LANGUAGE AND COGNITIVE PROCESSES
2005, 20 (1/2), 43–73

Psychology Press
Taylor & Francis Group

Prosodic cues for morphological complexity in Dutch and English

Rachèl J. J. K. Kemps
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Lee H. Wurm
Wayne State University, Detroit, MI, USA

Mirjam Ernestus
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Robert Schreuder
Interfaculty Research Unit for Language and Speech, University of Nijmegen, The Netherlands

Harald Baayen
Max Planck Institute for Psycholinguistics, Nijmegen and Interfaculty Research Unit for Language and Speech, University of Nijmegen, The Netherlands

Previous work has shown that Dutch listeners use prosodic information in the speech signal to optimise morphological processing: Listeners are sensitive to prosodic differences between a noun stem realised in isolation and a noun stem realised as part of a plural form (in which the stem is followed by an unstressed syllable). The present study, employing a lexical decision task, provides an additional demonstration of listeners' sensitivity to prosodic cues

---

Correspondence should be addressed to R. J. J. K. Kemps, 4–32 Assiniboia Hall, Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada, T6G 2E7. Email: rkemps@ualberta.ca

Part of this work has been made possible by the support of a Major Collaborative Research Initiative (MCRI) grant of the Social Sciences and Humanities Research Council of Canada awarded to Gary Libben.

© 2005 Psychology Press Ltd

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960444000223

in the stem. This sensitivity is shown for two languages that differ in morphological productivity: Dutch and English. The degree of morphological productivity does not correlate with listeners' sensitivity to prosodic cues in the stem, but it is reflected in differential sensitivities to the word-specific log odds ratio of encountering an unshortened stem (i.e., a stem in isolation) versus encountering a shortened stem (i.e., a stem followed by a suffix consisting of one or more unstressed syllables). In addition to being sensitive to the prosodic cues themselves, listeners are also sensitive to the probabilities of occurrence of these prosodic cues.

In languages with a concatenative morphological system, such as Dutch and English, morphologically complex words consist of (combinations of) stems preceded by one or more prefixes and/or followed by one or more suffixes. The orthographic representations of morphologically complex words suggest that these stems, prefixes, and suffixes are strung together as beads on a string. Acoustically, however, the realisations of morphemes that are concatenated to form a morphologically complex word are different from the realisations of these morphemes when produced in isolation, even when the morphemes are phonemically unchanged after concatenation. One of the reasons for this is that, in stress-timed languages, the duration of a stressed vowel reduces as a function of the number of unstressed syllables that follow (Nooteboom, 1972, for Dutch; Fowler, 1977, Lehiste, 1972, for English; Lindblom & Rapp, 1973, for Swedish). In other words, the duration of the vowel in a syllable is shorter when this syllable is followed by one or more unstressed syllables than when it is produced in isolation. For example, the vowel in the first syllable of *walking* is shorter than the vowel in *walk*.

Previous studies have shown that listeners are very sensitive to such acoustic differences. It has been shown that listeners can use these differences as cues to distinguish strings that are initially phonemically ambiguous between a word and a morphologically unrelated continuation form of that word. Salverda, Dahan, and McQueen (2003) recorded participants' eye movements while they listened to Dutch sentences including a word with a morphologically unrelated onset-embedded word (e.g., *hamster* containing *ham*). The participants saw four pictures of objects on a computer screen and were instructed to use the computer mouse to move the picture of the object that was mentioned in the sentence. There were more fixations to a picture representing the embedded word (*ham*) when the first syllable of the target word (*hamster*) had been replaced by a recording of the embedded word than when it came from a different recording of the target word. This demonstrates that segmentally ambiguous sequences can contain acoustic cues (in this case, the duration of the embedded word (*ham*) relative to the duration of its corresponding syllable in the target word (*hamster*)), that modulate its lexical interpretation.

Similar results were obtained by Davis, Marslen-Wilson, and Gaskell (2002). In a gating task, participants were presented with sentence fragments. In one condition (long-word condition), the sentence fragments ended in a long carrier word of which the initial syllable formed an onset-embedded word (e.g., *captain* containing *cap*). In the other condition (short-word condition), the fragments ended in the short word corresponding to the initial syllable of the carrier word followed by a word with an onset that matched the continuation of the longer carrier word (e.g., *cap* **t***ucked* versus *cap***t***ain*). The first syllable in the short-word condition was significantly longer than the first syllable in the long-word condition, and there was a marginally significant difference in average fundamental frequency (average fundamental frequency was higher in the long-word condition than in the short-word condition). Significantly more short-word responses were made to gates from short-word stimuli than to gates from long-word stimuli, suggesting that listeners take advantage of the acoustic differences that exist between short and long word sequences. Similar results were obtained in a cross-modal priming task. The stimuli from the gating task were presented up to the offset of the first syllable of the target word (e.g., *cap* from either *cap* or *captain*) as auditory primes, and were followed by a visual target that was either the short word (*cap*) or the long word (*captain*). Greater facilitation occurred when prime syllables came from the same word as the target.

More recently, it has been shown that listeners are also sensitive to acoustic differences between phoneme strings that are initially ambiguous between a stem and a morphologically *related* continuation form of that stem, in particular, between a singular and a plural form of a noun (Kemps, Ernestus, Schreuder, & Baayen, in press). In Dutch, the regular plural form of many nouns consists of the noun stem and the plural suffix -*en*, which is usually realised as just a schwa (e.g., *boek* [buk] ‘book’—*boeken* [bukə] ‘books’). As a result of the addition of the schwa, the stem of the plural form is durationally and intonationally different from the stem realised in isolation (the singular form). In what follows, we will refer to such non-segmental differences in duration and intonation as prosodic differences. Such differences partly reflect differences in syllable structure. For instance, in the plural *boe-ken* [bukə], the suffix -*en* [ə] induces resyllabification of the stem-final obstruent ([k]) as onset of the next syllable and, as a consequence, the stem vowel is syllable-final in the plural [bukə] as opposed to syllable-medial in the singular *boek* [buk]. Listeners were presented with singular forms and with stems that were spliced out of plural forms. These stimuli were segmentally identical, but the stems of the plural forms carried mismatching prosodic information: The absence of a plural suffix pointed to the singular form, whereas the prosodic information pointed to the plural form. When presented with the mismatching forms, listeners were significantly delayed, both in a number decision task as well as in a lexical decision task. Similar results were obtained when listeners were presented with plural forms of which the stems carried either matching or mismatching prosodic information (i.e., plurals of which the stems originated either from another token of the plural form or from a realisation of the singular form), and also when listeners were presented with pseudowords of which the 'stems' carried either matching or mismatching prosodic information (i.e., pseudowords of which the stems were originally realised in isolation or in combination with a plural suffix). Importantly, the magnitude of this prosodic mismatch effect, that is, the magnitude of the delay in response latencies, correlated with the magnitude of the durational mismatch: The larger the durational difference between the stem realised in isolation and the stem realised as part of the plural form, the larger the delay. This correlation was stronger for words than for pseudowords.

The prosodic differences between uninflected forms and the stems of their corresponding inflected forms reduce the ambiguity between these forms. The observed sensitivity of listeners to these prosodic differences suggests that these acoustic cues help the perceptual system in determining early in the signal whether an inflected (bisyllabic) or an uninflected (monosyllabic) form is likely to be heard. Plurals are not singulars with an additional suffix. The precise acoustic realisation of the stem provides crucial information to the listener about the morphological context in which the stem appears.

The present study, employing a lexical decision task, aims at replicating these findings for different types of morphologically complex forms in Dutch, and at extending the investigation of listeners' sensitivity to prosodic cues for morphological complexity to another language, English. The morphologically complex forms under investigation in the present study are comparatives (inflection) and agent nouns (derivation). Studying the effects of prosodic mismatch in the processing of stems of agent nouns and of comparatives in both Dutch and English enables us to determine whether the effects observed in the processing of singular and plural forms in Dutch are specific to plural formation in Dutch, or whether they generalise to a different type of inflection, to derivation, and to a different language.

In Dutch and English, many agent nouns are formed by adding the suffix *-er* (Dutch: [əʀ]; English: [ɚ]) to the stem, which is a verb stem. For example, the English agent noun *worker* [wɝkɚ] consists of the verb stem *work* [wɝkə] and the deverbal agentive suffix *-er* [ɚ]. Similarly, the Dutch agent noun *werker* [wɛʀkəʀ] consists of the verb stem *werk* [wɛʀk] and the deverbal agentive suffix *-er* [əʀ]. The suffix *-er* is homonymous (see Booij, 1979, for the many meanings of the suffix *-er* in Dutch): Many comparatives are also formed by adding the suffix *-er* to the stem, which in this case is an adjective. Thus, the English comparative *fatter* [fætɚ] consists of the adjective *fat* [fætə] and the comparative suffix *-er* [ɚ]. The Dutch comparative *vetter* [vɛtəʀ] consists of the adjective *vet* [vɛt] and the comparative suffix *-er* [əʀ]. The affixation of the suffix *-er* leads to shortening of the preceding stem and to changes in syllable structure. In the present study, employing a lexical decision task, we investigated whether listeners are sensitive to such prosodic differences between monosyllabic stems and the stems of bisyllabic complex forms. We presented listeners with stems of agent nouns and comparatives that carried either matching or mismatching prosodic information. If listeners are sensitive to the prosodic cues in the stem, they are expected to be slowed down in their responses when there is a mismatch between the number of syllables on the one hand, and the prosodic information in the acoustic signal on the other hand. If not, in other words, if listeners attend to segmental information only, mismatching prosodic information should not affect response latencies. Note that information about the identity of the complex forms that the stems originated from was not available to our listeners. The stem *werk* (‘work’), for instance, originating from the agent noun *werker* (‘worker’) could just as well have originated from the infinitive verbal form *werken* (‘to work’). We were therefore not interested in potential effects of the type of complex form that the stems originated from, but purely in the question of whether the prosodic mismatch effect observed in earlier work would generalize to different materials, and to a different language.

