




# How morphological structure affects phonetic realisation in English compound nouns

Melanie J. Bell<sup>1,2</sup>  · Sonia Ben Hedia<sup>2</sup> · Ingo Plag<sup>2</sup>

Received: 11 February 2019 / Accepted: 8 January 2020  
© The Author(s) 2020

**Abstract** Many studies have shown that syntagmatic and paradigmatic aspects of morphological structure may have an impact on the phonetic realisation of complex words (e.g. Cohen 2014a,b; Kuperman et al. 2007; Lee-Kim et al. 2013; Lõo et al. 2018; Plag et al. 2017; Schuppler et al. 2012; Smith et al. 2012; Sproat and Fujimura 1993; Zimmermann 2016, among many others). The majority of these studies have been concerned with affixes, often focusing on the acoustic properties of segments at a morphological boundary. The present study extends this line of investigation to compounds, exploring the extent to which consonant duration at compound-internal boundaries in English is dependent on morphological structure. Three competing hypotheses about the relationship between fine phonetic detail and morphological structure are tested. According to the Segmentability Hypothesis, greater morphological segmentability, i.e. a stronger morphological boundary, leads to acoustic lengthening (Ben Hedia and Plag 2017; Hay 2003; Plag and Ben Hedia 2018). The Informativity Hypothesis, on the other hand, states that higher informativity leads to lengthening (e.g. Jurafsky et al. 2001; van Son and Pols 2003). Finally, the Paradigmatic Support Hypothesis says that stronger paradigmatic support leads to lengthening (Cohen 2014b; Kuperman et al. 2007). To test these hypotheses, an experimental study was carried out using 62 compound types taken from the British National Corpus. The compounds were spoken by 30 speakers, yielding more than 1500 acoustic tokens overall. The data provide no support for the Segmentability Hypothesis, and only limited support for the Informativity Hypothesis. In contrast, the Paradigmatic Sup-

---

✉ M.J. Bell  
[melanie.bell@aru.ac.uk](mailto:melanie.bell@aru.ac.uk)

S. Ben Hedia  
[sonia.ben.hedia@uni-duesseldorf.de](mailto:sonia.ben.hedia@uni-duesseldorf.de)

I. Plag  
[ingo.plag@uni-duesseldorf.de](mailto:ingo.plag@uni-duesseldorf.de)

<sup>1</sup> Anglia Ruskin University, Cambridge, UK

<sup>2</sup> Heinrich-Heine University of Düsseldorf, Düsseldorf, Germany

port Hypothesis makes correct predictions: consonant duration at compound-internal boundaries is positively correlated with the probability of the relevant consonant following the first noun, and the duration of compound-internal geminate consonants is negatively correlated with the family size of the first noun. In other words, longer durations are associated with lower paradigmatic diversity.

**Keywords** Morphological boundaries · Phonetic realisation of complex words · Paradigmatic support · Gemination · Compound words · Compound constituent families · Informativity · Morphological boundary strength · Morphological segmentability

## 1 Introduction

Many studies have shown that syntagmatic and paradigmatic aspects of morphological structure may have an impact on the phonetic realisation of complex words (e.g. Cohen 2014a,b; Kuperman et al. 2007; Lee-Kim et al. 2013; Lõo et al. 2018; Plag et al. 2017; Schuppler et al. 2012; Smith et al. 2012; Sproat and Fujimura 1993; Zimmermann 2016, among many others). By ‘syntagmatic’ we mean the relationship between elements that occur in linear order in a stretch of speech or writing, while by ‘paradigmatic’ we mean the relationship of a given element to elements in absentia. This notion of ‘paradigm’ covers not only the classical inflectional paradigm but also other morphologically-related sets of words, including morphological categories, such as all words with the suffix *-ness*, and morphological families, such as all derived words containing a certain base, or all compounds that share a particular left or right constituent.

The majority of such studies have been concerned with inflectional and derivational affixes, often focusing on the acoustic properties of the segments at a morphological boundary (e.g. Lee-Kim et al. 2013; Smith et al. 2012). Investigations of morphologically induced phonetic variation in compounds are still rare but studies like Kuperman et al. (2007) or Kunter and Plag (2016) suggest that these types of complex words show similar effects. The present study extends this line of research by investigating the question of how consonant duration at compound-internal boundaries in English is dependent on morphological structure.

Insights into the relationship between morphological structure and phonetic implementation have important implications for theories of the mental lexicon and for theories of speech production, perception and comprehension. Strictly feed-forward models of speech production (such as Levelt et al. 1999) and theoretical models of the interaction between morphology and phonology (e.g. Bermúdez-Otero 2018; Kiparsky 1982) rely on a distinction between lexical phonology on the one hand, and post-lexical phonology and phonetics on the other. These models exclude the possibility that information about morphological structure influences the phonetic realisation of words, since they posit that this information is not available at the articulation stage. Such theories are therefore incompatible with the findings mentioned above.

There have been several attempts to explain the unexpected phonetic effects of morphological structure and to reconcile these findings with established ideas in

various fields. The literature includes three main types of hypothesis, which we characterise as the Segmentability Hypothesis, the Informativity Hypothesis and the Paradigmatic Support Hypothesis. According to the Segmentability Hypothesis (originating in the work of Hay 2003), the strength of a morphological boundary has an effect on phonetic implementation: higher morphological segmentability, i.e. a stronger morphological boundary, leads to acoustic lengthening (Ben Hedia and Plag 2017; Hay 2007; Plag and Ben Hedia 2018). The Informativity Hypothesis, on the other hand, states that linguistic units that convey more information are longer than similar units that convey less information (e.g. Jurafsky et al. 2001; van Son and Pols 2003); this has been shown for different kinds of linguistic units, including morphological units (see Hanique and Ernestus 2012 for discussion). Finally, the Paradigmatic Support Hypothesis takes the structure of the morphological paradigm as its starting point and says that stronger paradigmatic support for an element in the paradigm (i.e. greater relative frequency compared to other members of the paradigm) leads to acoustic lengthening of that element (Cohen 2014b; Kuperman et al. 2007).

In this paper, we test the three hypotheses by studying the duration of consonants at compound-internal boundaries in English. An experimental study was carried out with 62 compound types taken from the British National Corpus, spoken by 30 speakers, yielding more than 1500 acoustic tokens overall. The data provide no support for the Segmentability Hypothesis, and only limited support for the Informativity Hypothesis. In contrast, the Paradigmatic Support Hypothesis makes correct predictions: consonant duration at compound-internal boundaries is positively correlated with the probability of the relevant consonant following the first noun, and the duration of double consonants across compound-internal boundaries is negatively correlated with the family size of the first noun. In other words, longer durations are associated with lower paradigmatic diversity.

The paper is structured as follows. In Sect. 2 we discuss in more detail the theoretical underpinnings of this study, developing the hypotheses to be tested. Section 3 introduces our methodology, which is followed by the presentation of our results in Sect. 4. Section 5 discusses the theoretical implications of our findings.

## 2 Morphological structure and phonetic realisation

As mentioned in the introduction, recent research on morphologically complex words has found evidence for correlates of morphological structure in the speech signal, i.e. in the way complex words are pronounced. Most such studies have focused on durational properties, but other aspects have also been investigated, including e.g. vowel formants, centre of gravity and velarisation. This line of research is important because it puts to the test theories in different areas of linguistics: morphological theory (in particular theories of the interaction between morphology and phonology), theories of the mental lexicon (i.e. the representation and processing of complex words), and theories of speech production and perception. For all these theories there is a rather difficult problem to solve: how can morphological properties (e.g. the size of an inflectional paradigm or the strength of a suffix boundary) influence articulation in such a way that these properties have reflexes in the acoustic make-up of complex words?

While the details of a solution to this problem are still largely unclear (and will probably be out of reach for some time to come), there are at least three approaches that have something to say about a possible relationship between morphological make-up and phonetic detail: morphological segmentability, informativity, and paradigm structure.<sup>1</sup> We will discuss each in turn.

## 2.1 Morphological segmentability

It is a general and well-established assumption across theoretical camps that there are weaker and stronger morphological boundaries. The strength of a boundary is usually diagnosed by a syndrome of structural, semantic and phonological properties. Weaker morphological boundaries are characterised by lower productivity of the category in question, more bound bases, greater semantic opacity and enhanced phonological integration. At the phonological level, words with weaker boundaries show morpho-phonological alternations such as stress shift, resyllabification or assimilation. One theory that attempts to account for these phenomena is Lexical Phonology (e.g. Bermúdez-Otero 2018; Kiparsky 1982), where different lexical strata are posited to account for observable differences in boundary strength, and boundary strength is taken to be categorical: level 1 or level 2.

The assumption that morphological boundaries vary in strength is in line with dual route models of morphological processing, i.e. with models that allow both whole-word storage and morphological decomposition. Hay (2003) argued that words with a strong boundary are more likely to be segmented and their constituent morphemes processed individually, while words with a weak boundary are more likely to be processed holistically. In contrast to Lexical Phonology, Hay's (2003) approach takes boundary strength to be gradient, influenced by parameters such as semantic transparency, phonological transparency, and the relative frequency of the complex word and its base. Phonetically, words with weaker boundaries are expected to show more phonetic reduction across the morpheme boundary than words that have a strong boundary. For example, in contrast to less frequent and more easily-segmentable derived words, such as *imprison-ment* or *compact-ly*, high-frequency words like *government* or *exact-ly* show stronger reduction effects, such as the loss of the second syllable in *government*, or of the /t/ in *exactly* (cf. Hay 2003).

A number of empirical studies have suggested that morphological segmentability systematically affects the acoustic realisation of complex words at the level of individual phonemes. Sproat and Fujimura (1993) and Lee-Kim et al. (2013), for example, showed that the realisation of English /l/ depends on the strength of the morphological boundary it occurs at. Stronger boundaries go together with longer duration and stronger velarisation of /l/. In a study of English *-ly*-suffixed words, Hay (2003) found less acoustic reduction of base-final /t/ in more segmentable derivatives than in less segmentable derivatives. Similarly, Ben Hedia and Plag (2017) showed that a higher degree of segmentability correlates with less reduction in prefixed words. In their study on three English prefixes, locative *in-*, negative *in-* and *un-*, they found

<sup>1</sup>There is yet a fourth approach, prosodic phonology. In this approach different types of prosodic boundaries may correlate with different phonetic properties. Since the present paper deals with only one type of prosodic structure, i.e. compounds, we do not test this approach, and refrain from a discussion.

that the least segmentable prefix, locative *in-*, features the shortest nasal, and the most segmentable prefix, *un-*, features the longest nasal.

