Assessing the effect of ambiguity in compositionality signaling on the processing of diphones

Andreas Baumann (corresponding author)

University of Vienna

Department of English and American Studies

Spitalgasse 2-4, Court 8.3

1090 Vienna

Austria

andreas.baumann@univie.ac.at

Kamil Kaźmierski

Adam Mickiewicz University in Poznań

Department of Contemporary English Language

aleja Niepodległości 4

61-874 Poznań

Poland

kamil.kazmierski@wa.amu.edu.pl

Assessing the effect of ambiguity in compositionality signaling on the processing of diphones

Abstract: Consonantal diphones differ as to their *ambiguity* (whether or not they indicate morphological complexity reliably by occurring exclusively either within or across morphemes) and *lexicality* (how frequently they occur within morphemes rather than across morpheme boundaries). This study empirically investigates the influence of ambiguity and lexicality on the processing speed of consonantal diphones in speech perception. More specifically, its goal is to test the predictions of the Strong Morphonotactic Hypothesis, which asserts that phonotactic processing is influenced by morphological structure, and to clarify the two conceptions thereof present in extant research. In two discrimination task experiments, it is found that the processing of cross-morpheme diphones decreases with their ambiguity, but there is no processing difference between primarily cross-morphemic and morpheme-internal diphones. We conclude that the predictions of the Strong Morphonotactic Hypothesis are borne out only partially, and we discuss the discrepancies.

Highlights:

★ Ambiguity in signaling morphological complexity affects diphone processing
★ Speakers have probabilistic knowledge of how often diphone types span morpheme boundaries
★ Diphones that occur prototypically within morphemes are processed as fast as prototypically cross-morphemic diphones
★ Processing of cross-morphemic diphones is slow if they are ambiguous
★ Participants can be primed for analyzing diphones in nonce words as spanning a morpheme boundary

Keywords: morphonotactics, compositionality signaling, ambiguity, perception

1 Introduction

The processing of sound sequences, and that of word-internal consonant sequences in particular, have been argued to depend, among other factors, on the morphology of words they are embedded in: some diphones, such as /ld/ or /nd/, occur across morpheme boundaries (*call+ed, wan+ed*) as well as morpheme internally (*cold, wand*), while others are restricted to a single morphological environment (/md/ as in *seem+ed*, and /mp/ as in *lamp*, respectively). This has been suggested in turn to affect their acquisition and diachronic development (Dressler et al., 2010; Korecky-Kröll et al., 2014; Leykum et al., 2015a; Zydorowicz, 2007).

This work aims at assessing the influence of morphological status of consonantal sequences on the ease of their processing in speech perception. We address this aim by means of two related experiments conducted with speakers of Polish. Our experimental setup will, more specifically, address two divergent propositions that have been drawn from - and sometimes equated with - a central hypothesis in the research focusing on the interaction of phonotactics and morphology, i.e. the morphonotactic research paradigm (see Table 1 for terminological clarification). This hypothesis in a nutshell asserts that sound sequences may have the function of signaling morpheme boundaries and triggering the decomposition of a complex word. Above the morphological level, it is well known that phonotactic knowledge helps listeners in the decomposition of the speech stream into words (McQueen, 1998; Mattys et al., 1999; Mattys and Jusczyk, 2001; Daland and Pierrehumbert, 2011; van der Lugt, 2001). Thus, sound sequences which rarely occur within words function as boundary signals and thus speed up the parsing process. In morphonotactics, this principle is transferred to the word-internal domain, i.e. the decomposition of words into morphemes.

Put into semiotic terms, sound sequences are hypothesized to function as *signifiants* for the *signifié* 'morphological boundary' (Dressler and Dziubalska-Kołaczyk, 2006). If sound sequences indeed fulfill this semiotic function, then the reliability of this function and by consequence the ease of processing of boundary spanning sequences should be diminished as soon as signaling becomes ambiguous (in the sense that the same consonant sequence can be additionally used within morphemes). The question is whether the latter condition holds true. This is what we test in our first experiment. In the second experiment we consider the question of whether the ambiguity of a sequence in general affects its processing. The subtle difference between these two questions, which have both been addressed but not always clearly distinguished from each other in morphonotactic research, is this: The former is about the effect of ambiguity on the quality of a sign, which as a consequence is expected to affect the processing of a sequence (the sign's *signifiant*; 'Is the boundary-signaling sequence /md/ in *seem+ed* processed faster than /nd/ in *wan+ed*?'). The latter considers the effect of ambiguity on the processing of a sequence without being restricted to denoting a morpheme boundary ('Is /md/ generally processed faster than /ld/, irrespective of whether /ld/ occurs in *call+ed* or in *cold*).

We will show that the central hypothesis is confirmed by our experiments, albeit only partially: boundary-spanning instances of diphone types (such as /ld/ in *called*) are processed most slowly if the type occurs across morpheme boundaries and within morphemes at roughly equal frequencies (e.g. /ld/). Thus, speakers have probabilistic knowledge of the morphological environment of diphones. We argue that this suggests a cognitive model of phonotactics in which memories of instances of sound strings are stored together with morphological information (Plag et al., 2017). We do, however, not find a general advantage of non-ambiguous (/md/) over ambiguous (/ld/) diphone types if cross-morpheme instances are not explicitly tested, nor did we detect a general advantage of primarily boundary spanning (/md/) over primarily morpheme-internal (/rl/) diphone types (or the reverse).

In our analysis, we employ different ways of measuring ambiguity of signaling morpheme boundaries, in particular differentiating between type and token frequencies. In order to detect potentially nonlinear effects of ambiguity, we use generalized additive models (Wood, 2006), a modeling technique which recently gained momentum in linguistic research (e.g. Wieling et al., 2011; Baayen, 2013; Fruehwald, 2017). Thus, in addition to providing results on the processing of sequences of sounds, this study, on a more theoretical level, seeks to highlight and clarify some of the argumentative vagueness that seems to be present in the morphonotactic literature, while at the same time featuring relatively novel analytical methods.

The cornerstones of morphonotactics shall be described together with our specific research questions in the remainder of this section and in Section 2. Afterwards, the two experiments together with their analyses (Section 3 and 4) shall be presented and finally discussed (Section 5).

Table 1. Phonotactic and morphonotactic terminology

term	meaning	example
diphone	sequence of two single sound segments	/hæ/, /æn/, and /nd/ in *hand* /hænd/
consonant cluster	sequence of consonants; sometimes restricted to sequences within syllables (not in this study)	/nd/ in *hand* /hænd/
morphonotactic instance of a cluster	token of a cluster which spans a morpheme boundary; sometimes referred to as *morphotactic*, *boundary spanning* or *cross-morphemic* cluster	/nd/ in *bann+ed* /bænd/
lexical instance of a cluster	token of a cluster which is morpheme internal; also referred to as *phonotactic*	/nd/ in *hand* /hænd/
primarily morphonotactic cluster	cluster type which has exclusively or almost exclusively morphonotactic instances; sometimes measured in type frequency rather than token frequency; also referred to as *morphonotactic strong default*, *prototypically morphonotactic*, or if token frequency is used *low probability*	word final /ts/ as in *bit+s* or *cut+s* (but also in a few items like *blitz*)
primarily lexical cluster	cluster type which has exclusively or almost exclusively lexical instances; sometimes measured in type frequency rather than token frequency; also referred to as *lexical strong default*, *prototypically lexical*, or if token frequency is used *high probability*	word final /lk/ as in *bulk* or *milk*
ambiguous cluster	cluster type which many morphonotactic as well as many lexical instances; also referred to as *mid probability* if token frequency is used	word final /ld/ in *call+ed* or *cold* or /nd/ in *bann+ed* or *bind*
lexicality of a cluster	Fraction of lexical instances of a cluster type; also *probability* of a cluster type if token frequency is used	Close to 0 if primarily morphonotactic (English /ts/); close to 1 if primarily lexical (English /lk/); close to 1/2 for a perfectly ambiguous cluster (English /ld/)
ambiguity of a cluster	Similarity of a cluster distribution with a 1:1 distribution of morphonotactic and lexical instances	Close to 1 for a perfectly ambiguous cluster (English /ld/); close to 0 for primarily lexical or morphonotactic clusters (English /lk/ or /ts/)

1.1 Phonotactics

The phonotactics of a given language consists in imposing limitations on or expressing preferences with regard to sound sequences in that language. In this article, we limit the scope of our investigation to consonantal diphones, that is, to sequences of exactly two consecutive consonants. One approach to phonotactics is to look for universal rules (or constraints), whose ordering (or ranking) accounts for cross-linguistic differences as to which sound sequences are licit in particular languages. These rules (or constraints) can be formulated with regard to the syllable as the domain of their application (Kahn, 1976), with regard to strings (Steriade, 1999), or with regard to both strings and syllables (Albright, 2015). A different approach (e.g. Vennemann, 1988; Dziubalska-Kołaczyk, 2014), whose predictions for processing in speech perception will be tested here, is not to determine the legal sound sequences for a language, but to formulate preferences for particular sound sequences. The point of departure is the assumption that all consonantal diphones are in general 'dispreferred'. Various observations are used to support this notion, including typology (e.g. consonantal diphones in syllable codas are allowed in fewer than 21% of the world's languages; Donohue et al., 2013), casual speech phenomena (diphones are reduced in fast speech; Dziubalska-Kołaczyk and Zydorowicz, 2014) and language acquisition (they are acquired late; Levelt and Vijver, 1998, 2004; Jarosz et al., 2016).

While all consonantal diphones are dispreferred compared to singletons, they are said to differ as to the degree to which they are so. This approach, then, does not categorize sound sequences as licit or illicit, but instead ranks the observed sequences with respect to their 'preferability'. This is done with regard to proposed universal preferences regarding the distance (in place and manner of articulation) between the members of the consonantal sequence and the neighboring vowel or vowels. For example, for medial diphones, which are the focus of this paper, a diphone is preferred if the distance between the two consonants is less than or equal to the distance between each consonant and its neighboring vowel (Dziubalska-Kołaczyk, Pietrala, & Aperliński, 2014). Crucially, while preferability and frequency are related, they cannot be equated. Preferred diphones are expected to be, or become, frequent, but there are are other, e.g. lexico-grammatical and pragmatic factors influencing a diphone's frequency, besides its preferability.

A considerable amount of research has been devoted to the influence of phonotactics on speech segmentation, i. e. on spotting words in the speech stream. Diphones with a high frequency of occurrence between words and a low frequency of occurrence within words have been repeatedly found to help segmentation, both when listeners are infants (Mattys et al., 1999; Mattys and Jusczyk, 2001) and adults (van der Lugt, 2001). Daland and Pierrehumbert (2011), having tested learning models on speech corpora, show that phonotactic knowledge (here: phrase-medial word boundaries) is learnable given the input that infants typically receive. In contrast to the research on segmentation, however, we are looking at word-internal diphones only, and taking into account the probability with which they occur within or across morphemes rather than within or across words.