Dutch and English differ in morphological richness, in particular in the number of continuation forms that are possible given a certain monomorphemic stem. For example, whereas the verbal inflectional paradigm of the Dutch word *wandelen* (‘to walk’) consists of the forms *wandel*, *wandelt*, *wandelen*, *wandelde*, *wandelden*, *gewandeld*, *wandelend*, and *wandelende*, the verbal inflectional paradigm of the English word ‘walk’ contains only *walk*, *walks*, *walked*, and *walking*. In other words, the stem *wandel* is followed by an unstressed syllable in five inflectional forms, whereas the stem *walk* is followed by an unstressed syllable in only one form. In general, the number of continuation forms in which a stem is followed by an unstressed syllable is considerably smaller in English than in Dutch: Besides the richer verbal paradigm, Dutch also exhibits prenominal contextual inflection of adjectives (which consists of the addition of a schwa to the stem, e.g., *een groot boek* ‘a big book (neuter gender)’ versus *een grote auto* ‘a big car (common gender)’), whereas English does not. Furthermore, in Dutch, most noun inflections consist of the addition of an unstressed syllable to the stem: Many plurals are formed by adding the suffix *-en* [ə(n)] to the stem. In English, on the other hand, many plurals are formed by adding the plural suffix *-s* ([s] or [z]) to the stem (no additional syllable, except for stems ending in sibilants). Finally, Dutch has more unstressed derivational suffixes than English. For example, diminutives in Dutch are formed by adding (an allomorph of) the diminutive suffix *-tje* [cə] to the stem, whereas diminutive derivation is not productive in English. It is conceivable that, as a consequence of these differences in the number of possible continuation forms in which a stem is followed by one or more unstressed syllables, Dutch and English listeners are not equally sensitive to prosodic cues in the stem that signal whether or not the stem will be followed by unstressed syllables. Possibly, English listeners are less sensitive to such prosodic cues, as, in English, a stem is relatively infrequently followed by an unstressed syllable.

We not only investigated the effect of prosodic mismatch on reaction times, but we also investigated the predictive value of two covariates that are word-specific indications of the prevalence of possible continuation forms: Syllable Ratio and Cohort Entropy.

Syllable Ratio gives a word-specific indication of the likelihood of observing an unshortened versus a shortened stem. It is defined as the log of the ratio which has as the numerator the Surface Frequency of a stem in isolation, and as the denominator the summed Surface Frequencies of words in which this stem is followed by an inflectional or derivational suffix consisting of one or more unstressed syllables (i.e., words in which the stem occurs in shortened form). We only considered inflectional and derivational suffixes that consist of one or more syllables containing schwa, so that the phonological shortening process in the stem is maximally comparable to that in the comparative stems and in the agent noun stems. For example, for the stem *strict*, the numerator of the Syllable Ratio would consist of the surface frequency of *strict* (i.e., 362), and the denominator would consist of the summed surface frequencies of *stricter*, *strictest*, and *strictness* (i.e., 69). All instances of the stem, irrespective of grammatical category, are included in the numerator of Syllable Ratio. Note that when the numerator is smaller than the denominator, the Syllable Ratio will be negative, as the log of reals between 0 and 1 is negative. Compounds were not included in the denominator, as little is known about phonological shortening within left constituents of compounds.

Syllable Ratio is the log odds ratio of observing an unshortened form versus observing a shortened form. All words occurred in monosyllabic form in the experiment. We therefore expected a facilitatory effect of Syllable Ratio: if Syllable Ratio was high for a given word (i.e., if a word occurs relatively often as a monosyllabic stem), faster response latencies were expected. A facilitatory effect of Syllable Ratio would constitute evidence for listeners' sensitivity to the likelihood of occurrence of a certain prosodic manifestation of a stem.

Syllable Ratio only considers specific types of continuation forms, namely, the continuation forms that are morphologically related to the stem and in which the stem has undergone a shortening process as a result of the addition of one or more unstressed syllables. However, given a certain stem, many types of continuation forms are possible, including continuation forms that are not morphologically related. In order to rule out the possibility that an effect of Syllable Ratio is in fact just an effect of whatever is still present in the cohort at the final position in the stem, we need an index of the latter. We therefore introduce another covariate: the Cohort Entropy. Entropy is an information-theoretical measure, indicating the amount of uncertainty about the outcome of a selection process (Shannon, 1948, see also Moscoso del Prado Martín, Kostíc, & Baayen, 2004). Cohort Entropy (H) is defined as:

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

in which *pi* is the probability of a word given the n words that are still present in the cohort at the point in time when the stem-final segment of the target word has been perceived. In other words:

$$p_i = \frac{\text{Surface Frequency of Word}_i}{\text{Summed Surface Frequencies of } n \text{ Cohort Members at stem-final segment of target word}}$$

To illustrate, suppose that by the time that the final segment of Stem X has been perceived, the cohort consists of two word candidates: Word X_a and Word X_b Word X_a has a surface frequency of 80 and Word X_b has a surface frequency of 20. For Stem Y, the stem-final cohort also consists of two word candidates (Word Y_a and Word Y_b), both of which have a surface frequency of 50. The Cohort Entropies for Stem X and Stem Y are calculated as follows (note that the Cohort Entropy is larger for Stem Y):

$$p_{X_a} = \frac{80}{80+20} = 0.80 \qquad p_{X_b} = \frac{20}{80+20} = 0.20$$

$$H_X = -(0.80^{*}log(0.80) + 0.20^{*}log(0.20)) = 0.50$$

$$p_{Y_a} = \frac{50}{50+50} = 0.50 \qquad p_{Y_b} = \frac{50}{50+50} = 0.50$$

$$H_Y = -(0.50^{*}log(0.50) + 0.50^{*}log(0.50)) = 0.69$$

Cohort Entropy is calculated at the stem-final segment as only stems (with either matching or mismatching prosodic information) were presented to our listeners. Included in the cohort are *all* possible continuation forms, that is, both morphologically related and morphologically unrelated continuation forms. For example, the cohort for the stem *bake* consists of *bake*, *bakes*, *baked*, *baking*, *baker*, *bakers*, *bakery*, *bakeries*, but also *bacon* and *bakelite*. Cohort Entropy is a non-phonologically and non-morphologically based measure, defined purely in terms of lexical competition. Note however that for monomorphemic stems (the type of stems used in the present study), morphologically related continuation forms (i.e., inflections, derivations, and compounds) are more prevalent than morphologically unrelated continuation forms, both type-wise and token-wise. (Counts are presented below.) We expect an inhibitory effect of Cohort Entropy: The more uncertainty, the longer the response latencies.

EXPERIMENT

Part A: Dutch

Method

Participants. Twenty participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch.

Materials. From the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1993) we selected all Dutch comparatives and agent nouns that contained a monomorphemic and monosyllabic stem, in which the stem ended in a voiceless plosive. In Dutch, underlyingly voiced obstruents are devoiced in syllable-final position and all stems realised in isolation therefore end in voiceless obstruents (final devoicing). The suffix *-er* [əʀ] induces resyllabification of the stem-final obstruent as onset of the next syllable, and hence an underlyingly voiced stem-final obstruent remains voiced before *-er* (e.g., Booij, 1995). As a consequence, stems ending in underlyingly voiced obstruents do not have the same segments in isolation as before *-er* (e.g., [ɦɑʀt]—[ɦɑʀdəʀ] 'hard'—'harder'). We therefore only selected agent nouns and comparatives with stems ending in an underlying voiceless plosive, so that there is no change of the voicing characteristics of the plosive when the stems occur in isolation.

Furthermore, the comparatives and agent nouns in our initial data set occurred with surface frequencies larger than zero. (Token counts in CELEX are based on a corpus of 42.4 million words of written text for Dutch, and on a corpus of 17.9 million words of written and spoken text for

English.) From this initial data set of comparatives and agent nouns, we selected those forms that could subsequently be matched to English comparatives or agent nouns that met all the above criteria, and that, in addition, carried the same onset and coda characteristics (simplex versus complex), and that carried the same vowel characteristics (long versus short). The English set of items was used in Part B of this experiment. This selection procedure resulted in a set of 35 Dutch agent nouns and 27 Dutch comparatives (see Appendix A for a list of all Dutch items). Pseudowords were created from these words by changing several phonemes in the stem, while largely respecting the status of onset and coda (simplex versus complex), the vowel length (long versus short), and the restriction that the stem-final consonant is a voiceless plosive.[1] Due to errors, one word (comparative) and one pseudoword eventually had to be removed from the design.

Separate reading lists were created for the comparatives (e.g., *vetter*), the agent nouns (e.g., *werker*), the stems of the comparatives (e.g., *vet*), the stems of the agent nouns (e.g., *werk*), and their pseudoword counterparts. The lists were recorded in a soundproof recording booth by a native male speaker of Dutch, who was naïve regarding the purpose of the experiment. Each pseudoword list was read aloud for practice once before recording. The recordings were digitised at 18.9 kHz.

The forms were spliced out of their list using the PRAAT speech editing software (Boersma & Weenink, 1996). The stems that were produced in isolation functioned as the first type of stimulus in the experiment ('normal' stems, see top panel of Figure 2 for an example). From the complex forms, a second type of stimulus was created: the 'constructed' stems. The constructed stem consisted of the stem of the complex form – in other words, it was the complex form without the suffix *-er* [əʀ]. The point of splicing was located at the onset of the voicing of the schwa following the stem-final consonant. The point of splicing was always located at a zero-crossing. Figure 1 shows an example of a complex form (top panel) and the stem spliced out of that complex form (bottom panel).

As a result of the splicing manipulation, the constructed stem's prosodic information mismatched its number of syllables: Its prosodic characteristics signalled a bisyllabic form, whereas in fact the acoustic signal contained only one syllable. In the normal stem, there was no such mismatch. Duration was measured for the two types of stems, for both words and pseudowords. As expected, the constructed stems were

---

[1] This word-pseudoword matching in our materials was not perfect: We failed to match for the status of the coda for two Dutch items, we failed to match for the status of the onset for one Dutch item, and we failed to match for the length of the vowel for one Dutch item. For one English item, we failed to match for the status of the coda.