Other studies have investigated effects of morphological segmentability on acoustic realisation at the level of morphemes. For example, Smith et al. (2012) investigated prefixed words with *dis-* and *mis-*. In both categories, there are highly segmentable words, such as *mistime*, *mistype*, *displeased* and *discoloured*, called ‘prefixed’ by these authors, and less easily-segmentable ones, such as *mistake*, *discovered* and *distorted*, which they called ‘pseudo-prefixed’. The analysis of different phonetic characteristics, namely duration, formant structure, amplitude and spectral moments, showed that the prefixes in the pseudo-prefixed words have shorter durations than in the prefixed words, and that segments straddling a weaker morphological boundary have phonetic characteristics that are closer to those of morpheme-internal sequences of the same type. Hay (2007) found similar results for the prefix *un-*, i.e. the prefix was less reduced in more segmentable derivatives than in less segmentable derivatives. The results of these studies were corroborated by Plag and Ben Hedia (2018), who found that the durations of both *dis-* and *un-* are positively correlated with the segmentability of the word in which the prefix occurs.

While there is evidence that a higher degree of segmentability is associated with less phonetic reduction in some complex words, the scope and nature of these effects is currently unclear. For example, although Plag and Ben Hedia (2018) found a segmentability effect for the affixes *dis-* and *un-*, they found no comparable effect for *in-* or *-ly*. Similarly, Hay (2007) observed a segmentability effect only for certain speakers, while others showed no such effect. Furthermore, a number of studies have failed to find any effect at all of segmentability on the acoustics of complex words (e.g. Bürki et al. 2011; Schuppler et al. 2012).

As discussed in the overviews by Hanique and Ernestus (2012) and Ben Hedia (2019), the seemingly contradictory findings across studies might be caused by the application of different segmentability measures, as well as by differences in the structures investigated. Some studies have looked at suffixed words (e.g. Hay 2003; Schuppler et al. 2012), others at prefixed words (Ben Hedia and Plag 2017); some have looked at effects at the morpheme level (Hay 2007; Plag and Ben Hedia 2018), others at the segment level (Ben Hedia and Plag 2017; Bürki et al. 2011; Hay 2003); and some have looked at pre-boundary reduction (Ben Hedia and Plag 2017; Hay 2003), while others have looked at post-boundary reduction (Schuppler et al. 2012). There have been studies investigating inflection (e.g. Bürki et al. 2011; Schuppler et al. 2012), and others investigating derivation (Ben Hedia and Plag 2017; Hay 2007). However, no previous studies have investigated the phonetic effects of segmentability in compounds.

Although there is no empirical work specifically investigating the effect of segmentability on the acoustics of compounds, there are a number of studies that provide evidence that the semantic transparency of a compound, which is a kind of segmentability measure, affects the way it is processed. For example, Ji et al. (2011) found that when meaning decomposition is encouraged, transparent compounds are processed faster than opaque compounds. For Dutch, Zwitserlood (1994) found that, unlike semantically transparent compounds, semantically opaque compounds do not prime the associates of their constituents. According to the author, these results indicate that on a semantic level, transparent compounds are linked to their constituents

while opaque compounds are not. In a similar vein, MacGregor and Shtyrov (2013) argued that semantically transparent compounds are processed via their parts while semantically opaque compounds are processed as a whole.

One might hypothesise that the alleged differences in processing between semantically transparent and semantically opaque compounds would be mirrored in their articulation. Transparent, more segmentable compounds, processed via their parts, would show less acoustic reduction than opaque, less segmentable compounds, processed as a whole. This would fit with the segmentability effects found on the acoustics of derived words, where a higher degree of segmentability leads to less acoustic reduction. Some indirect support for this hypothesis comes from a study by Kunter and Plag (2016). In their study of triconstituent English compounds (e.g. [*day care center*]), the authors investigated whether the internal bracketing of compounds affects the acoustic realisation of the constituents. It was found that the greater the bigram frequency of the complex constituent (e.g. *day care*), the longer the duration of the third constituent (e.g. *center*), and the shorter the duration of the embedded constituent next to it (e.g. *care*). Assuming both that more frequent bigrams have weaker internal boundaries and that the weaker the internal boundary of a complex constituent the stronger its boundary with the remaining constituent, Kunter and Plag (2016) argued that the durational properties of the constituents straddling the boundary at the immediate constituent level (e.g. between *day care* and *center*) are indicative of the strength of that boundary.

In the present study we tested the idea that factors facilitating morphological segmentation lead to phonetically longer pronunciations, using English compounds as our data. We focussed our attention on what happens at the internal boundary of a compound, specifically on the duration of consonants at this boundary. We tested the ‘Segmentability Hypothesis’, which we specify for our purposes as in (1):

(1) **Segmentability Hypothesis**

The more segmentable a compound, the longer the duration of consonants at the compound-internal boundaries.

## 2.2 Informativity

Many studies have shown that the amount of information conveyed by a linguistic unit, i.e. its informativity, affects its phonetic realisation. Speakers pronounce words faster, i.e. with shorter duration, when they are contextually expected and therefore add little information to the given context. The more informative a unit is, the less reduction one finds. This has been shown for different types of units: individual segments (van Son and Pols 2003; van Son and van Santen 2005), syllables (Aylett and Turk 2006), morphemes (Cohen 2014b; Torreira and Ernestus 2012) and words (Bell et al. 2009; Jurafsky et al. 2001; Seyfarth 2014). For example, in a study on Spanish word-final /s/, Torreira and Ernestus (2012) found that /s/ suffixes in redundant morphosyntactic contexts were more likely to reduce than other word-final /s/ segments, e.g. the /s/ in *cuatro cosas* ‘four things’ is shorter than the /s/ in *quiero cosas* ‘I want things’. Working on English final /s/, Cohen (2014b) found a similar effect, in that third person singular -s is pronounced with shorter duration if it is contextually more probable.

While the studies just mentioned looked at the syntagmatic dimension of informativity, i.e. at the information added by a string in its syntagmatic context, there is also some work that has looked at informativity on the paradigmatic axis. In their study of Dutch complex words ending in *-igheid*, pronounced /əxɦeɪt/, Pluymaekers et al. (2010) measured the informativity of the cluster /xh/ in terms of the extent to which it reduces the cohort of possible words given the preceding portion of the word in question. In cases where *-igheid* constitutes a single suffix, the preceding portion is itself a possible word, and so the /xh/ is highly informative in terms of signalling continuation. In contrast, where the final suffix is *-heid*, the portion before /əxɦeɪt/ is not a possible word; the only option is continuation with /əx(h)/ and the cluster is therefore relatively uninformative. Pluymaekers et al. (2010) found that the acoustic duration of /xh/ was shorter in the contexts where it was less informative.

The studies mentioned in the previous two paragraphs all used probabilistic measures of informativity in the spirit of Information Theory (Shannon 1948). Others have used a combination of probabilistic and semantic measures. For example, Ben Hedia (2019) claimed that the acoustic realisation of morphological geminates in English affixed words (e.g. /nn/ in *unnatural*) depends on the informativity of the affix, including the extent to which it has a stable transparent meaning (e.g. *un-* always has a negative meaning and is therefore taken to be more informative than *in-*, which can be negative or locative). She found that the more informative an affix is, the longer is the duration of the double consonant in relevant words with that affix.

In the specific context of compounds, there is no empirical work on the influence of informativity on acoustic duration. However, Bell and Plag (2012) found that the probability of perceptual prominence falling on a given constituent in an English noun-noun compound is correlated with the informativity of that constituent. Using both probabilistic and semantic measures of informativity, they showed that more informative constituents are more likely to be perceived as prominent. And since length is known to be one of the factors contributing to prominence, it seems reasonable to assume that the more prominent, more informative constituents are likely to be longer.

To sum up, there is compelling evidence that more informative linguistic units are less reduced than less informative units. For our study, this leads to the ‘Informativity Hypothesis’ spelled out in (2):

(2) **Informativity Hypothesis**

The more informative a compound constituent, the longer the duration of consonants in that constituent.

### 2.3 Paradigmatic support

There is ample evidence that paradigmatic structure plays an important role in the processing of inflected words, derived words and compounds (see, for example, Baayen et al. 1997; Milin et al. 2009b on inflection, Kuperman et al. 2010; Milin et al. 2009a; Schreuder and Baayen 1997 on derivation, Kuperman et al. 2008, 2010; van Jaarsveld et al. 1994 on compounds). The effects of paradigmatic structure on processing are usually measured using numerical predictors gleaned from a word’s paradigm. Such measures can be the size of the paradigm, the number of competing forms in the paradigm, the relative frequency of a given form, or the entropy

of the paradigm as a whole. Using such measures, several studies have found that paradigmatic structure also affects pronunciation, and two opposite effects have actually been reported: enhancement and reduction. ‘Reduction’ is the paradigmatic informativity effect discussed in the previous subsection (Pluymaekers et al. 2010). ‘Enhancement’, on the other hand, refers to effects in which greater paradigmatic support (i.e. higher paradigmatic probability) is associated with longer durations or more distinct pronunciations. Interestingly, this effect works in the opposite direction from the reduction effects associated with lower informativity.

An early report of paradigmatic support was that of Kuperman et al. (2007), who investigated the interfixes *-s-* and *-en-* in Dutch compounds; they found that the more probable an interfix is within the relevant paradigm, the longer its duration. Cohen (2014b) not only found the syntagmatic effect described in the previous subsection, but also discovered a paradigmatic enhancement effect on the duration of the third person singular suffix *-s* in English verbs. The more frequent the third person singular form of a verb relative to its plural (unsuffixed) form, i.e. the greater the paradigmatic probability of the suffixed form, the longer the suffix. Similarly, investigating vowel formants in Russian verb suffixes, Cohen (2015) demonstrated that with rising paradigmatic probability of the verb form in question, the vowels are pronounced more distinctly. For Estonian case-inflected nouns, Lõo et al. (2018) found that the duration of a word form was positively correlated with its paradigmatic probability, both in terms of the number of inflectional variants in use and in terms of the number of derivatives: nouns with fewer attested inflectional forms and smaller morphological families were associated with longer durations.

In the present study, the idea of paradigmatic enhancement was tested in terms of the ‘Paradigmatic Support Hypothesis’ in (3):

(3) **Paradigmatic Support Hypothesis**

The greater the degree of paradigmatic support for a compound, the longer the duration of consonants at the compound-internal boundaries.

### 3 Methodology

#### 3.1 Data

We investigated the duration of consonants at compound-internal boundaries, for example:

(4) *steam engine*

(5) *cream mini*

(6) *survey manager*

The consonant is either part of the first noun (‘N1’), as in (4), the second noun (‘N2’), as in (6), or of both, as in (5). This allowed us to test which factors affect which part of the boundary. In other words, if there is reduction, does it take place before the boundary, after the boundary or at both sides of the boundary?



