Differences in perception and memory for speech fragments in complex versus simple words

Two experiments

Anne Pycha
University of Wisconsin

Two experiments investigated how people perceived and remembered fragments of spoken words that either corresponded to correct lexical entries (as in the complex word *drink-er*) or did not (as in the simple word *glitt-er*). Experiment 1 was a noise-rating task that probed perception. Participants heard stimuli such *~~drink~~er*, where strikethrough indicates noise overlaid at a controlled signal-to-noise ratio, and rated the loudness of the noise. Results showed that participants rated noise on certain pseudo-roots (e.g., *~~glitt~~er*) as louder than noise on true roots (*~~drink~~er*), indicating that they perceived them with less clarity. Experiment 2 was an eye-fixation task that probed memory. Participants heard a word such as *drink-er* while associating each fragment with a visual shape. At test, they saw the shapes again, and were asked to look at the shape associated with a particular fragment, such as *drink*. Results showed that fixations to shapes associated with pseudo-affixes (*-er* in *glitter*) were less accurate than fixations to shapes associated with true affixes (*-er* in *drinker*), which suggests that they remembered the pseudo-affixes more poorly. These findings provide evidence that the presence of correct lexical entries for roots and affixes modulates people's judgments about the speech that they hear.

Keywords: perception, memory, eye-tracking, morphology

Background and motivation

Previous research has provided evidence for dual-route models of word recognition, whereby both whole-word and decompositional mechanisms are operative (e.g., Balling & Baayen, 2012; Bergman et al., 1988; Wurm, 1997). For complex words such as *drinker*, these models offer two potential processing mechanisms:

https://doi.org/10.1075/ml.19004.pyc
The Mental Lexicon 15:2 (2020), pp. 189–222. issn 1871-1340 | e-issn 1871-1375
© John Benjamins Publishing Company

the whole-word route provides access to the lexical entry for *drinker,* while the decompositional route provides access to the lexical entries for *drink* and *-er,* which are its component morphemes. Interestingly, although simple words would seem to be logical candidates for a single, whole-word route, there is evidence that these words, too, undergo decomposition (Longtin et al., 2003; Rastle et al., 2004). In this case, however, the consequences are less clear. For a word like *corner,* decomposition provides access to incorrect lexical entries, such as *corn* and *-er*. Furthermore, for a word like *glitter,* decomposition creates speech fragments that are not lexical entries at all, such as *glitt.* As described in a large body of research, lexical entries have an influential role in people's judgments about the speech that they hear (e.g., Ganong, 1980; Gow et al., 2008; Samuel & Pitt, 2003). Given that lexical entries exert such strong modulating effects, we may ask whether people judge speech differently when it occurs in words parsed into correct entries (as in complex words) versus into incorrect entries (as in simple words). In the current study, we investigated this question by comparing listener responses to speech fragments in complex versus simple words in spoken American English, using a noise-rating task to probe perception (Experiment 1) and an eye fixation task to probe memory (Experiment 2).

In a whole-word mechanism, a word is processed in a continuous, left-to-right manner and, depending upon the model, is recognized either when other potential candidates become inconsistent with the input (Marslen-Wilson, 1984; Marslen-Wilson & Welsh, 1978; Norris & McQueen, 2008) or when prediction error for upcoming segments approaches zero (Balling & Baayen, 2012; Gagnepain et al., 2012). In a decompositional mechanism, on the other hand, a complex word is processed in a discontinuous manner, via lexical access of the root, as in the "prefix-stripping" approach of Taft and colleagues (Taft, 1979a, 1979b, 1981; Taft et al., 1986; Taft & Ardasinski, 2006; Taft & Forster, 1975).

Several studies have shown simultaneous support for both approaches, supporting the notion of a "dual-route". In a gating and lexical decision study of spoken prefixed English words, Wurm (1997) found that full-word uniqueness points and full-word frequency predicted performance, which supports a whole-word mechanism; meanwhile, prefixedness ratings, semantic transparency ratings, and prefix likelihood also affected performance, which supports a decomposition mechanism. More recently, in a lexical decision study of spoken prefixed and suffixed Danish words, Balling and Baayen (2012) examined the role of the "complex uniqueness point," which is the point at which a complex word becomes uniquely distinguishable from all words that share the same affix. Their results demonstrated that both traditional and complex uniqueness points modulated lexical decision responses, also supporting dual-route models. These authors, as well as Wurm (1997), point out that a dual-route model has concrete advantages for complex words: for example, it allows them to be associated with shades of meaning that may not be predictable from the meanings of their parts, and it also provides a useful framework in which to analyze semantic drift and neologisms.

For simple words, a whole-word mechanism would seem to be entirely sufficient, precisely because such words do not possess any internal morphological structure. But studies have reported that they nevertheless undergo decomposition, at least at some stage. In a masked priming study of printed English words, Rastle, Davis, and New (2004) found that simple words like *corner* facilitated responses to spurious constituents like *corn;* furthermore, the priming effect was equivalent to that of complex words and real constituents, such as *cleaner-clean.* The presence of a pseudo-affix, such as *-er,* appeared to be crucial, because facilitation did not occur for pairs that contained overlapping segments but no pseudo-affix, as in *brothel-broth.* Longtin, Segui, and Halle (2003) reported similar results for a masked priming study of printed French words. Note that in both of these studies, masked priming was achieved via extremely rapid visual presentation (42 or 46 milliseconds) of a visual prime; when Longtin and colleagues used the same French stimuli but with auditory primes and visual targets, they found priming only for *cleaner-clean* type pairs, which suggests that decomposition of simple words occurs relatively early in the process of recognition. Overall, then, decomposition seems to be a possibility for any word that contains a potential affix. The problem is that, while a dual-route model offers clear advantages for complex words, it does no such thing for simple words: instead, decomposition allows for incorrect parses whose role is murky at best.

Consider *drinker* and *glitter* again. As noted above, for the complex word, decomposition provides two lexical entries, *drink* and *-er.* Meanwhile, for the simple word, decomposition provides one speech fragment, *glitt,* and one lexical entry that is not a morphological constituent of the word in question, *-er.* Given the fact that the first parse is correct while the second is spurious, it is perhaps no surprise that some authors have reported advantages for complex words in recognition (Balling & Baayen, 2008; Ji et al., 2011), as well as stronger responses in MEG studies (Ettinger et al., 2014). Importantly, this asymmetry may extend beyond word recognition into more basic processes of perception and memory. This is because lexical entries have strong modulating effects on people's judgments about what they hear in the speech stream, both in perception and in memory. In the Ganong effect, for example, a sound that is ambiguous between [g] and [k] is more likely to be identified as [g] in the context *__ift,* but as [k] in the context *__iss* (Ganong, 1980). This finding, and many others in the literature, suggests that lexical entries have the capacity to alter our perception of speech (see Norris et al., 2000 for a different perspective), an idea that was recently confirmed in brain research (Gow et al., 2008). As another example, multiple studies have shown that memory for words is better than for comparable non-words (Daly et al., 2005; Hulme et al., 1991, 1995; Saint-Aubin & Poirier, 2000). Hulme and colleagues (1995), for example, showed that serial recall for items like *school* and *radio* was substantially better than for items like *bim* and *bepavit*. These findings suggest that lexical entries have the capacity to modulate our memory of what we have heard.

In the current study, we applied this logic to the lexical entries of individual roots and affixes. We hypothesized that listeners would perceive and remember speech fragments that correspond to correct lexical entries, such as [dɹɪŋk] and [ɚ], better than fragments that do not, such as [glɪɾ] and [ɚ]. In Experiment 1, participants heard spoken stimuli such as *drink~~er~~* or *glitt~~er~~* (alternatively, *~~drink~~er* or *~~glitt~~er*) where strikethrough indicates the presence of white noise at a controlled signal-to-noise ratio (SNR). The task was to assign a rating, from 1 to 5, indicating the loudness of the noise. Following previous studies (Goldinger et al., 1999; Jacoby et al., 1988), we interpreted lower ratings as an indication of increased perceptual clarity of the speech signal. The key question was whether the location of noise (e.g., on the true root *~~drink~~* vs. on fragment *~~glitt~~*, or on the true suffix -*~~er~~* vs. on the pseudo-suffix *~~-er~~*) modulated the ratings.

In Experiment 2, participants were first exposed to a spoken stimulus such as *drinker* or *glitter* while associating each morpheme or pseudo-morpheme with a different shape on the computer screen. Immediately afterwards at test, they saw the associated shapes again, and they were asked to look at the shape associated with a particular morpheme or pseudo-morpheme. As they did this, we tracked the movement of their eyes. For example, after hearing *drink* (while looking at a blue triangle) and -*er* (while looking at a red circle), they were shown a blue triangle and a red circle and asked to "Look at -*er*." The correct response would be to move the eyes to the red circle. Following previous studies (Hannula et al., 2010; Richardson & Spivey, 2000), we interpreted greater accuracy and speed of eye fixations as an indication of stronger memory for the associated speech signal. The key question was whether the type of target (e.g., the true root *drink* vs. the fragment *glitt, or* the true suffix -*er* vs. the pseudo-suffix -*er,*) modulated accuracy and/or time-to-fixation.

To preview, results from Experiment 1 revealed that listeners perceive certain true roots from complex words with more clarity than the corresponding pseudo-roots from simple words. Experiment 2 revealed that listeners remember true affixes from complex words with greater accuracy than pseudo-affixes from simple words. Both findings are consistent with our hypothesis.