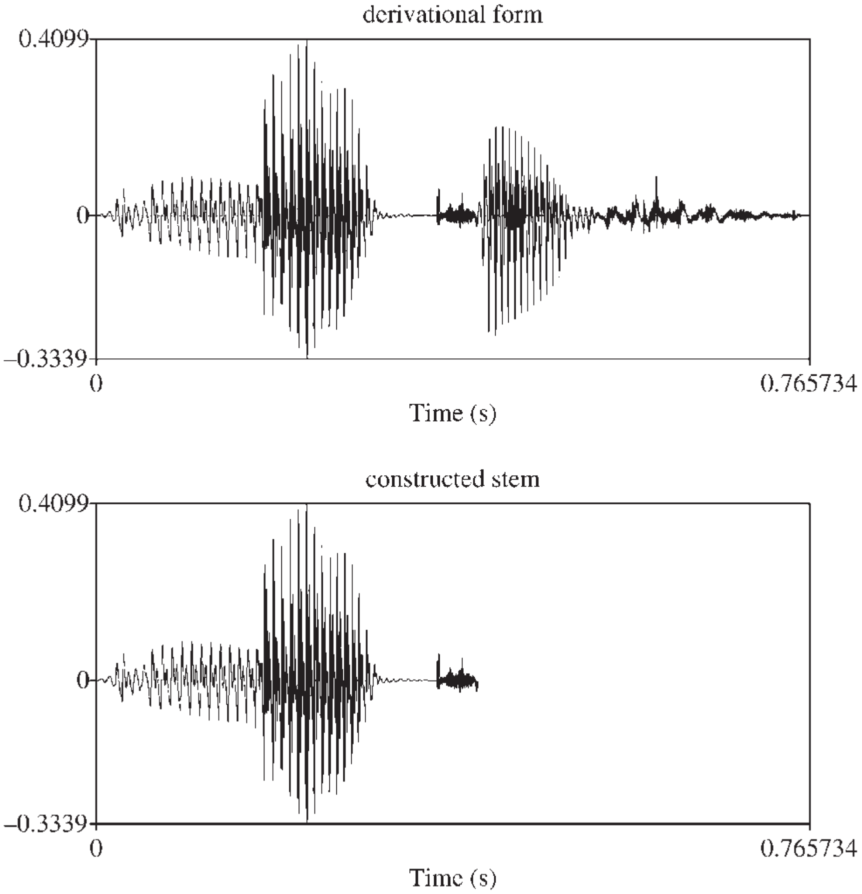

Figure 1. The complex form [nɑtəʀ] (top panel) and the constructed stem [nɑt] spliced out of the complex form (bottom panel).

significantly shorter (161 ms on average) than the normal stems, $F(1,\ 119) = 486.1$, $p < .0001$. The magnitude of this durational difference between normal and constructed stems was not significantly different for words and pseudowords (interaction of Stem Type (normal versus constructed stem) by Word Status (word versus pseudoword): $F(1,\ 119) = 1.6$, $p = .21$. For the words, we also measured the duration of the vowel, the duration of the closure of the stem-final obstruent, and the duration of the release noise of the stem-final obstruent. Analyses of variance with these durations as the dependent variable, and with Stem Type (normal versus constructed) and the Syllable Structure of the bisyllabic form (with an ambisyllabic stem-final obstruent, as in *gok-ker*, 'gambler'; with a syllable-initial stem-final obstruent and non-empty coda of the first syllable, as in *hel-per*, 'helper';

with a syllable-initial stem-final obstruent and an empty coda of the first syllable, as in *ma-ker*, 'maker') as predictors, revealed significant main effects of Stem Type and Syllable Structure for all three analyses ($p < .05$), but never an interaction of these factors ($p > .1$). Thus, the manipulation of Stem Type is independent of Syllable Structure.

The normal and constructed stems differed in prosodic structure. The normal and the constructed stems differed in yet another respect, however. The manipulation of interest (the manipulation of prosodic structure) was achieved through and therefore systematically confounded with a splicing manipulation: Splicing had occurred in the constructed stems (at the offset of the release noise of the stem-final consonant), whereas no splicing had occurred in the normal stems. We eliminated this confound by applying a splicing manipulation to the normal stems as well: We spliced away the last 25% of the release noise of the stem-final consonants (see Figure 2).

As a consequence, both stimulus types ended rather abruptly, the only difference remaining between normal and constructed stems being the difference in prosodic structure. Note that, by applying this splicing manipulation to the normal stems, we put the stimuli that we expected to be most easily processed at a disadvantage. This should make it harder for us to observe an effect of prosodic mismatch. The durational difference between the normal stems and the constructed stems after splicing away 25% of the release noises of the stem-final consonants of the normal stems was 131 ms on average, $F(1, 119) = 1391.3$, $p < .0001$. The interaction between Stem Type and Word Status remained non-significant, $F(1, 119) = 0.15$, $p = .70$. Table 1 lists the mean durations with their standard deviations for the two kinds of stems of words and pseudowords, before as well as after splicing away 25% of the release noise of the normal stems. In the following, the term 'normal stem' refers to the stem that carries matching prosodic information *and* of which 25% of the release noise of the stem-final consonant has been spliced away.

The total number of experimental trials amounted to 122 (35 agent noun stems and their matched pseudoword stems, and 26 comparative stems and

TABLE 1
Part A—Mean durations (in ms) with SD for normal stems and constructed stems in Dutch, before and after splicing away 25% of the release noise of the normal stems

	Before		*After*	
Type of stem	*Duration*	*SD*	*Duration*	*SD*
Normal word	635	91	597	91
Constructed word	465	79	465	79
Normal pseudoword	593	124	570	97
Constructed pseudoword	441	98	441	98

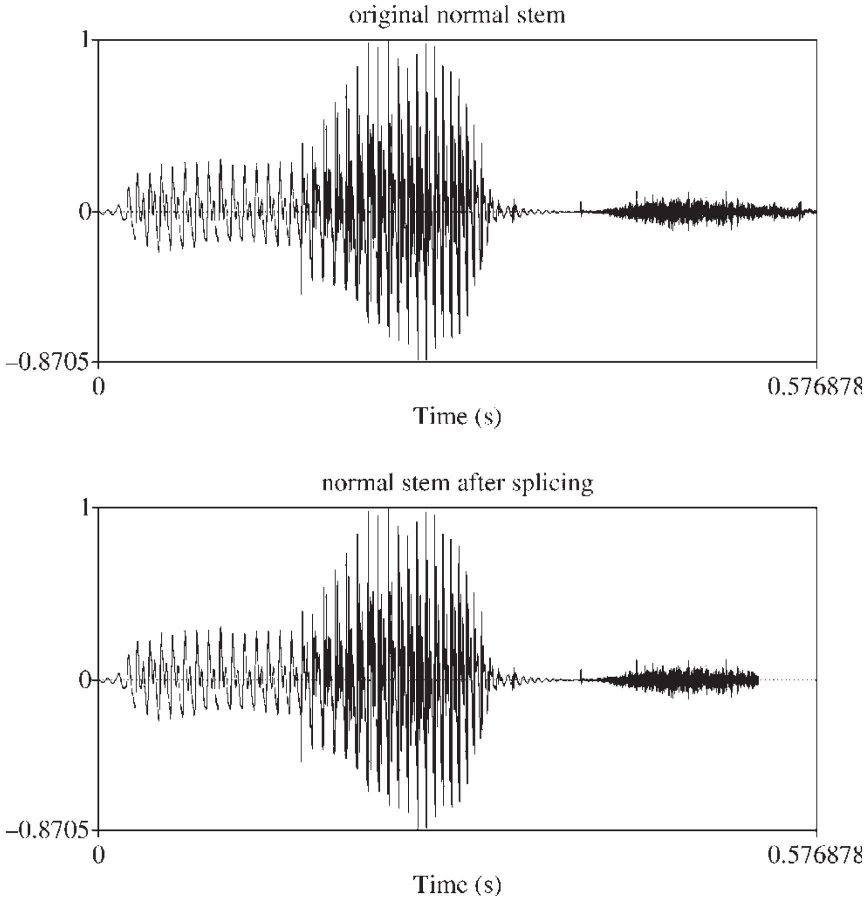

Figure 2. The original normal stem [nɑt] (top panel) and the normal stem [nɑt] after splicing away 25% of the stem-final release noise (bottom panel).

their matched pseudoword stems). So that participants would never be presented with both the normal and the constructed variant of a single stem, complementary versions of trial lists were created. If the normal form of a stem occurred in one version of a list, then the constructed form of that stem would occur in its complementary version. The composition of these lists (i.e., which items occurred in their normal stem variant and which items occurred in their constructed stem variant) was varied three times, resulting in six experimental trial lists (three 'compositions' with two complementary versions each). The order of presentation of the stimuli was pseudo-randomised within the three lists: no more than three words or pseudowords occurred successively. Orders were identical in the lists that were each other's complements. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the actual experiment. The practice set consisted of 16 trials: 4 normal pseudoword stems, 4 constructed pseudoword stems, 4 normal word stems (2 comparative stems and 2 agent noun stems), and 4 constructed word stems (2 comparative stems and 2 agent noun stems). None of the stems in the practice set was presented in the actual experiment.

Procedure. Participants performed a lexical decision task. They were instructed to decide as quickly as possible whether or not the form that they heard was an existing word of Dutch. They responded by pressing one of two buttons on a button box. Each trial consisted of the presentation of a warning tone (189 Hz) for 500 ms, followed after an interval of 200 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. Reaction times were measured from stimulus offset. Each new trial was initiated 2500 ms after offset of the previous stimulus. When a participant did not respond within 2000 ms post-offset, a time-out response was recorded. Prior to the actual experiment, the set of practice trials was presented, followed by a short pause. The total duration of the experimental session was approximately 10 minutes.

Part B: English

Method

Participants. Thirty-nine participants, students at Wayne State University, received course credit to participate in the experiment. All were native speakers of English.