We especially wanted to include compounds such as *cream mini*, with a double consonant at the boundary, to maximise our chances of finding a paradigmatic enhancement effect. The only previous report of such an effect for compounds is that of Kuperman et al. (2007), who found a paradigmatic enhancement effect on the duration of interfixes in Dutch compounds. Although English does not have interfixes, we reasoned that we might see a similar effect on the segments at compound-internal boundaries, perhaps especially on morphological geminates, since in such cases a single articulatory gesture straddles the boundary. Geminates may therefore be subject to influence by the lexical properties of both constituents, just as interfixes are. In the present study we focus on the consonants /m/, /n/ and /s/, since it has been shown (e.g. by Ben Hedia 2019) that these sounds may show clear phonetic effects of morphological gemination in English.

English compounds show considerable variation in orthographic representation between spaced, hyphenated and unspaced spellings. However, unspaced and hyphenated spellings tend to correlate with high frequency and lexicalisation (see discussion in Bell and Plag 2012). In order to find a sample of attested compounds with a wide range of frequencies, and at the same time avoid the complicating factor of varied spelling, we therefore decided to focus exclusively on spaced compounds.

The compounds used in the present study were selected from the spoken section of the British National Corpus. Using the spoken section of the corpus ensures that the resulting compounds have been spontaneously produced by a speaker at least once. The BNCweb interface (Hoffmann et al. 2008) was used to search for strings of two nouns, excluding strings that crossed a sentence boundary or that included a pause or any other form of interruption, such as a cough, between the two nouns. The corpus queries also specified that the word after the second noun should not be another noun, an adjective or a possessive. This restricted the searches to strings of exactly two nouns and excluded combinations that were part of a larger compound construction. The strings were subsequently checked in context to ensure that they represented constructions in which the first noun modified the second. We take the view, following e.g. Bauer (1998), Bell (2011), Plag et al. (2008), that all such constructions can be classed as compounds. At this stage, types in which the two nouns were identical or either noun was hyphenated, as well as proper names, appositive constructions and vocatives, were also excluded from the data.

Phonological transcriptions of the constituent nouns of the compounds were extracted from the CELEX lexical database (Baayen et al. 1995, henceforth CELEX) and in cases where a constituent did not appear in CELEX they were supplemented by manual transcription. These transcriptions were then used to identify types in which the first word ended with one of the consonants /s/, /m/ or /n/, and the second word started with the same phoneme. From this set, we selected only those combinations in which neither the word-final consonant nor the word-initial consonant formed part of a cluster. We also used the transcriptions to select compounds in which either the first word ended with /s/, /m/ or /n/ and the second word began with a vowel, or the second word started with one of these consonants and the first word ended with a vowel. Again, we excluded types with word-initial or word-final clusters. We further restricted ourselves to types in which, according to CELEX or our manual transcrip-

**Table 1** Examples of experimental items

<b>m</b>			
m#V	<i>calcium intake</i>	<i>problem area</i>	<i>bum ache</i>
m#m	<i>cream mini</i>	<i>pandemonium model</i>	<i>claim money</i>
V#m	<i>company money</i>	<i>media men</i>	<i>polo mint</i>
<b>n</b>			
n#V	<i>kitchen area</i>	<i>pavilion end</i>	<i>swan inn</i>
n#n	<i>pen knife</i>	<i>woman novelist</i>	<i>criterion number</i>
V#n	<i>polo neck</i>	<i>key note</i>	<i>mercury number</i>
<b>s</b>			
s#V	<i>surface area</i>	<i>bonus element</i>	<i>police act</i>
s#s	<i>dress sense</i>	<i>gas side</i>	<i>peace settlement</i>
V#s	<i>survey site</i>	<i>eye sore</i>	<i>tuna sandwich</i>

tions, the lexical stress of the second noun fell on the first syllable of that noun.<sup>2</sup> All compounds in the dataset therefore satisfy the following criteria: there is a single or a double /s/, /m/ or /n/ at the compound-internal boundary, and the relevant consonant(s) both follow a vowel and precede a stressed vowel. Some examples are shown in Table 1.

From the set of compounds described in the previous paragraph, we selected a subset to use in our study. In selecting the subset we aimed to achieve as wide and balanced a range as possible across the following criteria:

- Number of syllables in N1
- Number of syllables in N2
- Weight of final syllable of N1, strong or weak
- Expected position of compound stress, on N1 or N2
- Vowel phoneme preceding the consonant(s)
- Vowel phoneme following the consonant(s)

In other words, items were selected to enhance the diversity of the data with respect to these criteria, and avoid a bias towards any specific syllable structure, stress or vowel. On the other hand, items were excluded if they were unique in terms of any of these variables, since that would have introduced a confound between compound and condition. In cases where more than one compound satisfied all these constraints, the final selection was made randomly. These procedures resulted in a list of compounds with 19 compound types with /m/ at the boundary, 19 types for /n/, and 24 types for /s/.

### 3.2 Experimental set-up

Spoken tokens of all the compounds in our final dataset were elicited from 30 native speakers of British English, who read the compounds presented in carrier sentences

<sup>2</sup>This was not necessarily the main stressed syllable of the whole compound, only of the second constituent.

on a computer screen. Each compound was embedded in two different carrier sentences:

(7) *They talked about the [compound] again.*

(8) *She told me about the [compound].*

The two sentences differ with respect to whether or not the compound occurs in final position: this allowed for any lengthening or shortening effects of phrasal position to be included in the analysis. Each participant read each compound only once, either in sentence (7) or in sentence (8). However, overall each participant saw an equal number of both sentence types, and each compound was included in an equal number of tokens of each sentence type. The sentences were mixed with an equal number of unrelated filler sentences, which were the experimental items for another study. Since the filler sentences had a variety of different structures, they served to break up the repetitiveness of our carrier sentences and to reduce the risk of a list-like intonation developing. Each participant saw the items, including fillers, in a different randomised order.

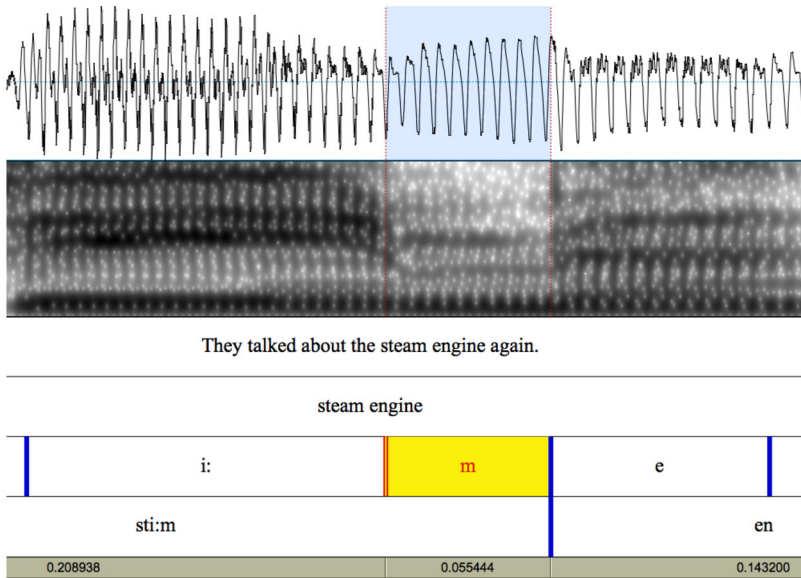
Each sentence was presented on two consecutive slides. The first slide of each pair asked the participant to read the sentence silently, while the second slide instructed them to read the sentence aloud. The silent reading phase was intended both to encourage semantic processing of the sentence and to reduce the risk of performance errors in the subsequent reading aloud. There was an initial training phase, and participants could move through the presentation at their own pace.

The recordings were produced in a sound-proof booth, digitised at 44.1 kHz using a Tascam HD-P2 digital recorder and a Sennheiser ME 64 cardioid microphone, with participants seated 15 cm from the microphone and recording levels set for each participant.

### 3.3 Acoustic measurements

After recording the sentences, we manually segmented the data and transcribed them phonetically using the software Praat (Boersma and Weenink 2014). We annotated the segments in question, as well as the preceding and following segments. The annotation for *steam engine*, for example, included the segmentation of /i:/, /m/ and /e/. The segmentation was carried out according to criteria that relied on the visual inspection of the waveforms and spectrograms of the items. These criteria were based on the segmentation criteria applied in Ben Hedia (2019), which in turn were based on the features of specific sounds as described in the phonetic literature (e.g. Ladefoged 2003).

As all the consonants occur in intervocalic position, we concentrated on the differences between the pertinent consonants and vowels. Like vowels, nasals have a regular waveform, but their formants are quite faint in comparison to those of vowels. This can be seen in Fig. 1, which shows a sample segmentation of the word *steam engine*. In contrast to vowels, fricatives have an aperiodic waveform and are therefore quite easy to identify in intervocalic position. All boundaries were set at the nearest zero crossing of the waveform. Double consonants (e.g. /mm/ in *cream mini*) were treated as one segment in the annotation when no boundary between the



**Fig. 1** Annotation of the compound *steam engine*

two identical consonants was discernible. If there was a visible boundary between the two consonants, both consonants were segmented. This was the case when the speaker produced a pause between the first and the second constituent. Such tokens were subsequently excluded from the analysis.

The reliability of the segmentation criteria was verified by a set of trial segmentations. In these trials, three annotators used the criteria to segment the same 20 items. If there was any discrepancy of more than 10 milliseconds in the placement of the boundaries, the annotators discussed the discrepancy and refined the criteria in order to reduce the amount of inter-annotator variation. These trial segmentations were repeated until all boundaries were reliably placed with only small variations (i.e. no greater than 10 milliseconds). For the final measurement, each annotator worked on a disjunct set of items. To facilitate consistency between annotators, regular meetings were held between the annotation team and the first two authors of this paper, at which we discussed any items where the annotator had a query. For these problematic items, the relevant boundaries were set by consensus and the annotation guidelines were updated to accommodate any previously unforeseen issues. As a further precaution against systematic inter-rater variability, we included annotator as a random effect in our models, though this proved not to be significant.

Tokens were excluded from further analysis if the speaker was deemed to have made a performance error, or if it was not possible to locate the relevant segment boundaries in the speech stream. This left a total of 1546 segmented compound tokens. For this set of tokens, a Python script was used to measure and extract compound duration, constituent durations, the duration of the consonants in question, and the duration of their preceding and following segments in milliseconds.