Experiment 1 on perception: Noise-rating task

Experiment 1 was designed to probe the perception of individual spoken fragments within complex words, where the fragments always corresponded to a morpheme, versus simple words, where they did not. Participants heard stimuli such as *~~in~~land*, *drink~~er~~*, *~~in~~dex*, and *glitt~~er~~*, where strikethrough indicates the presence of white noise at a SNR of +24, +17, or +10 dB (also included were stimuli such as *in~~land~~*, *~~drink~~er*, *in~~dex~~*, and *~~glitt~~er*). Their task was to assign a rating, from 1 to 5, indicating the loudness of the noise. Noise, of course, interferes with perception of the speech signal, and the logic of the task is that loudness ratings probe the extent to which the listener experiences this interference. Some studies using this task have focused on whole sentences or words, and have demonstrated a significant effect of prior exposure (Goldinger et al., 1999; Jacoby et al., 1988). That is, listeners gave lower loudness ratings to sentences or words that they had heard previously, compared to those that they had not. This is a perceptual illusion: the loudness of noise on old versus new stimuli was objectively the same, because SNRs were controlled across these conditions, but it appeared subjectively different to the participants, because they mis-attributed the ease with which they could perceive the old stimuli to a difference in noise loudness. In the current study, we were interested in whether the presence of a root or affix lexical entry also creates a perceptual illusion.

Other studies using this task have focused on morphological constituents, and demonstrated a significant effect of noise location. Specifically, in both American English and Spanish stimuli, listeners gave lower ratings for noise located on prefixed roots (e.g., *in~~land~~*) than on suffixed roots (*~~drink~~er*), as well as lower ratings for noise located on suffixes (*drink~~er~~*) than on prefixes (*~~in~~land*) (Pycha, 2015a, 2015b). These findings suggest that, like prior exposure, morphological constituency is also a cognitive variable that modulates perception.

Note that these previous studies contained some gaps that the current investigation sought to address. First, the previous studies focused only on complex words, whereas in current investigation, we were interested to compare complex versus simple words. Second, the previous studies used naturally-produced spoken stimuli. Thus, although SNRs were controlled across conditions, this noise was overlaid onto stimuli that contained naturally-occurring intensity and duration differences. As a result, we cannot be completely sure whether the results originate from morphological structure or, alternatively, from acoustic characteristics of the recorded stimuli. For example, if suffixes generally have lower amplitude than prefixes (e.g., -*er* < *in*-), then the noise overlaid on suffixes at a constant SNR would also have lower amplitude (-*~~er~~* < *~~in~~*-). As a result, participant judgments could conceivably be based on the objective criteria of loudness, when our goal is actually to probe the subjective criteria of morphological structure. A similar argument applies to duration, because longer stretches of noise may be perceived as louder. An important aspect of the current investigation, then, is that we implemented the noise-rating task with normalized stimuli. Thus, the speech fragments *in*, *land*, *drink*, *er*, *dex*, *glitt* all had the same amplitude and duration.

Method

As shown in Table 1, the stimuli were designed according to a 2x2 design with factors of morphological structure (complex vs. simple) and affix type (prefixed vs. suffixed).

Table 1. Schematic design of stimuli for Experiment 1

Structure	Prefixed	Suffixed
Complex	*inland*	*drinker*
Simple	*index*	*glitter*

The following sections describe the procedures we followed for selecting complex and simple words, and, since the same words were used in both experiments, they apply to both Experiments 1 and 2. For simplicity, we use the terms "root", "affix", "prefix", and "suffix" to refer to either true morphological constituents or to their pseudo-counterparts, distinguishing between them when necessary.

Overview of stimulus selection

A large number of variables affect word recognition, which is typically investigated either by analyzing reaction times to a lexical decision task or by evaluating brain responses to spoken words at a particular time point. Although our work obviously draws upon this literature, the current study is different because it investigated listener judgments of speech fragments embedded in different word types, and not the nature of the word recognition process or its time course. Furthermore, unlike previous studies, we focused on spurious, form-based decomposition of simple words, and the consequences for perception and memory. Our stimulus development process reflects these differences.

As noted in the introduction, decomposition of simple words only occurs when a pseudo-affix is present (Longtin et al., 2003; Rastle et al., 2004). For example, *corner* primes *corn* but *brothel* does not prime *broth*. The difference between these pairs is that the speech fragment [ɚ] is identical to that found in true affixes such as *-er* in *cleaner,* while the same is not true of the fragment [l̩]. As Rastle and colleagues note, this type of decomposition appears to disregard etymology and semantics, and operates according to form alone.

Because pseudo-affixes drive decomposition in simple words, we wanted to make comparisons between simple and complex words in which decomposition was triggered by the same speech fragment. Therefore, we selected words such that the surface form of pseudo-affixes for simple words always contained an exact match in the surface form of true affixes in complex words. In the suffixed condition, for example, [ɚ] occurs in our stimulus words *glitter* and *drinker,* while [n̩] occurs in both *linen* and *broaden.* In the prefixed condition, [bi] occurs in both *behave* and *befriend,* while [ɛn] occurs in both *enhance* and *enlarge.* This pairwise comparison was important because, in order to evaluate people's judgments of speech fragments that don't correspond to lexical entries, such as [glɪɾ] and [ɚ] in *glitter,* we needed a baseline. This baseline was provided by people's judgments of speech fragments that do correspond to lexical entries, such as [dɹɪŋk] and [ɚ] in *drinker.*

In addition to the basic the requirement to match the surface realizations of pseudo-affixes and affixes, our stimulus selection was based on (a) three variables that have been argued to modulate decomposition (type parsing ratio, phonotactic probability, and frequency ratio), and (b) balancing across prefixed and suffixed conditions.

Stimulus selection process

We started stimulus construction by selecting seventeen derivational prefixes and seventeen derivational suffixes that were matched for type parsing ratio (Hay & Baayen, 2002). This ratio indicates, for a given affix, the relative number of words with that affix which fall above the affix's parsing line (and therefore have a higher probability of being decomposed) versus the number that fall below it (and therefore have a lower probability of being decomposed). As closely as possible, we matched each prefix with a suffix, such that the mean type parsing ratio was 0.65 for prefixes (ranging from 0.43 to 0.93) and 0.65 for suffixes (ranging from 0.42 to 0.92). The selection process was limited to affixes that were monosyllabic.

We matched across prefix and suffix conditions because previous studies have reported differences between prefixed and suffixed words in word recognition (Colé et al., 1989; Feldman & Larabee, 2001) as well as in reading (Beauvillain, 1996). Such differences have also been reported for perception of speech fragments (Pycha, 2015a, 2015b), suggesting that the linear position of an affix can modulate people's judgments about what they hear.

For the complex condition, for each of the thirty-four true affixes, we used the CLEARPOND database (Marian et al., 2012) to search for complex words that contained the affix plus a root. In order that all of our target stimuli would be disyllabic, the search was confined to roots that were monosyllabic. This process yielded 916 complex candidates in total, although the candidates were not distributed equally across affixes. For some affixes, the number of candidates was very small: for example, there were only ten candidates for *mid-,* eight candidates each for *-fold, -most,* and *-ward,* seven for *sub-,* six for *-ette,* five for *-dom,* and three each for *trans-* and *-ize.* This placed constraints on final word selection.

For the simple condition, again for each of the thirty-four true affixes, we used CLEARPOND to search for simple words that contained a matching pseudo-affix plus a pseudo-root. The search was confined to monosyllabic pseudo-roots. To facilitate comparison with complex stimuli, we always matched stress patterns across complex and simple conditions. For example, *ˈin.land* and *ˈin.dex* both have stress on the first syllable, as do *ˈscream.er* and *ˈglitt.er.* Simple words whose stress pattern did not match a complex stimulus (e.g., simple *ˈdis.tal* vs. complex *dis.ˈband*) were excluded. Although pseudo-affixed words are not uncommon in English (Schreuder & Baayen, 1994), the requirements for monosyllabic pseudo-roots plus matching stress narrowed the pool of candidate words considerably, yielding 364 simple candidates. Again, for some pseudo-affixes, the number of candidates was particularly small: there were only ten candidates each for *en-* and *fore-,* seven for *-ette,* five each for *be-, im-, sub-, un-,* and *-ize,* three each for *trans-, -dom, -ward,* two each for *de-, mid-, out-, -ness,* and one each for *fore-, re-, -less, -ship.* Indeed, for a few pseudo-affixes, the number of candidates was zero, because the relevant speech fragment simply does not occur in monomorphemic English words. For each pseudo-prefix that did not occur, we eliminated the pseudo-suffix with the corresponding type parsing ratio, and vice versa, resulting in a total of eleven pseudo-prefixes and eleven pseudo-suffixes in the pool.

After we identified these initial pools of candidate words, we made our final selection of roots according to two criteria: the probability of the phonotactic transition that they created across the root-affix boundary, and, for complex words, the ratio that they created between the root frequency and whole-word frequency.

We controlled for transition probability because researchers have claimed that this factor modulates decomposition in complex words, with lower values generally facilitating decomposition (Hay, 2000, 2002). Given a phoneme occurring at a particular position in a word, this value indicates the probability that another specific phoneme will follow. For example, the probability of the [n-l] transition in *inland* is 0.0006, while the probability of the [m-ɚ] transition in *screamer* is 0.0005 (data are from the Phonotactic Probability Calculator, Vitevitch & Luce, 2004). From our pools of complex and simple candidate words, we selected stimuli such that transition probability was balanced across prefixed and suffixed conditions.

