Frequency and the emergence of prefabs

Evidence from monitoring

Vsevolod Kapatsinski & Joshua Radicke
Indiana University

Abstract

Native English speakers were instructed to detect instances of /ʌp/ in spoken sentences by pressing a button as soon as they hear /ʌp/ regardless of whether it is inside another word. We observe that detection of the particle *up* is slower when the frequency of the verb + *up* collocation is low or extremely high than when it is medium. In addition, /ʌp/ is more difficult to detect in high-frequency words than medium-frequency or low-frequency words. Thus word frequency has a monotonic effect on detectability of word parts while the effect of phrase frequency is U-shaped. These results support the hypotheses that lexical units compete with their parts during speech perception and that words and ultra-high-frequency phrases are stored in the lexicon.

1. Introduction*

There is much evidence that language users are sensitive to co-occurrence statistics between words in both perception and production. In perception, Lieberman (1963) finds that predictable words are more intelligible. McDonald & Shillcock (2004) and Underwood et al. (2004), using eye-tracking, find that words that are probable given the preceding word or words are fixated for a shorter time than words that are not probable. Bod (2001) finds that subjects are faster in deciding that a three-word subject-object-verb sentence is grammatical when the sentence is frequent (*I like it*) than when it is not (*I keep it*). Reali & Christiansen (2007) present self-paced reading data that shows center-embedded relative clauses to be read faster when the embedded clause consists of a frequent pronoun-verb combination (*I liked*) than when it consists of an infrequent one (*I phoned*). Thus the frequency with which words co-occur (or some other co-occurrence statistic) must be stored in memory. The question we address is what effect frequent co-occurrence has on the memory representation of a pair of words.

One hypothesis, which we shall call **the distributed account**, is that co-occurrence simply increases the strength of an associative connection between the co-occurring words. Another hypothesis, **the localist account**, is that the co-occurring words fuse into a larger unit, the prefab, which has its own separate representation in memory (e.g., Bybee 2002; Wray 2002; Solan et al. 2005). This does not mean that the representations for the component words are lost as a result of the fusion. They may well be retained and even used during the production and perception of the frequent phrase. However, under the localist account, the prefab has its own node in the lexicon. That is, the prefab is a lexical unit, just like the words and morphemes that it contains. As Wray (2002: 265) puts it, a formulaic sequence is morpheme-equivalent.

Both theories can account for the finding that high-frequency phrases are processed more easily. In a high-frequency phrase, the end is somewhat predictable given the beginning and will therefore be easier to perceive. Sensitivity to predictability does not necessarily imply that the predictor and the predicted fuse into a unit. Rather, co-occurrence may simply make the co-occurring words able to prime each other.

However, in order to predict that high-frequency phrases are processed more easily than low-frequency phrases, the distributed account must predict

*Many thanks to Joan Bybee, Jill Morford, David Pisoni and Rena Torres-Cacoullos for helpful comments. Work supported by NIH training grant DC-00012 and NIH Research Grant DC-00111 to David Pisoni.

that the more predictable a word, the easier it is to process and detect (due to contextual priming). In particular, the final word of a frequent phrase should be perceived more easily than the final word of a less frequent phrase because the final word of a frequent phrase is predictable given the rest of the phrase and is primed by it.

This is not necessarily the case under a localist account in which prefabs are processed more easily (in part) because they are stored in the lexicon. The predictions of the localist account depend on how the processing of lexical units is hypothesized to interact with the processing of the units' parts. If one assumes that recognition of the whole helps with recognition of the parts (as, for instance, in the Interactive Activation Model of McClelland & Rumelhart 1981), then the localist account makes the same prediction as the distributed one (Healy 1994). If, on the other hand, recognition of the lexical unit interferes with processing of the unit's parts (Healy 1976), parts of high-frequency lexical units (i.e., prefabs) are predicted to be more difficult to detect than parts of low-frequency lexical units.

The idea of between-level competition during lexical access has been proposed independently by Healy (1976), Hay (2003) and Sosa & MacFarlane (2002). Corcoran (1966) and Healy (1976) observed more letter detection errors on the ultra-high-frequency word 'the' than on other words, e.g., the low-frequency word 'thy'. Furthermore, frequency has an effect even when grammatical class is controlled: letters are more difficult to detect in high-frequency nouns than in low-frequency nouns (Healy 1976; Minkoff and Raney 2000). Healy proposed the Unitization Hypothesis to account for the result:

> We can [...] identify [...] syllables, words, or even phrases, without having to complete letter identification. The identification of these higher-order units is facilitated by familiarity [...] Once a larger unit is identified, the processing of its component letter units is terminated, even if the letters have not yet reached the point of identification. Instead, processing and attention are directed to the next location in the text. Because letter identification is not always completed for highly familiar words [...] many letter-detection errors are made on these words. (Healy 1994: 333)

A limitation of the work using orthographic stimuli is that the results could be due to the fact that readers are less likely to fixate low-frequency words than high-frequency words during reading (Corcoran 1966; Inhoff & Rayner 1986). High-frequency words can be perceived parafoveally, where visual acuity is lower, which may impair the reader's ability to identify individual letters within words. Consistently with this interpretation, Hadley & Healy (1991) found that letter detection is no harder in *the* than in other words when subjects can view only five letters at once while reading text and thus are forced to fixate every word.

In the auditory modality, Sosa & MacFarlane (2002) found that detecting the word *of* in spoken sentences taken from the Switchboard Corpus was more difficult when *of* occurred in an ultra-high-frequency phrase such as *kind of* or *sort of* than when it occurred in a lower-frequency phrase, such as *couple of* or *think of*. No difference between medium-frequency and low-frequency collocations was found. Sosa & MacFarlane (2002) argue that extremely frequent phrases (prefabs) are stored in the lexicon and thus detecting *of* in them entails the extra step of morphological decomposition.

A limitation of Sosa & MacFarlane's study is that *of* undergoes much articulatory reduction in high-frequency collocations, such as *kind of* or *sort of*, often appearing without the consonant. This introduces a dilemma for investigating detectability of *of* in such phrases: if a reduced token of *of* is used, it is acoustically non-salient and difficult to perceive as well as being difficult to perceive as an instance of *of*. If a non-reduced token is used, then one is presenting the subject with an instantiation of *of* that is not typical for the context in which it appears. In either case, reaction times may be slowed down for reasons other than the collocation being stored as a single unit.

Thus, in the present study we asked subjects to monitor spoken sentences for a stimulus that does not show much articulatory reduction, the particle *up*. As Sosa and MacFarlane did with *of*, we examine the influence of the frequency of the prefab in which *up* occurs on how easy *up* is to detect. Based on Sosa and MacFarlane's results, we would expect *up* to be more difficult to detect when it occurs in a high-frequency verb+*up* combination like *sign up* than in a less frequent one like *pin up* or *run up*. Using *up* should allow us to test the idea that "it is frequency of use itself that determines the units of storage [...]. The fact that the phrase is not (yet) reduced does not mean that it is not stored in memory as a unit" (Bybee 2001: 161). If high-frequency verb + *up* combinations are stored as lexical units, we would find evidence in support of the idea that abnormal phonological behavior is not a necessary precondition for storage.

Despite the fact that Sosa & MacFarlane did not find differences between low-frequency and medium-frequency phrases, there are reasons to suspect that *up* should be harder to detect in low-frequency phrases than in medium-frequency ones. Morton & Long (1976) and Dell & Newman (1980) found that phoneme detection was faster in words that were relatively predictable given the part of the sentence that preceded them relative to words that were not predictable, e.g., *book* vs. *bill* following *He sat reading a*; and *beer* vs. *brandy* following *He had a drink of* (from Morton & Long 1976). While at first glance this result appears to conflict with the results of Sosa & MacFarlane (2002), predictability of *beer* in *He had a drink of beer* is much lower than the predictability of *of* in *This was done kind of badly*. Conversely, *of* is still relatively predictable in the lowest-frequency collocations used by Sosa & MacFarlane (2002), e.g., *sense of, piece of, each of.* Thus, existing evidence points to a U-shaped effect of phrase frequency on detectability of the phrase's parts: parts of a low-frequency phrase should be harder to detect than parts of a medium-frequency phrase which should be easier to detect than parts of an ultra-high-frequency phrase.

One type of model that predicts a U-shaped effect of phrase frequency on part detectability is one that assumes that a collocation is likely to be stored in the lexicon only if its frequency is above a certain threshold. This type of model has been advocated by Alegre & Gordon (1999) who did not find whole-word frequency effects for regularly inflected English words with a frequency below 6 per million while finding frequency effects throughout the frequency range for monomorphemic controls. If, like regularly inflected words in Alegre and Gordon's model, phrases are stored in the lexicon only if they are frequent enough and, other things being equal, predictability improves detectability, we should find facilitatory effects of predictability in phrases whose frequencies are insufficient for the phrase to become a stored prefab. One version of the theory is depicted in Figure 1.