### 3.4 Predictor variables

#### 3.4.1 Overview

To test the three hypotheses under consideration, we extracted a number of frequency-based measures from ukWaC (<https://www.webarchive.org.uk/ukwa/>), a corpus of more than 2 billion words from the .uk internet domain. These included:

- **COMPOUND FREQUENCY:** the total frequency of the compound including all spelling variants (spaced, hyphenated and concatenated; British and American) and singular and plural forms of N2. We lemmatised N2 so that singular and plural forms of the same compound would be counted together, e.g. *tuna sandwich* and *tuna sandwiches*. However we did not include the plural form of N1 because we judged that plural modifiers are likely to represent different lemmas e.g. *arm chair* vs. *arms race*.
- **SPELLING RATIO:** the ratio of the number of tokens of the compound written unspaced, i.e. hyphenated or concatenated, to the number of tokens written with a space, calculated as:

$$\text{SpellingRatio} = (f(\text{concatenated}) + f(\text{hyphenated})) / f(\text{spaced})$$

- **N1 FREQUENCY and N2 FREQUENCY:** the total lemma frequency of each constituent including all spelling variants (British and American).
- **N1 FAMILY SIZE and N2 FAMILY SIZE:** the positional family size of each constituent, i.e. the number of compound types with the given constituent in the same position. For example, the N1 family of *problem area* consists of *problem behaviour*, *problem children*, *problem drinkers* etc., while the N2 family consists of *catchment area*, *subject area*, *surface area* etc. To estimate the family sizes we counted all spaced NN types with the relevant constituents, which occurred within a sentence and in which N1 was tagged as a singular/uncountable noun and N2 was tagged as a singular/uncountable or plural noun.
- **SYNTAGMATIC PROBABILITY OF N2 GIVEN N1:** the syntagmatic probability of N2 following N1, calculated as:

$$p(N2|N1)_{\text{syntagmatic}} = f(\text{compound}) / f(N1)$$

- **PARADIGMATIC PROBABILITY OF N2 GIVEN N1:** the probability of N2 following N1 when N1 is a compound modifier, calculated using the cumulative frequency of all compounds in the N1 family:

$$p(N2|N1)_{\text{paradigmatic}} = f(\text{compound}) / \sum_{i=1}^n f(N1\text{-compound}_i)$$

- **PARADIGMATIC PROBABILITY OF CONSONANT GIVEN N1:** the token-based paradigmatic probability of the relevant consonant following N1 within a compound. This was calculated as the cumulative frequency of N1 compounds in which N2 began with the consonant in question, divided by the cumulative frequency of the entire N1 family:

$$p(C|N1)_{\text{paradigmatic}} = \frac{\sum_{i=1}^n f(\text{N1-compound-with-N2-initial-}C_i)}{\sum_{i=1}^n f(\text{N1-compound}_i)}$$

We also calculated the type-based version of this variable: the number of N1 compounds in which N2 began with the consonant in question, divided by the family size of N1.

- **N1 ENTROPY** and **N2 ENTROPY**: the entropies of the constituent families. Constituent family entropy is a measure of the relative expectedness of the different compounds in the family, and the overall level of uncertainty in the family. It is highest in large families, and when compounds in the family are equally frequent, i.e. when it is difficult to predict the compound on the basis of the constituent. It is lowest in small families and larger families with a few very frequent members since in these cases, given the constituent, a small number of compounds are most likely. Entropy (H) for an Nx family of n compounds was calculated as:

$$H(\text{constituent family}) = -\sum_{i=1}^n p(\text{compound}_i) * \log_2(p(\text{compound}_i))$$

where

$$p(\text{compound}) = p(Ny|Nx)_{\text{paradigmatic}} = f(\text{compound}) / \sum_{i=1}^n f(\text{Nx-compound}_i)$$

Frequency measures, spelling ratio, family sizes and probabilities were log-transformed before entering them into the statistical analysis. Let us now see how these measures relate to the three hypotheses.

### 3.4.2 Segmentability

We used spelling ratio and N1 family size to estimate the segmentability of the compounds in our dataset. These variables are related to the Segmentability Hypothesis in the following ways:

- **Spelling ratio**: Kuperman and Bertram (2013) showed that English compounds are more likely to be written spaced when their constituents are more frequent or orthographically longer. They interpret these findings as evidence for a mediating effect of what they term ‘morphemic salience’: compounds whose constituents are more salient (by virtue of frequency or length) are more likely to be written spaced. We understand this notion of constituent salience as being related to segmentability, such that more segmentable compounds have more salient constituents. We assume that the space in a spaced compound is indicative of segmentation by the writer, and that writers are more likely to include a space the more segmentable they perceive a compound to be. Unspaced representations, on the other hand, are associated with lexicalisation and suggest that the writer perceives the compound as a single conceptual unit (cf. Bell and Plag 2012). Hence, a compound with a greater proportion of spaced tokens can be regarded as more segmentable than one with fewer spaced tokens, and spelling ratio can be taken to be negatively correlated with segmentability. If the Segmentability Hypothesis is correct, consonant duration will be longer at the internal boundaries of more segmentable compounds.

**Table 2** Summary of predictions made by the Segmentability Hypothesis

Measure	Consonant type		
	N1-final	Double	N2-initial
higher spelling ratio	shorter duration	shorter duration	shorter duration
larger N1 family size	longer duration	longer duration	longer duration

Assuming that writing and reading aloud both reflect the same construct of segmentability, the hypothesis therefore predicts that spelling ratio will be negatively correlated with consonant duration in our data.

- N1 family size: Here we assume that the larger the N1 family, the more productive is N1 as a compound modifier. Greater productivity has been shown to be associated with greater segmentability of complex words (cf. Hay and Baayen 2003), hence compounds with larger N1 families should be more segmentable than compounds with smaller N1 families. The Segmentability Hypothesis therefore predicts that N1 family size will be positively correlated with the duration of consonants at the compound-internal boundary.

Table 2 summarises the predictions made by the Segmentability Hypothesis. As described in Sect. 2.1, segmentability effects have been reported for linguistic elements occurring both before and after morphological boundaries. If the Segmentability Hypothesis is correct, we would therefore expect to find the relevant effects for all the internal-boundary consonants in our data: N1-final, double and N2-initial.

### 3.4.3 Informativity

Informativity is related to the concepts of probability and expectedness. A linguistic element that is less probable in any given context is less expected in that context, and in turn more informative. A linguistic element that is highly probable is more expected, and thus less informative. The Informativity Hypothesis therefore predicts that the less probable a consonant is in a given context, the longer its realisation will be.

We tested six different types of probability: compound frequency, constituent frequencies, conditional probability of N2 given N1, N1 family size, N1 entropy, and conditional probability of the consonant in question given N1. The first five of these variables relate to expectedness at word level. We assume that, if the hypothesis is correct, compound-internal consonants inherit informativity-related length effects from the constituent and from the compound in which they occur. In other words, the less probable the compound or constituent, the longer its realisation and hence the longer the realisation of each of its segments. In contrast, the final variable (conditional probability of the consonant given N1) measures the expectedness of the consonant directly. Furthermore, some of these variables measure the probability of N1 and/or N1-final consonants, while others measure the probability of N2 and/or N2-initial consonants. The double consonants are assumed to belong partly to N1 and partly to N2, and hence to reflect the probability of both N1 and N2. Relative to these various measures, the Informativity Hypothesis makes the predictions summarised in Table 3 and described in the following paragraphs.

**Table 3** Summary of predictions made by the Informativity Hypothesis

Measure	Consonant type		
	N1-final	Double	N2-initial
higher compound frequency	shorter duration	shorter duration	shorter duration
higher N1 frequency	shorter duration	shorter duration	–
higher N2 frequency	–	shorter duration	shorter duration
higher conditional probability of N2	–	shorter duration	shorter duration
larger N1 family size	shorter duration	–	longer duration
higher N1 entropy	shorter duration	–	longer duration
higher paradigmatic probability of C	–	shorter duration	shorter duration

- **Compound frequency:** The more frequent the compound, the more expected it is in the language generally, hence the shorter its realisation and that of any consonant within it. Thus all three types of consonant, N1-final, double and N2-initial, should show negative correlation between their duration and compound frequency. We might further expect that the slope of the correlation for doubles would be steeper than for the single consonants, since both the N1 and N2 components would be affected.
- **N1 frequency and N2 frequency:** The more frequent a constituent, the more expected it is in the language generally, hence the shorter its realisation and that of any consonant within it. Thus N1-final and double consonants should show negative correlation between their duration and N1 frequency, while double and N2-initial consonants should show negative correlation between their duration and N2 frequency.
- **Conditional probabilities of N2 given N1:** The higher the conditional probability of N2, either syntagmatically or paradigmatically, the less informative is N2 given N1, hence the shorter its realisation and that of its segments. So the duration of N2-initial and double consonants should be negatively correlated both with the syntagmatic probability of N2 given N1 and with the paradigmatic probability of N2 given N1.
- **N1 family size and N1 entropy:** The larger the N1 family and the greater its entropy, the less predictable is N2 given N1, so higher values of these variables indicate that N2 is more informative. Hence the duration of N2-initial consonants should be positively correlated with both N1 family size and N1 entropy. Conversely, the smaller the N1 family and the lower its entropy, the more informative is N1 regarding the possible values of N2. Hence the duration of N1-final consonants should be negatively correlated with N1 family size and N1 entropy. The duration of double consonants would be expected to show little or no overall effect of these variables, the positive correlation with duration of the N2 element being counterbalanced by the negative correlation with the duration of the N1 element.
- **Paradigmatic probability of consonant given N1 (type-based or token-based):** Since this is the probability of the consonant following N1 within the constituent family, i.e. at the start of N2, as far as the Informativity Hypothesis is concerned, it is only relevant to double and N2-initial consonants. The greater the token-based



**Table 4** Summary of predictions made by the Paradigmatic Support Hypothesis

Measure	Consonant type		
	N1-final	Double	N2-initial
larger N1 family size	?shorter duration	shorter duration	shorter duration
higher paradigmatic probability of C	?longer duration	longer duration	longer duration

or type-based probability of N2 starting with the consonant in question, the shorter these consonants should be.

#### 3.4.4 Paradigmatic support

With the exception of Lõo et al. (2018), who measured whole-word durations, most studies that have reported paradigmatic enhancement effects have found them in suffixes or compound interfixes. The affixes concerned mainly consist of single phonemes, and it is therefore unclear whether such effects operate at the level of the morpheme or the phoneme. Because of this, we included both these levels in our analysis of consonant duration, i.e. the paradigmatic probability of the consonant itself and of the compound constituent containing it. For inflected and derived words, the relevant paradigm consists of all words that share the same stem or affix. For compounds, the only study to have reported paradigmatic enhancement is Kuperman et al. (2007), who found that the relevant paradigm is the N1 positional constituent family, i.e. all compounds that share the same first element. To test the Paradigmatic Support Hypothesis, we therefore used N1 family size and the paradigmatic probability of the consonant given N1, as follows:

- N1 family size: The larger the N1 family size, the more possible values there are for N2, hence the lower the paradigmatic support for each compound in the family. The Paradigmatic Support Hypothesis therefore predicts that an increase in N1 family size will be associated with shorter consonant durations at the compound-internal boundary.
- Paradigmatic probability of the consonant given N1 (type-based or token-based): Higher values of these variables mean that when N1 occurs as the first element of a compound it is relatively more likely to be followed by the consonant in question. In other words, higher values indicate that compounds in which N2 starts with the relevant consonant are comparatively numerous and/or frequent within the N1 constituent family. The Paradigmatic Support Hypothesis therefore predicts that an increase in the paradigmatic probability of the consonant given N1 will be associated with longer consonant durations at the compound-internal boundary.

