# Constituent durations in English NNN compounds: A case of strategic speaker behavior?

Annika Schebesta[1], Gero Kunter

**Abstract**

This paper investigates the effect of morphological embeddedness and lexical frequency on the duration of constituents in left- and right-branching NNN compounds from a corpus of spoken English (Boston University Radio Speech Corpus, Ostendorf et al. 1997). Theories assuming that the phonetic signal is not affected by the internal structure of multimorphemic words are opposed by empirical studies on the morpho-phonetic interface which provide evidence that the phonetic signal is sensitive to different morphological boundaries. The analysis of 465 NNN compounds reveals that morphological embeddedness alone does not have the expected effect on constituent durations, however, we detected a complex interplay of the morphological structure of NNN compounds and the two involved bigram frequencies. For instance, the duration of N2 in left-branching compounds is affected by the frequency of N2N3 even though these two constituents do not form a morphological unit in this type of NNN compound. This interplay may be interpreted as a listener-oriented strategy employed by the speaker in order to resolve potential conflicts between the frequency of adjacent constituents and the morphological structure: In such an instance, speakers appear to use acoustic duration to signal the branching direction of the triconstituent compound.

***Keywords:*** English compounds, phonetic reduction, morphological embeddedness, lexical frequency

## 1. Introduction

Categorical approaches to the formation of multimorphemic words frequently propose the strict separation of morpho-phonological operations from the actual articulation of the words. According to morpho-phonological theories such as Lexical Phonology (Kiparsky 1982a,b) and its descendant Stratal Phonology (Bermúdez-Otero 2012, Kiparsky 2015, Bermúdez-Otero 2018) as well as the language planning model proposed by Levelt et al. (1999), the phonetic signal of morphologically complex words does not reflect their internal organization, because phonetics cannot access the morphological processes and phonological alternations that the complex word underwent. Consequently, such frameworks suggest that the word-internal morphological boundaries are invisible to the phonetic signal.

In contrast, a huge body of research provides empirical evidence that the phonetics of morphologically complex words reflect their internal structure: The duration of stems (Lehiste 1972, Blazej and Cohen-Goldberg 2015), affixes and affix-like units (Hay 2007, Smith et al. 2012), and segments

[1]corresponding author: Annika.Schebesta@uni-siegen.de, Universität Siegen, Seminar für Anglistik, Adolf-Reichwein-Str. 2, 57068 Siegen, Germany

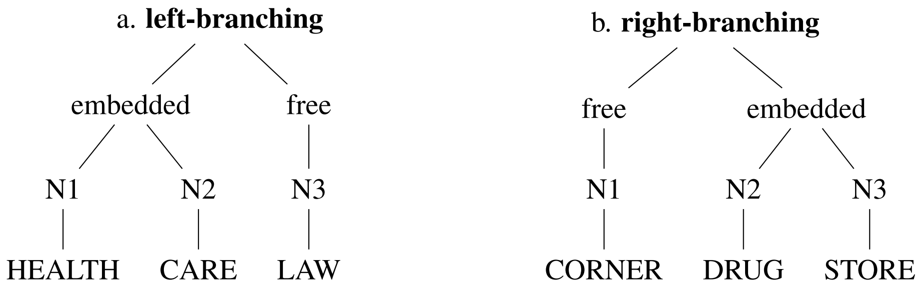
(Sproat and Fujimura 1993, Lee-Kim et al. 2013, Ben Hedia and Plag 2017, Plag et al. 2017) varies depending on the presence or absence of a morphological boundary as well as depending on the type of morphological boundary.

Such empirical findings suggest that different types of morphological boundaries affect the phonetic signal in different ways and, thus, seem to vary in boundary strength. In fact, Lexical Phonology (Kiparsky 1982a,b) adopts the binary distinction of weak and strong morphological boundaries put forward by Chomsky and Halle (1968). As laid out in Lexical Phonology, derivational affixes that attach to weak boundaries are grouped together under the label of Level I affixes, while the remaining derivational affixes attach to strong boundaries and constitute the group of Level II affixes. Likewise, since compounding also takes place on Level II in the theoretical framework, compound-internal boundaries are considered strong boundaries. According to Lexical Phonology, the order of word-formation processes is, thus, regulated by the Level I-Level II classification of derivational affixation and compounding.

Empirical studies provide evidence that the classification of morphological boundaries as either weak '+' boundaries (Level I affixation) or strong '#' boundaries (Level II affixation and compounding) as well as the fixed order of word-formation suggested in Kiparsky (1982a,b) is an inadequate description of word-formation processes. As reported in, for instance, Hay and Plag (2004), Plag and Baayen (2009) and Zirkel (2010), different affixes from one and the same group as well as across groups have different degrees of morphological boundary strength. The studies propose that the degree of morphological boundary strength is gradient across as well as within Level I and Level II. Similarly, Baayen (2010) shows that compound constituents are not freely combined, but instead, the order of constituents in a compound seems to be restricted by a particular preferred order. If the constraints on combinability of suffixes can be interpreted as differences in morphological boundary strength (Hay and Plag 2004, Plag and Baayen 2009, Zirkel 2010), and if compounds show similar constraints on the combinability of constituents (Baayen 2010), it follows that compounds may also show varying degrees of morphological boundary strength, similar to affixed words.

To our knowledge, Kunter and Plag (2016) is the only study that addresses the effect of morphological boundary strength on NNN constituent durations by taking into account that the compound-internal boundaries have different degrees of boundary strength. Due to the binary structure of compounds (see Bauer 2017, ch. 3), NNN compounds may be analysed as the result of a word-formation process involving three nominal lexemes as the constituents. These constituents undergo a recursive, two-level application of composition. At the earlier level, two constituents are combined to form a nominal compound NN, the 'embedded compound' that consists of two nominal 'embedded constituents'. Then, at the later level, the embedded compound is combined with the remaining, so-called 'free' constituent to form a triconstituent NNN compound. On each of these levels, a word-internal compound boundary is added to the morphologically complex word. The position of the embedded compound within the NNN compound determines the branching direction, as indicated in (1). Thus, [*health care*] *law* is a left-branching NNN compound because the embedded compound is in the left position, whereas *corner* [*drug store*] is a right-branching NNN compound.

(1) Branching direction in NNN compounds




While theoretical frameworks such as Lexical Phonology do not predict an effect of morphological boundaries on the phonetic signal at all, Kunter and Plag (2016) suggest that morphological embeddedness in NNN compounds coincides with different degrees of morphological boundary strength insofar as the boundary between embedded constituents is weaker than the boundary between the embedded compound and the free constituent. Their findings agree with the prediction that free constituents are overall longer than embedded constituents. An additional result of their analysis is that the two bigram frequencies (i.e. the frequency of the bigram N1-N2 and the frequency of the bigram N2-N3) also have a systematic effect on the acoustic duration. While the authors do not offer a consistent explanation for these effects, the fact that the acoustic realization of the NNN compound is affected by the frequency of co-occurrence even of parts that are structurally unconnected, such as N1-N2 in a right-branching compound or N2-N3 in a left-branching compound, is a serious challenge to theoretical approaches such as Lexical Phonology in which compounding is regarded a recursive process that does not really allow for frequency effects between constituents that do not belong to the same structural level. However, the analysis from Kunter and Plag (2016) is restricted by several properties of their data set, and hence the generalizability of their result is somewhat limited. The aim of the present paper is to build on Kunter and Plag (2016) by replicating and extending this previous research with another data set which is not limited in the same way. One of our primary interests is to add to the current understanding of the effect of frequencies in complex morphological structures as summarized in section 2.

The remainder of this paper is structured as follows. Section 2 discusses the interplay of morpho-phonological processes and the phonetic signal from a theoretical perspective and from the angle of empirical evidence. Section 3 presents a detailed survey of linguistic correlates of constituent duration. In section 4, we describe the methodology that we used to investigate constituent durations, including the description of the data set and the type of analysis we performed. The findings of this analysis are presented in section 5 followed by a discussion and concluding comments in section 6, which put the results into a wider perspective.

## 2. Does the phonetic signal of multimorphemic words reveal its internal structure?

The endeavor of this paper is to investigate whether, and if so, how morphological boundary strength affects the duration of compound constituents. While the presence of weaker and stronger morphological boundaries appears to be rather uncontroversial, it is less clear whether and, if so, to what extent these differences affect the articulation of morphologically complex words. Lexical Phonology, as laid out e.g. in Kiparsky (1982a,b), adopts the distinction between '+' and '#'

boundaries introduced in Chomsky and Halle (1968) and formalizes the interplay of morphology and phonology in a layered organization of the lexicon, i.e. in a sequence of levels that a lexical unit goes through. Each of the levels in the lexicon is associated with a particular set of word-formation processes and phonological rules which apply to the lexical entry that was previously activated for production. The order of morphological processes is strictly regulated, and a lexical unit cannot freely move back and forth between the levels. On the first lexical level, the affixation of the so-called Level I affixes at '+' boundaries takes place. This type of boundary allows for a high degree of phonological integration, i.e. bases are prone to phonological alternations such as stress-shift or resyllabification across the morpheme boundary. According to Kiparsky (1982a,b), the second level is associated with derivational affixation of Level II affixes and compounding, and on the third level, inflectional affixation takes place. On both the second and the third level, the lexical units attach to '#' boundaries which allow for less phonological integration: For instance, Level II affixes do not shift stress in the base and are, thus, considered stress-neutral.

In this theoretical framework, the lexicon, together with the morphological and phonological alternations that the lexical units undergo, is strictly separated from the so-called post-lexicon. This separation manifests itself in the recurrent application of Bracketing Erasure before the lexical unit enters the post-lexical stage, which eliminates any trace of morphological structure and phonological alternations from the lexical unit. That is, post-lexical rules shape the phonetic signal of a lexical unit and prepare its embedding in a syntactic string without any access to the internal structure of the lexical unit. Thus, Lexical Phonology assumes the existence of different types of morphological boundaries, namely weak '+' and strong '#' boundaries, on the lexical stage, but proposes that the phonetic signal does not reflect these differences that result from a sequence of word-formation processes.

Classical Lexical Phonology as described here has been adapted and extended more recent versions such as Stratal Phonology or Stratal Optimality Theory (Bermúdez-Otero 2012, Kiparsky 2015, Bermúdez-Otero 2018). These developments of Lexical Phonology propose that the order of morphological and phonological operations in the lexicon are more variable, opposing the strict separation of lexical levels outlined in the classical version. Stratal Phonology, however, still adheres to the separation of the lexical and the post-lexical stage, postulating that the morphological structure and phonological alternations in a multimorphemic word are inaccessible to the phonetic signal. In consequence, according to Lexical Phonology and its descendants, the morphological structure of these complex words should not be traceable in the post-lexical encoding, which includes, for instance, phonetic reduction.

Likewise, formal models of language planning such as the model proposed by Levelt et al. (1999) suggest to strictly separate the preparation of a lexical unit for production from its actual articulation (roughly corresponding to the post-lexicon component in Lexical Phonology). According to Levelt et al. (1999), the preparation stage is organized in two distinct systems, namely the conceptual system and the articulatory motor system. The conceptual system retrieves the required entries from the mental lexicon and activates them for speech production (a component that does not have an equivalent in Lexical Phonology). As soon as the lexical entry is activated, the articulatory motor system regulates the encoding of morphological, phonological and phonetic features of the anticipated morphologically complex word: The involved morphemes are combined and processed, the phonological segments of each of the morphemes are retrieved, and the resulting complex word is resyllabified. In the last step of the planning, the gestural score, that is, a set of gestures needed for the articulation of the complex word, is created. After the planning is complete, the complex word is articulated. The processes assigned to the post-lexical stage in Lexical Phonology are located at this final articulation step in the model described in Levelt et al. (1999).

These summaries show that both Lexical Phonology and the Levelt model agree when it comes to the nature of the output: in neither approach should traces of the internal morphological structure or of the phonological alterations that have been applied to the involved elements be visible in the phonetic encoding of the morphologically complex word. That is, morphological boundaries are invisible to the phonetic signal according to both theories.

In contrast, empirical studies provide evidence that morphological boundary strength is gradient, and that these different degrees of morphological boundary strength seem to find their reflection, *inter alia*, in the phonetic signal. Considering the gradience of morphological boundary strength, empirical evidence is provided by Hay and Plag (2004), Plag and Baayen (2009) and Zirkel (2010) who find that the combinability of English affixes is regulated along a scale of morphological boundary strength. Hay and Plag (2004), for instance, show that the combination of the Level II suffixes *-hood* and *-less* such as in *childhoodless* is attested and, therefore, a possible suffix combination. However, the reverse order, that is, base + *-less* + *-hood*, is unattested and considered an impossible suffix combination. According to Hay and Plag (2004), this is evidence that the Level II suffixes *-hood* and *-less* differ in their morphological boundary strength, contradicting the binary distinction of weak and strong morphological boundaries as put forward by Lexical Phonology. Hay and Plag (2004) and Plag and Baayen (2009) demonstrate that the cline of suffix boundary strengths affects the range of attested suffix combinations in the case of multiple affixation, thus forming an acyclic directed graph: While suffixes with stronger morphological boundaries can be attached to bases that contain a suffix with a weaker morphological boundary, the reverse case is rarely, if ever, attested. It seems that the association of an affix with the group of Level I or Level II affixes alone is not necessarily informative of morphological boundary strength. Baayen (2010) finds an equivalent constraint on the combinability of compound constituents. Apparently, there is a preferred order, that is, a particular constituent may easily occur after another constituent, while the reversed order of the two constituents is dispreferred. Considering such findings, it is plausible to expect that internal boundaries within complex words which consist of more than two morphemes have different degrees of boundary strength.