1.2 Phonotactics vs. morphonotactics

It has been observed that the phonotactics of a language interacts with its morphology. For example, final consonantal sequences in English words allowing four members are exclusively non-monomorphemic, e.g. *six+th+s, glimpse+d* (Cruttenden, 2014, p. 262). The interaction of morphology and phonotactics has been the focus of a proposed theory of 'morphonotactics' (Dressler and Dziubalska-Kołaczyk, 2006). Here, the very fact that consonantal diphones spanning morphemic boundaries are ranked low on the preferability scale is actually argued to be their strength. Dispreferred diphones are claimed to signal morphological complexity better through their status as 'dispreferred'. A diphone which is dispreferred stands out in semiotic terms, the argument goes, and thus may be an indication of a morphological operation having taken place.

1.3 The Strong Morphonotactic Hypothesis

While some consonantal diphones - 'purely morphonotactic' ones - always occur across morpheme boundaries in a given language (e.g. ENG /md/ as in *seem+ed*) and others - 'purely lexical' ones - always occur within morphemes (e.g. ENG /mp/ as in *lamp*), yet others might occur both across and within morphemes, and so can be seen as ambiguous. Ambiguous diphones differ as to the degree of their ambiguity, i.e. the relative frequency with which they occur across and within morphemes. And so clusters such as ENG /ts/ are morphonotactic by strong default, as they usually (e.g. *cat+s*) occur across morpheme boundaries, though not always (e.g. *waltz*). There are diphones which act as morphonotactic and lexical roughly equally frequently (e.g. ENG /ld/ as in *call+ed* and *cold*). Finally, there are clusters such as ENG /nd/, which are lexical by strong default as they usually (e.g. *hand*) occur within morphemes, though also occur across morpheme boundaries (e.g. *bann+ed*). Table 1 provides an overview of the morphonotactic terminology adopted in this paper.

Thus, going from purely morphonotactic diphones, through the three categories of ambiguous diphones all the way to purely lexical diphones, a lexicality scale can be formed - see Table 2. We will consider diphones in category 3 as maximally ambiguous, diphones in category 2 or 4 as less ambiguous, and diphones in category 1 or 5 as least ambiguous. Note that the frequency measure used to determine ambiguity is not a priori clear. Indeed, Dressler and Dziubalska-Kołaczyk (2006) left the question of whether type frequencies (number of word types a diphone occurs in) or token frequencies (number of diphone instances) should be employed as an open question.[1] We will account for both frequency measures in our analysis.

Table 2. The lexicality scale (morphonotactic - ambiguous - lexical)

1	2	3	4	5
Morphonotactic	Strong default	Equally frequent	Strong default	Lexical
/md/ *seemed*	/ts/ *cats, waltz*	/ld/ *called, cold*	/nd/ *banned, hand*	/mp/ *lamp*

Dressler and Dziubalska-Kołaczyk (2006: 83) postulate the following hypothesis with regard to the relationship between the position of a given diphone[2] on the lexicality scale and its ability to signal morphological complexity, which has come to be known as the Strong Morphonotactic Hypothesis (SMH; cf. Korecky-Kröll et al., 2014; Calderone et al., 2014; [Authors]), although Dressler and Dziubalska-Kołaczyk (2006) did not actually coin this term in this very paper:

[1] We would like to thank an anonymous reviewer for raising this issue.
[2] Consonantal diphones are, as in (1) above, often referred to as 'consonant clusters', or simply 'clusters'. Since the question of whether clusters are restricted to being contained within syllables is under debate, we prefer the more neutral term 'diphone'. Whenever we use the term cluster in this paper it simply denotes 'consonantal diphone'.

(1) Strong morphonotactic hypothesis (SMH):

a) “Prototypical morphonotactic clusters [...] have the function of co-signaling the existence of a morphological rule [i.e. presence of a morphological operation],”[3]

b) “morphonotactic default clusters [...] fulfill this [signaling] function less adequately,”

c) “while phonotactic clusters [...] cannot fulfill this [signaling] function [...]” (Dressler and Dziubalska-Kołaczyk 2006: p. 83)

The SMH figures centrally in morphonotactic research, and numerous attempts have been made to test it drawing on data from language acquisition (Freiberger et al., 2011), diachronic linguistics (Dressler et al., 2010), experimental research (Korecky-Kröll et al., 2014; Leykum et al., 2015a), or by means of computational modeling (Calderone et al., 2014). The authors of these studies, however, have not always tested the same hypothesis, as it seems.

This deserves to be elaborated on in more detail. The SMH as phrased in the quote above implies (a) that clusters that span a morpheme boundary (i.e. morphonotactic clusters) have the semiotic function of signaling that boundary and (b) that the success at which morphonotactic clusters signal morpheme boundaries decreases in their degree of lexicality as shown in Table 2 (cf. solid line in Figure 1a below). Morpheme-internal clusters (phonotactic or lexical clusters) - trivially - lack this function, i.e. (c).

Notably, part (a) and (c) of the SMH do not directly assert anything about whether or not this signaling function exhibits some beneficial effect on the processing of morphonotactic clusters as opposed to their lexical counterparts (nor does (b), obviously). This is interesting, because previous studies such as Korecky-Kröll et al. (2014: p. 57) have experimentally[4] investigated the following operational hypothesis:

(2) Operational hypothesis associated with SMH:
“[I]f a certain sequence occurs only over a morpheme boundary and is thus a prototypical morphonotactic sequence, it should be processed more easily than a purely phonotactic sequence”

According to Korecky-Kröll et al. (2014: 57) this operational hypothesis is meant to shed light on “[t]he Strong Morphonotactic Hypothesis, which assumes that phonotactics helps in the decomposition of words into morphemes”. Arguments to the same effect can also be found elsewhere. Leykum et al. (2015b: p. 1) who propose that “as an extension of the Strong Morphonotactic Hypothesis [...] morphonotactic clusters are more robust and more highlighted in speech production than phonotactic clusters”.[5] Similarly, Calderone et al. (2014: pp. 59-60) state that

[3] In this definition, morphonotactic clusters are those which arise from any morphological operation rather than just morphological concatenation (e.g. morphologically induced vowel drop between two consonants). In this paper, however, we restrict ourselves to morphological concatenation, so that morphonotactic clusters are equated with clusters spanning a morpheme boundary.

[4] In a series of experiments, participants were asked to find a particular substring of triconsonantal clusters. Response times were significantly lower if a morpheme boundary was present.

[5] They fail to show that this is the case in an experimental reading task assessing duration and intensity of word final diphones. Note that in their study, the respective articulation of morphonotactic versus phonotactic instances of the very same cluster type is compared against each other.

"[a]ccording to the strong morphonotactic hypothesis [...], speakers use morphonotactic consonant clusters as morphological boundary signals. Morphonotactic clusters are thereby assigned a morphological function in processing [...], which is assumed to facilitate processing and acquisition of complex consonantal structures."

The underlying rationale is this: morphonotactic diphones have the burden of signaling morpheme boundaries in a confident way in order to be of any help in morphological decomposition. To this end, they must be easily detected (cf. 1.1), and hence they are required to have properties that make them being easily processed in perception (such as beneficial perceptual contrast between segments or longer duration). Notably, these properties must outweigh any cognitive costs that are imposed by the process of morphological decomposition (otherwise there is no reason to expect hypothesis (2) to hold). Consequently, as we infer, ease of processing is expected to be a decreasing function of a diphones lexicality (see solid line in Figure 1b, below).

The argument contrasts with findings from the research on phonotactic signaling of word boundaries outlined in 1.1. It has been shown that diphones which occur word internally (i.e. which are 'lexically licensed' and thus belong to the phonotactic inventory) are perceived more easily, produced more accurately, and less likely subject to repair processes than diphones which occur only across word boundaries (e.g. Moreton, 2002, and Berent et al., 2007). If the same mechanisms also apply to morpheme boundaries within words, ease of processing must be an increasing function of a diphones lexicality (dotted line in Figure 1b).

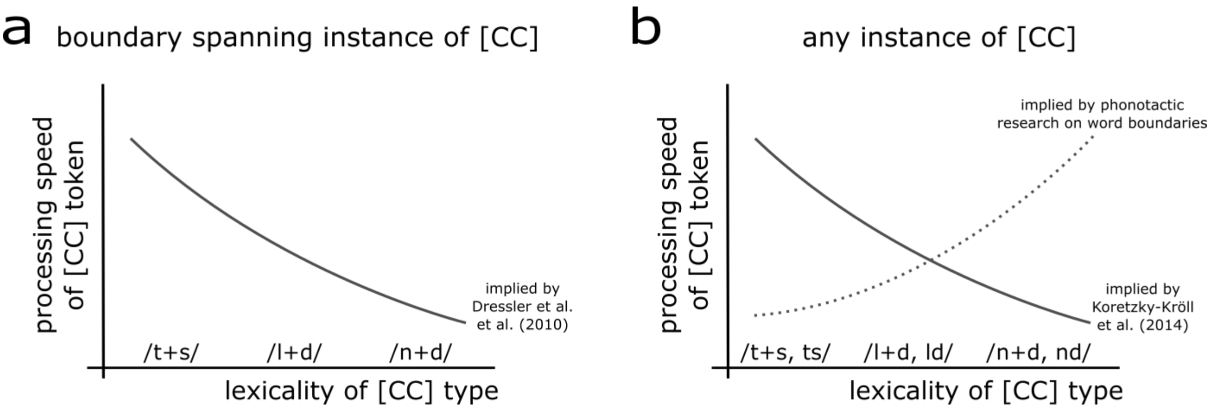

Figure 1. Schematic representation of the hypotheses addressed in this study. (a) Effect of lexicality on the processing of morphonotactic instances of a diphone type. Under one interpretation of SMH (1a), ease of processing is a decreasing function of lexicality (solid line), because diphones signal boundaries less reliably if they also occur morpheme internally. The question is addressed by our first Experiment 1. (b) Under a second interpretation of SMH (2), ease of processing of consonant diphones in general is a decreasing function of lexicality (solid line), because morpheme-boundary signaling diphones need to stand out to be detected easily. Phonotactic research on word boundaries suggests the opposite (dotted line). The question is subject to Experiment 2.

Clearly, interpretation (2) of the SMH lacks part (1b) in the formulation of Dressler and Dziubalska-Kołaczyk (2006). Ambiguity with respect to signaling morphological complexity is not relevant under this interpretation; it is only lexicality that seems to play a role.

In contrast, although not explicitly referring to the SMH, Freiberger et al. (2011) investigate in a series of visual-target experiments if prototypical morphonotactic clusters are processed faster than morphonotactic default clusters, thus explicitly covering (1b). Likewise, diachronic studies such as Dressler et al. (2010) and [Authors] focus on the relevance of ambiguity with respect to signaling morphological complexity to the diachronic stability (i.e. resistance against deletion processes) of clusters. In a (neuro-)computational simulation study, Calderone et al. (2014) find differences between the representational setup of purely morphonotactic and ambiguous clusters. In doing so, these studies tackle (1b) but they do not test whether exclusively boundary-spanning clusters are processed more easily (or acquired earlier or diachronically more stable), than exclusively morpheme-internal clusters, i.e. (2).