We controlled for frequency ratio because, again, it has been argued to modulate decomposition in complex words (Hay, 2000; Hay & Baayen, 2002). Frequency ratio is the frequency of the standalone root divided by the frequency of the derived word; values above 1 generally facilitate decomposition. For example, the frequency of *land* is 88.11 occurrences per million words, while the frequency of *inland* is 1.35 (frequency data are from CLEARPOND, Marian et al., 2012). After transforming each of these values by adding 1 and taking the base-10 log, we calculate a frequency ratio of 5.24 for *inland*. Similarly, the frequency of *scream* is 26.42, while the frequency of *screamer* is 1.02, and we calculate a frequency ratio of 4.71. Frequency ratio is only applicable to complex words; from our pools of complex candidates, we selected stimuli such that this ratio was balanced across prefixed and suffixed conditions. For simple words, we selected stimuli such that bare frequency was balanced across prefixed and suffixed conditions.

The literature on word recognition has identified a long list of additional variables that affect the process of word recognition, including part-of-speech, semantic transparency, judged prefixedness, length, and bigram frequency, to name a few. It seems feasible to think that some of these variables might also affect the behaviors under investigation in the current study, namely perceptual clarity and memory for speech fragments. As described above, however, our pools of candidate words were small – indeed, for many of the affixes, exceptionally small – which prevented us from including any of these additional variables. In interpreting our findings, this is a caveat to bear in mind.

For complex words, stimulus statistics are displayed in Table 2 and the full list is available in Appendix A. For simple words, stimulus statistics are displayed in Table 3 and the full list is available in Appendix B. As noted above, for simple words, certain pseudo-affixes only produced one or two real words, instead of our target of three. Therefore, four pseudo-affixed words were repeated across two lists, and four were repeated across all three lists. The number of repetitions was identical across pseudo-prefixed and pseudo-suffixed conditions.

Table 2. Stimulus characteristics for Experiments 1 and 2, complex words

Characteristic	Prefixed words	Suffixed words
n	51	51
Frequency ratio	16.82 (18.33)	16.85 (17.51)
Phonotactic probability	0.0013 (0.0021)	0.0011 (0.0011)

Table 3. Stimulus characteristics for Experiments 1 and 2, simple words

Characteristic	Pseudo-prefixed	Pseudo-suffixed
n	26	26
Frequency	0.715 (0.53)	0.727 (0.52)
Phonotactic probability	0.0027 (0.0028)	0.0020 (0.0020)

Note that although we aimed to keep phonotactic probability as low as possible, it was not possible to obtain a closer match for this value across complex and simple words. Indeed, the difference in our stimuli reflects a larger pattern found throughout the lexicon, namely that such probabilities are higher when they occur within morpheme boundaries rather than across them (e.g., Zipf, 1935). In addition, we could not match frequency ratio across complex and simple words, because this value, while important for complex words, does not apply to simple words; in these cases, its value is always 0, because the pseudo-root does not have a standalone frequency. And we could not match frequency across complex and simple words, because the small size of the candidate pools made this impossible. For these reasons, phonotactic probability, frequency ratio, and frequency were included as variables in our statistical models.

Finally, a reviewer pointed out that our simple stimulus set contains words with different types of pseudo-root, e.g., *sect* in *insect* happens to correspond to an actual root while *glitt* in *glitter* does not. We address this point in the statistical analyses of our results for Experiments 1 and 2. A reviewer also argued that our use of *self-* as a prefix and *-proof* as a suffix was not justified, and that these morphemes are better analyzed as elements in compounds. The reviewer also pointed out that the word *inward,* which was originally included in our list of complex words, does not consist of a prefix plus stem. As a post-hoc step to address these concerns, our final statistical analyses excluded the word *inward,* plus all words with *self-* and *-proof.* (Note that these two affixes have nearly identical type parsing ratios, 0.76 and 0.80 respectively, so no further adjustment to the stimulus list was necessary). In all cases, the results of this final statistical analyses were nearly indistinguishable from an analysis that included all of our original stimuli.

Fillers

We selected eighty-five filler words, which included thirty-nine words with two syllables and forty-six with four syllables. Of the two-syllable words, seven were prefixed and seven were suffixed, using affixes not included in the complex stimuli, e.g., *bypass* and *action.* In addition, seven of these words contained matching a pseudo-prefix and seven contained a matching pseudo-suffix, e.g., *bias* and *lotion.* And, eleven were simple words with no internal structure, e.g., *hazel.* Of the four-syllable words, eleven were affixed (e.g. *superhuman* and *communism*), and thirty-five were simple (e.g., *harmonica*).

Recording and segmentation

Each stimulus word was placed into one of four carrier sentences: *She said ____ to me, I said __ to you, He said __ to them, We said ___ to her.* The order of the sentences was randomized and five buffer sentences, which were later discarded, were placed at the beginning of the list. A female phonetically-trained speaker of American English, who was not aware of the purpose of the experiment, read each sentence aloud in a careful but fluent pronunciation. She wore a high-quality, unidirectional head-mounted microphone connected to a pre-amp. The recording was digitized at a sampling rate of 44.1 kHz.

Each target word was then segmented into two portions, which were saved into an individual WAV file using the Praat program (Boersma & Weenink, 2018). For complex words, the portions corresponded to root and affix, e.g., *inland* was segmented into [ɪn] and [lænd], and *drinker* was segmented into [dɹɪŋk] and [ɚ]. For simple words, the portions corresponded to pseudo-affix and the remaining portion of the word; e.g., *index* was segmented into [ɪn] and [dɛks], and *glitter* was segmented into [glɪɾ] and [ɚ]. For fillers that did not contain any affixes or pseudo-affixes, the division into parts was made according to syllable boundaries; e.g., *hazel* was segmented into [heʲ] and [zl̩] and *harmonica* was segmented into [hɑɹmɑ] and [nɪkə].

The acoustic measurements for these naturally-produced stimuli, before normalization, are listed in Appendix C.

Normalization and stimulus creation

We used Praat to normalize the stimuli (Boersma & Weenink, 2018). A script opened each sound file that corresponded to a word portion. First, it used the "Scale intensity" function to scale its average intensity to 70 dB. Next, following procedures outlined in section 8.2 of the Praat instructions, it created a "Manipulation object" and added two targets to its "Duration tier". One target was specified as the point 0,500/(dur*1000) and another target as the point dur,500/(dur*1000), where dur is the duration of the word portion in seconds. We then replaced the sound's old duration tier with the new one. For word portions whose initial duration was less than 500ms, then, the result was a lengthened sound. For word portions whose initial duration was greater than 500ms, the result was a shortened sound.

The resulting stimuli sounded somewhat distorted, which was not unexpected. To ensure that listeners could still perceive the intended word, we played the stimuli to three naïve listeners, and asked them to write down what they heard. They were able to do so with no errors.

To create the final stimuli, we used the Akustyk program (Plitcha, 2012) to add white noise to either the first portion (*~~in~~land*) or the second portion (*in~~land~~*) of a normalized word at a SNR of +24, +17, or +10 dB. Thus for each word, we created a total of six separate stimuli.

List construction

Three lists of 141 words each were constructed. Each list contained thirty-four complex words (seventeen prefixed and seventeen suffixed), twenty-two simple words (eleven pseudo-prefixed and eleven pseudo-suffixed, with eight words repeated across lists as described earlier), and eighty-five fillers (the same fillers for all three lists). Affixes were not repeated across lists; e.g., if a list contained *drinker,* it did not contain *bowler* or *screamer.* The same was true for pseudo-affixes; e.g., if a list contained *glitter,* it did not contain *caper* or *boulder.*

Procedure

Each participant was randomly assigned to a list and the order of the stimuli was randomized for each participant. On each trial, listeners saw a blank screen and heard a stimulus that had been cross-spliced into one of the carrier phrases, such as *I said ~~in~~land to you*. After a 500 millisecond (ms) delay, a screen with buttons labelled from 1 to 5 appeared, and printed instructions asked the participant to "Rate the loudness of the noise." The word "softest" appeared underneath the 1 button, and the word "loudest" appeared underneath the 5 button. Participants used the computer mouse to click one of the buttons, at which point the trial ended and, after a short delay, the next trial began.

The experimenter verbally provided each participant with brief instructions, which were reiterated with printed instructions on the computer screen. Participants were informed that some of the speech sounded distorted, and that they should simply do their best to understand it. To ensure that the participants attended to speech signal, and not just the noise itself, we informed participants to listen carefully to each word, because the experimenter would ask about them later. Each procedure began with five practice trials.

In order to avoid confounding effects of prior exposure, each participant heard a given word from their list only one time (thus completing 141 trials). For each word, then, the experiment randomly selected a location for the noise (e.g., either *~~in~~land* or *in~~land~~*) and randomly selected a SNR of +24, +17, or +10 dB. This procedure created an effective, although not strictly perfect, balance of the factors of noise location and SNR within each participant and also across the data set as a whole.

Each participant was seated in an individual carrel in a quiet laboratory, in front of a computer workstation. Auditory stimuli were delivered through Sennheiser HD 280 headphones. The experiment was delivered with E-Prime 2.0 software from Psychology Software Tools, Inc.

Participants

Participants were forty-seven adult native speakers of American English, 30 females and 17 males, ranging in age from 18 to 40, with a mean age of 22.02 (4.47). In return for their participation, they received extra credit in a linguistics course or small cash compensation.