Since paradigmatic enhancement has previously been reported mainly for suffixes and interfixes, i.e. for linguistic elements that follow a morphological boundary, it is unclear whether we should expect the effect for all our internal-boundary consonants, or just for doubles and N2-initial cases. The predictions of the Paradigmatic Support Hypothesis are summarised in Table 4.

Note that the different hypotheses make conflicting predictions about the effects of certain variables, especially N1 family size and the paradigmatic probability of the

consonant given N1. These predictors can therefore be used to test the hypotheses against one another.

### 3.4.5 Control variables

In addition to our predictors of interest, we also included a number of control variables in our models. These were:

- **BOUNDARY TYPE (C#C, C#V or V#C):** We included this variable for two reasons. Firstly, phonetic studies have shown that the duration of consonants may be influenced by the phonetic context in which they occur (e.g. Umeda 1977). Secondly, our hypotheses make different predictions for consonants in the different positions, so we expected to find interactions between boundary type and the other predictors.
- **CONSONANT (/m/, /n/ or /s/):** This variable controls for the inherent duration differences between the three consonants.
- **SPEECH RATE:** This is the local speech rate, measured as the number of segments per second. It was computed for each compound token by dividing the number of segments in the compound by the total duration of the compound in seconds. Obviously, a faster speech rate leads to shorter durations of individual segments.
- **NUMBER OF SYLLABLES IN N1 and NUMBER OF SYLLABLES IN N2:** It has been shown (e.g. by Lindblom 1963; Nootboom 1972) that segments may tend to be shorter if the words in which they occur have more syllables. This effect can be conceptualised as a kind of compression effect, where words with more syllables undergo reduction. We therefore included syllable counts of the two constituents in our set of covariates.
- **SPELLING:** This is a binary variable, coding whether or not the same orthographic consonant occurs on both sides of the constituent boundary. It takes the value ‘true’ if the same letter occurs on both sides of the boundary, e.g. *bus signal*. For all other compounds it has the value ‘false’, e.g. *peace settlement, media men, swan inn*. We included this variable because there is a well-established influence of spelling on pronunciation in literate speakers (see Damian and Bowers 2003 and references therein), so it is possible that consonants represented orthographically on both sides of the constituent boundary could see a different acoustic realisation than other alternatives.
- **PRESENTATION ORDER OF ITEMS:** This variable was included to control for effects of variability in attention or fatigue across the duration of the experiment.

## 3.5 Statistical analysis

We carried out mixed effects regression analysis using the `lme4` package in R (Bates et al. 2015). The dependent variable was **CONSONANT DURATION**, the duration of the consonant at the compound-internal boundary in milliseconds. Before analysis, we trimmed the data to remove outliers with very long or short durations and also removed outliers with respect to speech rate. This process resulted in a loss of 25 data points, about 1.6% of the data. The number of types and tokens in the dataset used for modelling is shown in Table 5.

**Table 5** Distribution of types and tokens

m	Number of types	Number of tokens
m#V	5	120
m#m	9	222
V#m	5	125
Total	19	467
n	Number of types	Number of tokens
n#V	5	101
n#n	9	246
V#n	5	120
Total	19	467
s	Number of types	Number of tokens
s#V	5	96
s#s	14	373
V#s	5	118
Total	24	587

Many of our variables of interest are highly correlated with one another, which means that they are likely to account for the same portion of variance in the dependent variable. Inclusion of collinear predictors can lead to unstable statistical models in which it is difficult to identify the effects of individual variables. Since we were primarily interested in the effects of specific predictors as a way of testing our hypotheses, we therefore needed to reduce the amount of collinearity in our models. To do this, we adopted the modelling procedure described in the following paragraphs.

In a first step, we built models with only random effects for participant, item, annotator and compound position (sentence-final or not). In the presence of random effects for participant and item, the effect of annotator was insignificant and this variable was therefore dropped from further analyses. Secondly, we added the control variables, including a three-way interaction between boundary type, consonant and speech rate. At this stage, neither the presentation order of items nor the number of syllables in either constituent turned out to be significant, and these variables were therefore also dropped. Thirdly, we modelled the effect of each predictor of interest on consonant duration in individual separate models. Each of these models also included the significant random effects and control variables, as well as three-way interactions between boundary type, consonant and the predictor, and between boundary type, consonant and speech rate. We included these interaction terms because, as described in Sect. 3.4.5, our hypotheses make different predictions for consonants at the different boundary types, and we also expected that the inherent duration differences between /m/, /n/ and /s/ might lead to differences in their slopes in relation to other predictors.

Amongst the variables listed in Tables 2 to 4, collinearity was especially high between compound frequency and the conditional probabilities of N2, between the

**Table 6** p-values of fixed effects in the final mixed-effects model fitted to the Box–Cox-transformed durations of consonants at the compound-internal boundaries

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
boundary type	0.057809	0.028905	2	67.08	62.7026	4.422e-16
consonant	0.055946	0.027973	2	421.60	60.6820	<2.2e-16
speech rate	0.067065	0.067065	1	828.70	145.4834	<2.2e-16
N1 family size	0.000295	0.000295	1	52.48	0.6407	0.4270620
N2 frequency	0.005827	0.005827	1	51.53	12.6405	0.0008191
paradigmatic prob. consonant	0.004128	0.004128	1	53.78	8.9543	0.0041733
boundary type:speech rate	0.028687	0.014344	2	998.37	31.1154	7.783e-14
consonant:speech rate	0.010461	0.005231	2	1071.44	11.3465	1.330e-05
boundary type:N1 fam. size	0.009240	0.004620	2	52.82	10.0221	0.0002042

type-based and token-based versions of consonant probability, and between N1 frequency, N1 family size and N1 entropy. In our full model, we therefore included only the variable from each of these groups that had the strongest effect on consonant duration in its individual model. The other variables from each group were dropped from the analysis. After this process, the variables of interest remaining in our full model were compound frequency, N1 family size, spelling ratio, N2 frequency and the token-based paradigmatic probability of the consonant given N1. This set of variables was checked for any remaining collinearity by using the `collin.fnc` function of the `LanguageR` package (Baayen and Shafaei-Bajestan 2019), which produced an acceptable condition number of about 27.15 (according to Baayen 2008, condition numbers of 30 and more may indicate potentially harmful collinearity).

Starting with the full model described above, and including the significant interactions from the individual models, we used the `step` function of the `lmerTest` package (Kuznetsova et al. 2017) to eliminate non-significant fixed effects and to select the optimal random effects structure. Inspection of the resulting model revealed that the residuals had an unsatisfactory, i.e. non-normal, distribution. To address this problem, starting again from the full model, we used Box-Cox transformation (Box and Cox 1964, Venables and Ripley 2002) to identify a suitable transformation parameter ( $\lambda$ ) for a power transformation of the dependent variable. The optimal value of  $\lambda$  was found to be  $\lambda = 0.5454545$ . This transformation was therefore applied, and non-significant effects were again removed using the `step` function. Finally, we removed data points whose standardised residuals had an absolute value greater than 2.5 standard deviations, which resulted in the loss of 1.9% of observations. The resulting final model had normally distributed residuals (Shapiro-Wilk normality test,  $W = 0.99818$ ,  $p = 0.1038$ ).

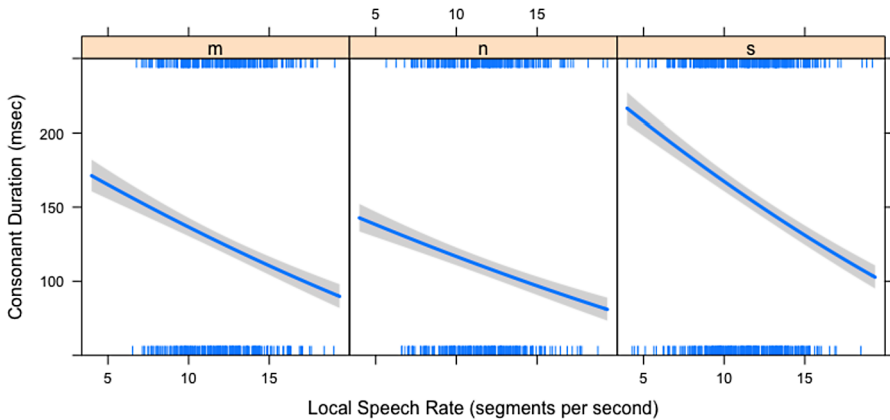
## 4 Results

The final model includes random intercepts for item and participant, plus five significant fixed effects, including three two-way interaction terms. The model is documented in Tables 6 and 7.

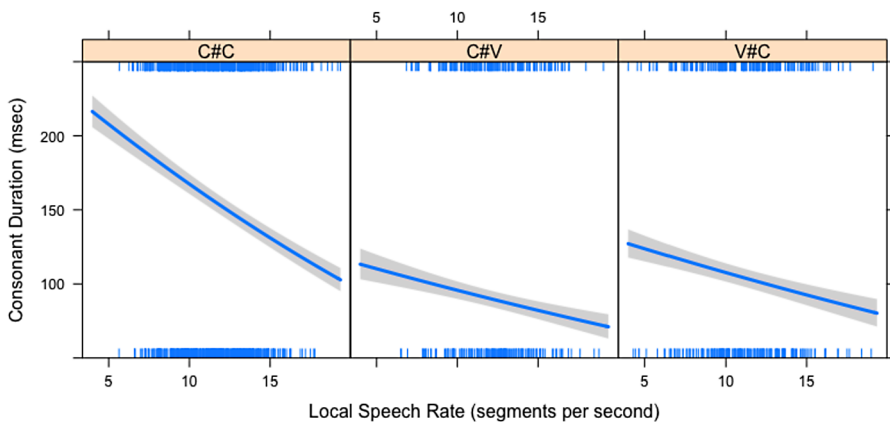
**Table 7** Fixed-effect coefficients and p-values in the final mixed-effects model fitted to the Box–Cox-transformed durations of consonants at the compound-internal boundaries

