probabilistic reduction effect in Kaqchikel at both the word and morpheme levels.

In the following sections, the effects of bigram probability, phonotactic probability, neighborhood density, and morphological structure are examined more closely.

A. Bigram probability for lexical and function words

Bell et al. (2009) examined the effect of bigram predictability on word duration in English, finding that backward bigram probability and forward bigram probability have different effects on function words and lexical words. For lexical words, only backward bigram probability was significant, while for function words, both bigram probability variables were significant, with backward bigram probability showing a slightly stronger effect.

However, these two bigram probability variables behaved differently in our study of Kaqchikel. Lexical words show a significant effect of forward bigram probability (probability given the previous word), but *not* backward bigram probability (probability given the following word) (see Table 4). In contrast, function words show a significant main effect of backward bigram probability (probability given the following word), but *not* forward bigram probability (probability given the previous word) (see Table 5).

We considered, first, whether the differences between our findings and the results of Bell et al. (2009) might reflect differences in how lexical and functional words are distributed in Kaqchikel and English. Figure 1 provides a density estimate plot of backward bigram probability for function and lexical words, and Figure 2 provides a comparable density estimate plot for forward bigram probability. Both figures show that function words are in general more predictable from their context than lexical words (in terms of both backward and forward bigram probability). This replicates the findings of Bell et al. (2009, Fig. 1, p.98) regarding the relative contextual predictabilities of lexical and function words in English. We conclude that differences between the present study and the results of Bell et al. (2009) are unlikely to reflect differences in the overall contextual predictability of lexical and function words in Kaqchikel vs. English.

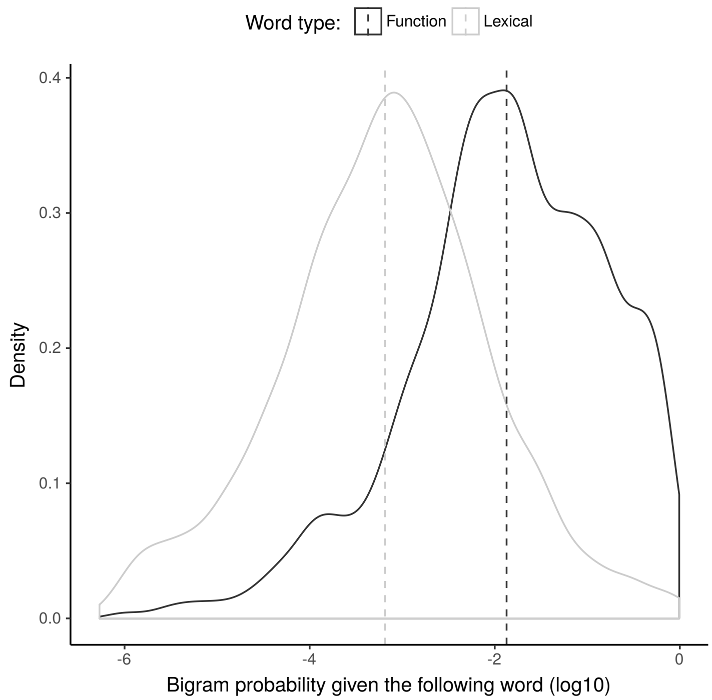

Figure 1: Density estimate plot of backward bigram probability for function words (in black) and lexical words (in grey). The mean probability value is plotted as a vertical dashed line for each word type.

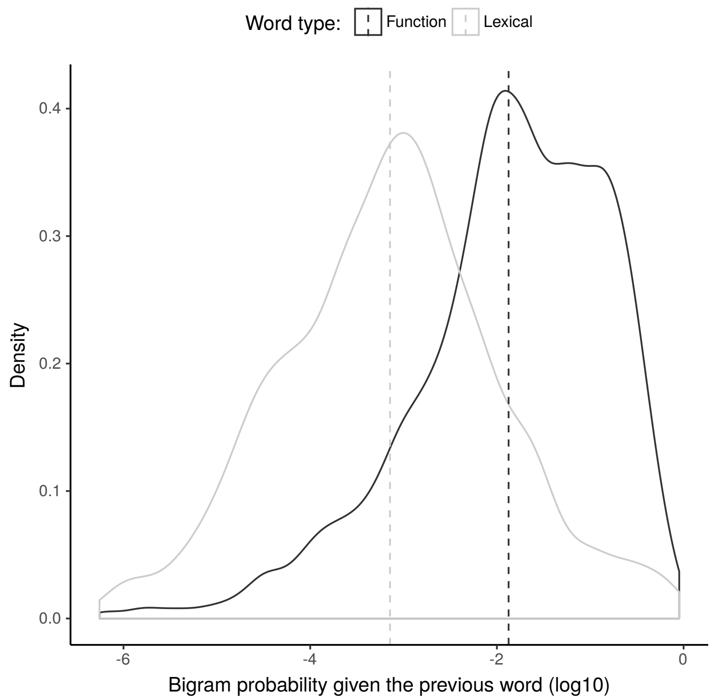

Figure 2: Density estimate plot of forward bigram probability for function words (in black) and lexical words (in grey). The mean probability value is plotted as a vertical dashed line for each word type.