Results

In the analyses that follow, we divided our data into a root set and an affix set. This facilitated our crucial comparisons between roots and pseudo-roots (e.g., *~~drink~~er* vs. *~~glitt~~er*) and between affixes and pseudo-affixes (e.g., *drink~~er~~* vs. *glitt~~er~~*).

Preliminary break-down by noise level

In order to ensure that participants performed the task as intended, we first examined participants' responses to the three different SNRs. Note that our experiment used positive SNRs (indicated with '+'). As such, +24 db SNR indicates noise that was the softest relative to the signal, and +10 db SNR indicates noise that was the loudest relative to the signal. Descriptive statistics are displayed in Table 4.

Table 4. Participant ratings (means, standard deviations) for loudness of noise in Experiment 1

SNR	Roots	Affixes
+24 db	1.83 (0.84)	1.76 (0.83)
+17 db	2.67 (0.84)	2.65 (0.84)
+10 db	3.93 (0.99)	4.01 (0.94)

To analyze the results, we used mixed-effects linear regression models as implemented by the lme() function in the nlme package of R, with one model for roots and one model for affixes. The outcome variable was the integer noise rating. The predictor variables were noise level (+24 dB vs. +17 dB vs. +10 dB SNR, sum coding), morphological structure (complex vs. simple, sum coding) and affix type (prefixed vs. suffixed, sum coding). We included participant, frequency ratio, and probability of phonotactic transition as random intercepts. Results for roots showed main effects of noise level (level 1: $\beta = 1.13$, std. error = 0.03, $DF = 35$, $t = 33.30$, $p = 0.00$, level 2: $\beta = -0.13$, std. error = 0.03, $DF = 35$, $t = -4.16$, $p = 0.00$), but no interaction with any other factor. Similarly, results for affixes showed main effects of noise level (level 1: $\beta = 1.18$, std. error = 0.03, $DF = 21$, $t = 34.73$, $p = 0.00$, level 2: $\beta = -0.13$, std. error = 0.03, $DF = 21$, $t = -3.85$, $p = 0.00$), but no interaction with any other factor. These results demonstrate that participants performed the basic task as intended, assigning the highest ratings to noise when it was objectively loudest, and lowest ratings when it was objectively softest. Since noise level exhibited the predicted main effect but did not interact with any of our variables of interest, subsequent analyses pooled the three noise levels together.

Main analysis

Descriptive results for Experiment 1, pooled across noise levels, are displayed in Table 5.

Table 5. Pooled participant ratings (means, standard deviations) for loudness of noise in Experiment 1

	Roots		**Affixes**	
Structure	**Prefixed roots**	**Suffixed roots**	**Prefixes**	**Suffixes**
Complex	2.78 (1.21)	2.70 (1.20)	2.90 (1.21)	2.73 (1.30)
Simple	2.69 (1.22)	2.95 (1.28)	2.85 (1.24)	2.91 (1.35)

To analyze the results, we used again mixed-effects linear regression models, with one model for roots and one model for affixes. The outcome variable was the integer noise rating. The predictor variables were morphological structure (complex vs. simple, sum coding) and affix type (prefixed vs. suffixed, sum coding). We included participant, frequency ratio, and probability of phonotactic transition as random intercepts. Table 6 displays the results.

Table 6. Results of linear regression analyses in Experiment 1, for roots (left) and affixes (right)

	Roots						**Affixes**				
Factor	**β**	**Std. Err.**	**DF**	***t***	***p***		**β**	**Std. Err.**	**DF**	***t***	***p***
Complexity	−0.03	0.03	356	−0.98	0.33		−0.04	0.04	358	−0.94	0.35
Affix type	−0.05	0.03	43	−1.39	0.17		0.03	0.04	29	0.78	0.44
Complexity* Affix type	0.09	0.03	43	2.54	0.01	*	0.05	0.04	29	1.36	0.18

Post-hoc testing revealed a significant difference between prefixed and suffixed roots in the simple condition ($\beta = -0.27$, $SE = 0.10$, $df = 43$, $t = -2.60$, $p = 0.01$), and a significant difference between suffixed roots in the complex versus simple conditions ($\beta = -0.24$, $SE = 0.09$, $df = 43$, $t = -2.53$, $p = 0.02$). These differences are depicted in Figure 1. No other pairwise comparisons were significant.

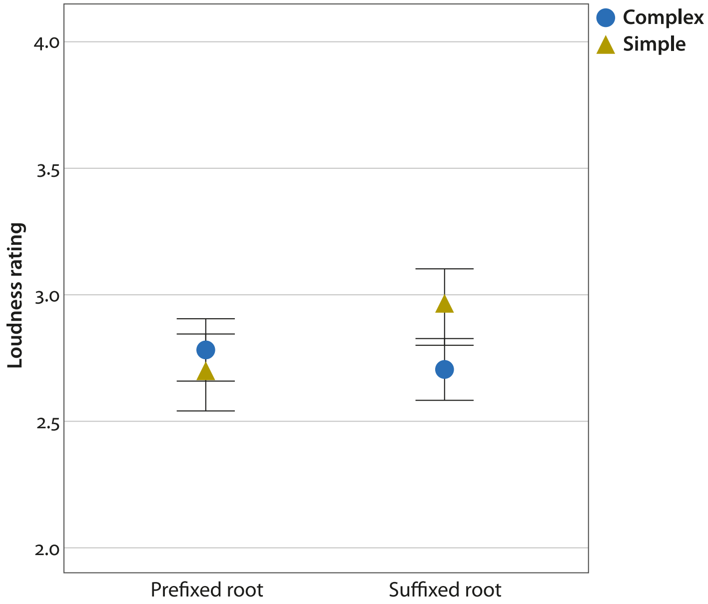


Figure 1. Loudness ratings for noise overlaid on roots in Experiment 1

Post-hoc power analyses using the `pwr.f2.test()` function from the pwr package in R yielded power values over 0.80 for small, medium, and large effect sizes, which suggests that Experiment 1 was sufficiently powered.

As mentioned earlier, a reviewer had pointed out that our simple stimuli actually contained three different types of words, which can be characterized according to the pseudo-root produced by decomposition: (a) items such as *insect* and *caper,* in which the pseudo-root is an actual root, albeit an incorrect one (e.g., *sect* and *cape*), (b) items such as *index* and *glitter,* in which the pseudo-root is a non-word fragment (e.g., *dex* and *glitt*), and (c) items such as *ingrate,* where the pseudo-root is an actual bound root (cf. *grateful*).

To investigate the effect of these three types, we performed a post-hoc coding of our simple stimuli, which produced nineteen words of type (a) (ten pseudo-prefixed and nine pseudo-suffixed), twenty-nine words of type (b) (thirteen pseudo-prefixed and sixteen pseudo-suffixed), four words of type (c) (three pseudo-prefixed and one pseudo-suffixed). As can be seen from the counts, the pseudo-prefixed versus pseudo-suffixed words were relatively evenly distributed across both types (a) and (b), less so for (c). The mean ratings are displayed in Table 7.

Table 7. Participant ratings (means, standard deviations) for loudness of noise in Experiment 1

	Roots		Affixes	
Structure	**Prefixed roots**	**Suffixed roots**	**Prefixes**	**Suffixes**
Complex	2.78 (1.20)	2.70 (1.20)	2.90 (1.21)	2.73 (1.30)
Simple-a	2.65 (1.27)	2.91 (1.23)	2.72 (1.26)	2.86 (1.36)
Simple-b	2.75 (1.18)	2.97 (1.30)	2.89 (1.25)	2.95 (1.34)
Simple-c	2.47 (1.22)	2.80 (1.64)	3.15 (1.09)	2.50 (1.52)

We ran regression analyses that were identical to those described above, except that we used four levels for the predictor variable of morphological structure: complex, simple-a, simple-b, and simple-c. Results for noise on roots revealed a significant interaction between simple-b and affix type ($\beta = -0.16$, $SE = 0.08$, $df = 43$, $t = -2.03$, $p = 0.04$). Post-hoc revealed no significant pairwise comparisons. No other results were significant. Results for noise on affixes revealed no significant effects.

Summary of Experiment 1

The primary finding from Experiment 1 is that, in the suffixed condition, participants gave noise on pseudo-roots in simple words (e.g., *~~glitt~~er*) significantly greater loudness ratings than noise on true roots in complex words (*~~drink~~er*), which we interpret to mean that they perceived the psuedo-roots with less perceptual clarity. As discussed in the Introduction, pseudo-roots in simple words do not correspond to a correct lexical entry, while true roots in complex words do. Therefore, this finding provides evidence to support our hypothesis that lexical entries – here, lexical entries for individual morphemes – modulate people's judgments about the speech that they hear.

Tentatively, we can state that this modulating effect seems to arise specifically from pseudo-morphemes that do not correspond to any lexical entry (e.g., *glitt* in *glitter*), rather than from those that do not correspond to a correct entry. Two aspects of the Experiment 1 results support this notion. First, our analysis according to pseudo-root type showed a significant effect only for words of type (b), where the pseudo-root did not correspond to any actual root (*glitt* in *glitter*). Meanwhile, no significant effect was found for words of type (a), where the pseudo-root did correspond to an actual root, albeit an incorrect one (*cape* in *caper*), or for words of type (c), where the pseudo-root is arguably a true bound root (*grate* in *ingrate*). Second, the significant effect was found only for suffixed roots, and not for suffixes themselves. Loudness ratings for noise on pseudo-suffixes in simple words (*glitt~~er~~*) did not differ significantly from those for noise on true suffixes in complex words (*drink~~er~~*). In both of these cases, *-er* coresponds to a lexical entry, even if that entry is not correct for simple words. Thus, it seems that lexical entries might facilitate perception, even when they are incorrect. This conclusion is tentative, however, because our breakdown of pseudo-root types was post-hoc, and also because the lack of effect for affixes constitutes a null result.