The presence or absence of morphological boundaries, but also the degree of boundary strength has been shown to affect the duration of, for instance, words, morphemes and segments, which is explicitly ruled out by the theoretical frameworks discussed above. For instance, Lehiste (1972) finds that stem durations in English vary depending on the following morphological boundary: In cases where the stem is conjoined with a suffix (e.g. *shaking*) its duration is shorter than in cases where the stem forms a compound with another lexical unit (e.g. *shake-down*). Such stem duration differences can be perceived by speakers, as Blazej and Cohen-Goldberg (2015) illustrate in a series of forced-choice and mouse-tracking experiments with English native speakers. With regard to segment durations, Sproat and Fujimura (1993) and Lee-Kim et al. (2013) examined the acoustic and articulatory properties of the phoneme combination /il/ across different types of boundaries in English, ranging from non-morphemic boundaries over suffix boundaries to compound boundaries, and eventually word boundaries. Both studies indicate that different boundary strengths lead to gradient realizations of the segments in focus. In particular, with increasing boundary strength, the degree of velarization of /l/ also increases, while the duration of the rime decreases.

A considerable number of studies does not only provide evidence for an effect of morphological boundary strength on the acoustic signal, but they also illustrate that presumably homophonous units differ phonetically. For instance, Hay (2007) and Smith et al. (2012) investigated English prefixes and non-morphemic, phonologically identical syllables and report that the prefixal sequences are systematically longer than the non-morphemic sequences. Ben Hedia and Plag (2017) present evidence that the English locative prefix *in-* and the negative prefix *in-* differ systematically in the duration of the nasal, thus suggesting that the two prefixes held to be homophonous e.g. in Lexical Phonology are indeed realized in different ways. Besides, Smith et al. (2012) analysed the formant frequencies of the sequence-central vowel and find that the vowels in the prefixes are less reduced than those in the non-morphemic sequences. Similarly, Plag et al. (2017) show that the duration of word-final /s/ varies according to the morphological boundary it occurs at. Contrary to Hay (2007) and Smith et al. (2012), however, their results indicate that non-morphemic /s/ is significantly longer than morphemic /s/.

While previous investigations of the interface of morphological structure and the phonetic signal have to a large extent focused on affixed words, only little is known about the morpho-phonetic interface in compounds. Only a few studies, such as Bell et al. (2020) for NN compounds and Kunter and Plag (2016) for NNN compounds, investigate the impact of compound boundaries on the phonetics of compound constituents. On the basis of experimental data, Bell et al. (2020) test the hypothesis that the degree of boundary strength in English NN compounds affects the duration of boundary-adjacent consonants. The underlying assumption is that less segmentable lexical units have a weak morphological boundary while more segmentable lexical units are set at a strong morphological boundary. Bell et al. (2020) predict that less segmentable NN compounds with a weak compound boundary show shorter consonant durations than more segmentable NN compounds with a strong compound boundary. However, their analysis does not reveal an effect of morphological boundary strength on segment duration.

Building on the observation that different degrees of morphological boundary strength in affixed words lead to different acoustic realizations, Kunter and Plag (2016) test the prediction that in English NNN compounds, the embedded compound with a weaker morphological boundary shows more phonetic reduction than the free constituent at the stronger morphological boundary. That is, in a left-branching compound [N1 N2] N3, the morphological boundary between the embedded constituents N1 and N2 is assumed to be weaker than the morphological boundary between the embedded constituent N2 and the free constituent N3. Simultaneously, a right-branching compound N1 [N2 N3] is considered to have a weaker morphological boundary between the embedded constituents N2 and N3, and a stronger morphological boundary between the free constituent N1 and the embedded constituent N2. Using the experimental data from Kösling et al. (2013), who analyse prominence patterns in NNN compounds, Kunter and Plag (2016) find that constituent durations systematically differ across branching direction. Yet, the statistical analysis incorporates an interaction of branching direction and bigram frequency which reveals a substantial effect on the durations of the compound constituents, and it remains unclear how much of the variation can be explained by branching direction alone.

In addition, a considerable number of studies investigated the effect of lexical frequency on the phonetic signal. For instance, Wright (1979), Pluymaekers et al. (2005b), Aylett and Turk (2006), and Bell et al. (2009) report that words with higher lexical frequencies are typically phonetically shorter than corresponding low-frequency words. The analysis of lexical frequency effects on homophonous words in Gahl (2008) reveals that frequent homophones are produced shorter than their infrequent counterparts. Beyond single words, Arnon and Cohen Priva (2013) find that the frequency of strings of words occurring together can affect the acoustic duration in speech production, as well. Their results from corpus data and experimental data reveal that higher n-gram frequencies lead to shorter word durations. This is a finding that is also supported by the results reported in Kunter and Plag (2016) for English NNN compounds. As described in more detail below, they find a complex interplay between the constituent durations and the bigram frequencies in their data. Summing up, there is strong empirical evidence that high-frequency words – or strings of words – have comparatively shorter durations than low-frequency words or strings of words. With regard to NNN compounds specifically, there are three different types of lexical frequency that are of interest: (a) the constituent frequency, i.e. the frequency of each constituent alone, (b) the bigram frequency of two neighboring compound constituents, and (c) the trigram frequency, i.e. the frequency of the whole compound.

The bigram frequencies are particularly interesting as they may be interpreted as a measure of the representation of the involved constituent as a lexical unit in the mental lexicon. Generally speaking, an N-N combination with a high bigram frequency such as *credit card* is more likely to be perceived and processed as one complex lexeme than a combination with a low bigram frequency such as *credit chain*, even if the latter may easily occur as a compound in a passage such as *Still, you could experience billing and credit errors - and potentially harm your credit record - if any of the links in the credit chain succumbs to a Y2K glitch* (from COCA: MAG 1999, Davies 2008). It seems reasonable to assume that the interpretation of the internal structure of a compound is affected by these representations. If the bigram frequency is particularly high for one N-N combination, this combination may be interpreted as the embedded compound, potentially overriding the branching based on semantics. However, there is no empirical evidence as yet for the use of bigram frequencies as cues to branching in NNN compounds. We will discuss below if the present study can at least partially fill this gap.

Considering the growing body of empirical studies that challenge categorical approaches such as Lexical Phonology (Kiparsky 1982a,b) and the language planning model by Levelt et al. (1999), and considering the somewhat contradictory and inconclusive findings from studies of the phonetic signal in compounds, we will provide more empirical evidence that tests the predictions made by theory. On the basis of corpus data, we will investigate the impact of morphological boundary strength on constituent durations in English NNN compounds. We decided to use the data set from Kösling and Plag (2009) that was originally generated from the Boston University Radio Speech Corpus

(Ostendorf et al. 1997). Since this study has already yielded systematic differences in the phonetic signal with regard to prominence patterns in NNN compounds using this data set, it is reasonable to expect that the data are suitable to investigate another phonetic aspect, namely constituent durations.

We will extend the investigation in Kunter and Plag (2016) by considering not only morphological embeddedness as a potential determinant of constituent durations, but also the representation of neighboring constituents in the mental lexicon as reflected in their bigram frequencies. Our study investigates the impact of the morphological structure on the constituent durations in NNN compounds, and it examines closely the complex interaction between branching direction and the involved lexical frequencies that has been described in Kunter and Plag (2016). Doing so, we address the following questions: Can we detect systematic differences in the phonetic signal that are related to differences in morphological boundary strength? What is the impact of branching direction and lexical frequency on the phonetic signal? Can categorical approaches to speech production, such as Lexical Phonology (Kiparsky 1982a,b) or the language planning model by Levelt et al. (1999) account for our findings?

## 3. Correlates of constituent durations

This section provides an overview of factors that, in addition to morphological and lexical frequency, have been shown in the literature to have an impact on the duration of words in general, and of compound constituents in particular. As the literature review will show, the bulk of previous research has focused on English, but there is evidence that similar patterns can be observed in other, related languages such as Dutch and German as well, e.g. in Pluymaekers et al. (2005b) and Bergmann (2018), respectively. These studies suggest that the systematic effects on the durational patterns of morphological units in complex English reported in the literature may act on a more universal level. One variable that can almost certainly be expected to show an effect on the duration of spoken words regardless of the language being spoken is an extra-linguistic variable: the overall speech rate of the speaker. Speech rate has been found to affect vowel duration (Johnson et al. 1993, Smith 2002) and segment reduction (Raymond et al. 2006) in such a way that a higher speech rate leads to an overall shorter duration of utterances, while a lower speech rate results in longer durations. Speech rate is usually defined as the number of linguistic elements per time interval, for example words, syllables, or segments per second. Different speakers will have different speech rates, and the speech rate of a given speaker is variable as well (Uhmann 1992). Thus, the same word will have different durations for different speakers, and durational differences due to intra-speaker variation is also possible for the same word, if speakers vary their speech rate.

The number of phonemes is a fairly obvious phonological factor determining word duration: other things being equal, a word that contains more phonemes will be longer than a word with fewer phonemes. In addition, the average duration of the phonemes interacts with the number of syllables in a word. In their review of the effect of word boundaries on duration, Turk (2010) illustrates that polysyllablic shortening, a term coined by Lehiste (1972), is a frequent phenomenon. In polysyllabic shortening, the duration of a base word (e.g. *shake*) decreases when additional syllables are added to it, either by suffixation (*shake+ing*) or in a compound (*shakedown*).

The number of phonologically similar words is another measure that is known to affect word durations. Gahl et al. (2012) show that the phonological neighborhood density, i.e. the number of words in the lexicon of a speaker that differ from the target word in only one phoneme, affects both word recognition and speech production. They find that target words in a dense neighborhood with a high number of phonological neighbors have shorter durations than target words in a sparse neighborhood. This speaks in favor of a training effect: lexemes with many phonological neighbors are, in Gahl et al. (2012)'s opinion, easy targets for production, and are therefore subject to articulatory reduction that manifests itself in vowel centralization and a shortened duration of the target. They conclude that these mechanisms are ultimately speaker-centric even though this reduction is detrimental to the intelligibility of the target. In contrast, Wright (2004) and Munson and Solomon (2004) show that a high number of phonological neighbors leads to less phonetic reduction in a lexical unit as compared to a lexical unit with fewer phonological neighbors. Both studies argue that targets show less shortening when the identification of the target is already challenging due to a large number of phonological neighbors. However, speakers tend to reduce more in cases when the target need not be distinguished from a high number of phonologically similar lexical units. This interpretation is in line with Lindblom (1990)'s H & H theory which proposes that speakers adapt their articulation to balance the articulatory economy of the speaker with the discriminatory needs of the listener: words with only few phonological neighbors will be more prone to hypoarticulation and hence, acoustic reduction than words with many phonological neighbors where the intelligibility benefits from hyperarticulation.

This need to balance intelligibility and articulatory effort is also reflected in listener-oriented frameworks such as the principle of Uniform Information Density proposed in Jaeger (2010) and Jaeger and Tily (2011). This approach is based on the assumption that the information content of linguistic units is linked to the predictability of the unit given their context, and that this information content will have an effect on the organization of the units. With regard to neighborhood density, a word with relatively few neighbors is easily predictable based on few segments as there is only little competition. In information-theoretic terms, this high predictability is equated to low information content. In contrast, a word with many neighbors is more difficult to predict, and its information content is therefore higher. Crucially, the principle of Uniform Information Density predicts that in order to facilitate the processing of a linguistic signal, speakers will attempt to balance the information content in such a way that high information is spread over longer linguistic units (or a higher number of units). As a consequence, a word with many phonological neighbors (i.e. high information content), is expected to be pronounced longer than a similar word with few phonological neighbors (i.e. low information content). Thus, the Uniform Information Density principle is a listener-oriented explanation for the lengthening effects of increasing neighborhood densities described in Wright (2004) and Munson and Solomon (2004).

The articulation of words without phonological neighbors, so-called lexical hermits, cannot be accounted for with phonological neighborhood density. Instead, Suárez et al. (2011) suggest that the average phonological Levenshtein distance (PLD20) is more appropriate in such cases. This metric computes the mean number of segments that need to be changed in the target word (i.e. deleted, substituted or added) to yield the 20 nearest neighbors. Suárez et al. (2011) find that words with low PLD20 values have slow recognition times, as there is a high number of phonologically similar lexemes in the vicinity of the target word. Due to the low predictability of these words with many competitors surrounding them, a low PLD20 would also be an indication of a high information content. Therefore, the effect of PLD20 should be similar to that of phonological neighborhood density: a low PLD20 value should lead to longer duration than a high PLD20 value. However, as Suárez et al. (2011) investigated word recognition only, this effect has not been empirically tested yet for production.

The correlation of duration and accentuation has been examined extensively in previous research. The presence of a pitch accent on a syllable has a lengthening effect on the syllable and, by extension, on the word containing it (Turk and Sawusch 1996, Turk and White 1999, de Jong 2004, Kunter 2011, Morrill 2012). With regard to English compounds in particular, it has been argued in Kunter (2011) that the two possible prominence patterns in English NN compounds ('left-prominent' as in *watch maker* and 'right-prominent' as in *apple pie*, which corresponds to what has traditionally been called 'compound stress' and 'phrasal stress', respectively; see Liberman and Prince 1977, Giegerich 1999) are phonologically realized by different accentuation patterns. Accordingly, the first constituent in left-prominent compounds was found to be longer than the first constituent in right-prominent compounds in Kunter (2011), a finding similarly reported in Farnetani et al. (1988).