Materials. The selection procedure described above for the Dutch materials resulted in a set of 35 English agent nouns and 27 English comparatives (see Appendix B for a list of all English items). Also for these words, pseudowords were created by changing several phonemes in the stem, while respecting the status of onset and coda (simplex versus complex), the length of the vowel (long versus short), and the restriction that the stem-final consonant is a voiceless plosive.

Reading lists were created in the same manner as in Part A of the experiment. The lists were recorded in a soundproof recording booth by a native male speaker of English.[2] Each pseudoword list was read aloud for practice once before recording. The recordings were digitised at 20 kHz.

---

[2] In American English, a stem-final /t/ typically becomes flapped in intervocalic position. Our speaker retained the non-flapped pronunciation in intervocalic position, which may be considered overly careful speech. Note, however, that the presence of unflapped stimuli in our experiment should work against our effect, as the unflapped /t/ in the constructed stem might be considered a strong cue for the monosyllabic form.

Normal and constructed stems were created in the same manner as in Part A of the experiment. As expected, the constructed stems were again significantly shorter (146 ms) than the normal stems, $F(1, 121) = 937.0$, $p < .0001$. The effect of Stem Type on duration was significantly larger for words than for pseudowords: Interaction of Stem Type by Word Status, $F(1, 121) = 7.3$, $p < .01$. Recall that, for Dutch, this interaction of Stem Type by Word Status was not significant, although it did show the same pattern (larger effect of Stem Type for words than for pseudowords). In the overall analysis, the interaction of Stem Type by Word Status was significant, $F(1, 141) = 6.5$, $p < .05$, and there was no significant three-way interaction of Stem Type by Word Status by Language, $F(1, 241) = 0.18$, $p = .67$. We will return to this issue below. Furthermore, the effect of Stem Type on duration was marginally smaller in English than in Dutch: interaction of Stem Type by Language, $F(1, 242) = 3.2$, $p = .07$. As for the Dutch words, we also measured the duration of the vowel, the duration of the closure of the stem-final obstruent, and the duration of the release noise of the stem-final obstruent for the English words. Analyses of variance with these durations as the dependent variable, and with Stem Type (normal versus constructed) and the Syllable Structure of the bisyllabic form as predictors, revealed only a main effect of Stem Type for the duration of the vowel ($p < .01$) and no effect of Syllable Structure nor an interaction of Syllable Structure with Stem Type ($p > .1$). None of these factors was predictive for the duration of the release noise. For the duration of the closure, Stem Type was predictive ($p < .01$), and there was an interaction of Syllable Structure with Stem Type ($p < .01$): For words such as *hel-per*, the difference in closure duration was somewhat less pronounced than for words such as *ma-ker* and *cut-ter*. Thus, the manipulation of Stem Type was independent of Syllable Structure, except for a small difference for one syllable type with respect to closure duration.

The difference in duration between normal and constructed stems remained significant after splicing away 25% of the release noise of the stem-final plosive for the normal stems (121 ms on average, $F(1, 121) = 837.7$, $p < .0001$. Table 2 lists the mean durations with their standard deviations for the two kinds of stems of words and pseudowords, before as well as after splicing away 25% of the release noise of the normal stems. The interaction of Stem Type by Word Status was now only marginally significant, $F(1, 242) = 2.9$, $p = .09$, and the three-way interaction of Stem Type, Word Status, and Language remained non-significant, $F(1, 242) = 1.4$, $p = .24$. The effect of Stem Type on duration was still marginally smaller in English than in Dutch: Interaction of Stem Type by Language, $F(1, 242) = 2.0$, $p = .09$.

Three experimental trial lists and their complements were created in the same manner as in Part A of the experiment. The total number of

TABLE 2

Part B—Mean durations (in ms) with SD for normal stems and constructed stems in English, before and after splicing away 25% of the release noise of the normal stems

	Before		*After*	
Type of stem	*Duration*	*SD*	*Duration*	*SD*
Normal word	506	101	475	97
Constructed word	347	84	347	84
Normal pseudoword	497	91	478	91
Constructed pseudoword	364	89	364	89

experimental trials amounted to 124. The practice set consisted of 16 trials: 4 normal pseudoword stems, 4 constructed pseudoword stems, 4 normal word stems (2 comparative stems and 2 agent noun stems), and 4 constructed word stems (2 comparative stems and 2 agent noun stems). None of the stems in the practice set was presented in the actual experiment.

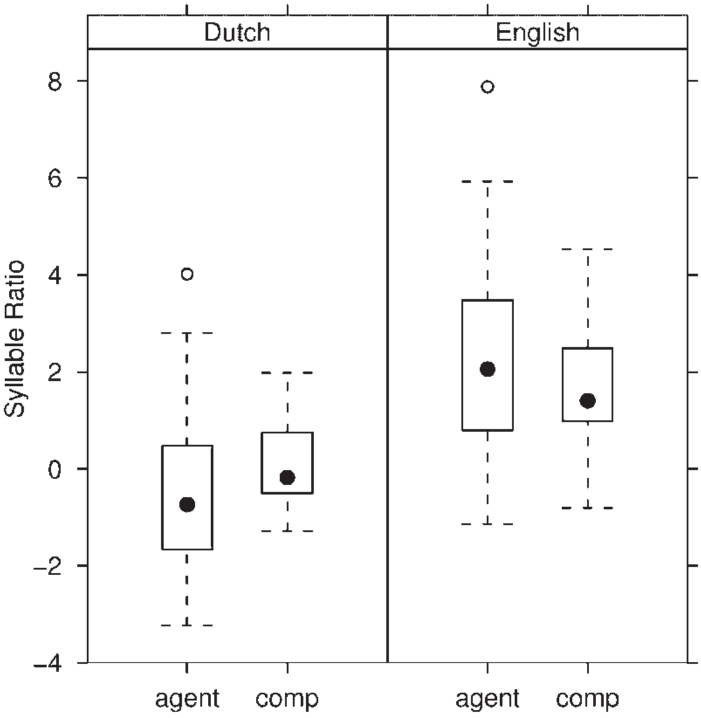

Figure 3. Syllable Ratio as a function of Word Type (stem of agent noun versus stem of comparative) and Language (Dutch versus English).

Syllable Ratio and Cohort Entropy were calculated for both the Dutch and the English words. Figures 3 and 4 summarise the distributions of Syllable Ratio and Cohort Entropy for the agent noun stems and the comparative stems in the Dutch and English part of the experiment, by means of boxplots. Each box shows the interquartile range, the filled circle in the box denotes the median, and the 'whiskers' extend to the observations within 1.5 times the interquartile range. Outliers beyond this range are represented by individual open circles.

Syllable Ratio was significantly higher for English than for Dutch, $F(1, 119) = 68.9$, $p < .0001$. This is what we expected, as there are fewer continuation forms with unstressed syllables in English than in Dutch. Word Type (agent noun versus comparative) had a stronger effect in Dutch than in English (with slightly higher Syllable Ratios for comparative stems than for agent noun stems), but this effect failed to reach significance in both languages; Dutch, $F(1, 59) = 2.1$, $p = .15$; English, $F(1, 60) = 2.2$, $p = .15$; interaction of Word Type by Language, $F(1, 119) = 10.1$, $p < .05$. Cohort Entropy was significantly lower for English than for Dutch, $F(1,$

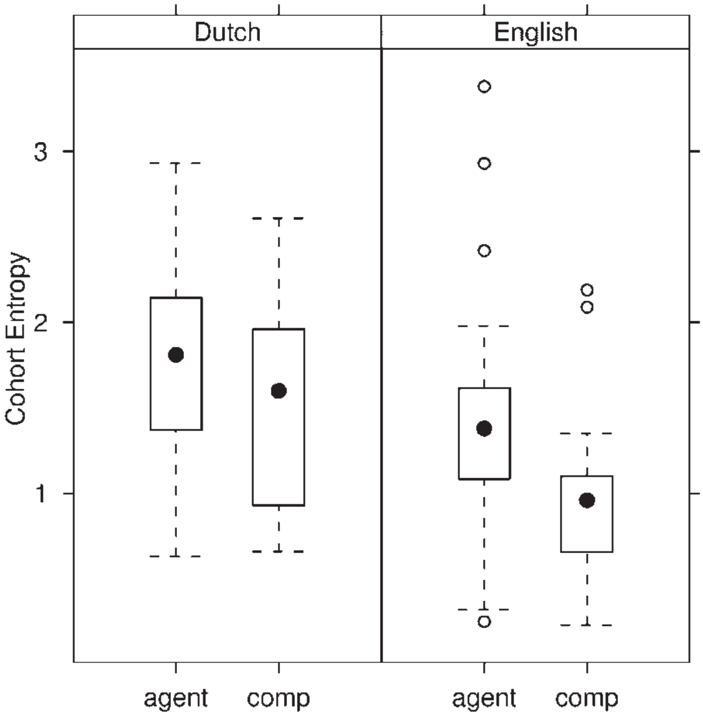

Figure 4. Cohort Entropy as a function of Word Type (stem of agent noun versus stem of comparative) and Language (Dutch versus English).

119) = 20.2, $p < .0001$. This was also expected, since there are fewer continuation forms in general in English than in Dutch. Cohort Entropy was significantly lower for comparative stems than for agent noun stems, $F(1, 119) = 11.0, p < .01$. This effect of Word Type on Cohort Entropy was similar for English and Dutch: Interaction of Word Type by Language, $F(1, 119) = 0.6, p = .42$. Furthermore, it turned out that Syllable Ratio and Cohort Entropy were correlated in English (Pearson's $r = -.24$, $p = .06$), but not in Dutch (Pearson's $r = -.14$, $p = .29$). Apparently, Cohort Entropy and Syllable Ratio consider largely the same continuation forms in English, but not in Dutch. In English, most continuation forms have unstressed syllables, whereas, in Dutch, many types of continuation forms are possible.

Procedure. Participants performed English lexical decision. The same procedure was followed as in Part A of the experiment.