Indeed, the logical relationship between (1a), i.e. that clusters facilitate morphological decomposition, and the operational hypothesis (2) tested in many of the above mentioned studies - as relevant and interesting as it may be in itself - is not entirely clear. For instance, it can be argued that, even though morphonotactic clusters fulfill their function of signaling morpheme boundaries, they are not acquired earlier (cf. Freiberger, 2007) or produced more accurately (cf. Leykum et al., 2015b) than their morpheme internal counterparts, just because morphological processing takes its cognitive toll. Likewise, the very fact that morphonotactic clusters are generally less preferred than lexical clusters from an articulatory and perceptual perspective (cf. Marecka and Dziubalska-Kołaczyk, 2014 and section 1.2) could mean that the processing of morphonotactic clusters is hampered, so that any advantage due to boundary signaling is immediately overridden. Conversely, showing that morphonotactic clusters are detected faster than phonotactic clusters (cf. Korecky-Kröll et al., 2014), i.e. (2), does not immediately entail that the former assist the speaker/hearer in recognizing a morpheme boundary, because it could in principle be the case that morphonotactic clusters are detected earlier just because they are located at very prominent positions (in that sense, facilitated processing of a cluster would be an epiphenomenal consequence of morphological parsing rather than the reverse).

We conclude that comparing the processing of morphonotactic against that of phonotactic clusters does not *a priori* allow for immediate conclusions about the presence of a signaling function in morphonotactic clusters, as proposed in (1a). Rather, we suggest that the ambiguity of morphonotactic clusters in signaling morpheme boundaries as originally proposed in (1b) should be taken as a more reliable diagnostic tool for testing the existence of their signaling effects, i.e. (1a). Clearly, if morphonotactic clusters exhibit a signaling function, then this function is expected to be diminished by ambiguity, so that clusters with differential degrees of ambiguity should also show differential degrees of ease of processing. This follows from basic principles of semiotics (Peirce, 1965). By contraposition, the absence of differences in processing among clusters with differential degrees of ambiguity renders compositionality signaling as dominant factor in the processing of morphonotactic clusters unlikely.

This stresses the relevance of an approach that explicitly incorporates ambiguity with respect to signaling morpheme boundaries as an explanatory factor. We do so in two slightly different experiments. Importantly, the differential design of these experiments allows for addressing both the hypothesis that the processing of morphonotactic clusters is diminished by ambiguity (Figure 1a), as well as the hypothesis that the processing of clusters in general is influenced by their degree of lexicality (Figure 1b). In the analysis of our experiments, we will employ both type and token frequencies of diphones to assess their degree of ambiguity.

2 Research questions

We will make use of the following terminology (see also Table 1). For our study, we conceptualize the 'degree of lexicality' of a cluster type as the amount of phonotactic instances among all instances of that type (i.e. phonotactic instances plus morphonotactic instances). The higher the degree of lexicality, the more phonotactic instances there are for a given cluster type. Note that we focus on the investigation of morphonotactic clusters with variable degrees of lexicality. This is why we obviously do not include purely phonotactic clusters, although our data set includes cluster types that surface morphonotactically extremely rarely (cf. Table A1). As a consequence, 'ambiguity' is largest if there are as many lexical as morphonotactic instances of a cluster type. We investigate two slightly different research questions:

(3) Research questions:
   a) Are there differences in how quickly cross-morphemic consonant-diphone instances with variable degrees of lexicality are processed in speech perception?
   b) Are there differences in how quickly instances of consonant-diphone types with variable degrees of lexicality are processed in speech perception?

The difference between these two research questions is very subtle. Question (3a) is about the processing of morphonotactic clusters (cf. Figure 1a), while (3b) is about the processing of clusters in general (Figure 1b). While the items tested in the latter question are cluster types that could be classified as primarily morphonotactic clusters (e.g. ENG /md/), ambiguous clusters (/ld/) or primarily lexical clusters (/nd/), the former question is about the processing of cross-morphemic instances of primarily morphonotactic clusters (/md/ in *seem+ed*), ambiguous clusters (/ld/ in *call+ed*) and of primarily lexical clusters (/nd/ in *bann+ed*).

The reason why we have chosen this set of research questions is twofold. First and foremost, answering (3b) will allow us to evaluate whether the experiment that addresses (3a) really captures the processing of morphonotactic instances of cluster types. In this sense, (3b) functions as a control hypothesis. If there is no difference between the respective outcomes of the experiments, then our experiment obviously failed to address morphonotactic clusters specifically. The second reason is that while (3a) directly refers to (1b) and is thus of major relevance to our study, (3b) relates to the operational hypothesis (2) mentioned in 1.3, namely that morphonotactic cluster tokens should on average be processed faster than phonotactic ones. Clearly, (3b) is not exactly the same since by design we do not actually consider purely phonotactic cluster types. However, we think that (3b) is nevertheless an interesting extension of what has been tested frequently in morphonotactic research (Korecky-Kröll et al., 2014). Moreover it relates to within-type comparison studies of the processing of morphonotactic and phonotactic clusters (cf. Freiberger, 2007; Leykum et al., 2015b) because it can be argued that if morphonotactic instances of some cluster type are on average processed faster than phonotactic instances of that type (i.e. hypothesis (2)), then, everything else being equal, (a random representative of) a less lexical cluster type should be on average processed faster than (a random representative of) a more lexical cluster type (i.e. the scope of (3b)). Whether or not (2) or (3b) are actually related to the SMH (1) is a different question, although not an uninteresting one, as we have pointed out in the previous section.

We are able to compare these two research questions with the differing experimental setups in Experiment 1 and Experiment 2 below, the former addressing (3a) and the latter addressing (3b). They shall be described in more detail in the following.

3 Experiment 1

3.1 Methods

3.1.1 Materials

A set of token-wise equally frequent Polish consonant diphones differing in their variability of occurrences within morphemes and across morpheme boundaries have been selected. These were the following medial sequences: /ʂk, lk, ɕm, vn, ʐn, kw, ɕɲ, lɲ, ɕl, zn/. Variability was operationalized by determining the fraction of morpheme internal occurrences for each diphone type (in terms of token frequency, see 3.1). Frequency counts were taken from a corpus of spoken Polish collected by one of the authors [Authors]. The particular set of consonantal diphones was chosen for several reasons:

a) They are existing Polish consonantal diphones in order to ensure the familiarity of our Polish participants with them;
b) They differ in their ambiguity (both, in terms of types and tokens), as it is the influence of ambiguity on processing speed that we set out to test;
c) They are roughly equally frequent in order to avoid effects of variable entrenchment (frequency effects);
d) They are all reasonably frequent to ensure that the participants are familiar with both their morphological and lexical instances;
e) The set is sufficiently large in order to include a range of different consonants so as to exclude articulatory bias;
f) We wanted to exclude the length of the sequence as an additional variable, hence only diphones were considered.

The only category of consonantal diphones fulfilling all of the above criteria are word-medial consonantal diphones listed above (Table 1 below lists the 9 cluster types together with their lexicality scores, frequency counts, and other properties).

It is worth noting that the comparably large size and diversity of the set of consonant diphones used in this study could only be achieved in the first place since Polish in general features a huge number of consonant-diphone types (138 initial, 382 medial, and 34 final consonant-diphone types in our underlying corpus of spoken Polish, of which the above 9 types fit the criteria (a-f) above; see also Zydorowicz et al., 2016).

The diphones were embedded in nonce words in order to prevent the token frequency of actual lexical items, as well as the relationship between the frequency of the base and the derived word (cf. Hay, 2001) from affecting the results. The stimuli were recorded by a native speaker of Polish (one of the authors) in the anechoic chamber of the Centre for Speech and Language Processing at Adam Mickiewicz University in Poznań through a head-mounted condenser microphone (Sennheiser HSP2) plugged into a Roland Duo Capture USB interface. The audio interface was connected to a laptop computer running the Speech Recorder program (Draxler and Jänsch, 2015), used to display the stimuli and automatically save the recorded words to individual, and uniquely named sound files. All recordings were normalized with respect to duration: 600ms.

Table 1. Cluster types together with measures considered in the experiments. For a given cluster, lexical type ratio is the fraction of word types featuring the cluster morpheme internally in medial position among all word types featuring that cluster medially. Lexical probability is the fraction of medial and morpheme internal tokens of that cluster among all medial instances. Numeric measures are derived from [Authors]. Note that some diphones feature fractions of 1.00 based on this corpus; however, they do in fact rarely occur across morpheme boundaries. MoA and PoA denote manner and place of articulation, and 1 and 2 the first and second segment of the diphone, respectively. NAD denotes whether the cluster is preferred according to the net auditory distance metric. Preferred items show high articulatory intersegmental contrast (see 3.2.1 for details).

cluster	Lexical type ratio	Lexical probability	Token frequency	MoA1	MoA2	PoA1	PoA2	NAD
ɕɲ	1.00	1.00	84	fricative	nasal	coronal	coronal	yes
ɕl	1.00	1.00	71	fricative	liquid	coronal	coronal	no
ʂk	0.95	0.90	100	fricative	stop	coronal	dorsal	yes
kw	0.88	0.96	85	stop	glide	dorsal	dorsal	no
zn	0.86	0.96	71	fricative	nasal	coronal	coronal	yes
lɲ	0.56	0.94	87	fricative	nasal	coronal	coronal	yes
vn	0.54	0.57	93	fricative	nasal	labial	coronal	yes
lɲ	0.45	0.24	84	liquid	nasal	coronal	coronal	yes
ɕm	0.02	0.01	95	fricative	nasal	coronal	labial	no

3.1.2 Participants

Twenty-two participants took part in Experiment 1. They were native speakers of Polish, undergraduate students at Adam Mickiewicz University in Poznań. They signed consent forms and filled out personal questionnaires. None of the participants reported any speech disorders.

3.1.3 Design

The materials were used for an AX discrimination task. Instructions were presented on a computer screen, auditory stimuli were presented through headphones, and participants responded by pressing keys on a keyboard.[6] The test phase was preceded by a training phase with additional, unrelated items which were superficially similar to the test items that followed. The experiment was implemented in PsychoPy (Peirce, 2007), an open-source Python-based software.

Altogether, the test phase consisted of 110 trials: 90 with test items and 20 with distractors. Each trial began with a pair of actual Polish words in which the respective diphone spans a morpheme boundary (‘priming pair’), and the participant had to make a decision as to whether the two items were the same or different. These responses were not recorded, as the sole purpose of the priming pairs was to induce the processing of the diphone as morphonotactic: the participants were primed with words in which the test diphone spans a morpheme boundary. This was meant to ensure that the processing of morphonotactic items is evaluated, as formulated in research question (3a). Afterwards, a pair of nonce words with the same diphone (‘test pair’) was presented, and the participant had to make a decision as to whether the two items were the same or different. Accuracy and reaction times of correct responses to the nonce word pairs in the test phase were recorded. The reaction time clock was started at the onset of X of AX in each test pair. Thus, participants were exposed to 220 word-pair stimuli including primes. Table 2 illustrates the procedure for one token.

Table 2. An illustration of the experimental procedure for one token of one diphone: /ɕm/ in Experiment 1. The primes were meant to induce the treatment of the diphone in the nonce word as spanning a morpheme boundary, but we do not want to prejudge the issue of whether the diphone really was processed as morphonotactic. In this example, the correct response is ‘different’.