Kösling and Plag (2009) and Kösling et al. (2013) investigate prominence patterns in English NNN compounds from corpus data and experimentally elicited data. They test the Lexical Category Prominence Rule (LCPR) proposed by Liberman and Prince (1977) which predicts two different prominent patterns for NNN compounds, that is, one for left-branching and one for right-branching compounds. The LCPR relies on the assumption that English is a left-prominent language with a sequence of strong and weak syllables as the default. As a consequence, the embedded NN compound in a NNN compound is assumed to be left-prominent, resulting in a prominence pattern of [ŃN] N for left-branching and N [ŃN] for right-branching NNN compounds. In accordance with the findings from Kunter (2011) and Farnetani et al. (1988), Kösling and Plag (2009) and Kösling et al. (2013) reject the predictions made by the LCPR for a portion of NNN compounds: NNN compounds with a left-prominent embedded NN compound show the predicted patterns in both branching directions, whereas NNN compounds with a right-prominent embedded NN do not. Instead, the findings indicate that the prominence pattern of the embedded NN compound determines the prominence pattern of the NNN compound, resulting in two different prominence patterns for left-branching NNN compounds and two different prominence patterns for right-branching NNN compounds.

Prosodic boundaries have been shown to have an impact on the phonological structure and the phonetic signal of words (e.g. Selkirk 1980, Hall and Kleinhenz 1999, Raffelsiefen 1999). Prosodic phonology assumes different prosodic domains, which are assigned to units as small as the syllable and as large as the whole phonological utterance. Boundaries of prosodic domains higher than the phonological word, e.g. the intonational phrase, have been shown to affect the acoustic signal. Studies on different phenomena such as phrase- and word-final lengthening as well as initial strengthening have illustrated that prosody is predominant in durational variation (see Turk and Shattuck-Hufnagel (2000, 2007), White (2014) for overviews), and that investigations of this variation are insufficient without taking into account the potential impact of these higher-ranked prosodic domains. For example, the results in Turk and Shattuck-Hufnagel (2007) reveal that the duration of word-final syllables and, in certain conditions, main-stressed syllables is lengthened the closer the material is located to a prosodic phrase boundary. Similarly, Fougeron and Keating (1997) found lengthened durations in phrase-initial position, and Cho (2005) reports a stronger articulation of linguistic material at the beginning of prosodic phrases. In the case of NNN compounds, which form one word out of three lexemes, it can be assumed that the material closest to prosodic phrase boundaries, i.e. N1 in phrase-initial position and N3 in phrase-final position, is affected in its duration.

This overview implies that the constituent durations of NNN compounds may be affected by a range of factors. Section 4 provides an overview of the variables that are typically described in the literature to affect the durational patterns in morphologically complex structures, and which we therefore incorporated in the present analysis of NNN compounds. That section also explains how the values were obtained from the acoustic signal, from appropriate reference corpora and databases, or, in the case of some categorical variables, on which basis the items were classified.

## 4. Methodology

### *4.1. Data*

In order to test which factors affect constituent durations, we used the data set analysed by Kösling and Plag (2009) in their analysis of prominence assignments in English NNN compounds. The tokens used in that study were extracted from the Boston University Radio Speech Corpus (Ostendorf et al. 1997), a collection of speech recordings of professional radio news speakers (3 female and 4 male speakers). The data set includes 471 compounds of both branching directions, i.e. left-branching [N1 N2] N3 and right-branching N1 [N2 N3]. One advantage of the speech data from the Boston corpus is that this corpus offers recordings of high sound quality, and the speakers, for the most part, provide error-free and fluent speech, which facilitates the segmentation of the recordings. The tokens cover a wide range of different bigram frequencies for both branching directions, and are therefore better suited for investigating bigram frequencies and boundary strength as independent factors than the data used in Kunter and Plag (2016) (see section 2). The authors of Kösling and Plag (2009) provided us with the NNN tokens, the classification of the branching direction of each token as well as an acoustic annotation of the compound boundaries and the compound-internal constituent boundaries.

All NNN compounds meet the following criteria: none of the three constituents is a word from a language other than English (this excludes, for instance, *Hillside hacienda*); contains a possessive clitic (for instance, *tenant's right crisis*); is an initialism (for instance, *U.S. district judge*), or is the name of an individual (for instance, *Thomas Crown affair*). All compounds are unique in the data set: as discussed in Kösling and Plag (2009), only the first token of any NNN type was sampled, thus avoiding a situation in which particularly frequent NNN types might skew the data set. This sampling procedure may have an unexpected side effect: the distribution of NNN types across speakers may be skewed in such a way that the highly frequent compounds are most likely to occur in the data obtained from the first speaker that was sampled. We will return to this point in the discussion of trigram frequencies below.[2]

The branching direction of about 20 percent of the compounds was determined by Kösling and

[2]Thanks to one anonymous reviewer who emphasized this potential consequence of the sampling procedure.

Plag (2009) on the basis of the orthography used in the Boston corpus: If two constituents were spelled in the transcripts either as one word or were conjoined by a hyphen, these constituents were assumed to form the embedded constituent of the NNN compound. For example, in the NNN compound *weekend series*, the one-word spelling of the first constituents *week* and *end* was considered by the authors to indicate a left-branching compound [*weekend*] *series*, while the hyphen between *whistle* and *blower* was considered a signal of right-branching in *company* [*whistle-blower*]. While the authors acknowledge that there is a non-negligible amount of between-speaker variation as far as compound spelling is concerned, they argue that the existence of hyphenated or one-word spellings is already sufficient for a “more word-like status of that combination” (Kösling and Plag 2009, 201). The branching direction of the remaining 80 percent of tokens, in which the orthography of the transcripts did not suggest a branching direction, was determined by means of a semantic analysis by Kösling and Plag. Only those compounds were included in the data set where both authors unambiguously agreed in their assessment whether the interpretation of the compound suggested left- or right-branching.

The reliability of the acoustic annotations by Kösling and Plag (2009) was tested by the present authors by selecting a random subset of the whole data (about 10%) using the speech analysis software Praat (Boersma and Weenink 2017). For these tokens, the beginning and the end of the compound as well as the two compound-internal constituent boundaries were identified using the common segmentation criteria (see Ladefoged and Johnson 2011, Ladefoged and Ferrari Disner 2012). Only the annotation of the compound-final boundary, i.e. the right boundary of N3, showed a significant difference to the annotations provided by Kösling and Plag (possibly because in their study of the intonational contour, the duration of the last constituent was of secondary relevance). Thus, this boundary was re-annotated for the complete data set. All remaining boundaries (i.e. the compound-initial boundary as well as the boundaries between N1 and N2, and between N2 and N3), were retained from Kösling and Plag.

The random sample revealed an inconsistency in the original annotation of identical consonants at constituent boundaries, as, for instance, *Massachusetts state house*, where N1 ends in /s/ and N2 begins with /s/. We set the boundary between the first and the second sound in a doublet (85 items in total) on the basis of the acoustic signal in the spectrogram and the waveform. For instance, a decrease in the activated energy in the spectrogram between two /s/ consonants indicates the end of the previous and the beginning of the following fricative. In cases where this was not possible, we decided to set the boundary at the exact midpoint of the sound doublet.

We excluded those constituents from the Kösling and Plag (2009) data for which not all variables used in our statistical models as predictor variables were reliably available. For instance, if the constituent was not listed in the data base of the *English Lexicon Project* (henceforth ELP, Balota et al. 2007), lexical information such as the number of phonological neighbors or the PLD20 could not be obtained. In addition, as the model used pitch measurements to incorporate effects of intonation, we excluded those constituents in which Praat failed to determine a reliable pitch, or in which the pitch contour appeared to feature extraordinarily high pitch ranges (larger than 20 semitones) which we considered invalid outliers. In this way, 78 constituents of the original total of 1413 constituents from 471 NNN were eliminated (a reduction by 5.5 percent). The final data set comprises

| Containing NNN is… | F1 | F2 | F3 | M1 | M2 | M3 | M4 | TOTAL |
|---|---|---|---|---|---|---|---|---|
| … left-branching | 93 | 112 | 230 | 115 | 137 | 45 | 219 | 951 |
| … right-branching | 45 | 48 | 96 | 41 | 89 | 6 | 59 | 384 |
| TOTAL | 138 | 160 | 326 | 156 | 226 | 51 | 278 | 1335 |

Table 1: Number of constituents by speaker and branching direction of the NNN containing the constituent (female speakers: F1–F3, male speakers: M1–M4).

1335 constituents from 465 NNN compounds. The distribution of compound constituents in terms of branching direction and speakers is illustrated in Table 1.

*4.2. Dependent Variable and Predictor Variables*

The dependent variable DURATION was measured for the 1335 compound constituents as the time interval covered by each constituent. As described above, the boundaries were based on the acoustic annotations provided by Kösling and Plag (2009), with the exception of the final boundary that delimits the compound. This boundary was manually set by one of the authors of this study. The following variables are included as predictors in the subsequent model. While some of these variables are central to the research questions outlined above, we also cover variables that may be considered as 'noise' variables that explain as much of the variance of the dependent variable as possible. This will make it easier for the model to detect a 'signal', i.e. effects of morphological boundary strengths, branching direction, and lexical frequency on the constituent durations.

*Extra-linguistic predictor.* One speaker-related variable is the overall SPEECHRATE, a measure that expresses the pace at which a speaker produces speech. The speech rate is determined on the basis of each of the paragraph-sized recordings in the Boston corpus. Each of these recordings spans about 175 words on average, typically without pauses (but occasionally including breaths, which were treated as word-like units). SPEECHRATE was calculated by obtaining the total number of words and breaths in each recording containing the NNN compound, subtracting the number of words of the compound itself (which was 2 or 3, depending on the orthography of the compound), and dividing this number by the total duration of the recording minus the duration of the NNN compound. In this way, the speech rate is not affected by unusual productions of the compound itself.

*Lexical predictors.* Our analysis included three types of lexical frequency: the token frequency of constituents (FREQ), the token frequency of two neighboring constituents in the NNN compound (N1N2FREQ and N2N3FREQ), and the token frequency of the NNN compound as a whole (TRIGRAMFREQ). In order to determine the frequency measures, the DVD version of COCA (1990–2012, Davies 2008) was queried using the corpus query tool Coquery (Kunter 2017).

Previous research such as Gregory et al. (1999), Jurafsky et al. (2001), Bell et al. (2003) or Tremblay and Tucker (2011) point out that the degree of predictability of a word given the preceding context affects the duration of the word in production: other things being equal, a word that is highly predictable given the preceding word will have a shorter average duration than the same word in an unpredictive context (see Seyfarth (2014) for an overview of probabilistic reduction). Similarly, Pluymaekers et al. (2005a) use mutual information with the neighboring words in order to account for predictability effects on the phonetic signal, and report that stem durations are shorter when mutual information with the neighboring words is higher. Findings such as these are compatible with the principle of Uniform Information Density (Jaeger 2010, Jaeger and Tily 2011) introduced above: highly predictable words have a low information content and are therefore more prone to reduction than words with low predictability. While such an effect may be expected both for the whole NNN compound as well as for the individual constituents of the NNN, we decided against including in our model a predictor that represents the degree of predictability such as conditional probabilities or mutual information scores. There are both theoretical as well as practical reasons for this decision. On a structural level, there is a huge difference between the type of words preceding an NNN and the words used as constituents. If we calculate, for instance, the conditional probability of N1 given the preceding word, the structure of English noun phrases makes it highly likely that the preceding word is either a determiner or possibly an adjective. In the case of the conditional probability of N2 given N1, and of N3 given N2, the preceding word is by definition always a noun. Consequently, the conditional probability for N1 will express a very different lexical environment than the conditional probabilities for N2 and N3, and it is unclear if they are still comparable to each other. After all, the probability of one noun out of the class of all nouns in the English lexicon following a determiner like, for instance, *the* may be reasonably expected to be much smaller than the probability of one certain noun out of the class of all nouns in the English lexicon following another certain noun. On a practical note, calculating conditional probabilities or mutual information scores involves the bigram frequency of the target word and the preceding word(s), as well as the frequencies of the individual words. Yet, if either one of the words or the bigram is not attested in the reference corpus, the predictability measure cannot be calculated anymore, and the observation would have to be excluded from the data set. In the present case, this would amount to a considerable number of observations: about 20 percent of the unique bigrams occurring in the data set are not attested in COCA (Davies 2008) which is used as the reference corpus. For the present analysis, we have to acknowledge that the data set is not well-suited to represent the effects of predictability based on linguistic context.

However, we included two measures that view the information content of the involved linguistic units from a different perspective. As discussed above in section 3, the density of the phonological neighborhood and the average distance between entries in the lexicon can also be interpreted as measures of predictability and hence, of information content. We assume that, based on the principle of Uniform Information Density, a constituent with few competitors in the mental lexicon is highly predictable and therefore less informative than a constituent with many competitors. Thus, in line with the findings reported in Wright (2004) and Munson and Solomon (2004), a word with a low neighborhood density is expected to have a shorter duration than an equivalent word with a high neighborhood density. Similarly, words with a high average phonological Levenshtein distance (PLD20) may be expected to be relatively short, as their more isolated position within the mental lexicon makes them more predictable than equivalent words with lower PLD20 values and many competitors. We gathered phonological neighborhood density (PHONON) and the phonological Levenshtein distance (PLD20) of each constituent from the ELP in order to incorporate this potential effect of competition in our model.