We speculate that the discrepancy between our results and the results of Bell et al. (2009) reflects, instead, syntactic differences between English and Kaqchikel. Kaqchikel is a head-initial language with basic V(O)S order in verb phrases (e.g. [*Xtutz'ët*]$_{\mathrm{V}}$ [*ri tz'i'*]$_{\mathrm{O}}$ [*Juan*]$_{\mathrm{S}}$ 'Juan will see the dog'). English, while also head-initial, has basic SV(O) order, and further differs from Kaqchikel in that verbs are often preceded by functional auxiliaries like *will, might, can, should*, etc. Additionally, in Kaqchikel possessors follow rather than precede their possessums (e.g. *rutz'i' Juan* 'Juan's dog').[6] Another major difference between Kaqchikel and English is that subjects, objects, and possessors may actually be omitted when recoverable from the context: for example, the single-word verb phrase *Xtutz'ët* 'he (i.e. Juan) will see it (i.e. the dog)', is a perfectly acceptable sentence in Kaqchikel, despite the absence of an overt object or overt subject. Possessive phrases are similar in that the possessor need not be expressed overtly when recoverable from the context, e.g. *rutz'i'* 'his (Juan's) dog'. Lastly, subjects, objects, and possessors can all be fronted (and often are) for discourse-related reasons involving topic and focus (3) (e.g. Ishihara 2016; Aissen 2017).

(3) [Ri a Juan$_{\text{POSS}}$]$_{\text{TOP}}$, [ja ri Pedro$_{\text{S}}$]$_{\text{FOC}}$ x-tz'et-o'$_{\text{V}}$ [ru-tz'i' ___ $_{\text{POSS}}$]$_{\text{O}}$ ___ $_{\text{S}}$
the CLF Juan FOC the Pedro COMPL-see-AF 3SG.ERG-dog
'As for Juan$_i$, it was Pedro who saw his$_i$ dog.'

As a consequence, statistical dependencies which might be robust in English (e.g. backward bigram probability of a verb, given its following object) may be less stable in Kaqchikel, a language with different syntactic organization and greater syntactic flexibility than English.[7]

To get a sense of how much the syntax of Kaqchikel differs from the syntax of English, we can compare corpus frequencies for some representative syntactic constructions. Unfortunately, in-depth corpus statistics are not available for syntactic constructions in Kaqchikel, another hurdle posed by the fact that Kaqchikel is an under-studied and under-resourced language. As a rough proxy, we can consider corpus frequencies reported for syntactic patterns in other Mayan languages, which have similar (though certainly not identical) morphosyntactic systems. However, in drawing these comparisons it should be kept in mind that there are likely real differences between Mayan languages with respect to the frequencies of particular syntactic collocations (e.g. England and Martin 2003).

First, we consider argument drop, understood here as the omission (i.e. non-pronunciation) of the subject or object of a verb. Argument drop is ubiquitous in Kaqchikel and other Mayan languages (e.g. Brody 1984; Du Bois 1987; England 1991; England and Martin 2003 and work cited there). For Tojolabal, Brody (1984) reports that the most common realization of transitive clauses is VO, with omission of the subject. In a study of argument realization in five Mayan languages, England and Martin (2003) find that fewer than 3% of transitive clauses contain both an explicit subject and an explicit object (this figure is taken from Clemens and Coon pear). Vázquez Álvarez and Zavala Maldonado (2014) report similar values for transitive clauses in the Mayan languages Ch'ol, and further note that most clauses with intransitive predicates also have non-overt subjects (see Clemens and Coon pear for additional references). This is in clear contrast with English, where argument drop is sharply limited, albeit possible in certain highly restricted contexts (for details, see Haegeman 1987; Haegeman and Ihsane 1999, 2001; Nariyama 2004, among others).

Relatedly, English and Kaqchikel differ in their use of pronouns, a frequent type of functional item (e.g. Zipf 1949). Verbal arguments are typically pronominal in English: Gregory and Michaelis (2001); Michaelis and Francis (2007) report that 95% of subjects and 34% of objects in the SWITCHBOARD corpus are pronouns (Godfrey et al., 1992). Independent pronouns are much less common in Kaqchikel, their referential function being largely subsumed by agreement morphology on verbs, which indicates the person and number of both subjects and objects, thereby facilitating full argument drop (see again Brody 1984; Du Bois 1987; England and Martin 2003, and for Kaqchikel, Maxwell 2009).

With respect to word order, Kaqchikel is significantly more flexible than English. As noted above, the basic word order in Kaqchikel is V(O)S. However, this is not the most *frequent* order in Kaqchikel, or in other Mayan languages which have basic V(O)S or VS(O) order. More typical are constructions in which the subject or object has been fronted for reasons of topic or focus (3) (see Aissen 2017 and references there). Particularly prevalent is SV(O) word order, though all other permutations of {S,V,O} are attested with some regularity (Brody, 1984).

By way of illustration, England (1991) describes the results of an unpublished study of word-order in 16$^{\text{th}}$ century Kaqchikel conducted by José Obispo Rodríguez Guaján (see also Maxwell and Hill 2010). In that study, which examined two major colonial-era documents written in Kaqchikel, only 54 sentences were realized with both an overt subject and an overt object. Of these 54 examples, 43 (80%) had at least one fronted argument (all of SVO, OVS, SOV, and OSV occur in this corpus). Twenty-seven of these 54 examples (50%) had fronted subjects, and 16 of these (30% of the total) had SVO, the majority pattern (tied with OVS). This comparison with 16$^{\text{th}}$ century Kaqchikel likely underestimates the incidence of argument fronting in modern Kaqchikel, which tends toward SV(O) order more strongly than the older colonial variety (England, 1991).