<table>
<tr><td colspan="3">1. Priming pair</td><td colspan="3">2. Test pair</td></tr>
<tr><td colspan="2">Exposure</td><td>Decision</td><td colspan="2">Exposure</td><td>Decision</td></tr>
<tr><td>/ʂliɕ+mɨ/</td><td>/ʂliɕ+mɨ/</td><td rowspan="2">Same or different?</td><td>/iɕmi/</td><td>/ɛɕmi/</td><td rowspan="2">Same or different?</td></tr>
<tr><td>‘we went’</td><td>‘we went’</td><td>-</td><td>-</td></tr>
</table>

3.2 Analysis

Overall, N=1980 responses were collected. A single data point was deleted as it showed a reaction time of almost zero. Of the remaining responses, 1906 (96%) were correct. Mean reaction time was 1.153s (SD=0.349). See Figure 2a for the distribution of reaction times. In the following we describe the statistical analysis of the collected data.

3.2.1 Variables

Two outcome variables were considered in our analysis: reaction time (RT) and accuracy. Response time was measured in seconds and therefore implemented as a continuous variable, thereby considering only those word pairs featuring the same diphones that have

[6] To control for possible influence of participants’ handedness, it was included as a regressor in the analysis. See 3.2.1.

been identified correctly, whereas accuracy was measured as a binary variable assuming the values 1 (‘similarity correctly identified’) or 0 (‘similarity not correctly identified’, defined as baseline). We are interested in the way in which the ambiguity of a cluster with respect to signaling a morpheme boundary affects processing. As discussed before, there are several options of how ambiguity can be operationalized.

First, we compute `lexical probability` of a diphone type by calculating the fraction $tokens_{lex}/(tokens_{lex}+tokens_{mpt})$, where $tokens_{lex}$ and $tokens_{mpt}$ are the token frequencies of lexical and morphonotactic (i.e. boundary spanning) occurrences of that diphone, respectively. Token frequencies were taken from [Authors]. Thus, `lexical probability` assumes scores in the unit interval. Cluster types closer to 1 are high probability clusters, while cluster types closer to 0 are low probability clusters. If the production, perception and processing of a cluster primarily depends on the number of lexemes it occurs in rather than on its utterance frequency (cf. Pierrehumbert, 2016) type frequencies should be considered. Thus we compute the `lexical type ratio` as $types_{lex}/(types_{lex}+types_{mpt})$, where $types_{lex}$ and $types_{mpt}$ are the respective morpheme internal (lexical) and boundary spanning (morphonotactic) type frequencies.

Both measures range from primarily lexical or high probability (score close to 1) to primarily morphonotactic or low probability (score close to 0) with perfectly ambiguous diphones in the middle at 0.5. That is, if a listener is exposed to a perfectly ambiguous cluster she has a chance of 50% to correctly predict the presence of a morpheme boundary. Thus, we operationalize the ambiguity of a diphone in the narrow sense (`ambivalence`) by means of how close the diphone type is to being perfectly ambiguous, i.e. `ambivalence` is defined as $1-|p - 1/2|/2$ where p is either `lexical probability` or `lexical type ratio` (which shall be denoted as `token ambivalence` and `type ambivalence` respectively). A score close to 1 means that a diphone type is very ambiguous, while a score close to 0 means that it is not (i.e. either almost exclusively lexical or morphonotactic, respectively). We will refer to `lexical probability`, `lexical type ratio`, `token ambivalence` and `type ambivalence` as primary predictors. Table 3 gives an overview of these four measures.

Table 3. Four different ways of measuring ambiguity in signaling morpheme boundaries.

Measure	Involved frequency measure	Computation	Terminology (0 vs. 1)	Maximally ambiguous score
`lexical probability`	tokens	$tokens_{lex}/(tokens_{lex}+tokens_{mpt})$	Low probability vs. high probability	1/2
`lexical type ratio`	types	$types_{lex}/(types_{lex}+types_{mpt})$	Primarily morphonotactic vs. primarily lexical	1/2
`token ambivalence`	tokens	$1-\|p - 1/2\|/2$; p = `lexical probability`	Non-ambiguous vs. ambiguous	1
`type ambivalence`	types	$1-\|p - 1/2\|/2$; p = `lexical type ratio`	Non-ambiguous vs. ambiguous	1

Due to the experimental design (AX), the binary variable `condition` (`same/different`) was included as an additional categorical predictor. A number of (potentially) phonologically relevant factors entered our analysis as secondary predictor variables. First, preferability classification based on Net Auditory Distance (`NAD`) was included as a measure of well-formedness of a cluster (Dziubalska-Kołaczyk, 2014). In a nutshell, NAD measures the articulatory difference between segments involved in the composition of a cluster in terms of manner and place of articulation. If this difference is larger than the contrast between the consonantal segments and their neighboring vowels, then a cluster is assumed to be preferred, and dispreferred otherwise. Binary values (`preferable: yes/no`) were computed for all clusters with the 'NAD Phonotactic Calculator' (Dziubalska-Kołaczyk et al., 2014), also considering consonant voicing. Second, token frequencies (`frequency`) were retrieved from [Authors]. Third, articulatory features (manner of articulation and place of articulation) of the first (`MoA1`, `PoA1`) and the second consonant (`MoA2, PoA2`) of the diphone were determined. We opted for a rather rough articulatory classification due to the relatively small number of diphone types in our study: `fricative` (baseline), `liquid`, `nasal`, `stop`, and `coronal` (baseline), `dorsal`, `labial`. Since phonological proximity of nonce words to actual Polish words can influence reaction times, we included edit distance (`edit`) between both nonce words in each trial and their closest neighbor as additional covariate. Due to the way in which responses were recorded, `handedness (left, right, n/a)` was included as an additional factor. Finally, `participant` was included as a cluster variable (random effect) in our analysis. There were no repeated measures per test item (nonce-word pair) per participant. In hierarchical models, random effects are assumed to be nested (Baayen et al., 2008: 391; West et al., 2015). Consequently no additional random effect was considered.

3.2.2 Calculation

In order to assess the effect of ambiguity in compositionality signaling on reaction time, a generalized additive mixed model (GAMM, Wood, 2006) was fitted to the data. The choice of GAMMs as opposed to (generalized) linear models was crucial, since we did not want to limit our analysis to linear or, more generally, monotone dependencies between lexical probability and the ease of processing consonantal diphones. In GAMMs, continuous variables can be integrated as so-called smooth terms, i.e. curves, allowing for more complicated functional relationships (Wood, 2006).

In a nutshell, smooth terms are composed of several relatively simple functions (so-called 'basis functions') which are added up in order to yield a more complicated curved shape which fits well to a given set of data points (hence 'additive' model). The composed function is then fit to the data so that its deviation from the data points (i.e. residuals as in conventional regression models) and at the same time the overall curvature ('wiggliness') is minimized. The family the basis functions belong to can be specified by the modeler. In our case we selected so-called 'thin-plate regression splines' which have the advantage that the modeler does not have to bother about where to place the basis functions (the computational cost incurred by this function family can be neglected given the relatively small sample size in our case). In addition, we allowed smooth terms to vanish ('shrinkage smoother') so that they effectively drop out of a model. In our model selection procedure (to be described below) this is particularly useful since we deliberately kept smooth terms of the respective ambiguity measure in all models. The number of basis functions is then determined automatically during the modeling procedure based on an initial value which can be specified by the modeler. The selected initial number of basis functions was checked with the `gam.check` function in order to avoid overspecification (`mcgv` package; Wood 2006).

As in generalized regression modeling, various link functions and distributional families can be implemented into GAMMs (hence 'generalized'). In the present analysis, we opted for an exponential model with inverse link. First, this transformation (i.e. 1/RT) accounts for the slightly positively skewed distribution of reaction times (see Figure 2a), second, and more importantly, reciprocal reaction time 1/RT can be interpreted as reaction or processing speed (see also Kliegl et al., 2010; Balota et al., 2013; Lo & Andrews, 2015).

Finally, random effects can be implemented as well (hence 'mixed') in order to capture hierarchical data structure, i.e. clustered data such as multiple data points belonging to a single participant. GAMMs allow for complex mixed effects (smoothing over every single cluster in the data). In our case we opted for the GAMM analogue of random intercepts to model participant random effects. All calculations were done in R (R Development Core Team, 2013). GAMMs were computed with the `mcgv` package (Wood 2006, see appendix for details on the R code used).

We employed the following bottom-up nesting procedure, in order to derive the most parsimonious and at the same time the most informative model for each constellation (West et al., 2015), starting with a minimal model in which reaction time only depends on `lexical probability`.[7] Pairs of nested models differing in exactly one predictor variable were compared with the `compareML` function from the `itsadug` package (van Rij et al. 2015). If the larger model was preferred to the smaller model, the latter model was rejected, and retained otherwise. In case of multiple models scoring better, the one with the lowest AIC (also provided by `compareML`) was selected. This procedure was applied iteratively until the model could not be further improved by adding fixed and/or random effects (Model 1.1). With the same procedure, three additional models were computed, one in which `lexical probability` was replaced by `lexical type ratio` (Model 1.2), two in which `lexical probability` was replaced by `token ambivalence` computed via token frequencies (i.e. `lexical probability`; Model 1.3) and `type ambivalence` computed via type frequencies (i.e. `lexicality ratio`; Model 1.4), respectively.

Concerning `accuracy`, ceiling effects could be observed (the number of incorrect responses was extremely low at 4%), which rendered any statistical analyses of this variable unfeasible. Consequently, we will neglect `accuracy` scores for the remainder of this paper.

3.2.3 Results

Each model resulting from the optimization procedure described in the previous section only contains its respective primary predictor from the list presented in Table 4 (as smooth term), as well as `condition` (as expected, `same` word pairs throughout led to significantly faster reaction times than `different` word pairs, see Figure 2b) and `participant` (both reaching statistical significance in all cases). All remaining variables turned out not to

[7] We are aware that optimal models determined through model-optimization procedures are in general inferior to averaged models generated by multimodel-inference techniques (Burnham & Anderson 2002). However, averaging of generalized additive mixed models is still subject to ongoing research (Grueber et al. 2011). We thus stick with more traditional step-wise model optimization to identify the best model. Nevertheless, we will employ certain methods from the multimodel-inference paradigm in the post-hoc analyses of our results (see 3.2.3)

contribute to the predictive strength of the models and were excluded by the model-optimization procedure.

Let us focus on what is most interesting, namely the primary predictors. Table 4 gives an overview of the most relevant features of Models 1.1-1.4 (see appendix for more details). We find that the effect that `lexical probability` exerts on 1/RT (significantly non-trivial smooth term at *p*=0.039; Model 1) exhibits the shape of a U which is significantly different from the null-assumption (i.e. 0 baseline). Diphone types in the middle of the spectrum take longer to be processed than those that surface either within morphemes or across morpheme boundaries (Figure 2c, left). As can be seen from the confidence region, the difference between diphones on the lexical end and those in the middle of the `lexical probability` spectrum can be classified as more substantial than the difference between the latter and low probability diphones. In contrast, looking at type frequencies (i.e. Model 1.2, `lexical type ratio`) we find that processing speed mildly increases the more lexical diphones are (marginally significant smooth term at *p*=0.092, Figure 2c, right) with the steepest slope on the lexical end of the spectrum.