In the prefixed condition, there was no comparable effect. Unlike the findings for the suffixed condition, loudness ratings for noise on pseudo-roots in simple words (*in~~dex~~*) were not significantly different from ratings for noise on true roots in complex words (*in~~land~~*). It is not entirely clear why the modulating effect of lexical entries should occur only in the suffixed condition. However, as noted earlier, previous studies with the noise-rating task have reported that listeners gave lower ratings for noise located on prefixed roots (e.g., *in~~land~~*) than on suffixed roots (*~~drink~~er*), as well as lower ratings for noise located on suffixes (*drink~~er~~*) than on prefixes (*~~in~~land*), in both American English and Spanish (Pycha, 2015a, 2015b). This pattern suggests that speech fragments might enjoy a perceptual advantage whenever they are in non-initial position. If true, this advantage could conceivably compensate for the lack of correct lexical entries in fragments such as *dex* (from *index*). Further research would be required to pursue this line of thinking.

Experiment 2 on memory: Fixation task

In Experiment 2, we probed the strength of memory for speech fragments in spoken words. After a listener has heard a complex word such as *inland*, for example, how strong is the memory for *in-*, and how strong is the memory for *land*? And, after the listener has heard a simple word such as *index*, how strong is the memory for *in,* and how strong is the memory for *dex*? In the case of complex words, the speech fragments correspond to correct lexical entries, and we therefore hypothesize that memory will be relatively strong. In the case of simple words, the speech fragments correspond either to incorrect entries or to non-entries, and we therefore hypothesize that memory will be relatively weak.

To test this hypothesis, we used a fixation task with eye-tracking. We asked participants to listen to words such as *inland,* and then to react to a portion of that word, such as *in-,* by moving their eyes. The design of the experiment relied upon two basic principles already established by previous work. The first principle concerns the behavior of the human eye. Several recent experiments have demonstrated that memory for previously-experienced events can drive eye movements (Hannula et al., 2010; Ryan et al., 2007), even when those "events" consist primarily of spoken words rather than visual images (Altmann, 2004; Richardson & Spivey, 2000). For example, Richardson and Spivey (2000;

Experiment 2) presented participants with short spoken passages (e.g., "Shakespeare's first plays were historical dramas; his last was The Tempest"), each of which occurred simultaneously with a large asterisk displayed in one of four quadrants of a computer screen. At test, they asked for a true/false judgment related to the passage ("Shakespeare's last play was The Tempest"). Results showed that, while listening to the test question, participants were twice as likely to fixate on the quadrant where the associated asterisk had appeared at study, compared to the other quadrants. Following other authors (Altmann, 2004; Ferreira et al., 2008), we interpret these findings to mean that the spoken stimuli and spatial location of the asterisk were both encoded as part of the same event. When, at test, a new speech input re-activated the memory, it also activated the spatial location of the asterisk, which drove the eyes to that location.

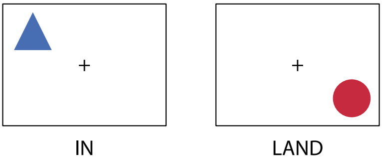
The second principle concerns reaction time. Research has repeatedly shown that the stronger the memory for a particular stimulus, the more quickly people react to that stimulus (e.g., Baddeley, 1998; e.g., Benjamin, 2001; Finnigan, 2002; Goldinger, 1998; Opitz, 2010 and many others). Combining these two principles, the basic idea of Experiment 2 is that memories for spoken fragments such as *in-* and *land* can drive eye movements, and that stronger memory will produce faster eye movements than weaker ones.

Method

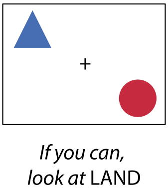
As with many memory experiments, Experiment 2 used a study-test paradigm. In our case, the study and test phases were encapsulated into a single trial, such that the the study phase for each word was followed immediately by the test phase for that same word. During the study phase of a trial, sketched in Figure 2a, participants heard a spoken word, and each portion of the word occurred simultaneously with the appearance of a visual shape in one quadrant of the computer screen. For example, when participants heard *inland,* the fragment *in* might occur simultaneously with a blue triangle in the upper left of the screen, and the fragment *land* might occur simultaneously with a red circle in the lower right of the screen. Thus, participants studied "events" that included both a spoken and a visual element. In this example, the two events are [spoken *in* + visual blue triangle] and [spoken *land* + visual red circle]. Each study phase contained three repetitions repetitions of the same word.

The study phase for a particular word was immediately followed by the test for that same word.

During the test phase, sketched in Figure 2b, the two same shapes appeared on the screen, in the same quadrants, and participants heard a prompt that contained one portion of the studied word, such as *If you can, look at land.* The



a. Study



b. Test

Figure 2. Sketch of study and test phases for a single trial in Experiment 2

task was to look at the shape associated with that portion of the word. In this example, the correct response would be to look at the red circle in the lower right quadrant. Using eye-tracking, we measured participants' accuracy (e.g., whether they looked at the red circle or not) and time-to-first-fixation (e.g, the amount of time it took them to move their eyes to the red circle).

Stimuli

For spoken stimuli, the complex words, simple words, and fillers were identical to those used in Experiment 1. The recording and normalization procedures were also identical to those used in Experiment 1.

Visual stimuli consisted of four graphics files in JPEG format, showing a red circle, a blue circle, a red triangle, and a blue triangle. As depicted in Figure 2 (not to scale), the diameter of the circle and the height of the triangle were equivalent.

Procedure

As in Experiment 1, the auditory stimuli were divided into three lists of 72 items, each of which contained 17 prefixed words, 17 suffixed words, 11 pseudo-prefixed words, 11 pseudo-suffixed words, and 16 fillers. Also as in Experiment 1, affixes were not repeated across lists; e.g., if a list contained *drinker,* it did not contain *bowler* or *screamer.* The same was true for pseudo-affixes; e.g., if a list contained *glitter,* it did not contain *caper* or *boulder.*

Each participant was randomly assigned to a list, and therefore completed 72 trials, each of which consisted of three study repetitions of a word followed by immediately by the test for that same word. The visual stimuli were divided into two sets: red circle and blue triangle, or blue circle and red triangle, and each participant was randomly assigned to a set. In addition, each participant was randomly assigned to an association pattern: circle shape with first spoken morpheme and triangle shape with second spoken morpheme, or triangle shape with first spoken morpheme and circle shape with second spoken morpheme.

Each participant was seated in front of an Eyelink 1000 Plus subject station, which consisted of a monitor with a 1920 x 1080 pixel resolution, infrared light, high-speed video camera, keyboard, mouse, and Sennheiser HD 280 headphones. To stabilize the head, participants placed their forehead and chin onto a headrest, and then adjusted the height of their chair to a comfortable position. The height of the headrest was not adjusted and remained constant across participants, such that the eyes were always approximately 98 cm from the center of the computer monitor and 61.5 cm from the camera lens. The experimenter then used the eye-tracker to obtain an image of the participant's right eye. Occasionally, if a problem occurred in tracking the right eye, the left eye was tracked instead. The experimenter used default threshold settings to obtain estimates of the pupil and corneal reflection, and performed twelve-point calibration and validation procedures. The eye-tracker was set to record at a rate of 1000 Hertz.

As mentioned earlier, each trial consisted of a study phase followed immediately a test phase. The study phase consisted of three repetitions of each word, such as *inland*, and its associated shapes. Pilot testing suggested that three repetitions were necessary in order for participants to successfully associate a spoken stimulus with a visual shape. With only one or two repetitions of the study word, participants exhibited low accuracy at test, often looking at the wrong shape or at no shape.

At the beginning of each repetition of the study phase, a fixation cross appeared in the center of the screen for 50 ms. Next, the first and second fragments of each word (e.g., *in* followed by *land*) were played in consecutive order, with no delay between the offset of the first portion and the onset of the second portion. At the onset of the first auditory fragment of the word, a visual shape appeared in one quadrant of the screen; at the offset of this fragment (i.e., 500 ms later), the shape disappeared. For example, a blue triangle might have appeared in the upper left quadrant at the onset of *in*, and disappeared at the offset *in*. Then, at the onset of the second auditory fragment of the word, a different visual shape appeared in a different quadrant of the screen; at the offset of this fragment (i.e., 500 ms later), the shape disappeared. For example, a red circle might have appeared in the lower right quadrant at the onset of *land*, and disappeared at the offset of *land*. Each word was quasi-randomly assigned to one of twelve possible quadrant locations (1–2, 1–3, 1–4, 2–1, 2–3, 2–4, 3–1, 3–2, 3–4, 4–1, 4–2, 4–3), with the constraint that for each participant, each of the twelve combinations occurred an equal number of times throughout the course of the experiment. The quadrant locations stayed the same across all three repetitions of a particular word on a given trial. An interval of 50 ms separated the end of one repetition, and the beginning of the next.

At the end of the study phase for a word, the fixation cross disappeared, and the entire screen displayed a single color (orange, green, yellow, or purple) for 2000 ms before proceeding to the test phase for that same word.