Discourse fronting is of course a feature of modern English as well (e.g. *Anchovies, I can't stand*; see Birner and Ward 1998, 2009; Huddleston and Pullum 2002; Miller 2008; Ishihara 2016, and many others). But statistically speaking, the fronting of arguments does not appear to be employed at the same rate in English as in Kaqchikel. Speyer (2010, p.27) observes that topicalization rates in English have declined sharply since the Old English period, and by ~1700 English texts show rates of object topicalization of about 5% or lower. Most topicalized objects in modern English (90.5%) are also pronouns (Speyer, 2010, p.84), while Mayan languages tend toward the topicalization of full nominals (Aissen, 1992, 2017). Lastly, Roland et al. (2007) find that clefting, a discourse fronting construction related to focus (e.g. *It's* ANCHOVIES *that I can't stand*), occurs in less than 0.1% of all sentences in English. While further corpus work is needed to firmly establish statistical differences in discourse fronting patterns in Kaqchikel and English, the available data suggests that discourse fronting is used in a qualitatively different way in the two languages.

There are of course many other syntactic differences between the two languages which could be relevant for conditioning the effects that backward and forward bigram probability have on the duration of lexical words. We highlight argument drop, clausal syntax, and possessive constructions here because (i) these phenomena typically involve multiple lexical words in sequence; (ii) the order of elements in these contexts often differs between Kaqchikel and English; and (iii) these are core aspects of the syntax of Kaqchikel and its use in discourse. As such, it may be that syntactic differences between these and other constructions account for the observed differences in how bigram probabilities condition the duration of lexical words in English vs. Kaqchikel. At present, however, this suspicion remains to be confirmed in a more empirically rigorous manner.[8]

B. Phonotactic probability and neighborhood density

As noted in Section II C 9, phonotactic probability and neighborhood density are known to be correlated, particularly for short words. Gahl et al. (2012) examined the effect of neighborhood density and phonotactic probability on the duration of /CVC/ words in English. They found that neighborhood density had a consistent, reductive effect on word duration, over and above the effect of phonotactic probability. On the other hand, the effect of phonotactic probability was less consistent in their study, and was highly sensitive to details of the statistical model used to analyze word duration.

In our study, the effect of neighborhood density differs depending on the word class (lexical vs. functional). In Study I, neighborhood density was a significant predictor ($\beta = -0.0486$, $p = .008$). However, neighborhood density did not emerge as significant in Study II and was dropped from the model. In both studies, phonotactic probability was insignificant and was dropped from the model.

Given the strong correlation between neighborhood density and phonotactic probability, we followed Gahl et al. (2012) by applying a statistical residualization technique on neighborhood density and phonotactic probability. Using simple linear regression, a residualized neighborhood density was created by regressing neighborhood density on phonotactic probability; similarly a residualized phonotactic probability was created by regressing phonotactic probability on neighborhood density. We repeated the same model fitting procedure as in Section II E and fitted two new models. In one model, the residualized neighborhood density measure replaced the original neighborhood density measure. Both residualized neighborhood density and phonotactic probability survived the nested model comparisons in this model, both with a negative estimate. The effects of other variables in this model (with residualized neighborhood density) were qualitatively the same as those in Model 1. In another model, the residualized phonotactic probability measure replaced the original phonotactic probability measure. Residualized phonotactic probability did not survive nested model comparison and was dropped, while neighborhood density survived, resulting in a best model identical to Model 1.

Our findings match those of Gahl et al. (2012) in a number of ways. Just as Gahl et al. (2012)'s found, phonotactic probability was only significant under a specific statistical model, namely when neighborhood density was residualized; it was not significant when phonotactic probability was residualized, and also when no residualization was applied at all. Neighborhood density was, however, a significant predictor under all three models, with or without residualized variables. This is similar to Gahl et al. (2012)'s finding that phonotactic probability is a less consistent and weaker predictor than neighborhood density.

C. Positional effects and disfluency

Study I found that lexical words are lengthened in utterance-final position, while Study II found that function words are lengthened in both utterance-final and utterance-initial position. We suspect that this difference reflects the fact that, on average, function words are shorter than lexical words in Kaqchikel. Previous work on lengthening at domain edges suggests that domain-initial lengthening has a smaller temporal scope than domain-final lengthening. Specifically, domain-initial lengthening primarily affects single *segments* (Byrd, 2000; Cho and Keating, 2001; Lehnert-LeHouillier et al., 2010), while domain-final lengthening has been found to extend over several *syllables* (e.g. Shattuck-Hufnagel and Turk 1998). On average, monomorphemic function words contain fewer syllables than lexical words in Kaqchikel (function words, mean = 1.25, sd = 0.52; lexical words, mean = 2.15, sd = 0.78). As a consequence, positional lengthening will have a proportionally greater effect on word duration for function words (shorter) than for lexical words (longer), which may explain why the effect of utterance initial vs. non-initial position was only observed for the function words in Study II.

Study II also found that function words were lengthened when adjacent to a disfluency (here, a silent pause in utterance-medial position). This effect was not replicated in the analysis of lexical words in Study I. We have not found any prior work that shows a difference between lexical words and function words in the extent of lengthening due to silent pauses. Bell et al. (2003) found that function words were lengthened when adjacent to silent pauses, but they did not investigate the effect of disfluency on lexical words. Bell et al. (2009) investigated the reduction of both function words and lexical words, but explicitly excluded any words adjacent to disfluencies (including silent pauses). We again speculate that the contextual lengthening of words adjacent to a disfluent pause has an effect for function words, but not lexical words, because function words tend to be shorter.