The models in which `type/token ambivalence` figures as primary predictor show even clearer results. In both cases, ambiguity decreases processing speed. In the case of we `token ambivalence` we find a significant decreasing linear effect (*p*=0.010, Figure 1d, solid line), while `type ambivalence` only yields a marginally significant linear to mildly concave effect on processing speed (*p*=0.056, Figure 2d, dashed line). Note that the latter curve is persistently less steep than the effect imposed by `token ambivalence`.

Given that all resulting models predict the same outcome variable (namely processing speed), we can apply post-hoc model-comparison techniques in order to assess which model, and in turn which primary predictor, accounts best for the differences in processing speed. Thus, we derive Akaike weights (Burnham & Anderson, 2002; Burnham et al., 2011) from the respective model AICs, assuming that the set of candidate models consists of Model 1.1 to 1.4. A model's Akaike weight can be interpreted as the probability of the model given the data and all other competing candidate models. Akaike weights for the four models are shown in Table 4 (see brackets in AIC column).

Table 4. Model overview for Experiment 1. For further details see appendix. Significance code: '*': $p < 0.05$; '°': $p < 0.1$.

	Primary predictor	Significance of primary predictor	Shape of primary predictor	AIC (Akaike weight)	Visualization
Model 1.1	`lexical probability`	*p* = 0.039 *	U shaped	176.12 (0.27)	Fig. 1b-c (left)
Model 1.2	`lexical type ratio`	*p* = 0.092 °	Convexly increasing	178.79 (0.07)	Fig. 1b-c (right)
Model 1.3	`token ambivalence`	*p* = 0.010 *	Linearly decreasing	174.71 (0.56)	Fig. 1d (dashed)
Model 1.4	`type ambivalence`	*p* = 0.056 °	Slightly concavely decreasing	178.19 (0.18)	Fig. 1d (solid)

It can be seen that Model 1.3 scores highest and that its probability is more than twice as large as that of the second-best Model 1.1. Models 1.4 and finally 1.4 show a much lower probability. This further corroborates what we have pointed out above: type frequencies are less relevant than token frequencies, and within each way of measuring frequency ambivalence is a better predictor than the fraction of boundary spanning items. Overall, it seems to be `token ambivalence` which captures ambiguity in signaling morpheme boundaries best.

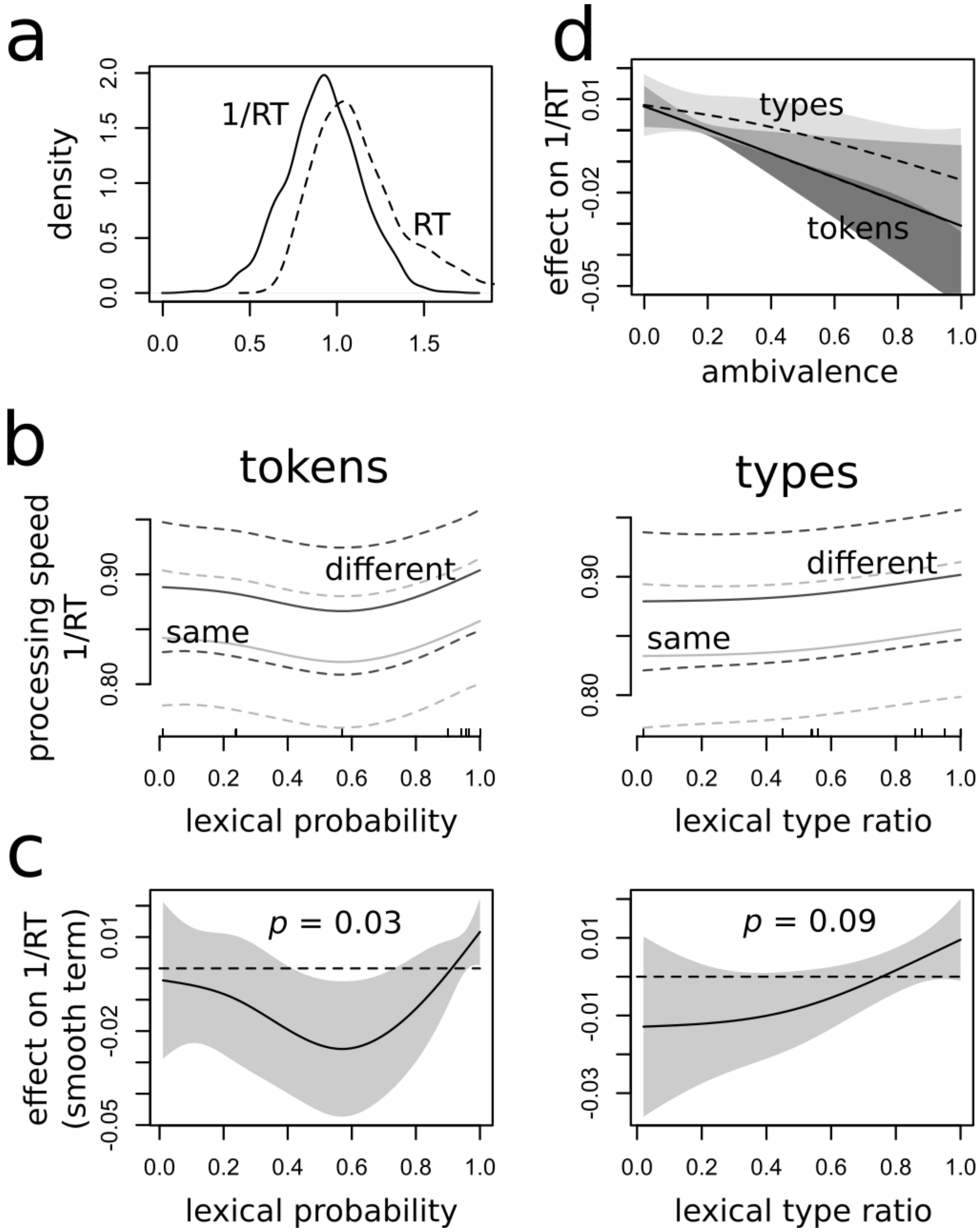

Figure 2. (a) Distributions of reaction time RT and processing speed 1/RT, respectively, the latter being less skewed. (b) Predicted GAMMs of processing depending on lexical probability (Model 1.1 on the left) and lexical probability (Model 1.2 on the right), respectively, as well as condition (same or different nonce words in the stimulus). Stimuli comprised of different nonce words are processed

faster. (c) Smooth terms of the effect of lexical probability (left) and lexical type ratio (right) on 1/RT (Model 1.1-2). The shape of the effect of lexical probability resembles a U with items in the mid range being processed significantly slower. (d) Smooth terms of two different ambiguity measures: `token ambivalence` (dashed, dark gray; Model 1.3) and `type ambivalence` (solid, light gray; Model 1.4) affecting 1/RT. Reaction speed decreases significantly in both measures.

3.3 Discussion

In this experiment, we were testing the effect of ambiguity in signaling morpheme boundaries on the processing of diphones that are in fact spanning a boundary. In order to encourage participants to analyze a diphone surfacing in a stimulus (nonce word) as spanning a boundary, primes were first presented to the participants in which diphones signal a morpheme boundary (we will see in the discussion of the second experiment lacking primes that the primes in the first experiment indeed have an effect).

There are two main findings to be discussed. First, ambiguity measures based on token frequency show larger effects on processing than ambiguity measures based on type frequency do. In fact, the latter effects were only marginally statistically significant. As a consequence, this means that the heuristic that listeners rely on in order to analyze whether or not a boundary is present is based on previously encountered utterance frequencies rather than on the number of word types a diphone occurs in. We will come back to this in the conclusion section.

Second, the ease of processing of boundary spanning diphones is a decreasing function of ambiguity rather than a decreasing function of lexical probability. This is evident from the U shape of the effect that lexical probability exerts on processing speed and becomes even clearer when the (linearly) decreasing effect of ambivalence on processing speed is considered. The result is surprising under the assumption that participants actually analyzed diphones as boundary-spanning instances. This is so because low-probability diphone types are expected to provide much worse boundary-signaling cues than those that signal a boundary more often (e.g. 50% of the time).

There are at least two possible explanations to this. First, it could be the case that positive effects on processing imposed by lexical licensing (i.e. the abundant presence of diphones within morphemes) overshadows the negative effects that result from deficient boundary-signaling properties of low probability diphones. We will see in the next section that this possibility can be ruled out because high probability diphones are not processed faster in the absence of boundary-signaling primes.

Second, it is possible that the participants in fact did not always analyze the diphones as spanning a boundary. That is, the prime did perhaps not trigger the decomposition of the subsequently presented nonce word into two parts. Under that interpretation, primes can be assumed to be most successful in low probability diphones. During perception, diphones are either categorized as signaling a boundary or occurring morpheme internally (in spite of the presented primes) but this categorization process is inhibited if the boundary-detection heuristic available to the speaker is not reliable. Consequently, reaction speed is lowest in maximally ambivalent diphones.

Finally, the robustness of our analysis is supported by the fact that neither phonological factors (manner and place of articulation), nor handedness or the proximity of nonce words to existing Polish lexemes (edit distance) contributed to the quality of our models. This indicates that the set of diphone types considered in this study is relatively balanced. Interestingly, the wellformedness metric NAD (Dziubalska-Kołaczyk, 2014) did not contribute to the explanation of differences in RT either. One would expect diphones which are preferred according to the NAD principle to show higher processing speed in our experiment than dispreferred diphones because NAD-preferred items are postulated to have advantages during perception. The possibility remains that the effects of NAD are obscured by that of morphological signaling (or frequency, see Experiment 2).

4 Experiment 2

4.1 Methods

4.1.1 Materials

The stimuli in Experiment 2 were the same as those in Experiment 1 with the sole exception that no existing Polish words were included.

4.1.2 Participants

Thirteen new participants took part in Experiment 2, all of them being native speakers of Polish. Again, none of the participants reported any speech disorders. The number of participants was determined in such a way that there are approximately equally many data points in both experiments. This helped to exclude sample size as a potential explanatory factor of the differences between Experiment 1 and Experiment 2.

4.1.3 Design

The design of Experiment 2 closely matches that of Experiment 1, with one major difference being that Experiment 2 does not include primes (the procedure is presented in Table 5). This control experiment is meant to show whether the priming implemented in Experiment 1 addressing research question (3a) had an effect, and to address research question (3b). Thirteen (new) participants took part in Experiment 2. Alltogether, Experiment 2 consisted of 192 word pairs, among them 150 test pairs and 42 distractor pairs. Recall that Experiment 1 featured 220 word pairs, hence both experiments took roughly the same time in total. Accuracy and reaction times of correct responses to the test pairs were recorded.

Table 5. An illustration of the experimental procedure for one token of one diphone, [ɕm], in Experiment 2.

<table>
<tr><td colspan="3">Test phase</td></tr>
<tr><td colspan="2">Exposure</td><td>Decision</td></tr>
<tr><td>[iɕmi]</td><td>[ɛɕmi]</td><td>Same or different?</td></tr>
</table>

4.2 Analysis

In total 1950 responses were collected, 1889 (97%) of which were correct. Thus, sample sizes are roughly equal in both experiments (1889 vs. 1906). Mean reaction time was 1.05s (SD=0.344s). The distribution of reaction times in shown in Figure 2a.