At the test phase of a trial, the fixation cross appeared in the center of the screen, and participants heard a spoken carrier phrase followed by a target portion of the word, e.g., *If you can, look at land.* At the offset of the target, the triangle and circle shapes appeared, with the same colors and in the same quadrants where they had appeared at study. For example, at the offset of *land*, a blue triangle appeared in the upper left quadrant and, simultaneously, a red circle appeared in the lower right quadrant. The shapes appeared at the offset of the spoken target, and not the onset, because we wanted to avoid potential confounding effects of uniqueness point. That is, if one morpheme could be uniquely identified after hearing just one or two of its phonemes, participants would look to its associated shape relatively quickly, but if another morpheme could only be uniquely identified after hearing all of its phonemes, participants would look to the associated shape relatively slowly. To mitigate the effect of such differences on participants' eye movements, we withheld the display of shapes until offset of the spoken target. As soon as the shapes appeared, the participants' task was to move their eyes to the associated shape. They had 3000 ms to respond, at which point the trial ended and the next trial began.

On each of the three trial lists, half of the studied words were followed by a test that prompted for a root target (*If you can, look at land*), while half were followed by a test that prompted for an affix target (*If you can, look at in*). Across lists, words with a particular affix (such as *-er* for *drinker, bowler, screamer*) were followed approximately half of the time by a root target, and approximately half of the time by an affix target. In addition, on each list, all sixteen fillers were followed by a test phase that prompted for a morpheme which did not occur at study. For example, after learning word-shape associations for a filler word like *by-pass,* participants would hear a test phase such as "*If you can, look at tion*". On any such trial where there was no target morpheme to look at, we instructed participants to keep looking at the fixation cross.

The experimenter provided verbal instructions to each participant before beginning, and there were also brief on-screen instructions. Participants were informed that some of the speech sounded distorted, and that they should do their best to understand it. The experiment began with eight practice trials, during which the experimenter verified that participants were looking at a shape target when prompted to do so at test. After practice, the participant was given the opportunity to ask questions, and then proceeded to the main experiment. Each participant took two brief breaks, primarily to close and rest their eyes, after the 24th item and again after the 48th item.

Participants

Participants were twenty-two college-aged native speakers of the midwestern variety of North American English. None of them reported difficulties with hearing, speech, or language. In return for their participation, they received extra credit in a linguistics course or small cash compensation.

Results

As in Experiment 2, we divided our data into a root set and an affix set, which facilitated our crucial comparisons between roots and pseudo-roots (e.g., target *drink* vs. target *glitt*) and between affixes and pseudo-affixes (e.g., target *er* from *drinker* vs. target *er* from *glitter*).

Accuracy results are displayed in Table 8. A trial was counted as "correct" if the participant's first eye fixation to an interest area was to the associated shape. For example, after hearing *If you can, look at land*, a correct response would be to fixate the eyes on the red circle in the lower right quadrant. An incorrect response would be to fixate the eyes anywhere else on the screen.

Table 8. Accuracy rates, calculated as the proportion of correct trials over all trials (mean, standard deviation) for Experiment 2

	Roots		**Affixes**	
Structure	**Prefixed roots**	**Suffixed roots**	**Prefixes**	**Suffixes**
Complex	0.93 (0.26)	0.93 (0.25)	0.95 (0.21)	0.97 (0.16)
Simple	0.92 (0.28)	0.95 (0.22)	0.94 (0.24)	0.91 (0.29)

To analyze accuracy, we used two mixed-effects logistic regression models using the glmer function from the lme4 package in R, one for root targets and one for affix targets. We used sum coding for the factors complexity (complex vs. simple) and affix type (prefix vs. suffix). We included random intercepts for participant and phonotactic transition probability; models that also included a random intercepts for frequency ratio and/or frequency, failed to converge. For trials with a root target, there were no significant results. For trials with an affix target, results showed a significant main effect of complexity ($\beta = 0.40$, std. error $= 0.20$, $z = 2.02$, $p = 0.04$). This effect is displayed in Figure 3.

Time-to-first-fixation results, which were calculated for correct responses only, are displayed in Table 9.

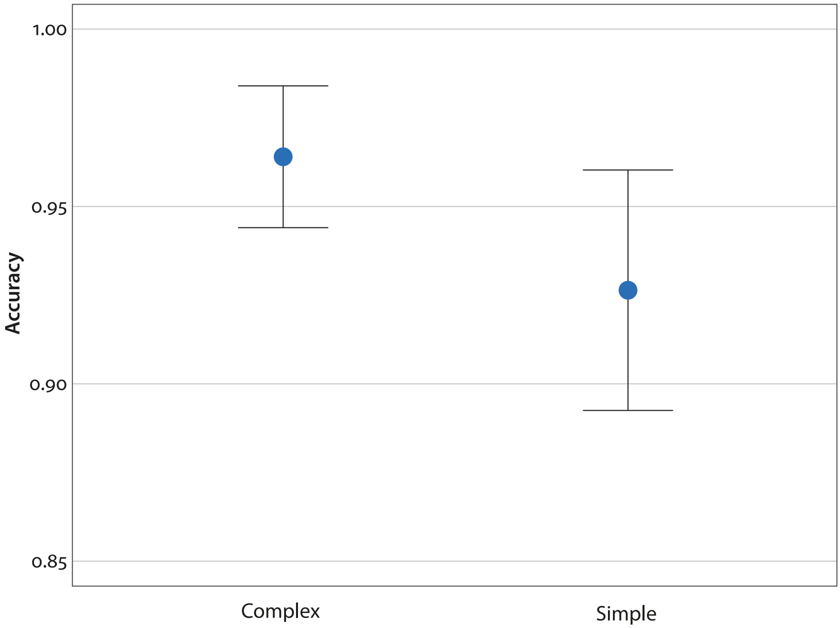


Figure 3. Main effect of morphological structure on accuracy Experiment 2, affix targets

Table 9. Time-to-first-fixation in milliseconds (means, standard deviations) for Experiment 2

	Roots		Affixes	
Structure	**Prefixed roots**	**Suffixed roots**	**Prefixes**	**Suffixes**
Complex	247.15 (281.99)	279.37 (387.94)	249.39 (311.11)	245.58 (349.21)
Simple	226.15 (275.31)	321.31 (433.91)	211.31 (245.26)	183.56 (181.79)

To analyze times-to-first-fixation, we used two mixed-effects linear regression models using lme, one for root targets and one for affix targets. We used sum coding for the factors complexity (complex vs. simple) and affix type (prefix vs. suffix). We included random intercepts for participant, phonotactic transition probability, frequency ratio, and frequency. The results of these analyses are displayed in Table 10. For both analyses, there were no significant results.

A post-hoc power analysis using `pwr.f2.test()` function from the pwr package in R yielded power values over 0.80 for both medium and large effect sizes. For a small effect size, it yielded a power value of 0.79. Given that 0.80 is often used as a cut-off value, it is possible that Experiment 2 was somewhat underpowered to detect small effect sizes.

Table 10. Results of linear mixed-effects models on times-to-first-fixation in Experiment 2, for roots (left) and affixes (right)

	Roots					Affixes				
Factor	β	**Std. Err.**	**DF**	t	p	β	**Std. Err.**	**DF**	t	p
Affix Type	−25.57	12.77	8	−2.00	0.08	7.01	9.40	105	0.75	0.46
Complexity	−7.59	12.77	410	−0.59	0.55	13.60	14.18	402	0.96	0.33
Affix Type *Complexity	12.27	12.77	8	0.96	0.36	−6.40	9.41	105	−0.68	0.50

As with Experiment 1, we also performed a post-hoc analysis after coding our simple stimuli for pseduo-root type (a) (e.g., *insect, caper*), type (b) (e.g, *index, glitter*), or type (c) (e.g., *ingrate*). For accuracy on trials with a root target, there were no significant results. For accuracy on trials with an affix target, results showed a significant main effect of morphological structure for complex words compared to all of the other levels of simple words ($\beta = 0.71$, std. error $= 0.33$, $z = 2.18$, $p = 0.03$). For times-to-first-fixation on trials with a root target or an affix target, there were no significant results.

Discussion of Experiment 2

The primary finding from Experiment 2 is that participants' eye fixations to shapes associated with pseudo-affixes in simple words (e.g., *-er* in *glitter*, *in-* in *index*) were significantly less accurate than their fixations to shapes associated with true affixes in complex words (e.g., *-er* in *drinker*, *in-* in *inland*), which suggests that they remembered the pseudo-affixes more poorly. As noted, pseudo-affixes in simple words do not correspond to a correct lexical entry, while true affixes in complex words do. Therefore, this finding provides evidence to support our hypothesis that lexical entries for individual morphemes modulate people's judgments about the speech that they hear.

It is not entirely clear why our results show a difference in accuracy, but not in time-to-fixation. Generally, the mean times-to-fixation in Table 9 are not very different from 200 ms, which is the approximate time required for a human to launch an eye fixation in response to a speech stimulus (Dahan et al., 2001; Fischer, 1987). It may be the case that our statistical analysis did not detect very small differences in times-to-fixation: as noted, a power analysis suggested that our experiment was sufficiently powered to detect medium and large effects, although it was on the borderline for detecting small effects. Or it may be the case that times-to-fixations simply are not modulated by the presence or absence of lexical entries.