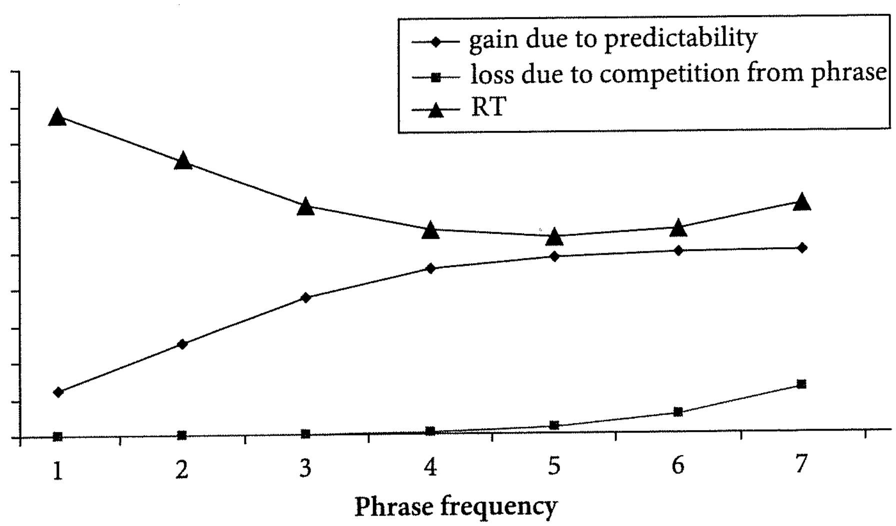

Figure 1. The theoretical relationship between phrase frequency and reaction time (RT) in detecting the second word in the phrase. Here $RT = A + loss - gain$ (predictability makes detection faster while competition from the prefab makes detection slower), where $Gain = \frac{1}{1 + B^{2\text{-}PhraseFrequency}}$ while. $Loss = \frac{1}{1 + B^{8\text{-}PhraseFrequency}}$.[1]

However, a U-shaped relationship between phrase frequency and word detectability is also expected in a model that assumes that the ease of detecting a word is a function of how easy it is to parse the word from the acoustic signal (parseability) and how surprising, and therefore salient, the occurrence of the word is.[2] If the more predictable a word, the easier it is to parse from the signal, words in high-frequency phrases should be easier to detect than words in low-frequency phrases. However, at the same time, the occurrence of a word is not surprising if it is predictable and thus is less likely to attract attention, which could in turn lead to lower detectability. If, as phrase frequency increases, parseability rises faster than salience falls and parseability reaches ceiling (i.e., *up* is always parsed out) before salience reaches floor (i.e., the occurrence of *up* is not paid any attention at all), a U-shaped relationship between phrase frequency and word detectability is expected. Before parseability reaches the ceiling, detectability increases with increases in phrase frequency. After the ceiling is reached, salience is the only factor influencing detectability, hence further increases in phrase frequency should decrease word detectability.

In order to distinguish between the two theories, we need to look at what happens when parseability is not at ceiling and when wholes at the low end of the frequency continuum are also likely to be stored. This can be accomplished by looking at stimuli in which the to-be-detected stimulus, /ʌp/, is not a word but instead occurs inside a word, e.g., *puppy*. In these cases, *up* is less likely to be parsed from the signal and parseability is not at ceiling (accuracy in *up* detection is not perfect). Hence, inhibitory effects of ultra-high-frequency should not be found for word-internal /ʌp/s if they are due to a parseability/salience tradeoff.

On the other hand, if the decrease in parseability of the parts is due to increased competition from the whole, /ʌp/ should be harder to detect in high-frequency words than in low-frequency words. Furthermore, since all words we examine are likely to be stored in the lexicon, there should be a negative correlation between /ʌp/ detectability and word frequency throughout the frequency range.

1. B and A are constants. The crucial feature is that the power to which B is raised is larger in the Loss formula than in the Gain formula. A processing interpretation of this mathematical formulation of the theory is that the word and the prefab are nodes with a sigmoid activation function. During recognition, the prefab and its parts compete for a limited amount of activation where the amount of activation received by a node is proportional to its resting activation level. The constant A represents the minimum time required to make a detection response.

2. This is Corcoran's (1966) idea that predictable words are skipped over/not attended to generalized to auditory perception.

2. Methods

2.1 Materials

The verb + *up* collocations were chosen for inclusion in the experiment based on having non-zero frequency in the British National Corpus (determined through the online interface at http://view.byu.edu/). The British National Corpus was chosen because of its size and the availability of part-of-speech tagging. To find all verb + *up* constructions, we searched for the following pattern: [v*] up.[avp]. We obtained the frequencies of the verb + *up* collocations from the corpus.