4.2.1 Variables

Variables were defined and analyzed precisely as for Experiment 1 (3.2.1).

4.2.2 Calculation

The statistical modeling procedure matches the one presented before in 3.2.2. That is, four models were computed, one for each primary predictor (`lexical probability`, `lexical type ratio`, `token ambivalence`, `type ambivalence`). The analysis of accuracy scores was omitted again due to clear ceiling effects.

4.2.3 Results

As in the analysis of Experiment 1, `condition` and the respective primary predictor survived the optimization procedure in all of the four models. However, none of the effects of the primary predictors reached statistical significance. Table 6 shows the main characteristics of the computed models.

Table 6. Model overview for Experiment 2. For further details see appendix ('n.s.' denotes 'not significant')

	Primary predictor	Significance of primary predictor	Shape of primary predictor	AIC (Akaike weight)	Visualization
Model 2.1	`lexical probability`	*p* = 0.940 (n.s.)	Flat	29.85 (0.30)	Fig. 2b (left)
Model 2.2	`lexical type ratio`	*p* = 0.219 (n.s.)	Flat	31.67 (0.12)	Fig. 2b (right)
Model 2.3	`token ambivalence`	*p* = 0.613 (n.s.)	Flat	29.92 (0.29)	-
Model 2.4	`type ambivalence`	*p* = 0.935 (n.s.)	Flat	29.92 (0.29)	-

As can be seen from Figure 3b-c, none of the ambiguity measures shows a significant impact on processing speed in the absence of boundary-signaling primes. In addition, token frequency turned out to contribute significantly to the quality of all models (in contrast to Experiment 1). That is, in the absence of primes, the effect of morphological structure is overshadowed by that of token frequency although diphone types with roughly equal frequency were selected for the experiments (cf. 3.1.1; notably none of the ambiguity measures reached statistical significance even in the absence of token frequency as a predictor in the model). Interestingly, the effect of frequency on reaction speed turned out to be non-monotonous (mid-frequency items scoring lower reaction speed) rather than strictly increasing. This is exemplarily shown for Model 2.1 in Figure 3c (the effect of frequency displays a similar shape in all other models, 2.2-2.4). A comparison of the respective Akaike weights (see 3.2.3) reveals that there is no clear single best model. Model 2.2 (featuring `lexical type ratio`) shows the lowest probability given the data and the set of four candidate models.

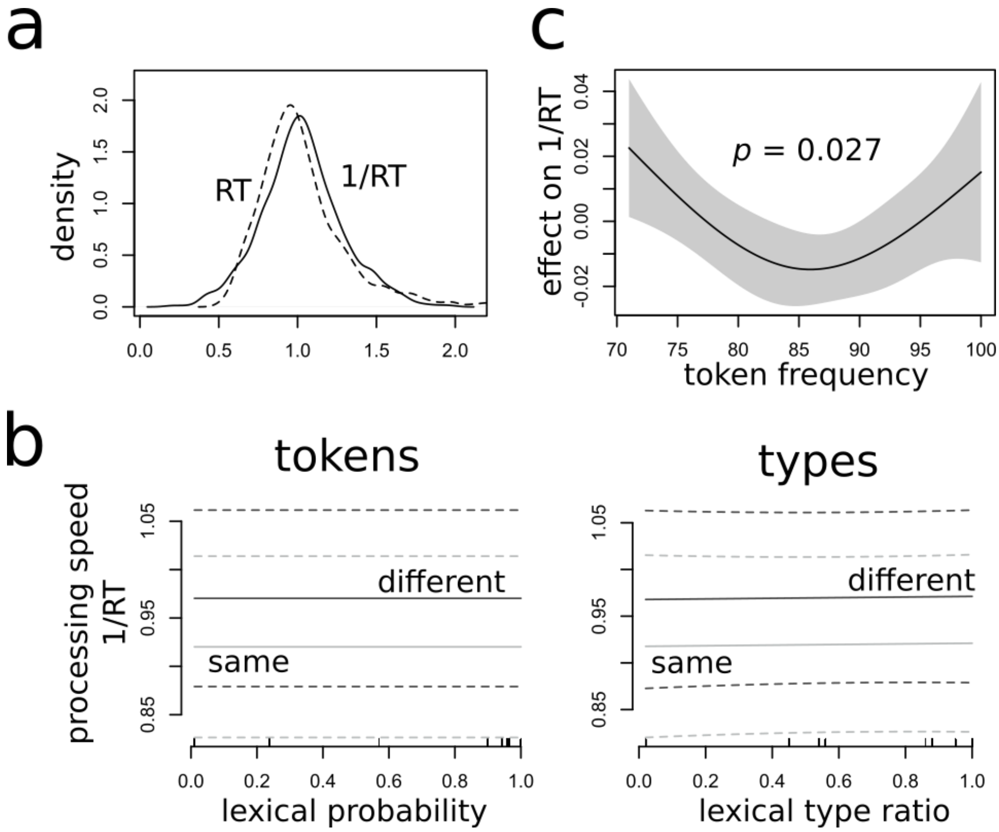

Figure 3. (a) Distributions of reaction time RT and processing speed 1/RT, the latter being slightly less skewed. (b) GAMMs of processing speed depending on ambiguity (lexical probability, Model 2.1, on the left; lexical type ratio, Model 2.2, on the right) and condition. Again, stimuli with different words are processed faster. Neither lexical probability nor lexical type ratio have a significant impact on 1/RT. Models 2.3-4 are not displayed as there is no statistically robust effect either. (c) Significant smooth term of the effect of token frequency on 1/RT in Model 2.1. It has a similar shape in Models 2.2-4.

4.3 Discussion

Our failure to detect any significant effects of ambiguity on processing speed in Experiment 2 can have different causes. First, the collected sample might have been simply too small (although we highlight that sample sizes in both experiments were roughly equal).
Second, and more interestingly, the lack of an effect of ambiguity on processing speed might result from the missing primes. Note that since we did not prime participants for detecting a morpheme boundary, we can assume that participants were free to analyze the diphone in the way they preferred. That is, diphones were analyzed more generally as sequences that may or may not span a morpheme boundary. If this is true, then it is not at all surprising that ambiguity does not affect phonotactic processing. Participants simply choose the diphone

category (boundary vs. morpheme internal) which suggests itself based on distributional grounds. What the result of Experiment 2 then suggests is that in Experiment 1 participants indeed were encouraged to analyze diphones as spanning a boundary. Thus, Experiment 2 functions as a control experiment, which supports our assumption that the primes in Experiment 1 worked in the way they were meant to.

The results of Experiment 2 do have another consequence, as they help to assess research question (3b). As discussed in Section 2, it has been hypothesized that morphonotactic diphones are processed faster than lexical diphones (Korecky-Kröll et al., 2014). The claim is, that since morphonotactic diphones must confidently signal morpheme boundaries they have the necessity to show properties that facilitate processing in order to be easily detected. Thus, we would expect low-probability/primarily morphonotactic diphone types to be processed faster in our experiment than high-probability/primarily lexical diphone types. Based on our results, this cannot be confirmed.

Any potential effects of boundary signaling are overshadowed by token frequency which turns out to predict reaction speed in a U-shaped manner. Mid-range items are processed more slowly than rare or frequent items. This is interesting, as the effect of frequency on processing speed, if there is one, is rather expected to be strictly positive. The effect of frequency on phonotactic processing, however, is not the focus of our study. Diphone types from a broader frequency range are needed to investigate this matter more thoroughly.

Finally, note that response times in Experiment 2 where on average shorter than those in Experiment 1, which could be seen as evidence against the hypothesis that morphonotactic instances of cluster types are processed faster than their homophonous lexical counterparts (Celata et al,. 2015; Leykum et al., 2014a). However, the differences in design between the two experiments render a direct comparison of response times difficult. We thus remain agnostic with respect to this question.

5 Conclusion

Two propositions that are related to or indeed part of what is generally referred to as the Strong Morphonotactic Hypothesis are present in the morphonotactic literature. The first one is that consonant diphones are processed faster the less lexical they are, and in particular that consonantal diphones which span a morpheme boundary (i.e. morphonotactic diphones) are processed faster than morpheme internal (i.e. lexical) consonant diphones (operational hypothesis (2), cf. 1.3). The second hypothesis is that the compositionality-signaling function of consonant-diphone types decreases the more frequently it is also used morpheme internally as this decreases the reliability at which a diphone signals morphological structure (hypothesis (1b)). In the morphonotactic literature, both hypotheses have been suggested to be linked with the hypothesis that clusters have the function of signaling morpheme boundaries (hypothesis (1a)). In this study, we experimentally addressed both hypotheses ((2) and (1b)).

In order to do so, it was necessary to operationalize ambiguity - and at the same time reliability - of signaling morphological structure. We proposed four different ways of doing so: lexical probability (the fraction of boundary-spanning diphone tokens); lexical type ratio (the fraction of word-types in which a diphone spans a boundary); token ambivalence (the extent to which lexical probability deviates from the most ambiguous configuration); type ambivalence (the extent to which the lexical type ratio deviates from the most ambiguous configuration). This allowed us to assess (a) which type of frequencies based on previous exposure and (b) which corresponding heuristic for measuring ambiguity most relevant to morphonotactic processing.

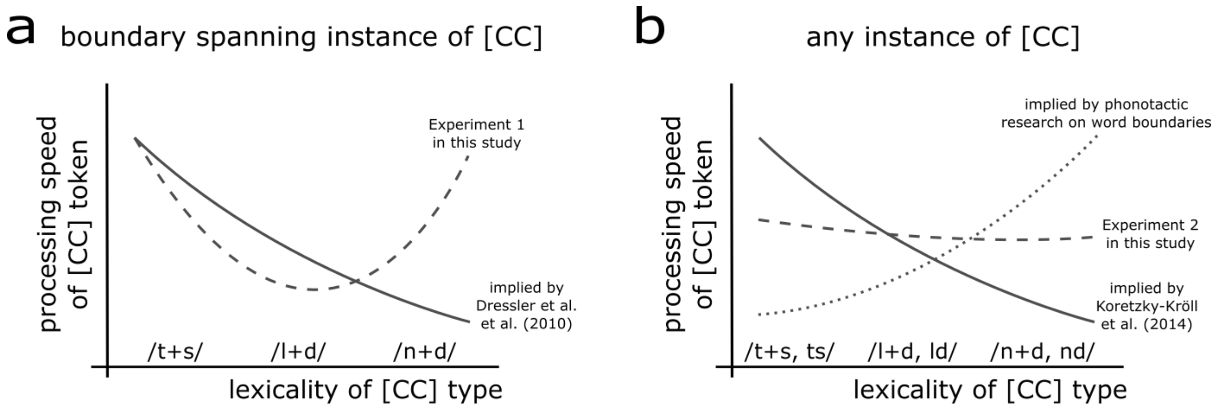

Figure 4. Schematic representation shown in Figure 1 extended by our results (dashed). (a) Research question (3a): from interpretation (1a) of the SMH follows that ease of processing of morphonotactic instances is a decreasing function of lexicality (solid line). Experiment 1 shows that the functional relationship is U shaped, maximally ambiguous clusters scoring the lowest processing speed (dashed line). (b) Research question (3b): interpretation (2) of the SMH implies, ease of processing of consonant diphones to decrease with lexicality (solid line), while phonotactic research on word boundaries suggests the reverse (dotted line). Experiment 2 does not reveal any clear non-trivial relationship (dashed line).