The significant effect for accuracy on affixes did not interact with word type, meaning that it was statistically equivalent across the prefixed and suffixed conditions. This result differs from what we saw in Experiment 1, where there was an effect only in the suffixed condition, but is consistent with our hypothesis, which proposes that lexical entries should exert modulating effects, without regard for their linear order within the word.

For the roots, there was no comparable effect. Unlike the findings for affixes, participants' eye fixations to shapes associated with pseudo-roots in simple words (e.g., *glitt* in *glitter*, *dex* in *index*) were not significantly less accurate than their fixations to shapes associated with true roots in complex words (e.g., *drink* in *drinker*, *land* in *inland*). It is not clear why this asymmetry occurs. One possibility, suggested by a reviewer, is that the study phase functions as a learning situation, such that three repetitions of a word – specifically, a simple word that has been artificially decomposed into two fragments – suffice for participants to begin treating fragments like *glitt* and *dex* as if they were actual roots. While this scenario would be consistent with other studies demonstrating rapid learning of novel morphemes (Gagnepain et al., 2012; Kapnoula et al., 2015), it does not make clear why the participants nevertheless treated pseudo-affixes differently from true affixes.

Alternatively, it is possible that the asymmetry is somehow due to pseudo-affixes' special status. Recall that decomposition of simple words only occurs when a pseudo-affix is present, and only under certain circumstances: with extremely rapid, masked visual presentation of the prime, *corner* primes *corn,* but *brothel* does not prime *broth;* with simple auditory priming, however, neither effect occurs (Longtin et al., 2003; Rastle et al., 2004). Thus, in the very early stages of recognition, a speech fragment such as [ɚ] functions as if it corresponds to a lexical entry, while in later stages, it does not. Possibly, the higher rates of inaccuracy for pseudo-affixes reflect this confusion. Since listeners do not treat speech fragments such as [glɪɾ] or [dɛks] as lexical entries at any stage, no such confusion occurs. Further work would obviously be needed to pinpoint what this "confusion" actually entails.

A final possibility is that memory for pseudo-roots was boosted by their relative infrequency. In the literature on recognition memory, many experiments have shown that studied, low-frequency words have significantly higher hit rates than studied, high-frequency words. That is, if you have studied both *abdication* and *word,* you are more likely to remember *abdication* (Glanzer & Adams, 1985; Glanzer & Bowles, 1976). Applying this logic to the current study, since fragments such as [glɪɾ] or [dɛks] result only infrequently from decomposition (i.e., only in cases of the specific words *glitter* or *index*), listeners are more likely to remember them, an advantage that may compensate for the fact that they do not correspond to lexical entries. By contrast, fragments such as [ɚ] and [ɪn] result relatively frequently from decomposition (i.e., in any complex word derived with these affixes, and also in some simple words), and therefore enjoy no such advantage. To pursue this idea further, we would need to construct experiments that explicitly manipulated frequencies of different pseudo-roots and affixes.

Discussion

We pursued the hypothesis that people perceive and remember speech fragments that correspond to correct lexical entries, such as [dɹɪŋk] and [ɚ] in the complex word *drinker,* better than those that do not, such as [glɪɾ] and [ɚ] in the simple word *glitter.* Our results show some support for this hypothesis. In Experiment 1, participants perceived speech fragments that corresponded to certain pseudo-roots with less clarity than those that corresponded true roots. And in Experiment 2, participants remembered fragments that corresponded to pseudo-affixes less accurately than those that corresponded to true affixes. Consistent with previous work, these findings show that lexical entries – in particular, those for roots and affixes – have facilitative effects on perception and memory, and suggest that the spurious decomposition of simple words could have consquences for people's judgments of speech.

Across Experiments 1 and 2, the results are broadly consistent with our hypothesis, but nevertheless exhibit asymmetries that raise questions. In Experiment 1, true roots facilitated perception, but only in suffixed position, and only when compared to "bad" roots such as *glitt* in *glitter.* Meanwhile, true affixes offered no such advantage. In Experiment 2, the pattern was different. True affixes facilitated memory, but true roots offered no such advantage. Previous research has shown that whole-word lexical entries exert consistent effects on both perception and memory (Daly et al., 2005; Ganong, 1980; Gow et al., 2008; Hulme et al., 1991, 1995; Saint-Aubin & Poirier, 2000; Samuel & Pitt, 2003), and so further research is required to understand why root entries versus affix entries diverge in their effects on these two processes. At the very least, we can say that the effect of a lexical entry seems to depend upon the nature of the entry itself.

Although this study did not focus on the process of word recognition per se, our findings have potential implications for it. Consider, for example, the research of Ettinger, Linzen, and Marantz (2014). In an auditory lexical decision task with simultaneous magnetoencephalograhy (MEG) recording, these authors asked participants to respond to Engish words that crossed morphological complexity (e.g., *bruis-er* vs. *bourbon*) with predictability of the word ending (*bourbon* vs. *burble*), which was calculated as "continuation surprisal". Their MEG results showed a significant interaction such that the main effect of surprisal was stronger for bimorphemic words than for monomorphemic ones, which they interpret to mean that morphological structure enhances the phoneme prediction process. Their behavioral results, however, showed no such interaction. Why might this be the case?

One answer, suggested by our results, is that the advantage for complex words may not derive from a single source. Experiment 1 shows that the roots of complex words – specifically, suffixed roots – enjoy an advantage in perception. Meanwhile, Experiment 2 shows that the affixes of complex words enjoy an advantage in memory. These are two very different effects, and lexical decision tasks do not distinguish between them. For example, to decide whether a stimulus such as *drinker* is a real word, participants must perceive the signal [dɹɪŋkɚ] and then access their memory for previously-heard instances of this signal, /dɹɪŋkɚ/. If they match, the participant will respond "yes", but this response reflects an interaction between both perception and memory whose nature is not well-understood. We do not know, for example, whether a "yes" response requires both clear perception and strong memory, just one of these components, or neither of them. Possibly, an interaction between morphological structure and surprisal, such as that found in MEG data (Ettinger et al., 2014), could be revealed in behavorial data by a task in which a single component – percpeption only, or memory only – was more isolated.

Beyond the question of morphological structure, our results also have certain implications for the link between perception and memory. Perception and memory often correlate with one another, such that clear perception creates the illusion of stronger memory and vice versa (e.g., Goldinger et al., 1999; Jacoby et al., 1988; Kelley & Jacoby, 1990; Whittlesea, 1993; Whittlesea et al., 1990; Whittlesea & Williams, 2000). According to many researchers, "familiarity" is the concept that links these results. That is, perceptual clarity creates a sense that a stimulus seems familiar, and familarity influences memory judgments. Similarly, previous exposure also creates a sense that a stimulus seems familiar, and familiarity influences perceptual judgments.

If this correlation between perception and memory is robust across diverse contexts, we would expect to also find evidence for it in the current study. According to the linking logic outlined above, greater familiarity for certain types of morphemes could be the single driving factor behind *both* perceptual clarity and strength of memory. In our study, however, we see no similarity between the perceptual results of Experiment 1, which show a facilitative effect for root entries, and the memory results of Experiment 2, which show a factiliative effect for affix entries. This suggests that perception and memory are actually two distinct processes whose effects can diverge from one another. Other examples of this divergence have occurred in the literature. For example, a frequent word is more likely to be correctly perceived in noise (Luce & Pisoni, 1998), but less likely to be remembered on a memory test (Glanzer & Adams, 1985; Glanzer & Bowles, 1976), suggesting that we might consider the relationship between roots and affixes as analogous to the relationship between frequent and infrequent words. We leave this intriguing question open for future research.

Acknowledgements

Dylan Pearson and Ho Eun Park provided valuable assistance in the development of Experiment 2. Ellen Abolt, Dylan Pearson, and Amara Sankhagowit helped to run participants. Editors and anonymous reviewers provided helpful feedback, as did the audience at 2019 Princeton Phonology Forum. Flaws are my own.

References

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm.' *Cognition*, *93*(2), B79–B87. https://doi.org/10.1016/j.cognition.2004.02.005

Baddeley, A. (1998). *Human memory: Theory and practice*. Allyn & Bacon.

Balling, L. W., & Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, *23*(7–8), 1159–1190. https://doi.org/10.1080/01690960802201010

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, *125*(1), 80–106. https://doi.org/10.1016/j.cognition.2012.06.003

Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *Journal of Memory and Language*, *35*(6), 801–820. https://doi.org/10.1006/jmla.1996.0041

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 941–947. https://doi.org/10.1037/0278-7393.27.4.941

Bergman, M. W., Hudson, P. T. W., & Eling, P. A. T. M. (1988). How Simple Complex Words Can Be: Morphological Processing and Word Representations. *The Quarterly Journal of Experimental Psychology Section A*, *40*(1), 41–72. https://doi.org/10.1080/14640748808402282

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Version 6.0.24) [Computer software]. www.praat.org

Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language*, *28*(1), 1–13. https://doi.org/10.1016/0749-596X(89)90025-9

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5–6), 507–534. https://doi.org/10.1080/01690960143000074

Daly, C., Fallon, A.B., Mak, E., & Tehan, G. (2005). Lexicality and phonological similarity: A challenge for the retrieval-based account of serial recall? *Memory*, *13*(3–4), 349–356. https://doi.org/10.1080/09658210344000215

Ettinger, A., Linzen, T., & Marantz, A. (2014). The role of morphology in phoneme prediction: Evidence from MEG. *Brain and Language*, *129*, 14–23. https://doi.org/10.1016/j.bandl.2013.11.004

Feldman, L.B., & Larabee, J. (2001). Morphological facilitation following prefixed but not suffixed primes: Lexical architecture or modality-specific processes? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 680–692.