The final sample of collocations used in the study was derived by keeping the 10 collocations closest to each end of the frequency continuum and randomly sampling the remaining collocations. In addition, we took all verbs that occurred with the particle *out* in the corpus and included a sample of such verbs that did not occur with *up* in the corpus but did occur with it on Google (the least frequent of these was *eke up*, as in *Tokyo's Nikkei slipped 0.9% and the FTSE 100 in London eked up 0.1%.*) paired with *up* to create the ultra-low-frequency end of the frequency distribution where *up* is not very predictable.

Most of the verb-particle phrases were presented using the past tense form of the verb. For regular verbs, this ensured that *up* was preceded by /d/ or /t/ (sometimes a flap). This was done to ensure that the location of the vowel onset in *up* can be reliably measured and to minimize the influence of phonological context on detectability of *up*.

The first author created 240 experimental sentences containing the particle *up* and 240 control sentences that were identical to the experimental sentences except for containing a different particle. The sentences were presented to the second author, a native English speaker, in a randomized order. The second author read the sentences aloud, having a fixed amount of time (5 seconds) to produce each sentence.

Thirty-five of the control sentences contained the particle *out*. Since experimental and control sentences were syntactically identical, prosody was not a cue to whether *up* occurs in the sentence. In most sentences, *up* was located immediately after the verb. However, to ensure that the subjects process the entire sentence, there were control sentences in which *up* either followed the direct object (*He brought it up*) or was sentence-initial (*Up he goes*). A verb occurring in these control sentences also occurred in an experimental sentence. The control sentences containing *up* were paired with control sentences of the same syntactic structure that contained a different particle so that the number of sentences containing *up* was equal to the number of sentences not containing *up*. The control sentences in which *up* is not immediately after the verb are not included in the analyses presented in this paper because the frequency of verb+*up* combinations was determined only for the most frequent location of *up*, which is immediately after the verb. The subject of the sentence was almost always a pronoun to ensure lack of co-occurrence-based priming between the subject and the particle. Twenty sentences containing noun-phrase subjects occurred in both the experimental and the control set to increase variability in particle location. Previous research has suggested that the greater the variability in location of the to-be-detected unit, the greater the likelihood of obtaining context effects (Lively & Pisoni 1990)

In addition to stimuli in which *up* is a particle, we included a set of sentences in which /ʌp/ was inside another word. These sentences increase variability in target location and allow us to examine how word frequency influences detectability of parts of the word. We can then compare the influence of word frequency to the influence of phrase frequency. The words used were found in the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm, Coltheart 1981). For the experimental sample, we excluded compounds (e.g., *buttercup*), verb-particle constructions, words in which /ʌp/ was followed by a stop (e.g., *interrupt*), and Internet terms, whose frequency would be elevated in Google counts relative to overall use (*pop-up, lookup, setup*). We did not exclude nouns and adjectives derived from verb-particle constructions (e.g., *holdup*). If a noun could be used in the plural, we created two sentences, one containing the noun in the plural and one containing it in the singular.

It was ensured that /ʌp/ was equally likely to occur word-finally (e.g., *holdup*, *cup*), word-medially (e.g., *puppy, hiccups*) and word-initially (e.g., *upholstery, upper*). Morphological and syllabic constituency of /ʌp/ was manipulated. For instance, /ʌp/ is a syllabic constituent (the rime) but not a morphological constituent in *cup* while it is a morphological constituent that crosses a syllable boundary in *upper*. There were 96 /ʌp/-containing words used in the experiment. Each sentence with an /ʌp/-containing word was paired with a control sentence in which the /ʌp/-containing word was replaced by a word containing /aut/. The /aut/-containing words were also found using the MRC Psycholinguistic Database using the same exclusion criteria as for /ʌp/-containing words.

2.2 Subjects and procedure

Twenty adult native English speakers were recruited from among introductory psychology students. They participated to fulfill a course requirement. The subjects were asked to press the 'present' button as soon as they hear *up*, regardless of whether it is a separate word or is inside another word. If the sentence did not contain *up*, they needed to press the 'absent' button to go on to the next sentence. They were encouraged to respond as soon as they hear *up* without waiting until the end of the sentence. The experiment lasted approximately 25 minutes.

2.3 Measurement of frequency and duration

For the purposes of deriving frequency-detectability correlations, we obtained phrase frequency estimates from the spoken portion of the British National Corpus (BNC) and Google. While a U-shaped phrase frequency- word detectability relationship was observed with both counts, the Google-based results exhibited both a larger facilitatory effect on the low-frequency end of the continuum and a larger inhibitory effect at the high-frequency end. Furthermore, the spoken portion of the BNC did not allow us to distinguish between many frequency classes at the low-frequency end of the continuum. Thus only Google results are reported in this paper.