*Phonological predictors.* The number of phonemes (NPHON) and syllables (NSYLL) for each constituent were also gathered from the ELP (Balota et al. 2007).

The Boston corpus contains prosodic annotations using the ToBI system, but only for a relatively small subset of texts. In order to incorporate the presence or absence of pitch accents in our data set, we used an approach similar to e.g. Seyfarth (2014) or Kunter (2017) and used pitch measurements to approximate the pitch contour in a numerical way. As discussed in detail in Kunter (2011, ch. 5), there are several different types of pitch-related measures that may be used to incorporate accentuation in a statistical analysis such as maximum pitch or pitch range. We obtained these pitch measures separately for each constituent from the speech recordings using Praat (Boersma and Weenink 2017) with floor and ceiling settings adjusted for the gender of the speakers (female speakers: 100 Hz and 500 Hz; male speakers: 75 Hz and 300 Hz, respectively) using a semitone scale (ST) relative to 100 Hz (e.g. a pitch measurement of 100 Hz corresponds to 0 ST, 50 Hz to -12 ST, and 200 Hz to +12 ST). While these pitch measures do not capture all details of the intonation of a speaker (for instance, it may be difficult to distinguish a high high pitch accent from a high boundary tone if only the maximum pitch is considered), previous studies suggest that such an approach is suitable to account for the lengthening effect of intonation on words and signals. In order to decide which pitch measure is most suitable to account for the effect of accentuation on the constituent durations, we fitted a set of preliminary models with different pitch measures included as predictors. An inspection of the AIC (Akaike 1974) revealed that PITCHRANGE, which was calculated as the difference between the highest and the lowest point of the pitch contour of a compound constituent, appears to be the most suitable variable. Constituents with a high pitch range are constituents with a great amount of movement in the pitch contour, i.e. constituents that can be expected to carry a pitch accent. A low pitch range indicates a relatively flat pitch contour; the corresponding constituent therefore is likely to be unaccented.

The lengthening effect that occurs at the end and, to a lesser extent, at the beginning of a prosodic phrase was addressed by coding each NNN compound on the basis to its position in the prosodic context. With regard to a phrase boundary that preceded the compound (PRECBOUNDARY), a compound was labeled 'clause-initial' if it occurred in clause-initial position, otherwise it was labeled 'phrase-medial'. Additionally, we coded for a preceding pause by adding '<pause>' to the label, and added '<no pause>' in case of a phrase boundary without an audible pause. Similarly, for the following phrase boundary FOLLBOUNDARY a compound was labeled 'NP-medial' if more linguistic material followed which belonged to the same phrase, otherwise it was labeled 'phrase-final'. We added '<pause>' to the label if there was an audible following pause present, while '<no pause>' indicates the absence of a following pause. In our data set NNN compounds with the labels 'clause-initial' and 'phrase-final' were always accompanied by a pause. Therefore, only the six combinations shown in Table 2 were used as classification in our data set. The rightmost three columns of the table indicate the distribution of the total number of constituents on the types of boundary as well as the distribution of constituents on the types of boundary by branching direction.

*Morphological predictors.* The variable MEMBER corresponds to the constituents in the compound, and has therefore the three levels N1, N2, and N3. By including this variable (and the interactions involving MEMBER, see below), it becomes possible to fit a single model that estimates the durations

| PrecBoundary | | N |
|---|---|---|
| clause-medial <no pause> | ... *at* [$_{\text{NP}}$ *the* ***Boston University school*** *of public health*]. | 1284 |
| clause-medial <pause> | ... [$_{\text{NP}}$ *Those* ***<pause> litmus test supporters***] *backing Weld...* | 18 |
| clause-initial <pause> | [$_{\text{NP}}$ ***<pause> Massachusetts school children***] *are improving ...* | 33 |
| **FollBoundary** | | **N** |
| NP-medial <no pause> | ... *for* [$_{\text{NP}}$ *immediate* ***emergency state aid*** *for hospitals*] . | 744 |
| NP-medial <pause> | ... *preoccupied with* [$_{\text{NP}}$ *the* ***news headlines <pause>***] *than with ...* | 95 |
| phrase-final <pause> | ... *in charge of* [$_{\text{NP}}$ *the* ***Mattapan Roxbury area <pause>***]. | 496 |

Table 2: Prosodic boundaries: three types of boundary with number of tokens (the light and dark bars indicate the proportion of left-branching and right-branching tokens for each boundary type, respectively).

of all constituents at the same time instead of fitting a separate model for each constituent. This has the advantage that the model has access to the information that some constituents belong to the same compound, while the interactions provide enough flexibility to allow for systematic differences between the three constituents.

The morphological structure of the compound is coded in the variable Branching. In the analysis, this predictor provides information on the type of NNN compound with the levels *left* and *right*. The branching direction of an NNN compound is naturally related to the variable Member. In left-branching compounds, constituents N1 and N2 constitute the embedded compound, and N3 is the free constituent, whereas in right-branching compounds, N1 is the free constituent, and N2 and N3 form the embedded compound (see (1)).

*Interactions.* An interaction of Branching × Member in the statistical model reveals whether the effect of Branching on the duration of NNN constituents is the same for all compound constituents N1, N2, and N3, or whether the three constituents of an NNN compound are affected differently. The interaction is therefore inevitable in the analysis of morphological structure and its impact on constituent duration in the statistical model.

In addition, Kunter and Plag (2016) have shown in their analysis of NNN compounds that bigram frequency affected the duration of embedded constituents and free constituents differently. In order to account for these potential differences, two three-way interactions (N1N2Freq × Branching × Member and N2N3Freq × Branching × Member) were incorporated into our analysis. Doing so, we can investigate the effect of bigram frequencies on embedded constituents (in cases where the bigram corresponds to the embedded constituents in the NNN compound) and on free constituents (in cases where the bigram corresponds to the free constituent and its embedded neighbor in the NNN compound) at the same time.

With regard to the prosodic structure of NNN compounds, we may also assume that the position of the compound with regard to prosodic boundaries affects embedded constituents and free constituents differently. Therefore, we incorporated the two-way interactions PrecBoundary × Member and FollBoundary × Member. While embedded constituents form one complex unit, free constituents stand on their own. If we consider a prosodic phrase boundary preceding the NNN compound, i.e. which is to the left of the complex word, the branching direction allows us to investigate whether both embedded constituents of a left-branching compound are as affected as the free constituent of a right-branching compound by the prosodic boundary. Simultaneously, we may expect that a prosodic phrase boundary that follows the NNN compound has an effect on the embedded constituents of a right-branching compound, and on the free constituent of a left-branching compound. In order to account for this complex interplay, we could include three-way interactions which are capable to capture these effects. Since the data set does not contain enough observations for all the interaction terms, only the three-way interaction FollBoundary × Branching × Member was included in the model.

| Numeric predictors | Mean | SD | Median | [Min, Max] |
|---|---|---|---|---|
| SpeechRate (words per second) | 3.16 | 0.281 | 3.31 | [2.66, 4.24] |
| Freq | 64500 | 73200 | 37400 | [207, 695000] |
| N1N2Freq | 2640 | 7460 | 193 | [0, 41200] |
| N2N3Freq | 838 | 3500 | 35 | [0, 35700] |
| trigramFreq | 304 | 1860 | 0 | [0, 29700] |
| PhonoN | 12.4 | 15.3 | 5.00 | [0, 58.0] |
| PLD20 | 2.05 | 1.07 | 1.80 | [1.00, 7.10] |
| NPhon | 5.30 | 2.14 | 5.00 | [2.00, 14.0] |
| NSyll | 1.95 | 1.02 | 2.00 | [1.00, 6.00] |
| PitchRange (semitones) | 6.18 | 3.09 | 5.74 | [0.0535, 19.4] |
| **Categorical predictors** | **N** | | | |
| PrecBoundary | | | | |
| clause-initial <pause> | 33 | 2.5% | | |
| phrase-medial <no pause> | 1284 | 96.2% | | |
| phrase-medial <pause> | 18 | 1.3% | | |
| FollBoundary | | | | |
| NP-medial | 750 | 56.2% | | |
| NP-medial <pause> | 94 | 7.0% | | |
| phrase-final <pause> | 491 | 36.8% | | |
| Member | | | | |
| N1 | 439 | 32.9% | | |
| N2 | 452 | 33.9% | | |
| N3 | 444 | 33.3% | | |
| Branching | | | | |
| left | 951 | 71.2% | | |
| right | 384 | 28.8% | | |
| **Interactions** | | | | |
| PrecBoundary × Member | | | | |
| FollBoundary × Member | | | | |
| N1N2Freq × Branching × Member | | | | |
| N2N3Freq × Branching × Member | | | | |

Table 3: Predictor variables in the regression model.

Table 3 summarizes the values of the numeric and categorical predictors that were included in the present analysis, as well as the interactions featured in the model. Note that each three-way interaction also entails all corresponding lower-order interactions. The gray bars for the values of the

| | NPHON | NSYLL | PHONON | PLD20 |
|---|---|---|---|---|
| NPHON | 1.000 | 0.883 | -0.886 | 0.913 |
| NSYLL | 0.883 | 1.000 | -0.826 | 0.872 |
| PHONON | -0.886 | -0.826 | 1.000 | -0.934 |
| PLD20 | 0.913 | 0.872 | -0.934 | 1.000 |

Table 4: Spearman correlations for NPHON, NSYLL, PLD20, and PHONON.

categorical predictors illustrate the proportion of observations containing the respective value.

### *4.3. Statistical Analysis*

The multifactorial analysis was run with the statistical analysis software R (version 3.6.3, R Core Team 2017). The linear mixed-effects regression model reported below was fitted with the lme4 (version 1.1, Bates et al. 2015) and lmerTest (version 3.1, Kuznetsova et al. 2017) packages. The random effects structure for the model was determined using the cAIC4 package (version 1.0, Säfken et al. 2021, see below for a description of the procedure).

The list of factors in Table 3 in the previous section includes several closely related predictors. On the one hand, there are the phonological predictors NPHON and NSYLL, which express the phonological length of a constituent in different units (phonemes and syllables, respectively). On the other hand, there are two predictors that relate to the density of the lexical neighborhood of a constituent (PLD20 and PHONON). Furthermore, phonological length and lexical density are bound to have a negative correlation: due to the increasing entropy of a longer phonological string, the lexical density is likely to decrease. Simply speaking, the probability of lexical neighbors decreases with increasing word lengths (see also Yarkoni et al. (2008) for a discussion). Indeed, as Table 4 shows, all pairwise Spearman correlations between the four variables exceed $r_s > 0.8$. Such a strong correlation between predictors can introduce highly undesirable side effects in regression models. In particular, multicollinearities such as the one between the four predictor variables may cause problems with regard to the interpretability of the involved variables (see Baayen (2008, ch. 6); Wurm and Fisicaro (2014) for detailed discussions).

There are several strategies available to address collinearity in regression models (see Dormann et al. (2013), Tomaschek et al. (2018) for overviews). For the present data set, we opted for a principal component regression. In this approach, a Principal Component Analysis (PCA) is used to transform the data space opened by the input variables. The goal of the transformation is to reduce the number of dimensions required to describe the data space with a minimal loss of information (see James et al. (2013, ch. 10.2)). The PCA yields as many principal components as there are input variables, but each lower-ranked principal component contributes less to the variation in the data space than any higher-ranked principal component. The principal components are by definition orthogonal to each other, and can therefore be used in a regression model instead of the original input variables without a risk of collinearity.

The resulting rotation vectors are represented in Table 5. In this table, each column is one principal component reflecting one dimension in the rotated data space. Each row shows the loadings of the respective input variable for each component. These loadings can be interpreted as a degree of contribution of each input variable to the dimension described by the respective principal com-

| | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| NPHON | -0.526 | 0.178 | -0.060 | 0.829 |
| NSYLL | -0.509 | 0.324 | 0.721 | -0.340 |
| PLD20 | -0.513 | 0.272 | -0.689 | -0.434 |
| PHONON | 0.448 | 0.888 | -0.040 | 0.091 |

Table 5: Loadings in principal component analysis of NPHON, NSYLL, PHONON, and PLD20.

| | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| Standard deviation | 1.825 | 0.647 | 0.399 | 0.303 |
| Proportion of Variance | 0.833 | 0.105 | 0.040 | 0.023 |
| Cumulative Proportion | 0.833 | 0.937 | 0.977 | 1.000 |

Table 6: Importance of principal components in PCA of NPHON, NSYLL, PHONON, and PLD20.

ponent. Higher absolute values indicate a stronger contribution of the input variable to the principal component, while the sign shows the direction of the contribution.

All in all, the principal component analysis shows very similar results to the values reported in Yarkoni et al. (2008) for the correlation between length and lexical neighborhood density. The first principal component PC1 consists in absolute terms to approximately equal terms of information contributed by all four input variables (absolute loadings between 0.448 and 0.526). With the exception of PHONON, all loadings are negative. In contrast, the second component PC2 is predominantly determined by PHONON with a loading of 0.888, while the other input variables have loadings below or equal 0.324. One interpretation of these loadings is that PC1 represents a data dimension in which phonological length and lexical density are roughly equally represented, but PC2 represents a dimension that represents mostly the lexical density information encoded in PHONON, i.e. the number of phonological neighbors.

Table 6 shows the contribution of each principal component to the variance in the data space opened by the input variables. Apparently, PC1 alone explains 83.3% of the variance in the data space, while PC2 contributes another 10.5% percent of the variance. The remaining two principal components PC3 and PC4 do not reach the threshold of 5.0% that is typically assumed for principal component regressions. Consequently, they were not considered for the subsequent model, and only PC1 and PC2 were included as predictors instead of the four input variables.