Random effects:					
Groups	Name	Variance	Std.Dev.		
item	(Intercept)	0.0001042	0.01021		
participant	(Intercept)	0.0001978	0.01407		
residual		0.0004610	0.02147		
Number of obs: 1491, groups: item, 62; participant, 30					
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>  t )
(Intercept)	3.685e-01	2.084e-02	6.990e+01	17.679	<2e-16
boundary type-C#C	2.192e-01	2.320e-02	6.882e+01	9.447	4.73e-14
boundary type-C#V	-8.345e-04	2.649e-02	6.319e+01	-0.032	0.974967
consonant-m	-6.010e-02	9.739e-03	4.209e+02	-6.172	1.59e-09
consonant-n	-1.017e-01	9.276e-03	3.212e+02	-10.964	<2e-16
speech rate	-4.686e-03	7.441e-04	5.351e+02	-6.297	6.33e-10
N1 family size	3.914e-03	2.546e-03	5.192e+01	1.537	0.130295
N2 frequency	-3.081e-03	8.667e-04	5.153e+01	-3.555	0.000819
paradigmatic prob. consonant	5.298e-03	1.771e-03	5.378e+01	2.992	0.004173
boundary type-C#C:speech rate	-4.735e-03	7.212e-04	9.068e+02	-6.565	8.72e-11
boundary type-C#V:speech rate	2.324e-04	8.567e-04	1.013e+03	0.271	0.786259
consonant-m:speech rate	2.046e-03	7.511e-04	9.414e+02	2.725	0.006558
consonant-n:speech rate	3.438e-03	7.270e-04	1.017e+03	4.729	2.58e-06
boundary type-C#C:N1 family size	-1.258e-02	3.055e-03	5.265e+01	-4.119	0.000135
boundary type-C#V:N1 family size	-2.746e-03	3.591e-03	5.201e+01	-0.765	0.447906

We will first discuss the effects of the control variables. Neither number of syllables in N1, number of syllables in N2, spelling nor presentation order of items survive in the final model. However, there are significant interactions of speech rate with both consonant and boundary type, which are shown in Figs. 2 and 3 respectively. In these and subsequent figures, the dependent variable has been back-transformed so that the y axis represents consonant duration in milliseconds. Figure 2 illustrates the effect of speech rate on consonant duration, modulated by consonant. As expected, the fricative is longer than either of the nasals, and /m/ is slightly longer than /n/. Also as expected, higher speech rates lead to shorter consonants; this shortening effect is most pronounced for /s/ and least pronounced for /n/. Figure 3 also shows the general effect of speech rate on consonant duration, as well as clear evidence of morphological gemination: double consonants are almost always longer than their singleton counterparts. However, there was no effect of spelling in the final model, suggesting that double orthographic consonants have no effect over and above phonological gemination. Initial consonants in the second constituent are marginally longer than



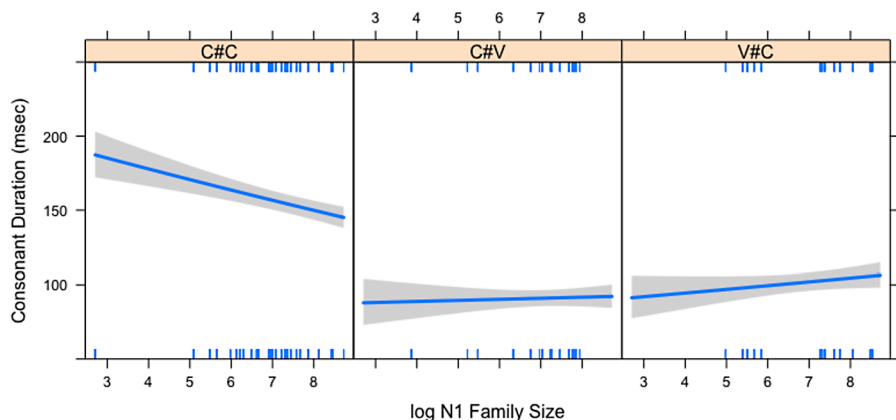
**Fig. 2** Partial effect of local speech rate by consonant in final model



**Fig. 3** Partial effect of local speech rate by boundary type in final model

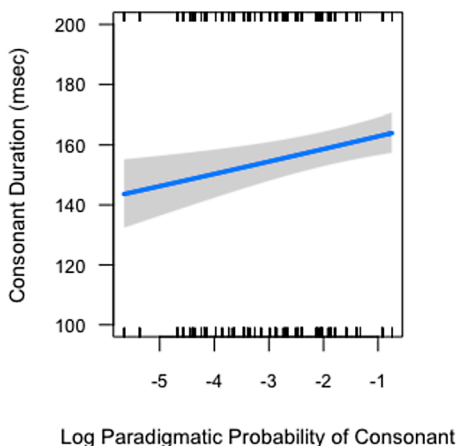
final consonants in the first constituent, possibly reflecting an effect of post-boundary lengthening similar to that reported by e.g. White et al. (2015).

We turn now to the predictors of interest, that is to those variables which are of immediate relevance to our three hypotheses. There are small but significant main effects of N2 frequency and the paradigmatic probability of the consonant given N1, as well as a significant interaction between N1 family size and boundary type. This interaction is shown in Fig. 4: it is clear that the family size effect is only really significant for the double consonants, and that the duration of these consonants is negatively correlated with N1 family size. Remember that N1 family size is the predictor variable that most clearly enables us to differentiate between the three hypotheses. According to the Segmentability Hypothesis, consonant duration would be expected to be positively correlated with N1 family size across all boundary types: the greater the family size of N1, the more productive it is as a modifier, hence the more segmentable the compound, the stronger the boundary and the longer the consonant(s).



**Fig. 4** Partial effect of N1 family size by boundary type in final model

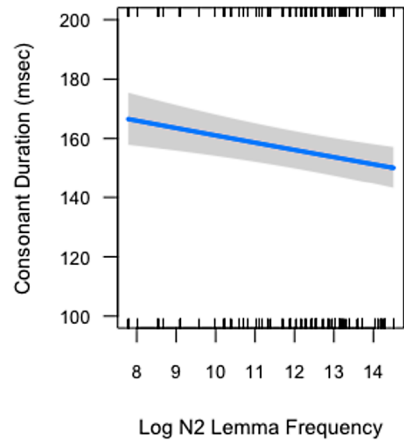
**Fig. 5** Partial effect of paradigmatic probability of consonant in final model



This is not what we find. According to the Informativity Hypothesis, on the other hand, N1 family size should be positively correlated with the duration of N2-initial consonants and negatively correlated with the duration of N1-final consonants, the two effects cancelling one another out in the case of doubles. But again, this is not what we find. On the contrary, we find that the duration of only the double consonants falls with increasing N1 family size; only the Paradigmatic Support Hypothesis predicts this effect.

The main effect of the paradigmatic probability of the consonant given N1, shown in Fig. 5, also goes in the direction predicted by the Paradigmatic Support Hypothesis: the more likely the consonant in question to follow N1 within the constituent family, the longer its duration. This result therefore lends further support to the idea that higher paradigmatic probability of a linguistic element can lead to enhanced articulation of that element. In contrast, the effect of N2 frequency, shown in Fig. 6, goes in the direction predicted by the Informativity Hypothesis: more frequent, hence less informative nouns in N2 position are associated with shorter consonant durations at

**Fig. 6** Partial effect of N2 frequency in final model



the boundary. Interestingly, this effect is found for consonants at all three boundary types, including N1-final consonants, for which we did not predict the effect.

## 5 Discussion and conclusion

This study has tested three hypotheses that seek to explain phonetic correlates of morphological structure, focussing on consonants at the boundary of noun-noun compounds in English. We modelled the duration of three different consonants in three different environments, using various predictors that operationalise effects of segmentability, informativity and paradigm complexity, in addition to a number of control variables whose effects are well established.

The control variables showed the expected effect of speech rate, as well as the expected variation between different consonants: higher speech rate led to shorter consonants, the fricative was longer than the nasals, and /m/ was longer than /n/. In addition, we found clear evidence of morphological gemination, with double consonants approximately twice as long as singletons across speech rates; this finding is in line with other studies on double consonants that straddle morphological boundaries (Ben Hedia 2019; Ben Hedia and Plag 2017; Kotzor et al. 2016; Plag and Ben Hedia 2018). We also found some evidence of word-initial consonant lengthening, since N2-initial consonants were slightly longer than those in N1-final position (cf. White et al. 2015).

### 5.1 Segmentability and informativity

Regarding the three hypotheses, we found no evidence for the Segmentability Hypothesis. The effect of N1 family size on consonant duration did not go in the direction predicted by the hypothesis, and the other relevant measure, spelling ratio, did not survive in the final model. There are several possible explanations for our failure to find an effect of segmentability. It could be that the segmentability of compounds is not related to the phonetic detail of their realisation, even though it has been found



to be related to processing measures such as reaction times in priming experiments (e.g. Zwitserlood 1994). Or it could be that the type of compound segmentability reflected in semantic transparency – the only kind for which effects have previously been reported – is different from the type of compound segmentability reflected in productivity and orthography – as used in the present study. In either case, it could be that the segmentability of compounds has different effects from the segmentability of other complex words. But it could also be that segmentability more generally is not a unitary property of morphological constructions, but rather a collection of semantic and structural attributes which, though they tend to correlate, actually have different effects. This could partly explain the apparently contradictory findings reported in the literature.

We found some limited evidence in support of the Informativity Hypothesis. Consonant duration was negatively correlated with N2 frequency across all boundary types, i.e. boundary consonants were shorter in compounds with more frequent, hence less informative heads. The hypothesis predicts this relationship for N2-initial and double consonants, whose duration we take to reflect the overall duration of N2; but we had not expected the effect to extend to N1-final consonants, which might be seen as independent of N2. One possibility is that the effect on N1 consonants reflects the cost of planning the right constituent. Overall however, as shown in Fig. 6, the informativity effect in our data was very small.

## 5.2 Paradigmatic support

Of the three hypotheses, we found most evidence for the Paradigmatic Support Hypothesis. The strongest relevant effect was a negative correlation between the duration of geminate consonants and the family size of N1. Although we predicted that we might see this effect, we did not expect that it would be restricted to geminate consonants. However, there are several possible explanations for this. It could be related to the fact that there are more tokens of double consonants in the data than singletons, so there is simply more statistical power to enable the effect to reach significance. Or it could be an acoustic effect, with the longer duration of morphological geminates allowing more scope for variation in that duration. Or, as suggested in Sect. 3.1, it may be that the nature of morphological geminates, straddling as they do a morphological boundary in a single articulatory gesture, makes them especially susceptible to paradigmatic effects such as paradigmatic enhancement.

We hypothesise that paradigmatic enhancement, such as the N1 family size effect we see for geminates, arises through a process of spreading activation within the mental lexicon: more paradigmatically probable elements receive greater activation and thus stronger articulation, in this case manifest as longer duration. We assume that, when a noun is produced as a compound modifier, activation spreads to all compounds that share that modifier, i.e. to all compounds in the N1 family. A possible explanation of our result involves the further assumptions that the total available activation is roughly constant, and that it is shared equally between all compounds in the relevant family. Hence the greater the number of N1 compounds known to a speaker, the less activation each one will receive. In other words, the greater the family size of

N1, the less activation spreads to each family member and the more weakly the compound will be articulated, with consequent shortening of any consonants that straddle the internal compound boundary.