Hypothesis (2) (Figure 4b, solid line) cannot be confirmed by our results. In Experiment 2 we did not reveal any significant effects of a diphone's ambiguity in signaling boundaries on diphone processing (Figure 4b, dashed line). This result is independent of how ambiguity is operationalized. Neither the fraction of boundary spanning word types nor the fraction of boundary spanning tokens showed an effect on reaction speed in our experiment. Thus, as long as there is no morphological processing involved, speakers do not differentiate between diphones which occur always, sometimes, rarely, or never across morpheme boundaries. This contrasts with the findings of Korecky-Kröll et al. (2014). The property of being prone to signaling a boundary alone does not significantly promote a diphone's processing during perception. This goes in line with reported differential effects of morphological structure on the acquisition of consonant diphones (Zydorowicz, 2007; Freiberger et al., 2011). At the same time, our results do not support the hypothesis that low-probability diphones are generally less preferred (i.e. processed more slowly) than their high-probability counterparts if the morphological level is taken into account. If, in contrast, lexical probability is defined as the fraction of word internal items (vs. crossing a word boundary) different pressures seem to apply. On the lexical level, words composed of high-probability diphones are less likely subject to repair processes and hence assumed to be processed faster than words composed of low-probability diphones (Moreton, 2002; see also Vitevich & Luce, 1998, for similar results in nonce words; Figure 4b, dotted line). Thus, it seems that there is a subtle difference between phonotactically guided decomposition of the speech stream into words, and that of words into morphemes, respectively.

Hypothesis (1b) shown in Figure 4a (solid line), and as a consequence likely also (1a), has been partially corroborated. When participants were primed for analyzing a diphone as morphonotactic, they took longer to identify the difference between two words containing that diphone, if it is commonly used ambiguously in speech (rather than either predominantly morpheme internal or spanning a boundary, respectively; Figure 4a, dashed line). Semiotically, this is plausible as signs, and consequently also markers of compositionality, in general tend to avoid ambiguity. In our study, it was token ambivalence which turned out to be the most reliable predictor. Our result has multiple consequences.

First and most fundamentally, it indicates that listeners rely on previously encountered diphone instances when parsing diphones. That is, individuals have detailed memories of diphone instances which also include morphological information, i.e. information about whether or not an encountered diphone has spanned a morpheme boundary. Otherwise, the differential effects of ambiguity on diphone processing observed in our study are hard to explain. Thus, sensitivity to phonotactic probability applies not only on the lexical level as demonstrated previously (Saffran et al. 1998; Vitevich & Luce 1998; McQueen 1998; Adriaans & Kager 2010) but also on the morphological level. This goes very much in line with an exemplar based approach to language perception and production (Pierrehumbert 2001; Wedel 2006; Bybee 2013; Ernestus 2014). Since the diphones in our experiments were embedded into nonce-words this suggests that speakers indeed store exemplars of sublexical units (pace Välimaa-Blum 2009) which carry morphological information, namely whether or not a morpheme boundary is present. This converges with Plag et al. (2017) who demonstrate that properties of sounds (such as duration) indicate the morphological structure of a word. This, as they argue, requires a phonological model which builds on detailed memories of sublexical items, such as those provided by exemplar theory. Finally, this notion also conforms with Calderone et al. (2014) and Celata et al. (2015) who argue that morphonotactic and lexical diphones show differential cognitive representations.

Second, we demonstrated that it is token frequency rather than type frequency which determines whether a diphone is ambiguous. This is interesting as it means that individuals do not necessarily differentiate between multiple lexical types in phonotactic processing. Rather it is overall exposure to a certain sound string which is more relevant. Again, this speaks for relatively self-contained sets of exemplars of sublexical items. However, the relationship between sublexical exemplars and (strings of) lexical exemplars is complicated. On the one hand, the former draw on information about the distributional properties of the latter (because phonotactic knowledge is based on which diphones occur within lexical items). On the other hand, word types are abstracted away during phonotactic processing in favor of a more general classification (boundary vs. no boundary). This entails that mental representations of phonotactic items must be subject to both abstractionist and episodic effects (Adriaans & Kager 2010; Ernestus 2014; Pierrehumbert 2016).

Third, coming back to our primary research question, it is the extent to which the distributional pattern of a diphone type deviates from the most ambiguous configuration (which amounts to tossing a coin) which matters for processing morpheme boundaries. This is interesting, because one would actually expect the probability of spanning a boundary to correlate with processing speed in the presence of boundary spanning primes (cf. solid line in Figure 4a). Why might that be the case? We suggest that the boundary spanning primes in our experiment indeed increased the likelihood of analyzing a subsequently perceived diphone as spanning a boundary as well. However, this likelihood decreases if an encountered item (a boundary spanning diphone) only vaguely fits the previously experienced instances of that diphone type. In the case of a high probability diphone (i.e. a prototypically morpheme internal diphone) this set of previously experienced instances largely consists of morpheme internal items. The consequence of this conflict must be that the listener classifies the encountered item as a member of the more prototypical category, i.e. as being morpheme internal. It is plausible then, that the analysis of diphone types which are maximally ambiguous by representing a nearly equal amount of boundary spanning and morpheme internal instances incurs the highest cost.

Fourth, it is interesting to see that processing speed decreases linearly with ambivalence (i.e. the degree to which an item is ambiguous). This indicates that every additional instance of a diphone type which does not fit to the prototypical usage of that type decreases the quality of the type's signaling function to the same degree. We have demonstrated elsewhere [Authors] by means of computational simulations that the exact shape of the functional relationship between ambiguity and quality of the signaling function of a diphone type can have consequences for the diachronic development of diphone inventories. Strongly convex functional relationships lead to diphone inventories almost lacking any ambiguous diphones while strongly concave relationships promote the stable establishment of some ambiguous diphones. Polish rather belongs to the former category. The hypothesis is that strongly convex relationships characterize languages that make much use of morphology while strongly concave relationships belong to less synthetic languages. The results of the present experiment at least do not contradict this hypothesis in that we did not detect a strongly concave relationship between ambivalence and processing speed in speakers of Polish. Running a similar experiment with speakers of languages that accommodate a larger amount ambiguous diphones, e.g. English (which is at the same time less synthetic than Polish), would be interesting (the conjecture being that having English speakers in the experiment leads to a more concave shape; cf. [Authors]).

One question we cannot address is if there is a general bias towards analyzing ambiguous diphones as morpheme internal. Neither does our second experiment shed any light on this issue (because we cannot be certain as to whether participants classified an encountered item as morpheme internal or boundary spanning), nor does our first experiment do so (because we only provided boundary spanning primes). What would be needed in order to assess whether there is such a bias, is an experimental setup similar to our first one but with morpheme internal instead of boundary spanning primes. We leave this question open for further experimental research on the interaction between phonotactics and morphology.

Acknowledgements
xxx
Funding
xxx

Appendix

Supplementary data: All collected data together with phonological characteristics and ambiguity measures can be found in the supplementary data files `Experiment_1.csv` and `Experiment_2.csv`.

Notation: In all models reported below, the following significance code applies: '***': $p < 0.001$; '**': $p < 0.01$; '*': $p < 0.05$; '°': $p < 0.1$. Here, the null-hypothesis always corresponds to a trivial (zero) effect (or zero intercept). Relevant abbreviations: '*edf*': estimated degrees of freedom (see below); '*SE*': standard error; '*AIC*': Akaike information criterion; '*N*': sample size (correct instances only). A description of the variables involved can be found in 3.2.1 and 4.2.1.

Remark on GAMM code: Smooth terms are coded as `s()` in the `mgcv` package (Wood 2006), and the specification `bs="ts"` refers to thin-plate spline modeling of terms which may vanish (i.e. shrunk to zero). The parameter `k` specifies the initial number of basis functions ('knots'). For modeling random intercepts in `mgcv`, the procedure of implementing random effects as penalized regression terms with `s(...,bs="re")` in `gam()` was chosen (for that reason, `participant` is listed as smooth term, although this variable is clearly categorical). For details on this see Wood (2015). The abbreviation '*edf*' in the `gam()` output refers to estimated degrees of freedom. If a smooth term corresponding to a one-dimensional continuous predictor shows high overall curvature (i.e. if it is highly nonlinear) then *edf* is high, while *edf* is close to 1 if the term is effectively linear. See 3.2.2 for more information on GAMMs.

Model 1.1: Exponential GAMM with inverse link of RT depending on `lexical probability` and `condition`; `participant` as random effect. $R^2 = 0.228$; $AIC = 176.12$; $N = 1907$. R code: `frm = RT ~ s(lexical_probability, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(frm, data = Experiment_1, family = Gamma(link=inverse)).`

Parametric terms			
`intercept`	0.89±0.03*SE*	$t = 32.4$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
`lexical probability`	$edf = 2.4$	$F = 1.7$	$p = 0.039$ *
`participant`	$edf = 20.3$	$F = 30.4$	$p < 0.001$ ***

Model 1.2: Exponential GAMM with inverse link of RT depending on `lexical type ratio` and `condition`; `participant` as random effect. $R^2 = 0.226$; $AIC = 178.79$; $N = 1907$. R code: `frm = RT ~ s(lexical_type_ratio, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(frm, data = Experiment_1, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.89±0.03*SE*	$t = 32.5$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
`lexical type ratio`	$edf = 1.4$	$F = 1.8$	$p = 0.09$ °
`participant`	$edf = 20.3$	$F = 30.4$	$p < 0.001$ ***

Model 1.3: Exponential GAMM with inverse link of RT depending on `token ambivalence` and `condition`; `participant` as random effect. $R^2 = 0.228$; $AIC = 174.71$; $N = 1907$. R code: `frm = RT ~ s(token_ambivalence, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(formula, data = Experiment_1, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.89±0.03*SE*	$t = 32.4$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
`token ambivalence`	$edf = 0.88$	$F = 1.4$	$p = 0.010$ *
`participant`	$edf = 20.3$	$F = 30.4$	$p < 0.001$ ***

Model 1.4: Exponential GAMM with inverse link of RT depending on `type ambivalence` and `condition`; `participant` as random effect. $R^2 = 0.226$; $AIC = 178.19$; $N = 1907$. R code: `frm = RT ~ s(type_ambivalence, k = 5, bs = "ts") + condition + s(participant, bs = "re"); gam(formula, data = Experiment_1, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.89±0.03*SE*	$t = 32.4$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -4.2$	$p < 0.001$ ***
Smooth terms			
`type ambivalence`	$edf = 1.2$	$F = 1.0$	$p = 0.055$ °
`participant`	$edf = 20.3$	$F = 30.3$	$p < 0.001$ ***

Model 2.1: Exponential GAMM with inverse link of RT depending on `lexical probability`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.196$; $AIC = 29.85$; $N = 1889$. R code: `frm = RT ~ s(lexical_probability, k = 5, bs = "ts") + condition + s(frequency, k = 5) + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse)).`

Parametric terms			
`intercept`	0.98±0.05*SE*	$t = 21.3$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -3.5$	$p < 0.001$ ***
Smooth terms			
`lexical probability`	$edf < 0.1$	$F = 0.0$	$p = 0.94$

`participant`	*edf* = 11.8	*F* = 43.7	$p < 0.001$ ***
`frequency`	*edf* = 2.0	*F* = 3.4	$p = 0.027$ *

Model 2.2: Exponential GAMM with inverse link of RT depending on `lexical type ratio`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.195$; *AIC* = 31.67; *N* = 1889. R code: `frm = RT ~ s(lexical_type_ratio, k = 5, bs = "ts") + condition + s(frequency, k = 5) + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse))`.