Results and discussion

For Dutch (Part A), no participants were excluded from the analyses, since they all showed error rates below 20%. Appendix A lists the mean reaction times and the error rates for the Dutch words and pseudowords. Fifteen items (10 existing words and 5 pseudowords) were excluded from subsequent analyses, as they showed error rates above 20%. Of these 15 items, 6 items had high error rates in both stem variants (i.e., normal versus constructed), 6 items had high error rates in the normal variant, and 3 items had high error rates in the constructed variant. Furthermore, trials eliciting incorrect responses were excluded (3% of the trials that remained after removal of the 15 items with high error rates), as well as trials eliciting reaction times faster than 150 ms (3% of all remaining correct trials).

For English (Part B), two participants were excluded from the analyses, since they performed with error rates above 20%. Appendix B lists the mean reaction times and the error rates for the English words and pseudowords, calculated over the trials remaining after removal of the two participants with high error rates. Twenty-five items (8 existing words and 17 pseudowords) were excluded from subsequent analyses, as they showed error rates above 20%. Of these 25 items, 6 items had high error rates in both stem variants, 9 items had high error rates in the normal variant, and 10 items had high error rates in the constructed variant. Finally, trials eliciting incorrect responses (5% of the trials that remained after removal of the two participants and the 25 items with high error rates) and trials eliciting reaction times faster than 150 ms were also excluded (4% of all remaining correct trials).

TABLE 3
Mean reaction times from word offset (in ms) with SD and error percentages for normal stems and constructed stems in Dutch and English

Type of stem	*Reaction time*	*SD*	*Error*
Dutch normal word	464	230	6%
Dutch constructed word	515	218	8%
Dutch normal pseudoword	526	238	6%
Dutch constructed pseudoword	596	226	6%
English normal word	335	160	2%
English constructed word	403	184	4%
English normal pseudoword	428	200	7%
English constructed pseudoword	488	215	7%

The mean response latencies (measured from word offset and calculated over the remaining correct trials only), their standard deviations, and the error percentages for the different types of stems for English and Dutch are summarised in Table 3. In general, incorrect responses occurred more often for pseudowords than for words ($z = -6.8$, $p < .0001$), and more often for constructed stems than for normal stems ($z = -3.0$, $p < .01$). The effect of Word Status on performance interacted with Language, however ($z = 4.6$, $p < .0001$): It was significant for English ($z = -6.8$, $p < .0001$), but not for Dutch ($z = 1.1$, $p = .29$).

In the following, we will report on an overall analysis, as well as on analyses of several subsets of the data. We will start with the overall analysis of the dataset including words as well as pseudowords, for Dutch as well as for English. Next, we will report on an analysis of only the pseudoword data for Dutch and English, and on a similar analysis of only the word data for Dutch and English. Finally, we will report on separate analyses for the Dutch and the English word data. The reasons for analysing each of these different subsets of the data will be clarified as we proceed.

In an initial, overall analysis, the data for Dutch and English words and pseudowords were analysed together. We fitted a multi-level covariance model (Pinheiro & Bates, 2000) to the data, with log reaction times[3] as the dependent variable, and Stem Type (normal versus constructed stem), Word Status (word versus pseudoword), Duration (the duration of the

---

[3] Here and in the following analyses, reaction times were logarithmically transformed in order to normalise their distribution.

form that was actually presented to the participants),[4] and Language (Dutch versus English) as predictors.[5] Note that in this analysis, we used only a subset of the available predictors. Syllable Ratio was not included as a predictor as it is not possible to calculate this ratio for pseudowords. It is possible to calculate Cohort Entropy for both words and pseudowords, but because Cohort Entropy exhibited very different distributions for words and pseudowords, we did not include Cohort Entropy as a predictor in this overall analysis. We will return to this issue below.

This analysis revealed significant effects of all predictors: Constructed stems were responded to slower than normal stems (56 ms on average for Dutch and 64 ms on average for English), $F(1, 5149) = 306.9$, $p < .0001$; pseudowords were responded to slower than words (91 ms on average for Dutch and 89 ms on average for English), $F(1, 5149) = 529.1$, $p < .0001$; duration was facilitatory (the longer the word, the faster the response latencies), $t(5149) = -8.7$, $p < .0001$; and the Dutch participants were slower than the English participants (100 ms on average), $F(1, 55) = 10.8$, $p < .01$. Furthermore, there were significant interactions of Word Status by Language (the effect of Word Status was less strong in Dutch than in English), $F(1, 5149) = 4.2$, $p < .05$; and of Stem Type by Duration (Duration was more facilitatory for the constructed stems), $t(5149) = 2.9$, $p < .01$. To understand the latter interaction, consider that the longer a given constructed stem is, the more it resembles its normal stem variant. Apparently, the less abnormal a form is, the faster listeners can respond to it.

---

[4] As reaction times were measured from word offset, we expect a facilitatory effect of Duration: At word offset, the listener has been exposed to more information when the duration of the word is long than when the duration of the word is short, facilitating the response. In order to establish whether Stem Type has an effect *independently* of Duration (normal stems have longer durations than constructed stems), we included Duration as a covariate in our analyses.

[5] In our multi-level covariance models, subject variability is accounted for by using subject as a grouping factor. In the analyses of word data exclusively, item variability is accounted for by including item-specific covariates in the regression model. However, in all our analyses involving both word and pseudoword data, and in all analyses involving pseudoword data exclusively, item variability has not been accounted for, as no item-specific covariates are available for pseudowords. Therefore, in all analyses involving pseudowords, Stem Type has been treated as a between-items factor even though we would have liked to treat it as a within-items factor. Nevertheless, even without the extra power of the within-items analysis, we obtained very robust effects of Stem Type. Furthermore, an analysis on Dutch and English words and pseudowords with *item* as the grouping factor yielded largely the same pattern of results as the analysis with *subject* as the grouping factor: Stem Type, $F(1, 203) = 117.3$, $p <$ 0001; Word Status, $F(1, 203) = 130.2$, $p < .0001$; Duration, $t(203) = -5.2$, $p < .0001$; Language, $F(1, 203) = 138.4$, $p < .0001$; Stem Type by Duration, $t(203) = 2.9$, $p < .01$. The interaction of Word Status by Language was not significant in this analysis, $F(1, 5149) = 4.2$, $p = .67$.

To conclude, we have replicated the prosodic mismatch effect for stems of agent nouns and comparatives, in both Dutch and English. The prosodic mismatch effect emerged both in words and in pseudowords. Now the question remains: Do Cohort Entropy and Syllable Ratio have any predictive value? This question calls for separate analyses for words and pseudowords, for two reasons. First, Cohort Entropy (calculated at the stem-final segment) turned out to be normally distributed for Dutch and English words, but not for Dutch and English pseudowords: For the majority of pseudoword items, the cohorts were empty at the stem-final segment, and thus, the Cohort Entropy for these items was zero. For only a small number of pseudoword items (14 out of 56 Dutch pseudowords, and 9 out of 45 English pseudowords), the cohort at the stem-final segment was not empty. Second, the predictor Syllable Ratio can not be calculated for pseudowords.

We first turn to an analysis of the pseudoword data only. Because of the non-normal distribution of Cohort Entropy, we decided to treat Cohort Entropy as a factor with two levels (Entropy Zero versus Entropy Non-Zero), instead of as a covariate. In a multi-level covariance analysis, log reaction times were analysed as a linear function of Stem Type (normal versus constructed stem), Cohort Entropy (Entropy Zero versus Entropy Non-Zero), Duration, and Language (Dutch versus English). This analysis revealed significant effects of all predictors: Constructed stems were responded to slower than normal stems, $F(1, 2486) = 152.9$, $p < .0001$; Duration had a facilitatory effect, $t(2486) = -6.5$, $p < .0001$; English reaction times were faster than Dutch reaction times, $F(1, 55) = 7.0$, $p < .05$; and, importantly, items with empty cohorts (Entropy Zero) were responded to faster than items with non-empty cohorts: Entropy Non-Zero, $F(1, 2486) = 41.8$, $p < .0001$. Furthermore, there was a significant interaction of Cohort Entropy with Language: The effect of Cohort Entropy was less strong for English than for Dutch, $F(1, 2486) = 4.4$, $p < .05$. The effect of Cohort Entropy was significant in both languages, however: Dutch, $F(1, 1040) = 36.7$, $p < .0001$; English, $F(1, 1444) = 10.5$, $p < .01$.

We now turn to the word data. Log reaction times to the words were predicted by the same variables as log reaction times to the pseudowords: Stem Type (normal versus constructed stem), Duration, Cohort Entropy, and Language (Dutch versus English). In addition, Word Type (agent noun versus comparative) and Syllable Ratio were introduced as predictors. For the words (as opposed to the pseudowords), the Cohort Entropy values were normally distributed. Therefore, Cohort Entropy was now treated as a covariate (as opposed to as a factor).

A multi-level covariance analysis revealed significant effects of Stem Type: Constructed stems were responded to slower than normal stems,

$F(1, 2597) = 194.9$, $p < .0001$; Duration, facilitatory effect, $t(2597) = -5.9$, $p < .0001$; Language (English participants were faster than Dutch participants), $F(1, 55) = 11.5$, $p < .01$; and Word Type (adjectives were responded to faster than verb stems), $F(1, 2597) = 9.4$, $p < .01$. Furthermore, there was a significant inhibitory main effect of Cohort Entropy, $t(2597) = 3.1$, $p < .01$, whereas there was no significant main effect of Syllable Ratio, $t(2597) = 1.7$, $p = .08$. In addition, however, there was a significant second-order interaction of Cohort Entropy by Language, $t(2597) = -2.1$, $p < .05$, and a significant third-order interaction of Syllable Ratio by Cohort Entropy by Language, $F(2, 2597) = 9.1$, $p < .0001$. We will return to this issue below. Finally, we observed a significant interaction of Stem Type by Duration: Duration was more facilitatory for the constructed stems, $t(2597) = 3.2$, $p < .01$. This interaction had already been observed in the overall analysis described above (words and pseudowords in Dutch and English): The longer a given constructed stem is, the more it resembles its normal stem variant, and the faster listeners can respond to it.