The explanation of the N1 family size effect outlined in the previous paragraph assumes that all members of an N1 family are activated equally whenever the relevant N1 is produced. Such an effect would therefore lead to differences in activation, hence articulation, between compounds in different N1 families, but not to differences between compounds in the same N1 family, since their level of activation would be equal. In other words, the effect would operate across paradigms, i.e. between different N1 families. This contrasts with the paradigmatic enhancement effects reported by previous studies, which operate within paradigms: lengthening and other kinds of enhanced articulation have been reported for individual forms based on their token frequencies relative to other forms in the same paradigm (e.g. Cohen 2014b; Kuperman et al. 2007). To explain these within-paradigm effects in terms of spreading activation, it is necessary to assume that activation flows more strongly to forms that have higher relative frequencies, which therefore get more extreme articulations, while those with lower relative frequencies get less activation, and hence less enhancement.

At the phoneme level, we did find evidence of a within-paradigm enhancement effect similar to those reported by previous studies, i.e. related to relative frequencies within the N1 family. The paradigmatic probability of the consonant occurring in N2-initial position was positively correlated with consonant duration across all our boundary types. This is comparable to the phonetic enhancement effects found by e.g. Kuperman et al. (2007) and Cohen (2014b), and can be conceived of as a sort of practice or entrenchment effect. A possible explanation goes as follows: if activation flows more strongly to the heads of more frequent compounds, then for example, production of *city* as a compound modifier is likely to strongly activate the very frequent compound *city centre*. With strong activation of *city centre*, the articulatory path from *city* to /s/ will also be strongly activated and N2-initial /s/ will be enhanced even in lower-frequency *city*-N2 compounds whose N2s begin with that sound. A similar explanation would apply to morphological geminates, where a very high-frequency compound such as *prime minister* would lead to strong activation of the articulatory path from *prime* to /m/ and thus enhancement of at least the N2-initial part of /mm/ in all relevant *prime* compounds.

Unlike the N1 family size effect, which is restricted to geminates, the effect of consonant probability applies to all three environments: geminate, N2-initial and N1-final. However, it is not immediately obvious how the paradigmatic probability of the consonant occurring in N2-initial position could affect the duration of N1-final single consonants, especially since such consonants in our data all occurred in compounds with an N2-initial vowel. To understand how this might happen, it is important to remember that the probability of the relevant consonant occurring in N2-initial position in the N1 families of these compounds is in fact the probability of there being a morphological geminate, since N1 always ends with the consonant. We therefore hypothesise that a high-frequency geminate compound such as *prime minister*, and the consequent strong activation of the articulatory path from *prime* to /m/, might lead to entrenchment of *prime* with an enhanced final /m/ even in cases where a second

/m/ does not actually follow. This assumes a model of the mental lexicon in which articulatory information is stored with lexical items and becomes reinforced with increasing frequency of use.

### 5.3 Type-based vs token-based effects

It is striking that, at constituent level, we found a paradigmatic enhancement effect for the type-based measure (N1 family size) but not for the corresponding token-based measure (paradigmatic probability of N2 given N1). At phoneme level, however, we found an effect for the token-based measure but not for the corresponding type-based measure. Neither the type-based paradigmatic probability of the consonant given N1, nor the token-based paradigmatic probability of N2 given N1, reached significance in their individual models. Furthermore, at first sight, the hypotheses put forward in the previous section to explain the two paradigmatic enhancement effects in our model may seem to be incompatible with one another. How can it be possible for all members of a compound family to be activated equally, as suggested by the type-based N1 family size effect, while at the same time receiving different levels of activation according to the form of N2, as suggested by the token-based consonant probability effect? One possibility is that the apparent discrepancy between the type-based and token-based effects in our model is a methodological artefact, arising from the fact that we used measures extracted from a large generic corpus to characterise the mental lexicon. In fact, the mental lexicon is not a constant across speakers, and the lexicons of the participants in our study will have varied in terms at least of the number of items known, the particular items known, and the relative resting activations of the different items.

Psycholinguistic studies on the nature of the mental lexicon often try to predict e.g. word recognition latencies or eye movement patterns on the basis of various paradigmatic measures extracted from corpora. Such studies have consistently found that token-based measures are most significant for inflected words while type-based measures work best for derivatives and compounds (Blevins 2016, p. 48). One possible explanation for this difference concerns the nature of the relevant paradigm and the methodology by which the measures are obtained. An inflectional paradigm constitutes a closed class: only certain forms are possible, even in theory, and in practice only a subset of these may actually occur. In contrast, the morphological family of a derivational affix or compound modifier constitutes an open class of indefinite and usually expanding size, as vocabulary continues to increase across the lifespan (Ben-David et al. 2015). This suggests that, whereas the size and constitution of an inflectional paradigm is likely to be relatively stable and consistent across the lexicons of different speakers, the make-up of a morphological family could vary widely between speakers depending on which words they happen to know.

Compound families may be especially prone to inter-speaker variation, since the majority of compounds occur with very low frequencies and many arise as situation-specific coinages, or within particular contexts. For example, the compound *constituent family* may occur quite frequently for some readers of this journal, but is probably completely unknown to the majority of English speakers. Furthermore, if frequency effects are indeed related to activation levels in the brain, then it seems

likely that recency of exposure and saliency or attention-related factors will also be relevant, over and above simple cumulative frequency of encounters. For example, the compound *withdrawal agreement* occurs only once in ukWaC, but a Google search of the UK web domain in October 2019 produced 18,800,000 results for the same compound (this was a time when discussion of the UK's withdrawal from the European Union dominated the UK media, and *withdrawal agreement* was likely to be occurring with relatively high frequency in the experience of many speakers of British English). Overall, it is unlikely that a set of relative compound frequencies extracted from a large corpus would closely reflect the mental lexicon of any particular individual at any given time.

Unlike individual compound frequencies, N1 family size reflects the general productivity of N1 as a modifier; it is therefore likely to be much more consistent both between speakers and between speakers and corpora, since it is independent of the particular compounds that a speaker happens to have encountered or that happen to occur in a corpus. For example, in our data, the compounds *class survey* and *pandemonium model* have the highest and lowest N1 family sizes respectively. In each of these families, the particular compounds known, and their relative frequencies, will vary widely between speakers. Some very frequent compounds, such as *class teacher* are probably known to most speakers, whereas lower-frequency items may be more specialised; for example, *class settlement* may be very familiar to certain lawyers but unknown to other speakers. Likewise, *pandemonium model*, though low frequency overall, may be quite familiar to some psychologists, while *pandemonium level* may be frequent for players of certain online games. It seems likely that, despite such differences in detail, the relative sizes of these two families will be fairly consistent, with most speakers of English knowing many more *class* compounds than *pandemonium* compounds. Because of this, we hypothesise that N1 family sizes taken from a corpus will better approximate to the make-up of individual speakers' lexicons than token-based compound measures would, by virtue of capturing more general patterns less prone to idiosyncratic variation. However, this is an artefact of the methodology, and does not mean that the raw number of compounds sharing a modifier in the lexicon of an actual speaker is necessarily the most salient variable determining how activation spreads between representations of those compounds in that speaker's brain. It could be that the degree of activation of a compound is indeed related to its frequency, but that our very crude corpus-based measures are unable to capture the relative frequencies experienced by actual speakers.

Whereas the number of compounds in a family is potentially unlimited, the number of possible N2-initial phonemes is limited by the phoneme inventory of the language. In this case, it is the type-based (rather than token-based) measure that is more susceptible to idiosyncratic variation. The highest frequency compounds in a family are likely to be present in the lexicons of most speakers, but the lower-frequency compounds will vary between speakers, depending on many aspects of individual experience. The raw number of compounds with any particular N2-initial phoneme is therefore also likely to vary considerably. In contrast, the relative token frequencies of the different phonemes in N2-initial position are likely to be dominated by the high-frequency compounds, known to most speakers. Because of this, we would expect relative token frequencies of initial phonemes to be more stable than relative type

counts across individual lexicons and across corpora. We were therefore not surprised to find that the token-based rather than type-based measure of consonant probability was significant in our model.

#### 5.4 Mechanisms of enhancement

We have explained paradigmatic enhancement as resulting from spreading activation flowing more strongly to more frequent, more recently experienced or otherwise more salient constituents. But even if this is correct, a full explanation of our results still requires us to postulate two slightly different mechanisms operating together. We saw in Sect. 5.2 that, to explain enhancement of N1-final consonants, it is necessary to assume that a frequent geminate in an N1 family can lead to some entrenchment of the associated consonant lengthening, which thus becomes manifest even when a geminate does not actually occur; for example, all *prime*-N2 compounds might have a lengthened N1-final /m/ as a result of the very frequent compound *prime minister*. More generally, we hypothesise that, as a result of a learning process, production of a given N1 strongly activates the articulatory neural pathways for N2s that frequently follow that N1 and for N2-initial phonemes that frequently follow that N1 across different compounds. In other words, upon utterance of a given N1, we suppose that the spread of activation in the lexicon is at least partly determined by existing strengths of association between that N1 and possible values both of N2 and of N2-initial phonemes. The stronger the association of N1 with a particular N2 or N2-initial phoneme, the more strongly activation will flow to that particular word or sound, regardless of whether it actually occurs in the current context. Such learned associations could underlie most of the paradigmatic enhancement effects reported in the literature.

One way of establishing strengths of association, such as those postulated in the previous paragraph, is through the mechanism of discrimination learning. In discrimination learning, the association strength between a cue (in our case N1) and an outcome (in our case a particular N2 or N2-initial phoneme) is updated after each learning event (in our case every utterance of an N1-initial compound). If a particular outcome follows the cue, the strength of association between them is increased, so utterance of e.g. *claim money* would strengthen the association between *claim* and N2-initial /m/. However, if a cue occurs in the absence of an outcome, the strength of association between them is decreased, so utterance of e.g. *claim form* would weaken the association between *claim* and N2-initial /m/. Using computational models, it has been shown that discrimination learning can indeed account for a wide range of linguistic and acoustic phenomena, including consonant duration. For example, Tomaschek et al. (2019) used a computational implementation of discrimination learning (NDL; Arppe et al. 2018) to predict the duration of word-final /s/ in English. Rather than the relative paradigmatic frequency of /s/ in relation to other phonemes, these authors considered the morphological function of specific tokens (e.g. plural noun or genitive), as well as their phonological and immediate syntactic environments. They found that the greater the strength of association between these environmental variables and the morphological function of /s/, as learnt by NDL, the longer the duration of the consonant. Their results were therefore similar to ours, in

the sense that greater contextual support for an outcome led to enhanced articulation. This similarity furthers the idea that our findings might also result at least partly from learnt strengths of association within the lexicon.