An initial linear model using this new set of predictors as well as the three-way interactions described in Table 3 revealed that the residuals deviated from a normal distribution, which is usually interpreted as an indication that the model assumption of linearity is violated. In order to alleviate this deviation, the dependent variable DURATION was Box-Cox-transformed using an exponent of $\lambda = 0.\overline{58}$ (Box and Cox 1964). This transformation removed the non-normality of the residuals. Consequently, the transformed variable DURATIONBC was used as the dependent variable in the model below.

The random effects structure was determined using the stepcAIC function from the cAIC4 package (Säfken et al. 2021). This function provides an algorithm that determines the optimal choice of random slopes for a predefined set of grouping variables based on the Conditional Akaike Information Criterion (cAIC). For the present model, we suggested the grouping variables SPEAKER (the speaker ID from the Boston corpus) and CONSTITUENT (the orthographic representation of each constituent token), as well as the numeric variables PC1, PC2, log TRIGRAMFREQ, log FREQ, SPEECHRATE, and PITCHRANGE as potential random slopes. The algorithm detected the lowest cAIC for a model that contained a random intercept and a random slope for PC1 within the grouping variable CONSTITUENT, and random slopes for PC1, log TRIGRAMFREQ and PITCHRANGE but no random intercept within the grouping variable SPEAKER.

The random slope for log TRIGRAMFREQ by SPEAKER may reflect the possibly biasing sampling procedure mentioned above. As only the first token that was encountered during sampling was included in the data set, there is a possibility that compounds with high trigram frequencies (which are more likely to occur early in the corpus at least once) will be overrepresented in the subsets from the speakers that were sampled first. Indeed, a Kruskal-Wallis test reveals significant differences in the average rank of the NNN frequencies across the speakers. In particular, speaker F1 (who is a probable target to be the first speaker to be sampled) has a higher median NNN frequency than any other speakers, and consequently, compounds sampled from this speaker may be shorter on average than compounds sampled from other speakers. The random slope is expected to account for this imbalanced distribution of trigram frequencies.

The specification of the final model including this random effect structure and the principal components is shown in (2) using the lme4 notation for the random effects. In total, this fairly complex model estimates 36 coefficients (including all interaction terms and the different levels for interaction terms). However, with a data set of 1335 observations, there does not appear to be a significant risk of overfitting if the usual rules of thumb are applied (for instance, Harrell 2001, ch. 4.4 suggests about 10–15 observations per estimated term).

(2) DURATIONBC ∼ SPEECHRATE +
FREQ + TRIGRAMFREQ +
PITCHRANGE +
PC1 + PC2 +
N1N2FREQ × BRANCHING × MEMBER +
N2N3FREQ × BRANCHING × MEMBER +
PRECBOUNDARY × MEMBER +
FOLLBOUNDARY × MEMBER +
(1 + PC1 | CONSTITUENT) +
(0 + PC1 + log TRIGRAMFREQ + PITCHRANGE | SPEAKER)

## 5. Results

The model specification in (2) resulted in a mixed-effects model that was determined by R without any convergence issues. Table 7 summarizes the resulting fixed effects. The reference level for the Intercept of the regression model is FOLLBOUNDARY = medial <no pause>, PRECBOUNDARY = <no pause> medial, BRANCHING = left and MEMBER = N1. The coefficients indicate the estimated effect of the respective predictor on constituent duration. Positive numbers show that the variable in question increases the constituent duration, and negative numbers indicate a decrease. Note that all coefficients relate to the Box-Cox-transformed durations. For the convenience of the reader, we back-transformed the dependent variable to duration in seconds in Figure 1 to Figure 5. Note that due to this back-formation, the partial effects plots actually show a non-linear relationship between the predictor on the x-axis and the back-transformed response variable, as can be seen in the plot for PC1. For the other predictors, the curvature is too small to detect this non-linearity.

Not all of the predictors included in the statistical model affect constituent durations significantly in our data, as can be seen in Table 7. That is, log TRIGRAMFREQ does not have a significant effect on constituent durations. Besides, neither the two-way interaction PRECBOUNDARY × MEMBER nor PRECBOUNDARY alone reach statistical significance. Given the low number of tokens that occurred in the PRECBOUNDARY environments with a preceding pause, this result is not particularly surprising.

*Single effects.* Figure 1 illustrates the partial effect plots for those predictors that have a significant effect on constituent durations in our statistical model. In each plot, only the predictor under investigation is varied, while all other predictors are held constant either at their median (for numeric predictors) or at the most frequent category (for categorical predictors).

The top-left plot for FREQ reveals the expected effect: Other things being equal, higher constituent frequencies lead to shorter durations. The partial effect of PITCHRANGE is shown in the top-right plot. Constituents with a small pitch range, i.e. those that have a flat pitch contour, have a relatively short duration, whereas constituents with a large pitch range are relatively long. The left and right plots in the middle show the partial effects of PC1 and PC2. As discussed earlier, the input variables to the principal components are the same, but the loadings of each of the variables in PC1 and PC2 differ. The mostly negative correlation of the input variables with PC1 explains the direction of the effect in the left plot: With increasing PC1, that is, with a decrease of phonological length and a denser phonological neighborhood, constituent durations decrease. In contrast, PC2, which mostly informs about phonological neighborhood density, has a lengthening effect on constituent durations, as shown in the right plot. We will return to the implications of these partial effects below in the discussion. The bottom-left plot illustrates the partial effect of SPEECHRATE on the x-axis. As can be seen, constituent durations decrease with increasing speech rate, i.e. with a faster pace of a speaker to produce speech the constituents get shorter.

FOLLBOUNDARY × MEMBER. The regression analysis reveals that the two-way interaction FOLLBOUNDARY × MEMBER has a significant effect on constituent durations. The interaction is able to uncover the impact of the type of following boundary on each of the constituents separately. Figure 2 shows the effect of the three different boundary types on the constituents N1, N2, and N3 with back-transformed durations on the y-axis. The results indicate that the nature of the following prosodic boundary, that is, whether the NNN compound is in medial or in final position, and whether or not a pause follows the NNN compound, has a particular lengthening effect on N3 constituents: Overall, N3 durations in final position followed by a pause are longest. Thus, N3 constituents become longer with an increase in boundary strength.

Contrary to N3 constituents, the type of following boundary does not have a significant effect on

| | Est. | SE | df | $t$ | $p$ | |
|---|---|---|---|---|---|---|
| (Intercept) | 0.705 | 0.031 | 1238.797 | 22.956 | 0.000 | *** |
| **Main effects** | | | | | | |
| log N1N2Freq | -0.002 | 0.001 | 1202.268 | -1.709 | 0.088 | . |
| Branching=right | -0.015 | 0.016 | 1086.044 | -0.936 | 0.350 | |
| Member=N2 | -0.020 | 0.025 | 1073.095 | -0.825 | 0.409 | |
| Member=N3 | 0.006 | 0.026 | 1092.559 | 0.249 | 0.804 | |
| log N2N3Freq | -0.002 | 0.001 | 1075.030 | -1.582 | 0.114 | |
| FollBoundary=NP-medial <pause> | 0.012 | 0.010 | 877.068 | 1.143 | 0.253 | |
| FollBoundary=phrase-final <pause> | 0.010 | 0.005 | 982.643 | 1.916 | 0.056 | . |
| PrecBoundary=phrase-medial <no pause> | 0.010 | 0.016 | 918.078 | 0.613 | 0.540 | |
| PrecBoundary=phrase-medial <pause> | 0.034 | 0.028 | 1109.444 | 1.208 | 0.227 | |
| log TrigramFreq | -0.000 | 0.001 | 23.160 | -0.347 | 0.731 | |
| log Freq | -0.008 | 0.002 | 602.517 | -4.712 | 0.000 | *** |
| PC1 | -0.044 | 0.002 | 9.759 | -20.621 | 0.000 | *** |
| PC2 | 0.017 | 0.004 | 486.056 | 4.795 | 0.000 | *** |
| PitchRange | 0.006 | 0.001 | 16.969 | 7.381 | 0.000 | *** |
| SpeechRate | -0.027 | 0.006 | 936.745 | -4.588 | 0.000 | *** |
| **Two-way interactions** | | | | | | |
| log N1N2Freq:Branching=right | 0.002 | 0.003 | 1087.429 | 0.725 | 0.469 | |
| log N1N2Freq:Member=N2 | -0.005 | 0.002 | 1255.219 | -2.361 | 0.018 | * |
| log N1N2Freq:Member=N3 | 0.004 | 0.002 | 1259.538 | 1.977 | 0.048 | * |
| log N2N3Freq:Branching=right | 0.003 | 0.002 | 959.913 | 1.236 | 0.217 | |
| log N2N3Freq:Member=N2 | 0.005 | 0.002 | 1111.646 | 2.751 | 0.006 | ** |
| log N2N3Freq:Member=N3 | -0.001 | 0.002 | 1223.959 | -0.559 | 0.576 | |
| Branching=right:Member=N2 | -0.008 | 0.023 | 1173.032 | -0.355 | 0.722 | |
| Branching=right:Member=N3 | 0.043 | 0.022 | 1148.706 | 1.991 | 0.047 | * |
| Member=N2:FollBoundary=NP-medial <pause> | -0.014 | 0.014 | 881.804 | -0.948 | 0.343 | |
| Member=N3:FollBoundary=NP-medial <pause> | 0.003 | 0.015 | 1082.948 | 0.172 | 0.863 | |
| Member=N2:FollBoundary=phrase-final <pause> | -0.006 | 0.007 | 1020.309 | -0.788 | 0.431 | |
| Member=N3:FollBoundary=phrase-final <pause> | 0.024 | 0.008 | 1169.536 | 3.048 | 0.002 | ** |
| Member=N2:PrecBoundary=phrase-medial <no pause> | 0.018 | 0.022 | 983.184 | 0.817 | 0.414 | |
| Member=N3:PrecBoundary=phrase-medial <no pause> | 0.012 | 0.023 | 1022.913 | 0.517 | 0.605 | |
| Member=N2:PrecBoundary=phrase-medial <pause> | 0.005 | 0.039 | 1016.401 | 0.138 | 0.891 | |
| Member=N3:PrecBoundary=phrase-medial <pause> | -0.008 | 0.042 | 1174.539 | -0.186 | 0.853 | |
| **Three-way interactions** | | | | | | |
| log N1N2Freq:Branching=right:Member=N2 | 0.010 | 0.004 | 1156.812 | 2.524 | 0.012 | * |
| log N1N2Freq:Branching=right:Member=N3 | -0.004 | 0.004 | 1166.623 | -1.117 | 0.264 | |
| log N2N3Freq:Branching=right:Member=N2 | -0.006 | 0.003 | 1086.070 | -1.871 | 0.062 | . |
| log N2N3Freq:Branching=right:Member=N3 | -0.007 | 0.003 | 1154.482 | -2.187 | 0.029 | * |

Table 7: Fixed effects of regression model predicting transformed constituent durations ($N = 1335$, reference level of the Intercept: FollBoundary = NP-medial, PrecBoundary = <no pause> medial, Branching = left, Member = N1).

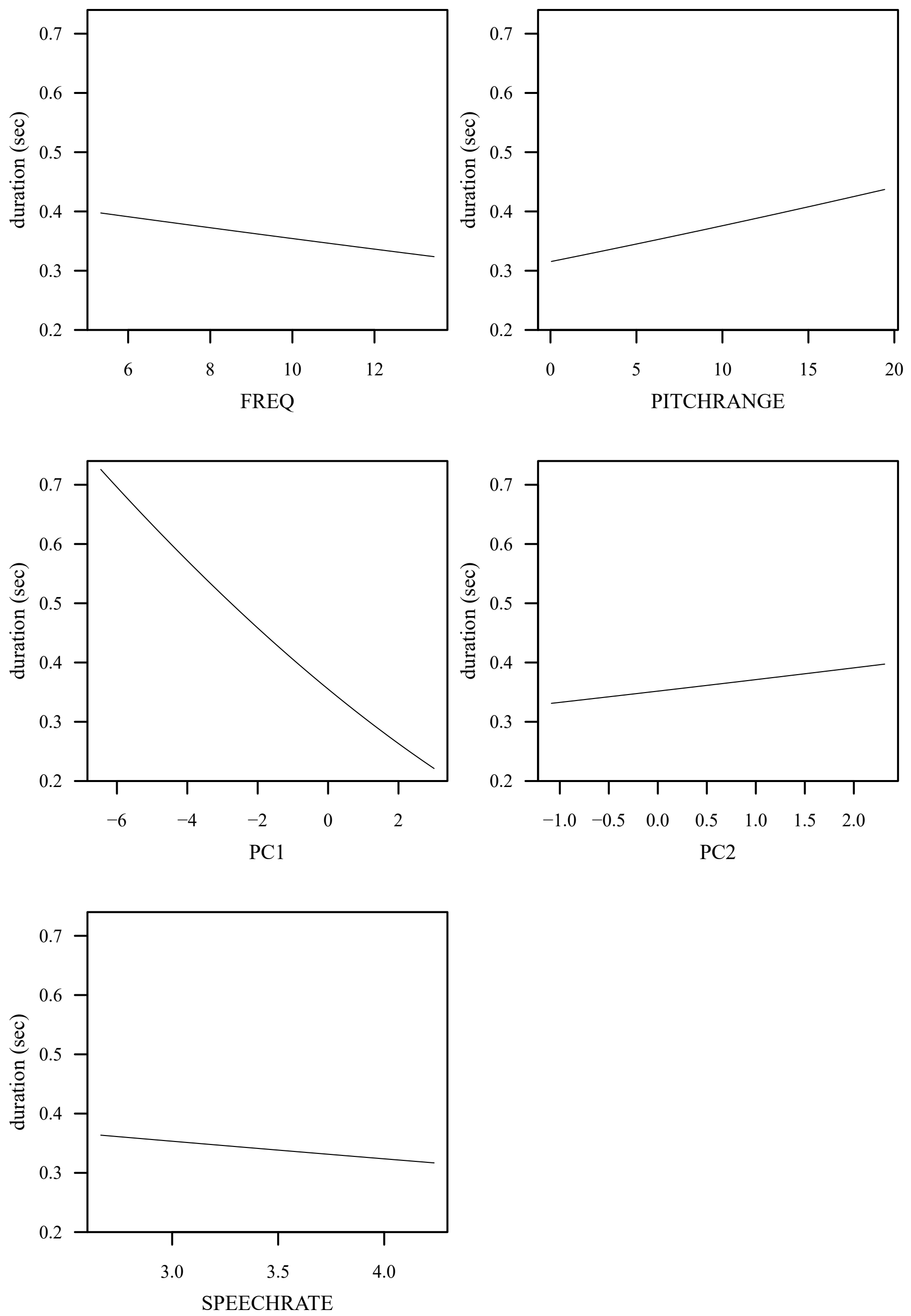



Figure 1: Partial effects of FREQ, PITCHRANGE, PC1, PC2 and SPEECHRATE on constituent duration (back-transformed (in sec), numeric predictors at median, categorical predictors at most frequent category).