As mentioned above, Syllable Ratio and Cohort Entropy were correlated in English (Pearson's $r = -.34$, $p < .05$), but not in Dutch (Pearson's $r = -.16$, $p = .26$).[6] This, in combination with the fact that we observed a second-order interaction of Cohort Entropy by Language, and a third-order interaction of Syllable Ratio by Cohort Entropy by Language, calls for separate analyses for the Dutch and the English word data. These separate analyses yielded the following results.

For Dutch, significant effects were again obtained for Stem Type, $F(1, 910) = 10.4$, $p < .01$; for Duration, $t(910) = -3.4$, $p < .001$; and for Word Type, $F(1, 910) = 4.6$, $p < .05$. Syllable Ratio had a significant facilitatory effect, $t(910) = -3.3$, $p < .01$, but there was no effect of Cohort Entropy, $t(910) = 0.1$, $p = .88$. Interestingly, there was a marginally significant interaction of Syllable Ratio by Stem Type, $t(910) = 1.9$, $p = .06$: The effect of Syllable Ratio was highly significant for the constructed stems, $t(458) = -3.5$, $p < .001$, but non-significant for the normal stems, $t(430) = -0.8$, $p = .40$. In other words, listeners only profited from a high Syllable Ratio when the monosyllabic form they were listening to was abnormal. This suggests that when the mapping of the acoustic signal on the representation of a stem is less effective as a result of the prosodic characteristics of the acoustic signal, the long-term probability of hearing an unshortened stem is more influential than when the bottom-up signal is unambiguous.

---

[6] The correlation coefficients reported here are calculated over the items that remained after removing the items with high error percentages, and are therefore numerically different from the correlation coefficients reported in the Materials section (which were calculated over all items that were presented to the participants).

For English, a different pattern emerged. We observed the usual effects of Stem Type, $F(1, 1686) = 159.8$, $p < .0001$; Duration, $t(1686) = -3.7$, $p < .001$; and Word Type, $F(1, 1686) = 4.7$, $p < .05$. Syllable Ratio, however, did not have a significant effect, $t(1686) = -0.8$, $p = .45$, whereas there was a significant inhibitory effect of Cohort Entropy, $t(1686) = 2.2$, $p < .05$. This effect of Cohort Entropy is an interesting finding, given the fact that the cohorts over which the Cohort Entropy values were calculated consist mainly of morphologically related continuation forms (i.e., inflections, derivations, and compounds). For English, of all 1,488 possible continuation forms (counted over all word stems), 1,280 forms (113,748 tokens) were morphologically related, and 208 forms (12,539 tokens) were morphologically unrelated. Of the morphologically related forms, 990 forms (110,234 tokens) were inflectional or derivational forms, and 290 forms (3,514 tokens) were compounds. In the cohort literature, it is generally assumed that morphological (inflectional and derivational) continuation forms should be excluded from the cohort (e.g., Marslen-Wilson, 1984; Tyler, Marslen-Wilson, Rentoul, & Hanney, 1988). Our finding shows that for a more realistic indication of the amount of competition in the mental lexicon, morphological continuation forms should be counted as cohort members. Unlike Syllable Ratio in Dutch, Cohort Entropy in English did not interact with Stem Type, $t(1685) = 0.17$, $p = .86$.

For completeness, we note that when Cohort Entropy is not included in the model, Syllable Ratio *is* predictive in English, $t(1687) = -2.0$, $p < .05$. However, when both correlated predictors Syllable Ratio and Cohort Entropy are entered into the model, only the latter is significant. In contrast, Cohort Entropy never showed an effect for Dutch, neither in a model with both Cohort Entropy and Syllable Ratio as predictors, nor in a model that included Cohort Entropy but not Syllable Ratio, $t(912) = 0.5$, $p = .61$.

To conclude, Syllable Ratio (a phonologically motivated measure) emerged as the superior predictor for Dutch reaction times, whereas Cohort Entropy (a non-phonologically motivated measure) emerged as the superior predictor for English reaction times. Apparently, in a language in which word stems are frequently followed by unstressed syllables, that is, in which stems frequently occur in shortened form, listeners develop a sensitivity for the likelihood of observing a shortened or an unshortened stem. In a language in which word stems occur relatively infrequently in shortened form, listeners are less sensitive to the likelihood of observing a shortened or an unshortened stem, but are instead sensitive to the contents of the cohort at stem-final position in general.

GENERAL DISCUSSION

In this study, we replicated the prosodic mismatch effect that was originally observed for plural inflection in Dutch (Kemps et al., in press) for another type of inflection (the formation of comparatives) and for derivation (the formation of agent nouns), in both Dutch and English. Listeners were presented with monosyllabic stems of comparatives (adjectives) and monosyllabic stems of agent nouns (verbs) that carried prosodic information that either matched or mismatched the number of syllables: The matching prosodic information pointed to a monosyllabic form, whereas the mismatching prosodic information pointed to a bisyllabic form. Lexical decision latencies were significantly slower for the items with mismatching prosodic information. This prosodic mismatch effect emerged for words as well as for pseudowords.

English is a morphologically less productive language than Dutch. As a consequence, a stem in English occurs less often in shortened form than a stem in Dutch. Nevertheless, our experiments show that Dutch and English listeners are equally sensitive to prosodic cues in the stem that signal whether or not the stem will be followed by one or more unstressed syllables. The difference in morphological richness between Dutch and English is however reflected in the predictive values of Syllable Ratio relative to Cohort Entropy. Dutch listeners are sensitive to Syllable Ratio, the log odds ratio of observing an unshortened form versus observing a shortened form: In the morphologically richer language, listeners are sensitive to the item-specific distribution of shortened and unshortened stems within the lexicon. In the morphologically poorer language, Cohort Entropy (the entropy of the distribution of cohort members at stem-final position) emerged as the superior predictor, and Syllable Ratio did not have any additional predictive value. Apparently, in a language such as English, in which stems occur relatively infrequently in shortened form, listeners are less sensitive to the item-specific distribution of shortened and unshortened stems within the lexicon. Instead, the contents of the (phonologically and morphologically non-restricted) cohort codetermine response latencies.

Our experiments also show that, in Dutch, Syllable Ratio is facilitatory for the constructed stems only. Apparently, when the mapping of the acoustic signal on the representation of a stem is less effective as a result of the prosodic characteristics of the acoustic signal, the long-term probability of hearing an unshortened stem has a larger role to play than when the bottom-up signal is unambiguous.

It might be argued that the prosodic mismatch effect arises purely due to a mismatch with syllable frame information. Consider the situation in which a listener hears the constructed form of *helper* (i.e., *help*). The prosodic cues of the stem might guide the listener to posit a syllable boundary before the stem-final plosive. Assuming that syllable frames are part of the lexical representations of *help* and *hel-per*, the inferred syllable boundary before the *p* in the constructed stem of *helper* would lead to a mismatch with the lexical representation of the stem (*hel-p* mismatches *help*). This line of reasoning predicts that a greater mismatch in syllabic structure should correspond with a greater prosodic mismatch effect. To test this prediction, we considered the three syllable structures exemplified by the words *ma-ker*, *hel-per*, and *cut-ter*. For words of the last type, the mismatch with a potential syllable frame is minimal, since the ambisyllabic stem-final plosive is both stem-final and syllable-final. Hence, the prosodic mismatch effect should be smallest for *cut-ter*, and larger for *ma-ker* and *hel-per* due to the misalignment of morphological and prosodic structure. Analyses of covariance of the response latencies in Dutch and English with Syllable Structure as an additional predictor revealed the following. In Dutch, an interaction of Syllable Structure with Stem Type emerged ($p < .05$), indicating that the words with an ambisyllabic stem-final plosive suffered most instead of least from the Stem Type manipulation, contrary to the above prediction. In English, no interaction was present ($p > .6$). We conclude that the prosodic mismatch effect cannot be reduced to a syllable frame mismatch effect.

The subsegmental durational effects documented in the present study probably arise during the mapping of the acoustic signal onto the lexicon. It is less clear at what level the effect of Syllable Ratio should be located. One possibility is to assume that it arises post-lexically. In that case, the inflected and derived words containing a given stem as the first constituent would form the sample space over which the (token-frequency based) probability for that stem of being followed by a syllable with a schwa would be estimated. This estimation, which can be conceptualised either as an on-line generalisation over stored exemplars (the inflectional and derivational types), or as an implicit generalisation represented in the weights of the connections between morphologically related lexical entries, would then take place after the mapping of the acoustic signal onto the lexical entries is completed. This is a way in which the present results might be incorporated in a model such as Shortlist (Norris, 1994).

To our mind, a post-lexical explanation of the effect of Syllable Ratio has the disadvantage that different aspects of what may well be the same morpho-phonological phenomenon are spread out over different levels of representation and processing. We view the subsegmental durational differences as providing subtle acoustic cues for the probability of particular syllable structures and of the likelihood of a following phonologically weak suffix. We interpret Syllable Ratio as a complementary frequency-based estimate of the same probabilities. Although it is technically possible to allocate the subsegmental and Syllable Ratio effects to different levels, we feel that this would lead to a generalisation being missed.

There is an alternative way in which the Syllable Ratio effect can be understood, namely, as an intrinsic part of the process mapping the acoustic input onto the lexicon. In this view, the fact that the inflectional and derivational types in the CELEX lexical database over which Syllable Ratio is calculated probably also have lexical representations would be irrelevant. What would be relevant is that the frequency with which the auditory system encounters these forms leaves its traces in the mapping of the acoustic input onto these lexical representations. This mapping operation would then be sensitive to both frequency and subsegmental duration.