As described in the previous two paragraphs, a process of learning strengths of association between compound modifiers (N1s) and compound heads (N2s) could explain why greater paradigmatic probability of the N2-initial consonant leads to consonant lengthening at the compound-internal boundary. However, such a process cannot fully explain the effect of N1 family size on the duration of morphological geminates. For example, in our data, *pandemonium model* not only has the smallest N1 family size but also the longest average duration for a geminate /mm/. But this compound occurs with very low frequency and is almost certainly unknown to most speakers. In fact, most speakers probably don't know any *pandemonium* compounds at all, so there is unlikely to be a strong association between *pandemonium* as N1 and any particular N2 or N2-initial phoneme. In general, it is not the case that small N1 families necessarily consist of high-frequency compounds; thus the effect of N1 family size, whereby smaller N1 families are associated with longer geminate duration at the compound-internal boundary, cannot rely exclusively on learned associations.

In the case of small N1 families containing only low-frequency compounds, we hypothesise that the enhancement effect arises online, with compounds in such families getting a large share of the available activation simply by virtue of having few competitors.<sup>3</sup> Our participants were asked to rehearse each sentence in silent reading, so they had an opportunity to access the meaning of each utterance before articulating it, as speakers presumably do in most situations. Once the speaker has in mind a particular N1, we assume that some activation starts to spread to the articulatory neural pathways for all known compounds in that N1 family. However, in the case of an unusual N1 like *pandemonium*, few if any compounds will be present in the lexicon; most or all of the total activation will therefore remain available for articulation of the intended N2, in this case *model*. It might be that a careful study comparing phonetic enhancement effects with eye tracking or reaction time data would be able to tease out the effects of learnt associations from online effects resulting from lack of competition. We would expect that learnt associations due to high relative frequency might be associated with faster reaction times for those items, whereas no such association is expected for low-frequency items in small N1 families.

There remains the question as to how phonetic reduction associated with higher syntagmatic or paradigmatic probability can be squared with phonetic enhancement associated with higher paradigmatic probability. Our results suggest that these two effects can operate simultaneously in the same data, a conclusion also drawn by Cohen (2014b) and Kuperman et al. (2007). One possibility is that these two types of effect operate at different linguistic levels, involving different neural pathways. Hall et al. (2018) suggest that predictability-associated phonetic effects are always based on the predictability of meaning-bearing units, i.e. words rather than phonemes. However, it is also possible that paradigmatic enhancement is essentially a form-driven articulatory phenomenon, whereas probabilistic reduction is more semantically driven. This hypothesis could perhaps be tested by exploring the production of non-words, i.e. forms without any semantic content.

<sup>3</sup>Compound frequency itself was not a significant predictor of consonant length in our final model.

In summary, we have found clear evidence for our Paradigmatic Support Hypothesis as well as more limited evidence for the Informativity Hypothesis. Paradigmatic enhancement is a robust effect in our data, surviving even in the presence of strong control variables and random effects for item and speaker. The results are compatible with a model of language in which lexical items are stored as weighted connections (i.e. strengths of association) in a neural network that includes connections both to articulatory pathways and to morphologically related items.

**Acknowledgements** Open Access funding provided by Projekt DEAL. We would like to thank the members of the DFG Research Unit FOR2373, the reviewers of this journal, and the audiences of the following conferences and colloquia for valuable input and comments on earlier versions of this paper: 40th Annual Meeting of the German Linguistic Society, Stuttgart 2018; LabPhon16, Lisbon 2018; Glasgow University Laboratory of Phonetics Lab Lunch 2018. We are grateful to Javi Sanz for collecting the data and to the University of Cambridge Phonetics Laboratory, especially Francis Nolan and Ian Smeeton, for allowing and helping us to collect the data there. This research was partially funded by the Deutsche Forschungsgemeinschaft (FOR2373, Projects PL 151/7-1 and PL 151/8-1), which we gratefully acknowledge.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., & Shaoul, C. (2018). *Package ndl: Naive discriminative learning*. R package version 0.2.18. <https://CRAN.Rproject.org/package=ndl>.
- Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119, 3048–3058.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R. H., Burani, C., & Schreuder, R. (1997). Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1996* (pp. 13–34). Dordrecht: Kluwer Academic Publishers.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium. Philadelphia: University of Pennsylvania.
- Baayen, R. H., & Shafaei-Bajestan, E. (2019). *languageR: Analyzing linguistic data: A practical introduction to statistics*. R package version 1.5.0. <https://CRAN.Rproject.org/package=languageR>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bauer, L. (1998). When is a sequence of two nouns a compound in English? *English Language and Linguistics*, 2(1), 65–86.
- Bell, M. J. (2011). At the boundary of morphology and syntax: Noun noun constructions in English. In A. Galani, G. Hicks, & G. Tsoulas (Eds.), *Morphology and its interfaces* (pp. 137–167). Amsterdam and Philadelphia: Benjamins.
- Bell, M. J., & Plag, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics*, 48(3), 485–520.

- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.
- Ben Hedia, S. (2019). *Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics. Studies in laboratory phonology: Vol. 8*. Berlin: Language Science Press.
- Ben Hedia, S., & Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, *62*, 34–49.
- Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). “Older is always better”: Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, *30*(4), 856.
- Bermúdez-Otero, R. (2018). Stratal phonology. In S. J. Hannahs & A. Bosch (Eds.), *Routledge handbook of phonological theory* (pp. 100–134). London: Routledge.
- Blevins, J. P. (2016). *Word and paradigm morphology*. London: Oxford University Press.
- Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer* [computer programme].
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, Methodological*, *26*(2), 211–243.
- Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., & Frauenfelder, U. H. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *The Journal of the Acoustical Society of America*, *130*(6), 3980–3991.
- Cohen, C. (2014a). *Combining structure and usage patterns in morpheme production: Probabilistic effects of sentence context and inflectional paradigms*. PhD thesis, UC Berkeley.
- Cohen, C. (2014b). Probabilistic reduction and probabilistic enhancement. *Morphology*, *24*(4), 291–323.
- Cohen, C. (2015). Context and paradigms: Two patterns of probabilistic pronunciation variation in Russian agreement suffixes. *The Mental Lexicon*, *10*(3), 313–338.
- Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language*, *49*(1), 119–132.
- Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard*, *4*(s2).
- Hanique, I., & Ernestus, M. (2012). The processes underlying two frequent casual speech phenomena in Dutch: A production experiment. In *Proceedings of the 13th annual conference of the International Speech Communication Association*.
- Hay, J. B. (2003). *Causes and consequences of word structure*. New York and London: Routledge.
- Hay, J. B. (2007). The phonetics of un-. In J. Munat (Ed.), *Lexical creativity, texts and contexts* (pp. 39–57). Amsterdam / Philadelphia: Benjamins.
- Hay, J. B., & Baayen, R. H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, *1*, 99–130.
- Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund-Prytz, Y. (2008). *Corpus linguistics with BNCweb – a practical guide. English corpus linguistics: Vol. 6*. Frankfurt an Main: Peter Lang.
- Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*(4), 406–430.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Typological studies in language: Vol. 45. Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: Benjamins.
- Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 131–176). Dordrecht: Foris.
- Kotzor, S., Molineaux, B. J., Banks, E., & Lahiri, A. (2016). “Fake” gemination in suffixed words and compounds in English and German. *The Journal of the Acoustical Society of America*, *140*(1), 356–367.
- Kunter, G., & Plag, I. (2016). Morphological embedding and phonetic reduction: The case of triconstituent compounds. *Morphology*, *26*(2), 201–227.
- Kuperman, V., & Bertram, R. (2013). Moving spaces: Spelling alternation in English noun-noun compounds. *Language and Cognitive Processes*, *28*(7), 939–966.
- Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7–8), 1089–1132.
- Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, *62*, 83–97.



- Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2007). Morphological predictability and acoustic salience of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, 121, 2261–2271.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
- Ladefoged, P. (2003). *Phonetic data analysis*. Malden and Mass: Blackwell.
- Lee-Kim, S.-I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, 4(2), 475–511.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35(11), 1773–1781.
- Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V., & Baayen, R. H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, 28(1), 71–97.
- MacGregor, L. J., & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language*, 126(2), 217–229.
- Milin, P., Đurđević, D. F., & del Prado Martín, F. M. (2009b). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, 60(1), 50–64.
- Milin, P., Kuperman, V., Kostić, A., & Baayen, R. H. (2009a). Words and paradigms bit by bit: An information-theoretic approach to the processing of paradigmatic structure in inflection and derivation. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 214–252). Oxford: Oxford University Press.
- Nooteboom, S. G. (1972). *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*. Utrecht: University of Utrecht.
- Plag, I., & Ben Hedia, S. (2018). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In S. Arndt-Lappe, A. Braun, C. Moulin, & E. Winter-Froemel (Eds.), *The dynamics of wordplay: Vol. 5. Expanding the lexicon: Linguistic innovation, morphological productivity, and the role of discourse-related factors* (pp. 93–116). Berlin, New York: de Gruyter Mouton.
- Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final s in English. *Journal of Linguistics*, 53(1), 181–216.
- Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language*, 84(4), 760–794.
- Pluymaekers, M., Ernestus, M., Baayen, R. H., & Booij, G. (2010). Morphological effects in fine phonetic detail: The case of Dutch -igheid. In C. Fougeron (Ed.), *Laboratory phonology 10, volume 4-4 of Phonology and phonetics* (pp. 511–531). Berlin, New York: de Gruyter.
- Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, 37, 118–139.
- Schuppler, B., van Dommelen, W. A., Koreman, J., & Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, 40(4), 595–607.
- Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140–155.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Smith, R., Baker, R., & Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics*, 40(5), 689–705.
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21, 291–311.
- Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, 1–39.
- Torreira, F., & Ernestus, M. (2012). Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica*, 69(3), 124–148.
- Umeda, N. (1977). Consonant duration in American English. *The Journal of the Acoustical Society of America*, 61(3), 846.

- van Jaarsveld, H. J., Coolen, R., & Schreuder, R. (1994). The role of analogy in the interpretation of novel compounds. *Journal of Psycholinguistic Research*, 23, 111–137.
- van Son, R. J., & Pols, L. C. (2003). Information structure and efficiency in speech production. In *Proceedings of Eurospeech-2003*, Geneva (pp. 769–772).
- van Son, R. J., & van Santen, J. P. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47, 100–123.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. New York: Springer.
- White, L., Mattys, S. L., Stefansdottir, L., & Jones, V. (2015). Beating the bounds: Localized timing cues to word segmentation. *The Journal of the Acoustical Society of America*, 138(2), 1214–1220.
- Zimmermann, J. (2016). Morphological status and acoustic realization: Is there a difference between Bra[d] Pitt and a grille[d] cheese omelet, or between Kate Mo[s] and killer robot[s]? In J. Wahle, M. Köllner, H. R. Baayen, G. Jäger, & T. Baayen-Oudshoorn (Eds.), *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics (QITL-6)*, Tübingen. Universität Tübingen.
- Zwitzerlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9(3), 341–368.