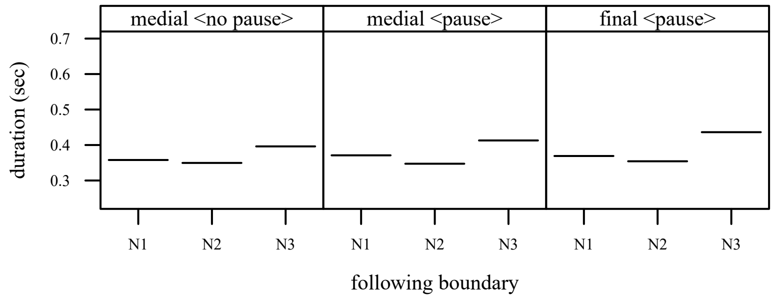



Figure 2: Constituent duration in different positions (back-transformed (in sec), numeric predictors at median, categorical predictors at most frequent category).

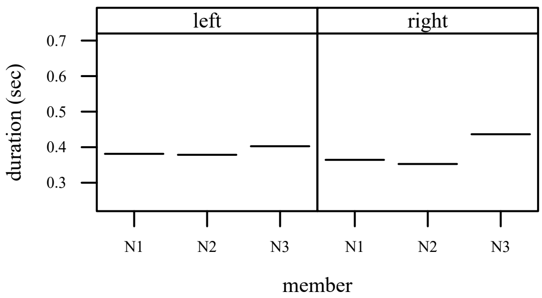



Figure 3: Constituent duration in left- and right-branching compounds (back-transformed (in sec), categorical predictors at most frequent category, numeric predictors at median except: log N1N2FREQ, log N2N3FREQ = 0).

the duration of N1 and N2 constituents in our analysis. Besides, N3 durations are overall significantly longer in phrase-final position followed by a pause than N1 and N2 durations.

N1N2FREQ × BRANCHING × MEMBER *and* N2N3FREQ × BRANCHING × MEMBER. Before examining the three-way interactions involving bigram frequencies, we will first take a glimpse at the constituent durations in an average left- and right-branching compound, i.e. one that occurs in the most frequent prosodic position in the data set, namely "medial <no pause>", and in which all numeric variables are set to their median, except for log N1N2FREQ and log N2N3FREQ which are kept at = 0 on purpose. As shown in Figure 3, N1 and N2 constituents appear to be overall shorter than N3 constituents in both branching directions. However for left-branching compounds, this difference is not statistically significant. In right-branching compounds, N3 constituents are significantly longer than N2 constituents, whereas N1 durations pattern in between. Across branching direction, constituent durations of N1, N2, and N3 do not significantly differ.

The two significant three-way interactions indicate that the average duration distribution illustrated in Figure 3 changes with increasing bigram frequencies, however, these changes affect the individual constituents in different ways. Furthermore, the effect is different depending on the branching direction. Figure 4 illustrates the effect of log N1N2FREQ on the three constituents in left- and right-branching compounds, while Figure 5 shows the corresponding effect of log N2N3FREQ.

Each bar in Figure 4 illustrates the average deviation of a constituent from the average duration of

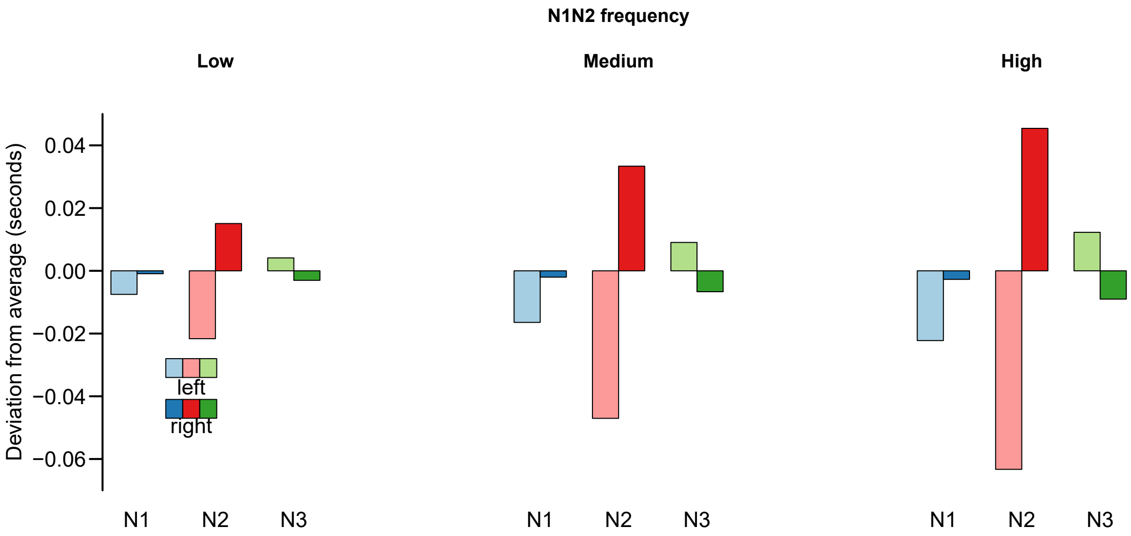



Figure 4: Effect of low, medium and high N1N2 frequencies on the deviation of constituent durations from an average constituent (back-transformed, in sec). Light and dark bars correspond to left-branching and right-branching compounds, respectively.

that constituent as shown in Figure 3 (N1 = blue, N2 = red, N3 = green). N1N2FREQ is equivalent to the frequency of the embedded compound N1N2 in left-branching compounds, while it is the bigram frequency of the free constituent N1 and the neighboring embedded constituent N2 in right-branching compounds. There are three panels in the plot indicating three levels of N1N2 bigram frequencies: low, medium and high N1N2 bigram frequencies, indicating the contribution of N1N2FREQ at the 25% quantile, the median, and the 75% quantile. The branching direction is indicated in the bars by lighter (left-branching) and darker (right-branching) shades of the respective colors. The zero line on the y-axis indicates the average duration of constituents from which two colored bars per constituent depart. In order to see the full picture of the effect of the three-way interaction involving N1N2FREQ, the deviation from an average constituent in Figure 4 is added to or subtracted from the constituent duration indicated in Figure 3, depending on the direction of the deviation.

The strongest effect of N1N2 bigram frequencies on constituent duration can be seen in the deviations of N2 from the duration of an average N2 constituent. In left-branching compounds, N2 significantly decreases in duration with higher N1N2 bigram frequencies, which is illustrated by the increasingly negative value of the light-red bars. For very high N1N2FREQ, left-branching N2 constituents are approximately 0.065 s shorter than for very low N1N2FREQ. At the same time, N2 durations increase significantly in right-branching compounds with increasing N1N2 bigram frequencies, as indicated by the dark-red bars going in the positive direction: with high N1N2FREQ, N2 constituents are longer than an average right-branching N2 (approx. +0.050 s). In other words, increasing N1N2FREQ has a shortening effect on N2 constituent durations in left-branching compounds, but a lengthening effect on N2 in right-branching compounds. The duration of left-branching N1 constituents decreases with an increase of N1N2 bigram frequency by -0.020 s, however, this effect is not statistically significant in our analysis. Increasing N1N2FREQ appears to hardly affect N3 durations in left-branching compounds, or N1 and N3 durations in right-branching compounds.

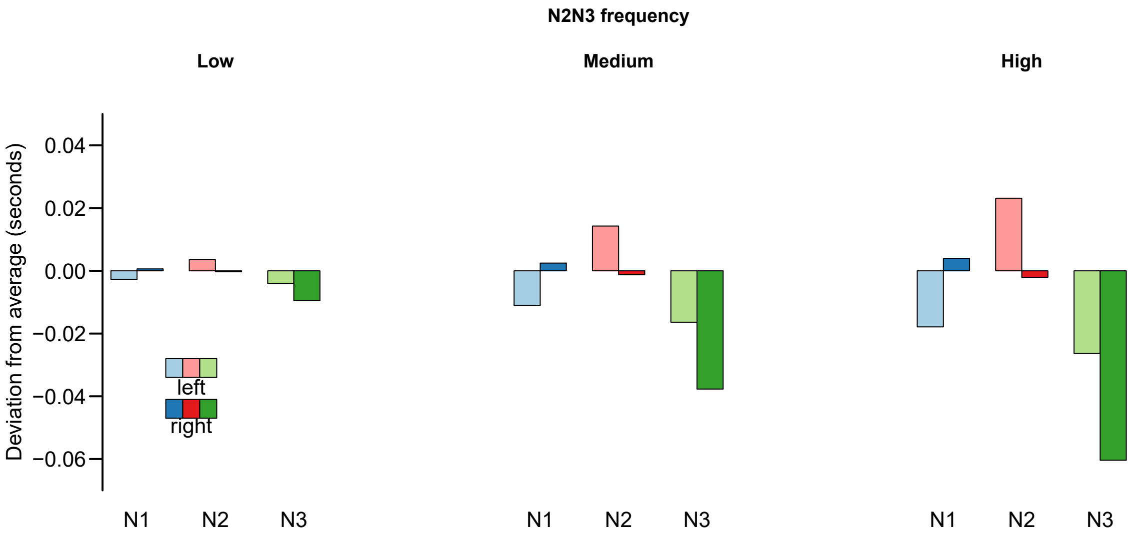



Figure 5: Effect of low, medium and high N2N3 frequencies on the deviation of constituent durations from an average constituent (back-transformed, in sec). Light and dark bars correspond to left-branching and right-branching compounds, respectively.

Figure 5 illustrates the effect of N2N3 bigram frequencies and branching direction on constituent durations. Similar to Figure 4, the zero line marks the average duration of constituents, and it is the point of departure for the colored bars indicating the three constituents N1 (blue), N2 (red) and N3 (green) in two different color shades indicating left-branching (light) and right-branching (dark) constituents. The three panels show three levels of N2N3FREQ, namely low, medium and high N2N3 bigram frequencies. N2N3FREQ is equivalent to the frequency of the embedded compound in right-branching compounds, while it is the bigram frequency of the embedded constituent N2 and the free constituent N3 in left-branching compounds.

The most striking effect of this interaction can be seen in N3 constituent durations. With increasing N2N3FREQ, right-branching N3 constituents decrease significantly in duration compared to an average right-branching N3 constituent (approx. -0.060 s). The effect on left-branching N3 constituents is similar, however, these constituents are not as affected as their right-branching counterparts (approx. -0.030 s). In left-branching compounds, N2 durations increase significantly with increasing N2N3FREQ (approx. +0.025 s), whereas the durations of right-branching N2 constituents as well as N1 constituents in either branching direction are not affected.

## 6. Discussion and Conclusion

Our statistical model provides a number of intriguing effects on constituent durations. The majority of noise variables that we incorporated in our model predict durations as anticipated. For instance, constituents with a pitch accent are longer, and a higher speech rate leads to overall shorter constituent durations. The type of following boundary affects N3 durations, which can be considered evidence for final lengthening of phrase-adjacent material (cf. Turk and Shattuck-Hufnagel 2007 on syllabic lengthening): While N1 and N2 constituents are hardly affected by the position of the NNN compound or the presence of a pause, N3 constituents are longest if they are followed by a phrase boundary and a pause. Overall, the statistical analysis yields findings which address our three research questions from section 2 and which provide evidence for evaluating our predictions.