D. Morphological effects

Morphological complexity had no influence on the probabilistic reduction effect in our study. This lack of an interaction is surprising considering the rich morphology of Kaqchikel. Several speculations can be made about the lack of an interaction. First, the failure to find any effect of morphological complexity might simply be due to a lack of statistical power, given the size of our data. Study I examined a mere 2745 tokens, which is very small compared to other similar studies on English (e.g. Seyfarth 2014 examined 41,167 word tokens from the BUCKEYE corpus, and 107,981 word tokens from the SWITCHBOARD corpus). Future examinations of our entire spoken corpus (about 40,000 word tokens) should be able to better assess the effect of morphological complexity on the probabilistic reduction effect. Second, as far as we are aware, no previous studies have reported an interaction between morphological complexity and probability measures when modeling phonetic reduction. It may simply be the case that probabilistic reduction effects do not interact directly with morphological complexity. Third, it could be that our definition of morphological complexity is too crude. In particular, our measure of morpheme count is derived from traditional linguistic analysis (e.g. Harris 1951), and ignores the possibility that speakers may store some morphologically complex words as unanalyzed wholes, or even just partially decomposed forms, in their mental lexicon (Hay 2001, Plag 2003, Chs. 3,4).

E. Inter-morpheme predictability

While we did not find an interaction between morphological complexity and the probablistic reduction effect in Section V D, we cannot immediately rule out a role for morphological structure in conditioning probablistic reduction. In Study III we addressed this question more directly by examining the effect of contextual morpheme predictability on morpheme duration.

Study III showed that, after controlling for word duration as well as segmental quality, the predictability of the aspect markers /ʃ-/, /n-/, /j-/ given the following morpheme has a significant, reductive effect on the duration of the aspect marker itself. This is consistent with the findings of Cohen (2014) regarding the English subject-verb agreement suffix *-s*, and Cohen (2015) on Russian verbal inflection suffixes. We therefore found contextual reduction effects at the level of morphemes (Study III) as well as the level of words (Study I and Study II).

VI. Conclusion

Our paper set out to examine the probabilistic reduction effect in Kaqchikel with several goals in mind. First, the general lack of research on the probabilistic reduction effect in languages with complex morphology motivated us to assess the effect in Kaqchikel, a language with relatively rich morphology when compared to well-studied majority languages such as English. Second, of all the factors previously shown to probabilistically condition word duration, we paid particular attention to contextual predictability at the word level (backward and forward bigram probabilities). This was motivated by the observation that many functional items which are realized as independent words in English are instead realized as affixes in Kaqchikel. We hypothesized that this difference might affect the distribution of contextual probabilities between words in the two languages. In addition, we examined a number of other predictability-related factors, essentially as controls (phonotactic probability, neighborhood density, and word frequency). Third, since most studies (with the exception of Bell et al. 2009, on English) have examined only lexical words in research on the probabilistic reduction effect, we evaluated whether the factors involved in the reduction effect differ by word class (lexical vs. function words). Fourth, given the rich morphology of Kaqchikel, and the fact that very few studies have examined the effect of morpheme probability on morpheme duration, we shifted our attention to contextual predictability at the morpheme level, with a focus on aspect markers.

We found, first, that contextual predictability (backward and forward bigram probability) had a significant effect on word duration. We also found the same type of effect for a measure of context-free predictability, namely neighborhood density (though only for lexical words). This finding is consistent with a large number of past studies that have found that both context-free and context-sensitive measures of predictability conspire to probabilistically reduce a word's duration. Most importantly, we replicated these effects in a morphologically complex language, in which we might expect contextual measures of predictabilty to behave differently than in English or Dutch. Furthermore, many of these effects seem to depend on word class, with some effects emerging as significant for lexical words but not function words, or *vice versa*. Lastly, we found that contextual predictability at the morpheme level has a significant effect on morpheme duration. This finding is consistent with the few existing previous studies on morpheme-level predictability. We therefore found effects at multiple levels (between words and between morphemes), and we think that investigating those findings and their relation to each other, especially in heavily affixing languages, will be important for understanding how contextual probability affects duration. We cannot do this at the level of detail we would like, because we are working with an under-resourced language and do not currently have the right type of corpus to ask these questions in a broader and/or more targeted way.

While our findings are broadly consistent with many previous studies of the probabilistic reduction effect (primarily on English), some of the details of our results are different. For instance, backward bigram probability was *less* robust than forward bigram probability with lexical words. Precisely these differences highlight the importance of examining the probabilistic reduction effect in languages beyond English, Dutch, and other standardly studied languages — particularly languages which, like Kaqchikel, have morpho-syntactic characteristics which distinguish them from the majority, Indo-European languages most commonly investigated in experimental and corpus linguistics.

Methodologically, we have demonstrated that even for languages with limited corpus resources (e.g. small amounts of digitized text), it is possible to examine the interplay between lexical statistics and the phonetic details of speech production in naturalistic contexts. Given that 'big data' is unavailable for the vast majority of the world's languages, we hope that this paper will inspire further examination of the probabilistic reduction effect in other minority languages, across a range of typological profiles, even if the size and quality of the data currently available for those languages is less than ideal.