Ferreira, F., Apel, J., & Henderson, J.M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Finnigan, S. (2002). ERP "old/new" effects: Memory strength and decisional factor(s). *Neuropsychologia*, *40*(13), 2288–2304. https://doi.org/10.1016/S0028-3932(02)00113-6

Fischer, B. (1987). The preparation of visually guided saccades. In *Reviews of Physiology, Biochemistry and Pharmacology* (Vol. *106*, pp. 1–35). Springer.

Gagnepain, P., Henson, R.N., & Davis, M.H. (2012). Temporal Predictive Codes for Spoken Words in Auditory Cortex. *Current Biology*, *22*(7), 615–621. https://doi.org/10.1016/j.cub.2012.02.015

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Glanzer, M., & Adams, J.K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20. https://doi.org/10.3758/BF03198438

Glanzer, M., & Bowles, N. (1976). An analysis the word-frequency effect in recognition memory. *Journal OF Experimental Psychology-Human Learning and Memory*, *2*(1), 21–31. https://doi.org/10.1037/0278-7393.2.1.21

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Goldinger, S.D., Kleider, H.M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition*, *27*(2), 328–338. https://doi.org/10.3758/BF03211416

Gow, D.W., Segawa, J.A., Ahlfors, S.P., & Lin, F.-H. (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage*, *43*(3), 614–623. https://doi.org/10.1016/j.neuroimage.2008.07.027

Hannula, D.E., Althoff, R.R., Warren, D.E., Riggs, L., Cohen, N.J., & Ryan, J.D. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*(166). https://doi.org/10.3389/fnhum.2010.00166

Hay, J. (2000). *Causes and Consequences of Word Structure*.

Hay, J. (2002). From speech perception to morphology: Affix ordering revisited. *Language*, *78*(3), 527–555. https://doi.org/10.1353/lan.2002.0159

Hay, J., & Baayen, H. (2002). Parsing and productivity. In *Yearbook of Morphology 2001* (pp. 203–235). Springer. http://link.springer.com/chapter/10.1007/978-94-017-3726-5_8. https://doi.org/10.1007/978-94-017-3726-5_8

Hulme, C., Maughan, S., & Brown, G.D.A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685–701. https://doi.org/10.1016/0749-596X(91)90032-F

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, *86*(4), 527–536. https://doi.org/10.1111/j.2044-8295.1995.tb02570.x

Jacoby, L.L., Allan, L.G., Collins, J.C., & Larwill, L.K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 240.

Ji, H., Gagné, C.L., & Spalding, T.L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*(4), 406–430. https://doi.org/10.1016/j.jml.2011.07.003

Kapnoula, E.C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. https://doi.org/10.1016/j.cognition.2014.09.007

Kelley, C.M., & Jacoby, L.L. (1990). The construction of subjective experience: Memory attributions. *Mind & Language*, *5*(1), 49–68. https://doi.org/10.1111/j.1468-0017.1990.tb00152.x

Longtin, C.-M., Segui, J., & Hallé, P.A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*(3), 313–334. https://doi.org/10.1080/01690960244000036

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and Performance: Vol. X. Control of Language Processes* (pp. 125–150).

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Norris, D., & McQueen, J.M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357. https://doi.org/10.1037/0033-295X.115.2.357

Norris, D., McQueen, J.M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–325. https://doi.org/10.1017/S0140525X00003241

Opitz, B. (2010). Context-dependent repetition effects on recognition memory. *Brain and Cognition*, *73*(2), 110–118. https://doi.org/10.1016/j.bandc.2010.04.003

Plitcha, B. (2012). *Akustyk* (Version 1.9.3) [Computer software]. Akustyk, Version 1.9.3. http://bartus.org/akustyk/index.php

Pycha, A. (2015a). Listeners perceive prefixes differently: Evidence from a noise-rating task. *Word Structure*, *8*(1), 53–83. https://doi.org/10.3366/word.2015.0073

Pycha, A. (2015b). Subjective perception of affixation: A test case from Spanish. *Lingua*, *159*, 47–69. https://doi.org/10.1016/j.lingua.2015.03.004

Rastle, K., Davis, M.H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10.3758/BF03196742

Richardson, D.C., & Spivey, M.J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

Ryan, J.D., Hannula, D.E., & Cohen, N.J. (2007). The obligatory effects of memory on eye movements. *Memory*, *15*(5), 508–525. https://doi.org/10.1080/09658210701391022

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review*, *7*(2), 332–340. https://doi.org/10.3758/BF03212990

Samuel, A.G., & Pitt, M.A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, *48*(2), 416–434. https://doi.org/10.1016/S0749-596X(02)00514-4

Schreuder, R., & Baayen, R.H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, *33*(3), 357–375. https://doi.org/10.1006/jmla.1994.1017

Taft, M. (1979a). Lexical access-via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 21–39. https://doi.org/10.1016/S0022-5371(79)90544-9

Taft, M. (1979b). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*(4), 263–272. https://doi.org/10.3758/BF03197599

Taft, M. (1981). Prefix stripping revisited. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 289–297. https://doi.org/10.1016/S0022-5371(81)90439-4

Taft, M., & Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *The Mental Lexicon*, *1*(2), 183–199. https://doi.org/10.1075/ml.1.2.02taf

Taft, M., & Forster, K.I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Taft, M., Hambly, G., & Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *The Quarterly Journal of Experimental Psychology*, *38*(3), 351–365. https://doi.org/10.1080/14640748608401603

Vitevitch, M.S., & Luce, P.A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481–487. https://doi.org/10.3758/BF03195594

Whittlesea, B.W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253.

Whittlesea, B.W., Jacoby, L.L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*(6), 716–732. https://doi.org/10.1016/0749-596X(90)90045-2

Whittlesea, B.W., & Williams, L.D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 547–565.

Wurm, L.H. (1997). Auditory Processing of Prefixed English Words Is Both Continuous and Decompositional. *Journal of Memory and Language*, *37*(3), 438–461. https://doi.org/10.1006/jmla.1997.2524

Zipf, G.K. (1935). *The psycho-biology of language.* http://psycnet.apa.org/psycinfo/1935-04756-000

Appendix A. Morphologically complex words used as stimuli in Experiments 1 and 2

Prefixed		Suffixed	
behold	nonfat	freedom	eastmost
bewitch	nonmeat	kingdom	westmost
befriend	nonword	stardom	northmost
conform	outshout	broaden	greatness
conjoin (verb)	outfox	weaken	shyness
consign (verb)	outclass	toughen	wellness
derail	prebake	bowler	catproof
degrade	prewash	screamer	dogproof
dewax	prejudge	drinker	fireproof
disband	refund	diskette	friendship
disown	rewrite	pipette	lordship
dislodge	regroup	rosette	kinship
enlist	self-doubt	tenfold	backward
enlarge	self-worth	eightfold	upward
enroll	self-love	sixfold	inward
forehead	subway	careful	earthy
forearm	subtext	painful	glassy
foresight	sublet	sinful	heady
imprint	transport	manhood	eastmost
impure	transplant	priesthood	westmost
impart	transform	boyhood	northmost
inbound	unlock	fiendish	greatness
inward	unzip	sluggish	shyness
inland	unhook	freakish	wellness
midtown		quantize	
midday		realize	
midwife		stylize	

Appendix B. Morphologically simple words used as stimuli in Experiments 1 and 2

Prefixed	Suffixed
benign	raven
behave	omen
beyond	linen
congeal	glitter
convey	caper
confide	boulder
debate	duet
detail	cassette
display	abet
dispatch	ruffle
discuss	shuffle
engage	punish
enhance	polish
endorse	rubbish
formant	apprise
immerse	despise
imbue	comprise
immune	trellis
ingrate	witness
insect	harness
index	worship
midget	awkward
middle	coward
resort	duty
transom	mercy
translate	copy

Appendix C1. Acoustic measurements (means, standard deviations) for the naturally-produced recordings used in Experiments 1 and 2, before normalization, for complex words

	Roots				Affixes			
Measure	**Prefixed**		**Suffixed**		**Prefix**		**Suffix**	
Duration	554.55	(109.15)	357.52	(90.34)	272.77	(112.82)	423.53	(103.08)
Amplitude	74.83	(2.62)	77.79	(1.99)	77.26	(2.59)	72.63	(3.08)
Fo	173.93	(21.25)	211.01	(19.87)	198.38	(17.12)	173.44	(42.38)

Appendix C2. Acoustic measurements (means, standard deviations) for the for the naturally-produced recordings used in Experiments 1 and 2, before normalization, for simple words

	Pseudo-Roots				Pseudo-Affixes			
Measure	**Prefixed**		**Suffixed**		**Prefix**		**Suffix**	
Duration	534.67	(126.39)	276.78	(77.03)	227.03	(101.09)	389.65	(131.40)
Amplitude	75.47	(1.77)	78.13	(2.34)	75.89	(2.42)	74.22	(3.19)
Fo	176.33	(37.81)	202.87	(29.04)	194.01	(13.39)	179.44	(42.27)

Address for correspondence

Anne Pycha
Department of Linguistics, Box 413
University of Wisconsin
Milwaukee, WI 53201-0413
USA

pycha@uwm.edu

Copyright of Mental Lexicon is the property of John Benjamins Publishing Co. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.