This way of thinking is compatible with the results of Goldinger's study (1998) which suggest that perceptual details of speech are stored in memory and are integral to later perception. In this study, shadowers showed a tendency to spontaneously imitate the acoustic patterns (speakers' voice characteristics) of words and nonwords. Goldinger simulated these data with the strictly episodic MINERVA 2 model (Hintzman, 1986). In this model, which includes a mechanism of random forgetting necessary to avoid an exponential increase in the costs of storage and retrieval, spoken words were represented by vectors of simple elements. Each vector (i.e., each word token) contained 200 elements, of which 50 elements coded details of the speaker's voice that had produced the word. The model correctly predicted the tendency for shadowers to imitate the idiosyncratic acoustic details of speech, and it successfully predicted the response times in the shadowing task. These results strongly suggest the storage of detailed episodes in the mental lexicon. In the Goldinger study, the vector elements coded – among other things – voice characteristics. Vectors with elements coding other acoustic details, like segment durations, fit well within this approach.

Another subsymbolic, exemplar-based model that allows perceptual detail to be stored in memory, is discussed by Johnson (1997). Word-specific prosodic information was implicitly incorporated in a connectionist model. Johnson trained his model on vector-quantized speech data, which contained—among other things—information regarding the durations of the segments. This model correctly anticipated whether the incoming syllable was followed by another (unstressed) syllable or not. The connection weights in this model applied to our data would be higher between relatively *long* stem exemplars and the stem node than between relative *short* stem exemplars and the stem node. In this model, a constructed stem (with relatively short segment durations) would therefore less effectively activate the stem node than a normal stem (with relatively long segment durations). Similarly, more frequently encountered patterns would lead to enhanced performance.

The probabilistic, exemplar-based framework by Pierrehumbert (2001, 2003) offers a symbolic account of the representation of word-specific phonetic detail in the mental lexicon. In this framework, phonetic categories have probability distributions over a parametric phonetic space. These probability distributions consist of memory traces (exemplars), and are gradually built up as speech tokens are encountered and encoded. Word-forms, in turn, are viewed as sequences of phonetic categories, and also have probability distributions over temporal sequences of events in the phonetic space: Individual words have exemplar clouds associated with them. Extending this approach, we might imagine that morphologically complex forms will be associated with exemplars with relatively short stem segments, whereas isolated stems will be associated with exemplars with relatively long segments. Constructed stems are further away from the center of the distribution of stem exemplars than normal stems, and will therefore less effectively activate the representation of the stem.

To conclude, the present study provides more evidence for the role of prosodic information in morphological processing: Detailed acoustic information in the stem reveals whether the stem is realised in isolation or as part of a morphologically complex form. In a morphologically rich language like Dutch (compared to English), listeners are in addition sensitive to the likelihood within the morphological paradigm of a word of encountering a specific prosodic manifestation of that word. Although the data that we have presented in the present paper do not allow us to force a choice between different rival theoretical explanations, the most parsimonious interpretations seem to point to theories in which the mapping of the acoustic input onto the lexical representations is sensitive to both duration and probability of occurrence.

REFERENCES

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Boersma, P., & Weenink, D. J. M. (1996). PRAAT, a system for doing phonetics by computer, version 3.4. Institute of Phonetic Sciences of the University of Amsterdam, *Report 132*. Available at http://www.fon.hum.uva.nl/praat/.

Booij, G. E. (1979). Semantic regularities in word formation. *Linguistics*, *17*, 985–1001.

Booij, G. E. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 218–244.

Fowler, C. A. (1977). *Timing control in speech production*. Bloomington, IN: Indiana University Linguistics Club.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Hintzman, D. (1986). 'Schema extraction' in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics*, *50*, 101–113.

Kemps, R. J. J. K., Ernestus, M., Schreuder, R., & Baayen, R. H. (in press). Subsegmental and suprasegmental cues for morphological complexity: The case of Dutch plural nouns. *Memory and Cognition.*

Lehiste, I. (1972). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *PILUS*, *21*, 1–59.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. A tutorial overview. In H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes*, (pp. 125–150). Hove, UK: Lawrence Erlbaum Associates Ltd.

Moscoso del Prado Martín, F., Kostíc, A., & Baayen, R. H. (in press). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*, 1–18.

Nooteboom, S. G. (1972). *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*. Unpublished doctoral dissertation, University of Utrecht, Utrecht.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper, (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157. Amsterdam: John Benjamins Publishing Company.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, and S. Jannedy, (Eds.), *Probability theory in linguistics* (pp. 177–228. Cambridge, MA: MIT Press.

Salverda, A., Dahan, D., & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.

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

Tyler, L. K., Marslen-Wilson, W., Rentoul, J., & Hanney, P. (1988). Continuous and discontinuous access in spoken word-recognition: The role of derivational prefixes. *Journal of Memory and Language*, *27*, 368–381.

APPENDIX A

Dutch materials

Dutch agent noun stem and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem		Constructed Stem		Pseudoword	Cohort Entropy	Normal Stem		Constructed Stem	
			RT	Error	RT	Error			RT	Error	RT	Error
1. beuk	2.10	-0.99	513	11.11%	489	0.00%	veek	0.43	730	18.18%	662	0.00%
2. bijt	1.95	-0.73	492	27.27%	664	11.11%	peut	2.06	690	11.11%	734	9.09%
3. breek	1.41	-2.66	507	30.00%	695	0.00%	krook	1.75	507	20.00%	562	0.00%
4. denk	1.33	-0.49	437	0.00%	556	10.00%	lart	0.00	578	20.00%	751	30.00%
5. doop	2.46	0.79	692	22.22%	647	54.55%	book	0.70	468	0.00%	560	0.00%
6. dop	1.73	0.64	922	36.36%	639	0.00%	tep	0.00	694	11.11%	638	0.00%
7. drink	1.12	-1.13	578	0.00%	456	0.00%	plink	0.00	487	18.18%	579	0.00%
8. duik	2.07	-1.00	430	0.00%	480	0.00%	ponk	0.00	464	0.00%	615	0.00%
9. dweep	2.49	-2.68	696	44.44%	663	54.55%	sniep	0.00	650	0.00%	591	0.00%
10. eet	2.41	-1.34	514	0.00%	726	11.11%	oot	1.65	642	0.00%	686	0.00%
11. fluit	1.89	-0.62	332	0.00%	447	0.00%	skoot	0.00	411	0.00%	512	0.00%
12. fok	2.31	-1.27	421	0.00%	732	30.00%	guk	0.00	379	10.00%	499	0.00%
13. gok	1.81	-0.37	390	0.00%	558	0.00%	fot	0.00	696	33.33%	734	36.36%
14. haat	1.05	0.75	378	0.00%	457	0.00%	gaak	0.00	467	0.00%	608	0.00%
15. help	0.97	-2.41	418	0.00%	425	0.00%	relt	0.00	671	22.22%	665	9.09%
16. kaart	1.45	2.69	299	0.00%	418	0.00%	peelt	0.00	553	10.00%	723	0.00%
17. kijk	1.31	-0.89	398	0.00%	584	18.18%	liek	2.25	637	0.00%	624	0.00%
18. kraak	2.20	-3.23	396	0.00%	537	11.11%	plook	0.00	472	0.00%	488	0.00%
19. kweek	2.07	-1.96	505	0.00%	534	0.10%	bleet	0.00	607	11.11%	557	0.00%
20. lok	1.81	-2.01	470	20.00%	825	20.00%	taft	0.00	497	10.00%	643	20.00%
21. maak	1.24	-2.85	612	11.11%	653	27.27%	naap	1.32	613	0.00%	599	0.00%
22. melk	1.75	2.81	292	0.00%	387	0.00%	bork	0.00	474	0.00%	551	0.00%
23. muit	1.66	-1.41	772	60.00%	813	90.00%	beep	0.00	712	11.11%	563	0.00%
24. plant	1.92	0.75	374	10.00%	448	0.00%	krint	0.00	642	18.18%	645	11.11%
25. pleit	0.98	-0.44	636	22.22%	652	9.09%	kleip	0.00	504	0.00%	602	0.00%
26. pluk	1.46	-1.92	435	0.00%	665	11.11%	klek	0.00	505	0.00%	572	0.00%
27. rook	1.79	0.36	461	0.00%	416	0.00%	biek	0.73	439	0.00%	634	10.00%
28. schep	2.34	-2.49	365	10.00%	521	0.00%	spik	1.49	727	9.09%	628	11.11%
29. slaap	2.19	-0.14	413	0.00%	421	0.00%	breep	0.00	395	0.00%	590	10.00%
30. spit	2.30	0.22	518	0.00%	694	0.00%	spep	0.00	500	0.00%	486	0.00%
31. sprint	1.47	0.14	366	0.00%	431	0.00%	skrump	0.00	516	0.00%	457	0.00%
32. strip	2.04	2.56	467	0.00%	504	0.00%	strok	0.00	444	0.00%	546	0.00%
33. vent	1.13	4.02	552	11.11%	426	0.00%	benk	0.00	581	9.09%	592	0.00%
34. werk	2.93	0.60	321	0.00%	492	0.00%	birk	0.00	441	10.00%	601	0.00%
35. zet	0.63	-1.41	642	20.00%	576	10.00%	wuk	0.00	541	10.00%	637	10.00%