Appendix A: Model structures

Study I and Study II

The regression structure for the initial model for Model 1 (fitted over lexical words) is shown below.

> Duration ~ Baseline duration + Syllable count + Speech rate + Word position (Initial vs. non-initial) + Word position (Final vs. non-final) + Disfluency + Word frequency + Neighborhood density + Phonotactic probability + Bigram probability (previous word) + Bigram probability (following word) + Morpheme count + Morpheme count:Word frequency + Morpheme count:Neighborhood density + Morpheme count:Phonotactic probability + Morpheme count:Bigram probability (previous word) + Morpheme count:Bigram probability (following word) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Participant) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Word)

The regression structure for the initial model for Model 2 (fitted over the monomorphemic function words) differs from the above structure in that it does not include any fixed or random effects which have Morpheme count as a term, because Model 2 is restricted to monomorphemic function words. The structure for Model 2 is shown below.

Duration $\sim$ Baseline duration + Syllable count + Speech rate + Word position (Initial vs. non-initial) + Word position (Final vs. non-final) + Disfluency + Word frequency + Neighborhood density + Phonotactic probability + Bigram probability (previous word) + Bigram probability (following word) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Participant) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Word)

The regression structure for the best model for Model 1 (fitted over lexical words) is shown below.

Duration $\sim$ Baseline duration + Syllable count + Speech rate + Word position (Final vs. non-final) + Neighborhood density + Bigram probability (previous word) + Bigram probability (following word) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Participant) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Word)

The regression structure for the best model for Model 2 (fitted over the monomorphemic function words) is shown below.

Duration $\sim$ Baseline duration + Syllable count + Speech rate + Word position (Initial vs. non-initial) + Word position (Final vs. non-final) + Disfluency + Bigram probability (previous word) + Bigram probability (following word) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Participant) + (1 + Bigram probability (previous word) + Bigram probability (following word) | Word)

Study III

The regression structure for the initial model for Model 3 is shown below.

Marker duration $\sim$ Word duration + Target segment + Following segment type + Morpheme bigram probability (following morpheme) + (1 + Morpheme bigram probability (following morpheme) | Participant) + (1 + Morpheme bigram probability (following morpheme) | Word)

The regression structure for the best model for Model 3 is shown below.

Marker duration $\sim$ Word duration + Target segment + Morpheme bigram probability (following morpheme) + (1 + Morpheme bigram probability (following morpheme) | Participant) + (1 + Morpheme bigram probability (following morpheme) | Word)

Notes

[1]Glossing conventions follow the Leipzig Glossing Rules (`https://www.eva.mpg.de/lingua/resources/glossing-rules.php`) and the Mayan-specific conventions set out in [ANON] (2016a).

[2]Note that the transcriptions of the spoken corpus were used to form part of the larger written corpus that was used to compute the language models. Since all of the bigrams in the spoken corpus were thus attested in the written corpus, the estimates of the backward and forward bigram probability do not depend on the smoothing parameters used to compute the language models.

[3]The question of whether we should be normalizing phonotactic probability by word length is both a philosophical issue (see Daland 2015) and an empirical issue. Bailey and Hahn (2001) compare different phonotactic probability measures, and find that a non-normalized measure of phonotactic probability (which penalizes longer words more harshly than shorter words) provides a modest but consistent gain in variance explained in a word-likeness judgment task. For this reason we adopt a non-normalized measure of phonotactic probability here, acknowledging that best practices have not yet been established on this point (see also Nerbonne et al. 1999).

[4]Note that the descriptive statistics for the continuous variables are based on values before z-score normalization to be maximally informative about the distribution of the variables, because z-scores have by definition a mean value of zero and a standard deviation of one.

[5]We thank Andrea Maynard for carefully hand-correcting these TextGrids.

[6]A reviewer observes that possessors can also follow possessums in English, as in *the tail of the dog.* There are many non-trivial differences between this construction and the corresponding construction in Kaqchikel. First, postnominal possession in English involves a prepositional phrase, while postnominal possession in

Kaqchikel does not. Second, post-nominal possession is the primary means of expressing possessive relations in Kaqchikel (Aissen 1999, Brown et al. 2010, 155-7), while English also makes frequent use of an alternative construction, the Saxon genitive *-s* (*the dog's tail*). (Grafmiller 2014 reports that the Saxon genitive *-s* is used for 22-45% of possessive constructions, depending on the corpus genre.) Third, postnominal possession in English is subject to a raft of semantic and pragmatic conditions which do not appear to condition post-nominal possession in Kaqchikel (Barker 1995; Rosenbach 2014; Grafmiller 2014 and references there). All of these grammatical differences could plausibly lead to substantial differences in word-level transitional probabilities between Kaqchikel and English.

[7]We assume here and elsewhere that statistical dependencies (such as high bigram probabilities between words) are more likely to hold between words which occur within the same syntactic constituent than between words which belong to different syntactic constituents (e.g. Saffran 2002, 2003).

[8]A nagging issue which we do not address here concerns the fact that speech production is essentially 'future-oriented'. For example, anticipatory coarticulation is typically stronger than perseveratory (hold-over) coarticulation, and anticipatory speech errors are more common than perseveratory speech errors. Such facts suggest that speech production is more strongly influenced by upcoming words than by previously uttered words (see Manuel 1999; Hyman 2002; Hansson 2010; Garrett and Johnson 2013 for discussion and further references). We might therefore expect that backward bigram probability should affect word duration in all languages, due to entirely general facts about speech planning and speech production. On this view, the lack of an effect of backward bigram probability on the duration of lexical words in Kaqchikel remains unexplained, despite the syntactic differences between English and Kaqchikel that we pointed to above. We leave a deeper investigation of this issue to future work.