*Can we detect systematic differences in the phonetic signal that are related to differences in morphological boundary strength?.* Kunter and Plag (2016) correlate the depth of embedding in a tri-constituent compound with morphological boundary strength. Consequently, they propose that since differences in boundary strength emerge from the branching order, the embedded constituents will be more prone to phonetic reduction including acoustic shortening than the free constituent. While their analysis of acoustic durations seems to support this claim, the present analysis does not find convincing evidence for such an effect of embedding on the constituent durations. As illustrated in Figure 3, the embedded constituents in left- and right-branching NNN compounds are not significantly shorter than the free constituent in compounds in which the two involved bigram frequencies are so low that they cannot possibly interfere with the suggested effect of morphological boundary strengths. In our model, the durations of left-branching constituents overall do not differ significantly from each other, although N3 as the free constituent should be notably longer than the embedded constituents N1 and N2. Likewise, in the right-branching compounds, where N1 is the free constituent and N2 and N3 are embedded, N3 is significantly longer than N1 and N2, which again does not agree with the effect proposed by Kunter and Plag (2016). Apparently, for compounds with minimal bigram frequencies, the analysis of our data set does not point to an impact of morphological boundary strength on the phonetic signal. But perhaps the picture changes if the bigram frequencies are not minimal because in that case, speakers may use them as cues to the actual branching structure of the compound, which in turn may affect the way they process these structures?

*What is the impact of branching direction and lexical frequency on the phonetic signal?.* We find a clear effect of branching direction, i.e. the indicator of compound-internal morphological structure in our analysis, but the effect is rather complex, as it is mediated by both bigram frequencies. Before we discuss this complex interaction in detail, we will first discuss the other frequency-related predictors. To start with, the unigram frequency of a constituent has a robust shortening effect on durations. This result supports findings from previous research that has investigated low-frequent and high-frequent linguistic units and their phonetic signal which is presumably shaped by their frequency (Wright 1979, Pluymaekers et al. 2005b, Gahl 2008), but it raises an interesting question: why can the lexical frequencies of constituents that are embedded into larger morphological structures, which in turn can be considered as lexical units, still affect the durations of the constituents in production? After all, NNN compounds are morphologically complex units which contain smaller units, namely free constituents and embedded compounds, which themselves consist of two constituents. Thus, it seems that speakers accomplish an intricate task when processing NNN compounds: token frequencies of the constituents themselves affect their durations even though the constituents are not just loosely ordered individual lexemes, but form an NNN compound with a certain internal organization. The finding that FREQ still has an effect even though the constituents are part of these complex units suggests that the individual constituents still have to be accessible during the production. This speaks against accounts that treat compounds or other morphologically complex words as monolithic items that are articulated without insight into their internal structure.

Our statistical analysis does not reveal a similar shortening effect of trigram frequencies on the phonetic signal. Based on previous research on n-gram frequency effects, one could expect a shortening effect of the frequency of the three compound constituents to occur together. However, we assume that the low range of trigram frequencies in our data set appears to prevent such an effect. In general, NNN compounds are rather rare in English, which suggests that speakers tend to choose different structures such as complex noun phrases over NNN compounds in spoken language. In our data, trigram frequencies range from 0 to 29700 tokens in COCA, with a median of 0 and a mean of 304. These figures imply a rather skewed distribution: the majority of NNN compounds is unattested in the corpus. Therefore, the fact that we do not find a significant effect of trigram frequency on duration does not contradict findings of, for instance, Arnon and Cohen Priva (2013), who report a shortening effect of n-gram frequencies on the duration of high-frequency phrases (*a lot of work*, mean frequency $= 12.04$ per million) and low-frequency phrases (*a lot of years*, mean frequency $= 2.25$ per million). Thus, the potential explanatory power of this lexical variable may be weakened by the particularly small trigram frequencies in our study.

There is yet another way to address the second question, namely by looking at the joint impact of branching direction and bigram frequencies on constituent durations. The interplay of N1N2FREQ × BRANCHING × MEMBER and N2N3FREQ × BRANCHING × MEMBER reveals some intriguing effects on particular constituents only, which we will interpret below as a strategic use of the acoustic signal by the speakers to disambiguate conflicting cues to the branching direction in NNN compounds.

Based on the findings reported in Kunter and Plag (2016), we predicted an effect of the morphological structure, i. e. branching direction, on the duration of compound constituents as well as a decrease of constituent durations with increasing bigram frequencies. The significant effect of the three-way interactions, however, indicates that the impact of morphological structure and lexical frequency can only be reasonably investigated when simultaneously taking both factors into account.

In left-branching compounds, where N1 and N2 constituents form the embedded compound, we expected shorter constituent durations for N1 and N2 and comparatively longer constituent durations for the free constituent N3. Likewise, with high N1N2 bigram frequencies, we predicted shorter N1 and N2 constituent durations while N3 constituents remain unaffected. As a consequence, both predicted effects are assumed to manifest themselves in this duration pattern: (N1, N2) < N3. In fact, this is more or less what we see: N2 durations decrease significantly with increasing N1N2 bigram frequencies, and also the duration difference between N1 and N3 grows so that the duration pattern is N2 < N1 < N3.

Similarly, for right-branching compounds with N2 and N3 constituents forming the embedded compound and N1 as the free constituent we predicted shorter constituent durations for N2 and N3 and comparatively longer constituent durations for N1. We predicted that high N2N3 bigram frequencies lead to shorter N2 and N3 constituent durations, while they have no effect on N1 constituents. Again, the duration pattern that follows from our predictions is the same for both factors: N1 > (N2, N3). This, however, is not what we find in our results. Instead, N3 constituents are overall longer than N1 and N2 constituents in right-branching compounds, although increasing N2N3 bigram frequencies have a very strong shortening effect on N3. This is, as a matter of fact, not only true for right-branching, but also for left-branching N3 constituents.

The effects that we observe for N1N2 and N2N3 bigram frequencies on constituent N2 in left- and right-branching compounds are more complex. Theoretically, we would assume that with increasing bigram frequencies, the duration of the involved constituents becomes shorter. However, this is not the case for all constituents – we find surprisingly different effects of the bigram frequencies on constituent N2 in left- and right-branching compounds. Our statistical analysis reveals that the N2 constituents in left-branching compounds show the expected reaction to increasing N1N2 bigram frequencies, that is, N2 durations decrease significantly. In contrast, right-branching N2 constituents become significantly longer when N1N2 bigram frequencies increase, which is unpredicted by previous research on effects of lexical frequency on the phonetic signal. A similar lengthening effect can be observed for the N2N3 bigram frequencies. Here, a frequency increase leads to a lengthening of left-branching N2 constituents, but there is no evidence for the expected shortening effect on the N2 constituents in right-branching compounds. This means that the duration of N2 constituents, which form part of the embedded compound regardless of branching direction, can also be affected by the bigram frequency that does not coincide with the branching direction of the NNN compound: in left-branching compounds, increases in N2N3FREQ coincide with an increased length of N2 even though N2N3FREQ is not the bigram frequency of the embedded compound in the left-branching structure. Likewise in right-branching compounds, the duration of N2 increases with increasing N1N2FREQ, but again N1N2FREQ does not correspond to the frequency of the embedded compound in the right-branching NNN.

This effect of the bigram frequencies is surprising for two reasons. First, in every NNN compound, there is competition between two possible semantic interpretations, either as a left-branching [NN]N in which the last constituent is modified by a preceding compound, or as a right-branching N[NN] in which the initial constituent is a modifier for the following compound. For the speaker, the intended branching direction is determined on the speaker's intention, which takes the larger context of the compound into account. Yet, despite this, the effect of N1N2 frequency in right-branching compounds and the effect of N2N3 frequency in left-branching compounds suggest that the production of the NNN compound is affected by the representation of both bigrams in the lexicon, even if only one of the bigram representations is relevant for the actual branching direction, and the other bigram representation actually competes with the intended meaning of the NNN. Second, lexical frequency is usually assumed to have a unidirectional effect on duration: If linguistic material is more frequent, its duration is shorter because access to the lexical entry in the mental lexicon is simplified. Yet, our findings show that the impact of morphological structure on the one hand and of lexical frequency on the other hand operate in different ways. Apparently, this frequency effect cannot easily be explained e.g. by a facilitating effect on lexical access or by articulation patterns learned through repeated practice.

Instead, we interpret these unexpected findings as an attempt of speakers to disambiguate the morphological structure of an NNN compound from the frequency-driven parsing of the NNN compound. If the branching direction coincides with a high bigram frequency of the embedded compound, the effect of the bigram frequency follows the expected pattern for high-frequent linguistic units. This is particularly true in left-branching compounds where a high N1N2 frequency has a strong shortening effect on constituent N2. Similarly in right-branching compounds, increasing N2N3 frequencies lead to a shortening effect of N3 (but somewhat unexpectedly, N2 is unaffected). However, in cases where the branching direction and a highly frequent bigram do not coincide, the effects on the constituent duration do not agree with the expected reductions. Instead, speakers seem to adjust the duration pattern in such a way that hearers are able to parse the intended branching direction of the NNN compound despite the bigram frequency. For example, the left-branching compound [*waste company*] *officials* has a N1N2 bigram frequency of 6 for *waste company*, and a N2N3 bigram frequency of 549 for *company officials*. Yet, the compound refers to the officials of a waste company, not to company officials who have to do with waste. Our analysis suggests that in these cases where the bigram frequencies suggest a different branching direction than the semantically intended direction, speakers emphasize the semantic branching by a lengthened N2. Conversely, the right-branching compound *consumer* [*credit chain*] has a N2N3 bigram of 0 for *credit chain*, and a N2N3 bigram frequency of 247 for *consumer credit*. Yet, the compound does not refer to a consumer credit having to do with chains, but rather to a credit chain having to do with consumers. Thus, the lengthening of N2 constituents both in right-branching compounds with high N1N2 bigram frequencies and in left-branching compounds with high N2N3 bigram frequencies may be interpreted as an attempt of the speaker to signal to the listener that the high bigram frequency may be misleading. In other words, if the duration of N2 is unexpectedly long, this may serve to indicate that the intended branching direction and the cues provided by the bigram frequencies do not match.

*Can categorical approaches to speech production, such as Lexical Phonology (Kiparsky 1982a,b) or the language planning model by Levelt et al. (1999) account for our findings?.* As discussed in section 2, neither Lexical Phonology nor the Levelt model implement an interface of morphological structure and the phonetic signal. Accordingly, constituent durations are not expected to differ across branching direction. Although our statistical analysis does not reveal a significant effect of morphological embeddedness in our data, and our own prediction is rejected by the results, we still observe different duration patterns between left-branching and right-branching NNN compounds. Lexical Phonology as well as the Levelt model do not take into account how lexical frequency could possibly affect constituent durations. While one could argue that, therefore, this finding does not contradict such categorical approaches, the significant impact of branching direction and bigram frequencies together on individual compound constituents is unpredicted. In particular, the disambiguation effect that we detected for N2 durations with increasing N1N2 bigram frequencies contradicts the strict separation of morphology and phonology on the one hand and the phonetic signal on the other hand.

*Effect of neighborhood density.* The discussion so far focused mostly on the variables BRANCHING, MEMBER, and the bigram frequencies N1N2FREQ and N2N3FREQ, while treating the other variables mostly as noise variables that serve only to reduce the variance in the signal. However, there is an interesting effect in the principal components that represent phonological length (in terms of number of phonemes and syllables) and the structure of the phonological neighborhood (operationalized by the PLD20 measure and the number of phonological neighbors) that warrants some attention. As described above in section 4, the loadings of the first principal component PC1 are fairly similar for all four involved variables. This means that the information contained in this principal component consists of information of the four input variables to roughly equal parts. As described above, PC1 has a negative coefficient: with increasing values of PC1, the duration of constituents decreases on average. Looking at the loadings for PC1 in Table 6, there is a straightforward interpretation of this effect: An increase in PC1 means a decrease of the number of phonemes, the number of syllables, as well as a decrease of PLD20 and an increase of the number of phonological neighbors PHONON – in other words, the fewer phonological material, and the more dense the phonological neighborhood, the shorter the duration of the constituent. However, it is plausible to assume that the neighborhood density effect is highly related to differences in potential information that words with different phonological material can express: the more phonological segments a word has, the higher its entropy will be, i.e. the information that can be expressed by this string of phonemes (see Shannon 1948). Hence, long words are *ceteris paribus* less likely to have a large number of phonologically similar words in the lexicon than short words. Conversely, a word that is phonologically short is more probable to have a large number of phonologically close entries in the lexicon than a long word simply due to its low entropy. We suggest, then, that PC1 can mostly be interpreted as a representation of phonological length and the impact that phonological length has on lexical density based on information that can be encoded in strings with different numbers of segments. This interpretation is in line with the effects described in Gahl et al. (2012). They report a negative correlation between the orthographic length and the neighborhood density (ND, expressed by the number of phonological neighbors), which reflects the lower entropy of shorter words, and hence the higher probability of phonological neighbors. Similar to PC1 in our model, which has a positive loading for PHONON but a negative coefficient overall, they find a negative effect of neighborhood density ND on the word durations, so the effect of the number of phonological neighbors is very similar in our model and in the one reported in Gahl et al. (2012). Yet crucially, while Gahl et al. (2012) attempt to disentangle the confounding effect of the phonotactic probability within the words from the effect of ND, they do not report a similar disentangling effort for the confounding effect of word length. The orthographic length predictor did not reach significance in their model, and was hence eliminated, while the neighborhood density predictor was retained as significant. In the light of the loadings of PC1 in our principal component analysis, which were similar for the number of phonological segments and the number of phonological neighbors but with reversed signs, it may be possible that ND simply suppressed the effect of word length (see Wurm and Fisicaro (2014), Tomaschek et al. (2018) for discussions of suppression effects), resulting in a stronger effect of ND than really warranted.