Parametric terms			
`intercept`	0.98±0.05*SE*	*t* = 21.3	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	*t* = -3.5	$p < 0.001$ ***
Smooth terms			
`lexical type ratio`	*edf* = 1.0	*F* = 0.0	$p = 0.88$
`participant`	*edf* = 11.8	*F* = 43.7	$p < 0.001$ ***
`frequency`	*edf* = 2.0	*F* = 2.4	$p = 0.022$ *

Model 2.3: Exponential GAMM with inverse link of RT depending on `token ambivalence`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.196$; *AIC* = 19.92; *N* = 1889. R code: `frm = RT ~ s(token_ambivalence, k = 5, bs = "ts") + condition + s(frequency, k = 5, bs = "ts") + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse))`.

Parametric terms			
`intercept`	0.98±0.05*SE*	*t* = 21.3	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	*t* = -3.5	$p < 0.001$ ***
Smooth terms			
`token ambivalence`	*edf* < 0.1	*F* = 0.0	$p = 0.61$
`participant`	*edf* = 11.8	*F* = 43.7	$p < 0.001$ ***
`frequency`	*edf* = 2.0	*F* = 2.0	$p = 0.016$ *

Model 2.4: Exponential GAMM with inverse link of RT depending on `type ambivalence`, `frequency` and `condition`; `participant` as random effect. $R^2 = 0.196$; *AIC* = 29.92; *N* = 1889. R code: `frm = RT ~ s(type_ambivalence, k = 5, bs = "ts") + condition + s(frequency, k = 5, bs = "ts") + s(participant, bs = "re"); gam(frm, data = Experiment_2, family = Gamma(link = inverse))`.

Parametric terms

`intercept`	0.98±0.05*SE*	$t = 21.3$	$p < 0.001$ ***
`condition (same)`	-0.05±0.01*SE*	$t = -3.5$	$p < 0.001$ ***
Smooth terms			
`type ambivalence`	$edf < 0.1$	$F = 0.0$	$p = 0.93$
`participant`	$edf = 11.8$	$F = 43.7$	$p < 0.001$ ***
`frequency`	$edf = 2.0$	$F = 2.0$	$p = 0.016$ *

References

Adriaans, F., Kager, R., 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. Journal of Memory and Language 62 (3), 311–331. 10.1016/j.jml.2009.11.007.

Albright, A., 2015. Cluster phonotactics require both syllables and strings: Evidence from Lakhota, Manchester.

Baayen, H., 2013. Multivariate Statistics. In: Podesva, R., Sharma, D. (Eds.) Research Methods in Linguistics. Cambridge University Press, Cambridge, pp. 337–372.

Balota, D.A., Aschenbrenner, A.J., Yap, M.J., 2013. Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. Journal of experimental psychology. Learning, memory, and cognition 39 (5), 1563–1571. 10.1037/a0032186.

Berent, I., Steriade, D., Lennertz, T., Vaknin, V., 2007. What we know about what we have never heard: Evidence from perceptual illusions. Cognition 104 (3), 591–630. 10.1016/j.cognition.2006.05.015.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Burnham, K.P., Anderson, D.R., Huyvaert, K., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol 65, 23–35.

Bybee, J.L., 2013. Usage-based Theory and Exemplar Representations of Constructions. In: Hoffmann, T., Trousdale, G. (Eds.) The Oxford Handbook of Construction Grammar. Oxford University Press.

Calderone, B., Celata, C., Korecky-Kröll, K., Dressler, W.U., 2014. A computational approach to morphonotactics: evidence from German. Language Sciences 46, 59–70.

Celata, C., Korecky-Kröll, K., Ricci, I., Dressler, W.U., 2015. Phonotactic processing and morpheme boundaries: wordfinal /Cst/ clusters in German. Italian Journal of Linguistics 27 (1), 85–110.

Cohen, J., 1992. A power primer. Psychological Bulletin 112, 155–159.

Cruttenden, A., 2014. Gimson's pronunciation of English, 8th ed. Routledge, London.

Daland, R., Pierrehumbert, J., 2011. Learning diphone-based segmentation. Cognitive Science 35, 119-155.

Donohue, M., Hetherington, R., McElvenny, J., Dawson, V., 2013. World phonotactics database. http://phonotactics.anu.edu.au. Accessed September 16th, 2015.

Dressler, W.U., Dziubalska-Kołaczyk, K., 2006. Proposing Morphonotactics. Wiener Linguistische Gazette 73, 69–87.

Dressler, W.U., Dziubalska-Kołaczyk, K., Pestal, L., 2010. Change and variation in morphonotactics. Folia Linguistica Historica 31, 51–68. 10.1515/flih.2010.003.

Dziubalska-Kołaczyk, K., 2014. Explaining phonotactics using NAD. Language Sciences 46, 6–17.

Dziubalska-Kołaczyk, K., Pietrala, D., Aperliński, G., 2014. The NAD Phonotactic Calculator – an online tool to calculate cluster preference in English, Polish and other languages. http://wa.amu.edu.pl/nadcalc.

Dziubalska-Kołaczyk, K., Zydorowicz, P., 2014. The production of high-frequency clusters by native and non-native users of Polish. Concordia Working Papers in Applied Linguistics 5, 130–144.

Ernestus, M., 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. Lingua 142, 27–41.

Freiberger, E.M., Abbrederis, C., Luckabauer, F., Stammler, M., 2011. Morphonotaktik in der Sprachverarbeitung: Online- und Offline-Tests mit deutschsprachigen Erwachsenen. Wiener Linguistische Gazette 75, 33–52.

Fruehwald, J., 2017. Generations, lifespans, and the zeitgeist. Lang. Var. Change 29 (01), 1–27. 10.1017/S0954394517000060.

Grueber, C.E., Nakagawa, S., Laws, R.J., Jamieson, I.G., 2011. Multimodel inference in ecology and evolution: Challenges and solutions. Journal of evolutionary biology 24 (4), 699–711. 10.1111/j.1420-9101.2010.02210.x.

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

Jarosz, G., Calamaro, S., Zentz, J., 2016. Input Frequency and the Acquisition of Syllable Structure in Polish. Language Acquisition. 10.1080/10489223.2016.1179743.

Kahn, D., 1976. Syllable-based generalizations in English phonology. Dissertation.

Kliegl, R., Masson, M.E.J., Richter, E.M., 2010. A linear mixed model analysis of masked repetition priming. Visual Cognition 18 (5), 655-681. 10.1080/13506280902986058.

Korecky-Kröll, K., Dressler, W.U., Freiberger, E.M., Reinisch, E., Mörth, K., Libben, G., 2014. Morphonotactic and phonotactic processing in German-speaking adults. Language Sciences 46, 48–58.

Levelt, C., van de Vijver, R. 1998/2004. Syllable types in cross-linguistic and developmental grammars. In: Kager, R., Pater, J., Zonneveld, W. (Eds.) Constraints in phonological acquisition. Cambridge: Cambridge University Press, pp. 204–218.

Leykum, H., Moosmüller, S., Dressler, W.U., 2015a. Homophonous phonotactic and morphonotactic consonant clusters in word-final position. In: Proceedings of Interspeech 2015, 1685-1689.

Leykum, H., Moosmüller, S., Dressler, W.U., 2015b. Word-final (mor-)phonotactic consonant clusters in Standard Austrian German, in: The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0701.pdf

van der Lugt, A. H. 2001. The use of sequential probabilities in the segmentation of speech. Perception and Psychophysics 63 (5), 811–823.

Marecka, M., Dziubalska-Kołaczyk, K., 2014. Evaluating models of phonotactic constraints on the basis of sC cluster acquisition data. Language Sciences 46, 37–47.

Martin, A., 2007. The evolving lexicon. PhD dissertation, Los Angeles.

Mattys, S. L., Jusczyk, P. W. 2001. Phonotactic cues for segmentation of fluent speech by infants. Cognition 78, 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., Morgan, J. L. 1999. Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology 38, 465–494.

McQueen, J., 1998. Segmentation of Continuous Speech Using Phonotactics. Journal of Memory and Language 39, 21-46.

Moreton, E., 2002. Structural constraints in the perception of English stop-sonorant clusters. Cognition 84 (1), 55–71. 10.1016/S0010-0277(02)00014-8.

Nakagawa, S., Cuthill, I., 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Reviews 82 (4), 591–605. 10.1111/j.1469-185X.2007.00027.x.

Peirce, C. S. 1965. Collected papers. Cambridge: Harvard University Press.

Peirce, J., 2007. PsychoPy: Psychophysics software in Python. Journal of Neuroscience Methods (162), 8–13.

Pierrehumbert, J., 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J., Hopper, P.J. (Eds.) Frequency effects and the emergence of lexical structure. Benjamins, Amsterdam [a.e.], pp. 137-157.

Pierrehumbert, J., 2016. Phonological representation: Beyond abstract versus episodic. Annual Review of Linguistics 2, 33-52.

Pitt, M.A., 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. Perception & Psychophysics 60 (6), 941–951.

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

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science (New York, N.Y.) 274 (5294), 1926–1928.

Steriade, D., 1999. Alternatives to the syllabic interpretation of consonantal phonotactics, in: Proceedings of the 1998 Linguistics and Phonetics Conference. The Karolinum Press, pp. 205–242.

Välimaa-Blum, R., 2009. The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? CogniTextes 2. 10.4000/cognitextes.211.

van Rij, J., Wieling, M., Baayen, H., van Rijn, H., 2015. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs: R package version 1.0.1.

Vennemann, Theo. 1988. Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin. Berlin: Mouton de Gruyter.

Vitevitch, M.S., Luce, P.A., 1998. When Words Compete: Levels of Processing in Perception of Spoken Words. Psychological Science 9 (4), 325-329.

Wedel, A., 2006. Exemplar models, evolution and language change. The Linguistic Review 23, 247-274.

West, B., Welch, K., Gałecki, A., 2015. Linear mixed models. CRC Press, Boca Raton, Florida.

Wieling, M., Nerbonne, J., Baayen, H., 2011. Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially. PloS one. 10.1371/journal.pone.0023613.

Wood, S., 2006. Generalized Additive Models: an introduction with R. Chapman & Hall/CRC, Boca Raton, Florida.

Wood, S., 2015. Random effects in GAMs. https://stat.ethz.ch/R-manual/R-devel/library/mgcv/html/random.effects.html

Zydorowicz, P., 2007. The acquisition of Polish morphonotactics. Wiener Linguistische Gazette Online 74, 24–44.