References

Aissen, J. (1992). Topic and focus in Mayan. *Language*, 68(1):43–80.

Aissen, J. (1999). External possessor and logical subject in Tz'utujil. In Payne, D. and Barshi, I., editors, *External Possession*, pages 451–485. John Benjamins, Amsterdam.

Aissen, J. (2017). Information structure in Mayan. In Aissen, J., England, N., and Zavala Maldonado, R., editors, *The Mayan Languages*, pages 293–324. Routledge, New York.

[ANON] (2016a). *Language and Linguistics Compass*, 10(10):1–14.

[ANON] (2016b). *Language and Linguistics Compass*, 10(10):469–514.

[ANON] (In press, 2018). *Laboratory Phonology.*

Arnon, I. and Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3):349–371. PMID: 24416961.

Arnon, I. and Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, 9(3):377–400.

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

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

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

Baayen, R. H., Vasishth, S., Kliegl, R., and Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, 94:206–234.

Bailey, T. M. and Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44(4):568–591.

Barker, C. (1995). *Possessive descriptions.* CSLI Publications, Stanford, CA.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255 – 278.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

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

Bell, A., Gregory, M. L., Brenier, J. M., Jurafsky, D., Ikeno, A., and Girand, C. (2002). Which predictability measures affect content word durations? In *ISCA Tutorial and Research Workshop (ITRW) on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology*.

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

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Birner, B. and Ward, G. (1998). *Information Status and Noncanonical Word Order in English*. John Benjamins, Amsterdam.

Birner, B. and Ward, G. (2009). Information structure and syntactic structure. *Language and Linguistics Compass*, 3(4):1167–1187.

Boersma, P. and Weenink, D. (2014). Praat: doing phonetics by computer (version 5.3. 64)[computer program]. retrieved february 12, 2014.

Brody, J. (1984). Some problems with the concept of basic word order. *Linguistics*, 22(5):711–736.

Brody, M. (2004). *The fixed word, the moving tongue: Variation in written Yucatec Maya and the meandering evolution toward unified norms*. PhD thesis, University of Texas Austin, Austin, TX.

Browman, C. and Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica*, 45(2-4):140–155.

Brown, R. M., Maxwell, J., and Little, W. (2010). *La Ütz Awäch?: Introduction to Kaqchikel Maya Language*. University of Texas Press, Austin, TX.

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

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., and Frauenfelder, U. H. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of french connected speech. *The journal of the Acoustical Society of America*, 130(6):3980–3991.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57(1):3–16.

Caselli, N. K., Caselli, M. K., and Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon. *The Quarterly Journal of Experimental Psychology*, 69(3):432–454.

Chacach Cutzal, M. (1990). Una descripción fonológica y morfológica del kaqchikel. In England, N. and Elliott, S., editors, *Lecturas sobre la lingüística maya*, pages 145–190. Centro de Investigaciones Regionales de Mesoamérica, Antigua, Guatemala.

Chen, S. F. and Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, 13(4):359–393.

Cho, T. and Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in korean. *Journal of Phonetics*, 29(2):155–190.

Clemens, L. E. and Coon, J. (to appear). Deriving verb-initial word order in mayan. *Language*.

Coetzee, A. and Pater, J. (2011). The place of variation in phonological theory. In Goldsmith, J., Yu, A. C., and Riggle, J., editors, *The Handbook of Phonological Theory*, pages 401–434. Wiley-Blackwell, Malden MA.

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement. *Morphology*, 24(4):291–323.

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

Cohen Priva, U. (2008). Using information content to predict phone deletion. In *Proceedings of the 27th west coast conference on formal linguistics*, pages 90–98. Citeseer.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, 6(2):243–278.

Coon, J. (2016). Mayan morphosyntax. *Language and Linguistics Compass*, 10(10):515–550.

Daland, R. (2015). Long words in maximum entropy phonotactic grammars. *Phonology*, 32(3):353–383.

De Jong, N. H. and Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, 41(2):385–390.

Demberg, V., Sayeed, A. B., Gorinski, P. J., and Engonopoulos, N. (2012). Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 356–367. Association for Computational Linguistics.

DiCanio, C., Nam, H., Whalen, D., Bunnell, H. T., Amith, J. D., and Castillo García, R. (2013). Using automatic alignment to analyze endangered language data: testing the viability of untrained alignment. *The Journal of the Acoustical Society of America*, 134(3):2235–2246.

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, 64(4):805–855.

England, N. (1991). Changes in basic word order in Mayan languages. *International Journal of American Linguistics*, 57(4):446–486.

England, N. (2003). Mayan language revival and revitalization politics: Linguists and linguistic ideologies. *American Anthropologist*, 105(4):733–743.

England, N. and Martin, L. (2003). Issues in the comparative argument structure analysis in Mayan narratives. In Du Bois, J. W., Kumpf, L. E., and Ashby, W. J., editors, *Preferred Argument Structure: Grammar as Architecture for Function*, pages 131–157. John Benjamins, Amsterdam.

Fischer, E. and Brown, R. M., editors (1996). *Maya cultural activism in Guatemala.* University of Texas Press, Austin, TX.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and speech*, 31(4):307–319.