The second principal component PC2 is also significant in our model. While the loadings for this principal component show that this variable includes information about the phonological length and the PLD20 measure, it is predominantly determined by the number of phonological neighbors. In our model, the coefficient for PC2 is positive: with increasing values of the variable, the word duration becomes longer. We interpret this effect as a response to the need to disambiguate a word with many phonological neighbors by strengthening the acoustic signal. Conversely, a word with only few phonological neighbors has little competition, and hence can be more safely reduced without the risk of ambiguity. This interpretation is in line with the information-based Uniform Information Density from Jaeger (2010) and Jaeger and Tily (2011). Their listener-oriented framework argues that speakers attempt to retain an even distribution of information content by manipulating the duration of the involved linguistic units. Generally speaking, a phonological neighbor may also be considered a lexical competitor: speakers have to use the available phonological material to decide which lexical entry they need to access. Consequently, this decision will be more difficult for a word with many phonological neighbors (and hence many competitors). In information-theoretic terms, such a word will have a higher information content than a word with fewer phonological neighbors (and hence fewer competitors) because it will be more difficult to predict based on the phonological material than the latter. As a result of this difference in information content, the information-based accounts predict that other things being equal, a word with many competitors will be longer than a word with few competitors, as speakers prefer to distribute the higher information content of the former over more acoustic material than in the case of the latter. This prediction agrees well with our second principal component.

Thus, while we find a very similar effect of neighborhood density as reported in Gahl et al. (2012), we interpret it very differently. Gahl et al. (2012) argue that words with dense phonological neighborhoods are more reduced as far as both durations and vowel centralizations are concerned. They ascribe this to the production-focused nature of casual speech, which assigns lesser weight to the intelligibility and distinctiveness of the acoustic signal. However, we propose that the neighborhood density effect is a lengthening effect that highlights the increased information content of words with a dense neighborhood, a pattern that is predicted by Jaeger (2011) and Jaeger and Tily (2011)'s principle of Uniform Information Density.

This listener-oriented explanation also agrees with the effect that we have observed for the compounds in which the bigram frequencies suggest a different branching direction than the intended branching. These compounds are cases in which it is increasingly difficult to decide between the two possible branching alternatives. This increase in unpredictability makes these compounds more informative than compounds in which both the bigram frequency and the intended branching are aligned. In agreement with the account in Jaeger and Tily (2011), the speaker appears to react to the increased information content of these mismatched compounds by increasing the acoustic duration in particular of N2.

*Generalizability of the results.* One issue that needs to be addressed is in how far the findings from this corpus study may be generalized to other speakers of English. After all, the recordings used in this analysis comprise a particular type of spoken language: they were produced by professional radio speakers who are trained to broadcasting their utterances as clearly as possible to their audiences, while at the same time maintaining an uninterrupted and connected flow of speech. In contrast, the utterances analysed in Kunter and Plag (2016), while elicited in an experimental setting, are from non-professional speakers whose speech may be more prone to phonetic reduction than that of the professional speakers. As discussed above, the different bigram frequency constellations are probably one reason why the conclusion of the present study with regard to the effect of morphological boundary strength differs from that in Kunter and Plag (2016), but it is to be assumed that the different speaking styles may affect these fine phonetic details in production. For example, Warner and Tucker (2011) find that plosive reduction rates differ clearly between careful and spontaneous speech. Consequently, Ernestus and Warner (2011) and Tucker and Ernestus (2016) argue that casual, spontaneous speech is most suitable for investigating effects that pertain to morphological processing and speech production. The radio speech from the Boston University Radio Speech Corpus that we used in this analysis may not meet these criteria in full. It is probably more casual and spontaneous than the experimentally derived utterances from Kunter and Plag (2016), but probably less natural than speech corpora such as the Buckeye corpus (Pitt et al. 2007) or the Switchboard corpus (Godfrey and Holliman 1997). Thus, the present research may need validation by data from sources like the latter, even if, due to the very casual nature especially of the latter corpus, triconstituent compounds may be much more rare than in the radio speech corpus, which is one of the strengths of the condensed radio style found in the Boston corpus. We expect that with an increasing level of casualness, the phenomena discussed in detail below especially with regard to the bigram frequency become more pronounced because the speakers are less likely to self-monitor themselves, and more likely to produce phonetically reduced forms. However, this expectation needs to be tested in future research.

*Conclusion.* Our investigation of the constituent duration in a set of NNN compounds has revealed systematic differences between left- and right-branching compounds. This is a finding that is difficult to reconcile both with phonological theories as well as theories of speech production, but it agrees well with other recent research that reveals a rich interplay between morphological structure and the phonetic signal. In particular, we find a complex interaction between the branching direction and the two involved bigram frequencies. This interaction can only partially be explained as the result of phonetic reduction, as we find clear evidence of a lengthening effect on the second constituent N2. We interpret this lengthening effect as a strategic attempt of the speaker to disambiguate those compounds in which the bigram frequency signals a different branching direction than the intended one. In addition, we also find that constituents with a dense phonological neighborhood are lengthened. Taken together, these findings are in line with approaches which argue that durations in the acoustic signal can be manipulated during production in order to react to changes in the information content.

## References


Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

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(Pt 3):349–371.

Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *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, Cambridge, UK and New York.

Baayen, R. H. (2010). The directed compound graph of English: an exploration of lexical connectivity and its processing consequences. In Olsen, S., editor, *New Impulses in Word-Formation*, Linguistische Berichte, pages 383–402. Buske, Hamburg.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459.

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.

Bauer, L. (2017). *Compounds and compounding*. Cambridge Studies in Linguistics. Cambridge University Press.

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., 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.

Bell, M. J., Ben Hedia, S., and Plag, I. (2020). How morphological structure affects phonetic realisation in English compound nouns. *Morphology*, 31(2):87–120.

Ben Hedia, S. and Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, 62:34–49.

Bergmann, P. (2018). The phonological word in german – insights from an acoustic-phonetic study of complex words. In Ulbrich, C., Werth, A., and Wiese, R., editors, *Empirical Approaches to the Phonological Structure of Words*, pages 13–44. De Gruyter.

Bermúdez-Otero, R. (2012). The architecture of grammar and the division of labor in exponence. In Trommer, J., editor, *The morphology and phonology of exponence*, Oxford studies in theoretical linguistics, pages 8–83. Oxford Univ. Press, Oxford.

Bermúdez-Otero, R. (2018). Stratal phonology. In Hannahs, S. J. and Bosch, A. R. K., editors, *The Routledge handbook of phonological theory*, Routledge handbooks, pages 100–134. Routledge, London and New York, NY.

Blazej, L. J. and Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of experimental psychology. Human perception and performance*, 41(1):50–68.

Boersma, P. and Weenink, D. (2017). Praat: Doing phonetics by computer. `http://www.praat.org/`.

Box, G. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, 26(2):211–256.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *The Journal of the Acoustical Society of America*, 117(6):3867–3878.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. MIT Press, Cambridge, Mass., 1. mit press paperback. ed., 4. print edition.

Davies, M. (2008). The Corpus of Contemporary American English (COCA). `https://www.english-corpora.org/coca/`.

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

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–46.

Ernestus, M. and Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3):253–260.

Farnetani, E., Torsello, C. T., and Cosi, P. (1988). English compound versus non-compound noun phrases in discourse: an acoustic and perceptual study. *Language and speech*, 31 ( Pt 2):157–180.

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

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

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

Giegerich, H. J. (1999). *Lexical Strata in English*. Cambridge University Press, Cambridge.

Godfrey, J. J. and Holliman, E. (1997). *Switchboard-1 Release 2*. Linguistic Data Consortium, Philadelphia, PA.

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

Hall, T. A. and Kleinhenz, U., editors (1999). *Studies on the Phonological Word*. Current Issues in Linguistic Theory. John Benjamins Publishing Company, Amsterdam.

Harrell, F. E. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer eBook Collection Mathematics and Statistics. Springer, New York, NY.

Hay, J. (2007). The phonetics of ‘un’. In Munat, J., editor, *Lexical creativity, texts and contexts*, volume 58 of *Studies in Functional and Structural Linguistics*, pages 39–57. Benjamins, Amsterdam.

Hay, J. and Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory*, 22(3):565–596.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

Jaeger, T. F. (2011). Corpus-based research on language production: Information density and reducible subject relatives. In Bender, Emily M., Arnold, Jennifer E., editor, *Language from a cognitive perspective: Grammar, usage, and processing ; studies in honor of Thomas Wasow*, CSLI lecture notes. CSLI Publ, Stanford, Calif.

Jaeger, T. F. and Tily, H. (2011). On language 'utility': processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews. Cognitive Science*, 2(3):323–335.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer, New York, 6th printing 2015 edition.

Johnson, K., Ladefoged, P., and Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94(2 Pt 1):701–714.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. L. and Hopper, P. J., editors, *Frequency and the emergence of linguistic structure*, volume 45 of *Typological Studies in Language*, pages 229–254. John Benjamins Publishing Company, Amsterdam.

Kiparsky, P. (1982a). From Cyclic Phonology to Lexical Phonology. In van der Hulst, H. and Smith, N., editors, *The structure of phonological representations (Part 1)*, pages 131–175. Foris Publications, Dortrecht.

Kiparsky, P. (1982b). Lexical morphology and phonology. In Linguistic Society of Korea, editor, *Linguistics in the morning calm*, pages 3–99. Hanshin Pub. Co, Seoul.

Kiparsky, P. (2015). Stratal OT: A synopsis and FAQs. In Hsiao, Y. E. and Wee, L.-H., editors, *Capturing phonological shades within and across languages*, pages 2–44. Cambridge Scholars Publishing, Newcastle upon Tyne.

Kösling, K., Kunter, G., Baayen, R. H., and Plag, I. (2013). Prominence in triconstituent compounds: Pitch contours and linguistic theory. *Language and speech*, 56(Pt 4):529–554.

Kösling, K. and Plag, I. (2009). Does branching direction determine prominence assignment? an empirical investigation of triconstituent compounds in English. *Corpus Linguistics and Linguistic Theory*, 5(2).

Kunter, G. (2011). *Compound stress in English: The phonetics and phonology of prosodic prominence*, volume 539 of *Linguistische Arbeiten*. De Gruyter, Berlin.

Kunter, G. (2017). Coquery: A free corpus query tool. `https://www.coquery.org`.

Kunter, G. and Plag, I. (2016). Morphological embedding and phonetic reduction: The case of triconstituent compounds. *Morphology*, 26(2):201–227.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.

Ladefoged, P. and Ferrari Disner, S. (2012). *Vowels and consonants*. Blackwell Publishing Ltd, Chichester, 3 edition.

Ladefoged, P. and Johnson, K. (2011). *A course in phonetics*. Wadsworth, Boston, 6 edition.

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

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America*, 51(6B):2018–2024.

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

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2):249–336.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the h&h theory. In Hardcastle, W. J. and Marchal, A., editors, *Speech Production and Speech Modelling*, pages 403–439. Springer Netherlands, Dordrecht.

Morrill, T. (2012). Acoustic correlates of stress in English adjective-noun compounds. *Language and speech*, 55(Pt 2):167–201.

Munson, B. and Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, 47(5):1048–1058.

Ostendorf, M., Price, P., and Shattuck-Hufnagel, S. (1997). *Boston University Radio Speech Corpus: Speech and annotations*. Linguistic Data Consortium, Philadelphia, PA.

Pitt, M., Dilley, L. C., Johnson, K., Kiesling, S., Raymond, W. D., Hume, E., and Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech*. Department of Psychology, Ohio State University, Columbus, OH, 2nd release edition. `https://buckeyecorpus.osu.edu/`.

Plag, I. and Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language*, 85(1):109–152.

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53(1):181–216.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (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. `https://www.R-project.org`.

Raffelsiefen, R. (1999). Diagnostics for prosodic words revisited. In Hall, T. A. and Kleinhenz, U., editors, *Studies on the Phonological Word*, volume 174 of *Current Issues in Linguistic Theory*, page 133. John Benjamins Publishing Company, Amsterdam.

Raymond, W. D., Dautricourt, R., and Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18(01):333.

Säfken, B., Rügamer, D., Kneib, T., and Greven, S. (2021). Conditional model selection in mixed-effects models with cAIC4. *Journal of Statistical Software*, 99(8):1–30.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11(3):563–605.

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

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423.

Smith, B. L. (2002). Effects of speaking rate on temporal patterns of English. *Phonetica*, 59(4):232–244.

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

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

Suárez, L., Tan, S. H., Yap, M. J., and Goh, W. D. (2011). Observing neighborhood effects without neighbors. *Psychonomic Bulletin & Review*, 18(3):605–611.

Tomaschek, F., Hendrix, P., and Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71:249–267.

Tremblay, A. and Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2):302–324.

Tucker, B. V. and Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon*, 11(3):375–400.

Turk, A. E. (2010). Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology*, 1(2):227–262.

Turk, A. E. and Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, 99(6):3782–3790.

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

Turk, A. E. and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35(4):445–472.

Turk, A. E. and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27(2):171–206.

Uhmann, S. (1992). Contextualizing relevance: On some forms and functions of speech rate changes in everyday conversation. In Auer, P. and Di Luzio, A., editors, *The contextualization of language*, pages 297–336. John Benjamins, Amsterdam.

Warner, N. and Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3):1606–1617.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, 63-64:38–54.

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.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In Local, J., Ogden, R., and Temple, R., editors, *Phonetic Interpretation: Papers in Laboratory Phonology VI*, Papers in Laboratory Phonology, page 75–87. Cambridge University Press.

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

Yarkoni, T., Balota, D., and Yap, M. (2008). Moving beyond Coltheart's N. a new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15:971–979.

Zirkel, L. (2010). Prefix combinations in English: structural and processing factors. *Morphology*, 20(1):239–266.