Dutch comparative stems and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem		Constructed Stem		Pseudoword	Cohort Entropy	Normal Stem		Constructed Stem	
			RT	Error	RT	Error			RT	Error	RT	Error
1. bleek	0.73	-0.50	552	0.00%	474	18.18%	vroot	0.00	473	11.11%	592	0.00%
2. bloot	1.77	-0.39	352	9.09%	464	11.11%	snook	0.00	460	0.00%	495	0.00%
3. dik	1.73	-1.00	436	0.00%	605	10.00%	det	0.00	695	0.00%	592	0.00%
4. flink	0.79	0.31	333	0.00%	427	0.00%	vrunt	0.00	428	0.00%	484	0.00%
5. gek	0.96	1.75	404	0.00%	476	0.00%	gip	1.12	458	10.00%	643	0.00%
6. groot	1.67	-0.80	411	9.09%	388	0.00%	glaat	0.00	535	0.00%	739	20.00%
7. heet	1.19	-0.08	475	0.00%	475	0.00%	wuut	0.00	419	0.00%	578	0.00%
8. juist	0.66	1.28	332	0.00%	346	0.00%	paagt	0.00	456	22.22%	407	54.55%
9. kort	1.67	-0.01	419	0.00%	474	9.09%	firt	0.00	414	0.00%	579	0.00%
10. laat	1.62	-1.28	429	0.00%	546	0.00%	voot	0.69	627	0.00%	768	11.11%
11. mat	2.51	-0.82	480	0.00%	578	20.00%	tup	0.00	619	0.00%	685	0.00%
12. nat	2.07	-0.26	281	0.00%	439	0.00%	plik	0.00	579	18.18%	695	0.00%
13. rank	1.58	-1.21	633	22.22%	594	9.09%	wink	2.12	764	22.22%	740	36.36%
14. rijk	2.44	0.91	416	0.00%	458	11.11%	leep	2.14	661	20.00%	584	10.00%
15. rijp	1.99	-0.41	405	0.00%	472	0.00%	tiek	1.06	562	0.00%	796	20.00%
16. scherp	1.50	-0.01	228	0.00%	429	0.00%	stimp	0.00	497	0.00%	519	0.00%
17. sterk	1.33	0.07	521	11.11%	383	0.00%	blask	0.00	410	0.00%	718	11.11%
18. stomp	2.07	-0.41	528	0.00%	487	0.00%	krunt	0.00	445	9.09%	467	0.00%
19. stout	1.80	-0.14	403	0.00%	477	0.00%	praak	0.00	501	0.00%	705	0.00%
20. strikt	0.68	0.76	527	0.00%	657	0.00%	sprent	0.00	434	0.00%	554	0.00%
21. vast	2.61	1.21	488	0.00%	394	0.00%	mork	0.00	397	0.00%	647	0.00%
22. vlak	0.83	1.21	336	0.00%	518	0.00%	blek	0.60	549	0.00%	751	10.00%
23. wit	0.93	-0.89	451	0.00%	537	0.00%	bup	0.00	482	0.00%	643	0.00%
24. zout	1.26	1.98	335	0.00%	427	0.00%	beut	0.00	508	0.00%	647	0.00%
25. zwak	1.96	-0.44	423	0.00%	445	0.00%	slek	0.00	535	10.00%	613	0.00%
26. zwart	0.89	-0.21	362	0.00%	461	0.00%	knesp	0.00	518	0.00%	633	9.09%

APPENDIX B

English materials

English agent nouns and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem		Constructed Stem		Pseudoword	Cohort Entropy	Normal Stem		Constructed Stem	
			RT	Error	RT	Error			RT	Error	RT	Error
1. look	1.28	7.88	298	0.00%	471	0.00%	nop	0.00	488	9.52%	531	25.00%
2. doubt	1.11	1.93	339	6.25%	386	28.57%	doyp	0.00	416	5.26%	419	0.00%
3. smoke	1.43	2.01	249	0.00%	296	0.00%	shneep	0.00	316	0.00%	421	4.76%
4. bank	1.38	2.24	200	0.00%	428	5.56%	gailt	0.00	487	0.00%	495	43.75%
5. reap	2.42	0.90	438	27.78%	493	10.53%	layp	0.00	449	33.33%	481	12.50%
6. hop	1.74	0.67	490	14.29%	397	6.25%	wep	0.87	541	44.44%	572	63.16%
7. drink	1.06	2.06	228	0.00%	426	15.79%	dromp	0.00	425	5.56%	433	36.84%
8. fake	1.12	2.55	288	0.00%	335	0.00%	veet	1.01	376	23.81%	430	6.25%
9. creep	1.54	0.70	315	10.53%	501	0.00%	klope	0.00	584	10.53%	505	16.67%
10. eat	1.19	3.52	281	5.26%	504	11.11%	ope	1.43	602	10.53%	624	11.11%
11. float	1.47	0.40	279	12.50%	471	33.33%	fruke	0.00	464	31.25%	549	4.76%
12. kick	1.48	2.32	321	0.00%	362	0.00%	kak	1.81	393	6.25%	513	9.52%
13. mock	1.98	3.61	370	16.67%	474	15.79%	nep	0.69	492	12.50%	471	9.52%
14. boat	0.81	4.83	331	0.00%	353	0.00%	doot	0.00	431	4.76%	534	6.25%
15. help	1.31	2.58	340	0.00%	402	0.00%	walp	0.00	460	36.84%	555	22.22%
16. sort	0.49	3.76	419	15.79%	513	5.56%	zaylt	0.00	412	0.00%	412	0.00%
17. make	1.00	4.23	409	0.00%	360	0.00%	neek	0.67	497	27.78%	468	15.79%
18. stalk	2.93	0.63	292	6.25%	403	0.00%	stip	1.67	532	4.76%	488	6.25%
19. break	1.60	3.43	239	0.00%	313	0.00%	ploot	0.76	403	0.00%	462	15.79%
20. pack	0.25	0.00	376	10.53%	609	16.67%	tep	0.24	545	16.67%	555	31.58%
21. bake	1.88	-1.13	354	0.00%	414	0.00%	keek	0.00	396	0.00%	514	0.00%
22. milk	0.55	5.93	269	0.00%	351	0.00%	malp	0.00	371	10.53%	523	16.67%
23. beat	1.67	0.24	391	0.00%	497	0.00%	toop	1.27	412	0.00%	550	0.00%
24. plant	1.20	1.33	287	0.00%	333	0.00%	krent	0.00	374	11.76%	465	10.53%
25. skate	0.94	1.75	219	0.00%	383	0.00%	spote	0.00	496	18.75%	505	0.00%
26. track	1.70	3.45	379	5.26%	523	44.44%	brip	0.00	361	6.25%	503	14.29%
27. hike	1.60	0.97	375	0.00%	523	16.67%	hewt	2.37	530	22.22%	583	63.16%
28. snap	1.56	3.44	285	5.56%	383	0.00%	smik	0.00	402	10.53%	487	11.11%
29. sleep	1.17	2.75	301	0.00%	340	12.50%	shrape	0.00	366	0.00%	420	6.25%
30. quit	0.64	3.90	366	0.00%	478	4.76%	kwop	0.00	383	6.25%	469	9.52%
31. sprint	0.32	0.25	296	4.76%	302	0.00%	strent	0.00	433	15.79%	500	11.11%
32. strip	3.38	1.94	392	0.00%	495	0.00%	splik	0.00	438	0.00%	500	26.32%
33. hunt	1.63	-0.41	299	0.00%	353	0.00%	yamp	0.00	370	0.00%	424	4.76%
34. work	1.48	1.39	267	0.00%	330	0.00%	yert	0.00	361	0.00%	453	0.00%
35. cut	1.15	3.97	345	0.00%	329	0.00%	gak	0.00	471	21.05%	488	11.11%

English comparative stems and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem		Constructed Stem		Pseudoword	Cohort Entropy	Normal Stem		Constructed Stem	
			RT	Error	RT	Error			RT	Error	RT	Error
1. bleak	0.54	2.83	441	28.57%	408	25.00%	gloop	0.00	402	5.56%	537	10.53%
2. cute	0.90	1.81	342	0.00%	512	36.84%	goip	0.00	338	12.50%	438	4.76%
3. thick	1.25	1.33	301	0.00%	394	5.56%	thep	0.00	431	5.56%	503	10.53%
4. drunk	1.13	-0.18	279	0.00%	282	0.00%	klunt	0.00	405	4.76%	528	0.00%
5. sick	2.19	1.27	304	0.00%	350	12.50%	zek	0.00	395	0.00%	375	5.56%
6. great	1.06	1.11	256	0.00%	519	4.76%	bluke	0.51	450	4.76%	625	0.00%
7. tight	2.09	0.53	475	0.00%	565	10.53%	powk	0.00	395	12.50%	470	0.00%
8. moist	0.23	1.10	242	0.00%	279	0.00%	newsk	0.00	369	11.11%	457	15.79%
9. dark	1.01	0.87	272	6.25%	386	0.00%	barp	0.00	431	12.50%	548	14.29%
10. late	0.95	-0.80	353	0.00%	446	5.56%	roke	0.00	504	33.33%	662	25.00%
11. fat	0.89	2.37	347	11.11%	572	42.11%	fik	1.73	585	26.32%	583	44.44%
12. wet	0.66	2.75	231	0.00%	400	0.00%	lup	0.00	509	15.79%	529	5.56%
13. pink	0.55	4.53	223	0.00%	319	0.00%	gont	0.00	532	14.29%	502	12.50%
14. weak	1.35	-0.28	348	0.00%	385	0.00%	yate	0.00	474	12.50%	515	14.29%
15. ripe	1.22	0.73	291	0.00%	360	0.00%	loik	0.00	432	0.00%	578	0.00%
16. sharp	1.26	1.32	231	0.00%	356	4.76%	shelk	0.00	549	5.26%	419	5.56%
17. stark	0.61	2.62	505	31.25%	619	38.10%	spelk	0.00	383	9.52%	540	0.00%
18. plump	0.65	1.78	325	4.76%	428	12.50%	krump	0.66	434	27.78%	453	52.63%
19. bright	1.07	1.30	252	0.00%	332	5.56%	kloik	0.00	304	6.25%	451	4.76%
20. strict	1.01	1.65	305	4.76%	403	12.50%	splekt	0.00	376	5.26%	362	0.00%
21. fast	0.79	0.67	175	0.00%	351	0.00%	hesk	0.00	427	5.26%	551	0.00%
22. black	0.96	3.23	241	6.25%	412	14.29%	kret	1.37	322	5.26%	501	0.00%
23. hot	0.39	3.21	242	0.00%	491	6.25%	yek	0.00	377	20.05%	583	0.00%
24. white	0.56	4.06	267	0.00%	428	5.56%	howk	0.00	461	0.00%	464	5.26%
25. quick	1.04	1.41	338	0.00%	410	5.56%	kwep	0.00	379	12.50%	508	0.00%
26. slick	0.71	2.10	384	0.00%	451	0.00%	slek	0.00	409	11.11%	487	15.79%
27. smart	0.98	1.78	193	0.00%	277	0.00%	snalk	0.68	379	6.25%	363	0.00%