Fowler, C. A. and Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26(5):489–504.

Fox Tree, Jean E.and Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62(2):151–167.

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

García Matzar, P. O., Toj Cotzajay, V., and Coc Tuiz, D. (1999). *Gramática del idioma Kaqchikel*. Proyecto Lingüístico Francisco Marroquín, Antigua, Guatemala.

Garrett, A. and Johnson, K. (2013). Phonetic bias in sound change. pages 51–97. Oxford University Press, Oxford, UK.

Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, volume 1, pages 517–520. IEEE.

Goldrick, M. and Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, 107(3):1155–1164.

Gorman, K., Howell, J., and Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, 39(3):192–193.

Gorman, K. and Johnson, D. E. (2013). Quantitative analysis. In Bayley, R., Cameron, R., and Lucas, C., editors, *The Oxford Handbook of Sociolinguistics*, pages 214–240. Oxford University Press, Oxford, UK.

Grafmiller, J. (2014). Variation in English genitives across modality and genres. *English Language and Linguistics*, 18(3):471496.

Gregory, M. and Michaelis, L. (2001). Topicalization and left-dislocation: A functional opposition revisited. *Journal of pragmatics*, 33(11):1665–1706.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Proceedings of the Chicago Linguistic Society (CLS 35)*, volume 35, pages 151–166, Chicago, IL.

Haegeman, L. (1987). Complement ellipsis in English: Or, how to cook without objects. In Simon-Vandenbergen, A., editor, *Studies in honour of René Derolez*, pages 248–261. Seminarie voor Engelse en Oud-Germaanse Taalkunde R.U.G., Ghent.

Haegeman, L. and Ihsane, T. (1999). Subject ellipsis in embedded clauses in English. *English Language and Linguistics*, 3(1):117–145.

Haegeman, L. and Ihsane, T. (2001). Adult null subjects in the non-pro-drop languages: Two diary dialects. *Language acquisition*, 9(4):329–346.

Hanique, I. and Ernestus, M. (2011). Final /t/ reduction in dutch past-participles: The role of word predictability and morphological decomposability. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)*, pages 2849–2852, Florence, Italy.

Hanique, I., Schuppler, B., and Ernestus, M. (2010). Morphological and predictability effects on schwa reduction : The case of dutch word-initial syllables. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010)*, pages 933–936, Makuhari, Japan.

Hansson, G. (2010). *Consonant harmony: long-distance interactions in phonology.* University of California Press, Berkeley, CA.

Harris, Z. (1951). *Methods in structural linguistics.* University of Chicago Press, Chicago.

Hawkins, S. and Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics.*

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

Hay, J. B. and Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in cognitive sciences*, 9(7):342–348.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440.

Henderson, R. (2016). Mayan semantics. *Language and Linguistics Compass*, 10(10):551–588.

Holbrock, M. (2016). *Mayan Literacy Reinvention in Guatemala*. University of New Mexico Press, Albuquerque.

Hsu, B.-J. (2009). MIT Language Modeling Toolkit 0.4.1. (accessed June 1, 2015).

Huddleston, R. D. and Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, UK.

Hyman, L. (2002). Is there a right-to-left bias in vowel harmony? In Rennison, J., Neubarth, F., and Pochtrager, M., editors, *Phonologica 2002*. Mouton, Berlin.

Ishihara, C. F. S., editor (2016). *The Oxford Handbook of Information Structure*. Oxford University Press, Oxford, UK.

Jaeger, F. (2009a). HLP Jaeger lab blog: Centering several variables. https://hlplab.wordpress.com/2009/04/27/centering-several-variables/ (April 27).

Jaeger, F. (2009b). HLP Jaeger lab blog: Some R code to understand the difference between treatment and sum (ANOVA-style) coding. https://hlplab.wordpress.com/2009/12/18/some-r-code-to-understand-the-difference-between-treatment-and-sum-anova-style-coding/ (December 18).

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4):434–446.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language*, 45:229–254.

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

Kaufman, T. (1990). Algunos rasgos estructurales de los idiomas mayances con referencia especial al K'iche'. In England, N. and Elliott, S., editors, *Lecturas sobre la lingüística maya*, pages 59–114. Centro de Investigaciones Regionales de Mesoamérica, Antigua, Guatemala.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221.

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

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

Lehnert-LeHouillier, H., McDonough, J., and McAleavey, S. (2010). Prosodic strengthening in American English domain-initial vowels. In *Proceedings of Speech Prosody 2010*. International Speech Communication Association (ISCA). Available online at http://speechprosody2010.illinois.edu/papers/100082.pdf.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). Multiple perspectives on word production. *Behavioral and Brain Sciences*, 22(01):61–69.

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

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

Luce, P. A. (1986). *Neighborhoods of Words in the Mental Lexicon.* PhD thesis, Department of Psychology, Indiana University, Bloomington, Indiana.

Manuel, S. (1999). Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In Hardcastle, W. and Hewlett, N., editors, *Coarticulation: theory, data, and techniques*, pages 179–198. Cambridge University Press, Cambridge, UK.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R. H., and Bates, D. (2015). Balancing Type I error and power in linear mixed models. *arXiv preprint arXiv:1511.01864*.

Maxwell, J. (2009). Stylistics of the second person singular independent pronoun in Kaqchikel. In Avelino, H., Coon, J., and Norcliffe, E., editors, *New Perspectives in Mayan Linguistics*, volume 59, pages 374–391. MIT Working Papers in Linguistics, Cambridge, MA.

Maxwell, J. and Hill, R. (2010). *Kaqchikel chronicles: the definitive edition.* University of Texas Press, Austin, TX.

Menzerath, P. and de Oleza, J. M. (1928). *Spanische lautdauer: eine experimentelle Untersuchung, mit 4 Abbildungen. 15 Figuren und 37 Tabellen.* W. de Gruyter & Company.

Michaelis, L. and Francis, H. (2007). Lexical subjects and the conflation strategy. pages 19–48. John Benjamins, Amsterdam.

Miller, J. (2008). The handbook of english linguistics. pages 670–691. Wiley-Blackwell, Malden, MA.

Nariyama, S. (2004). Subject ellipsis in English. *Journal of Pragmatics*, 36(2):237–264.

Nerbonne, J., Heeringa, W., and Kleiweg, P. (1999). Edit distance and dialect proximity. In Sankoff, D. and Kruskal, J., editors, *Time Warps, String Edits and Macromolecules: The theory and practice of sequence comparison*. CSLI Publications, Stanford, CA.

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

Pierrehumbert, J. (2002). Word-specific phonetics. In Gussenhoven, C. and Warner, N., editors, *Papers in Laboratory Phonology VII*, pages 101–139. Mouton de Gruyter, Berlin.

Plag, I. (2003). *Word-formation in English*. Cambridge University Press, Cambridge.

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

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken dutch. *The Journal of the Acoustical Society of America*, 118(4):2561–2569.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raymond, W., Dautricourt, R., and Hume, E. (2006). Word-internal t/d deletion in spontaneous speech: The effects of lexical, phonological, and extra-linguistic factors. *Language Variation and Change*, 18(1):55–97.

Reichenbach, H. (1947). Elements of symbolic logic.

Richards, M. (2003). *Atlas lingüístico de Guatemala*. Instituto de Lingüístico y Educación de la Universidad Rafael Landívar.

Robertson, J. S. (1992). *The history of tense/aspect/mood/voice in the Mayan verbal complex*. University of Texas Press, Austin, TX.

Roland, D., Dick, F., and Elman, J. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of memory and language*, 57(3):348–379.

Rosenbach, A. (2014). English genitive variation – the state of the art. *English Language and Linguistics*, 18(2):215262.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47(1):172–196.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current directions in psychological science*, 12(4):110–114.

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

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

Shattuck-Hufnagel, S. and Turk, A. (1998). The domain of phrase-final lengthening in english. In *Proceedings 16th Int. Cong. on Acoust. and 135th Meeting Acoust. Soc. Am*, volume 2, pages 1235–1236.

Speyer, A. (2010). *Topicalization and Stress Clash Avoidance in the History of English*. Mouton de Gruyter, Berlin.

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

Stemberger, J. P. (2004). Neighbourhood effects on error rates in speech production. *Brain and Language*, 90(1):413–422.

Tily, H. and Kuperman, V. (2012). Rational phonological lengthening in spoken dutch. *The Journal of the Acoustical Society of America*, 132(6):3935–3940.

Torreira, F. and Ernestus, M. (2009). Probabilistic effects on french [t] duration. In *10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, pages 448–451. Causal Productions Pty Ltd.

Turk, A., Nakai, S., and Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In Sudhoff, S., Lenertova, D., Pappert, R. M. S., Augurzky, P., Mleinek, I., Nicole Richter, N., and Schließer, J., editors, *Methods in Empirical Prosody Research*, pages 1–28. De Gruyter, Berlin.

Turk, A. and Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, 28(4):397–440.

van Son, R. J. J. H., Bolotova, O., Lennes, M., and Pols, L. C. W. (2004). Frequency effects on vowel reduction in three typologically different languages (Dutch, Finish, Russian). In *Proceedings of the 8th International conference on spoken language processing (Interspeech 2004)*, pages 1277–1280, Jeju Island, South Korea.

van Son, R. J. J. H. and Pols, L. C. W. (2003). How efficient is speech? In *Proceedings (Instituut voor Fonetische Wetenschappen, Universiteit van Amsterdam)*, volume 25, pages 171–184, Amsterdam, The Netherlands.

van Son, R. J. J. H. and van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47(1):100–123.

Vázquez Álvarez, J. J. and Zavala Maldonado, R. (2014). La estructura argumental preferida en el chol, una lengua agentiva. In *Conference on Indigenous Languages of Latin America (CILLA) VI*, Austin, Texas. The Center for Indigenous Languages of Latin America (CILLA) at the University of Texas at Austin. Available online at http://www-ailla.lib.utexas.edu/site/cilla6/Vazquez_Zavala_CILLA_VI.pdf.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech*, 40(3):211–228.

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

Vitevitch, M. S., Armbrüster, J., and Chu, S. (2004). Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):514.

Vitevitch, M. S. and Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics*, 2:75–94.

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

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

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

Wissmann, M., Toutenburg, H., and Shalabh (2007). Role of categorical variables in multicollinearity in the linear regression. Technical Report 008, Department of Statistics, University of Munich, Munich, Germany.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition*, 7(6):411–419.

Yao, Y. (2011). *The effects of phonological neighborhoods on pronunciation variation in conversational speech*. PhD thesis, University of California, Berkeley.

Zipf, G. K. (1949). Human behavior and the principle of least effort. *An Introduction to Human Ecology*.