The use of web-based frequency estimates of phrase frequency is supported by the results of Keller & Lapata (2003) who found that plausibility judgments for bigrams that are found only on the Web (and not in the BNC) are reliably predicted by Google frequencies, indicating that Google counts are capturing psychologically relevant variation on the low end of the phrase frequency continuum that the BNC counts are not. Furthermore, even for bigrams found both in the BNC and on Google, correlations with plausibility judgments were higher for web-based frequency counts than for corpus-based ones.

Both base and surface frequency estimates were derived. The surface frequency estimate is the frequency of the verb + *up* combination where the verb is in the particular inflected form used in the experiment. The base frequency estimate is the summed frequency of verb + *up* summed across all forms of the verb. The results did not differ depending on whether base or surface frequency estimates were used.

In analyzing the effect of phrase frequency, the frequency continuum was split into seven bins based on natural discontinuities in our sample of frequencies, as shown in Figure 2.

To investigate the effect of phonological reduction on detectability, we measured the durations of each occurrence of *up* in the materials. We also measured the distance between *up* and the beginning of the sentence. All measurements were done in Praat. The release of the stop closure was taken as the end of the particle. Following stops and fricatives, the beginning of the particle was determined by the beginning of the vowel formants on the spectrogram (since the preceding verb was almost always in the past tense, this was the usual case). When the vowel onset was not readily apparent on the spectrogram, we listened for cues to the identity of the vowel in the preceding speech signal. We took the onset of the vowel to be the latest point at which we could not yet detect cues to the identity of the upcoming vowel. In order to control for possible effects of phonological reduction and measurement error, we measured reaction time both from the onset and the offset of the particle.

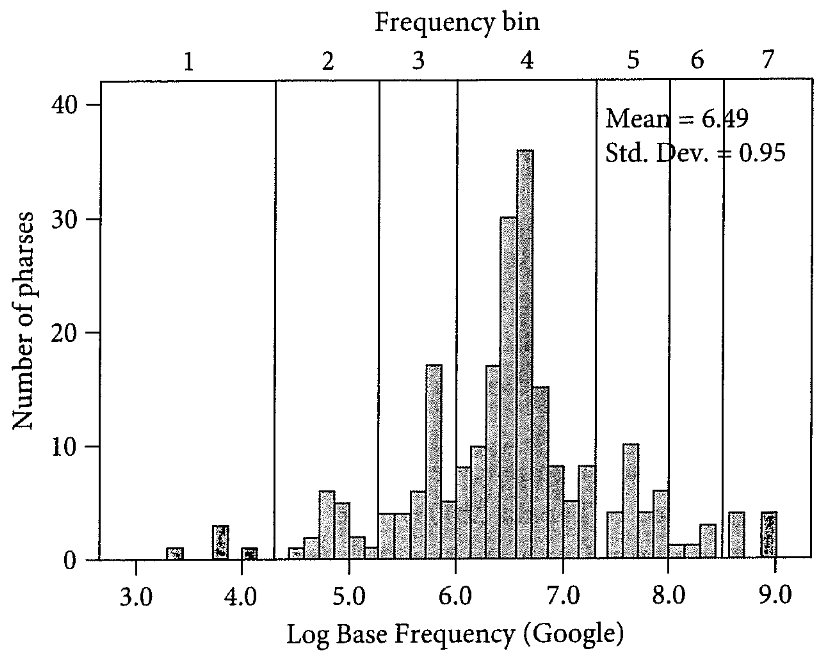

Figure 2. The frequency bins were derived based on discontinuities in the sample of frequencies.

3. Results

3.1 /ʌp/ as a particle

Unlike in Sosa & MacFarlane (2002), accuracy in particle detection in the present study was quite high. Sosa & MacFarlane report that accuracy of *of* detection was at 47% in the lowest-frequency phrases, 60% in medium-low-frequency phrases, 38% in medium-high-frequency phrases, and 37% in the ultra-high-frequency phrases. Results from the present experiment are shown in Table 1. Accuracy in the lowest-frequency group is significantly lower than in any other group (with all other groups combined $p < .0005$; according to one-way ANOVA). Frequency bins 5 and 6 exhibit higher accuracy than either bin 7 ($p = .038$), or bins 2, 3, and 4 ($p = .005$). These results indicate that *up* is easier to detect when it is somewhat predictable than when it is unexpected (Morton & Long 1976; Dell & Newman 1980). The data suggest a U-shaped relationship with accuracy steadily increasing with phrase frequency but then dropping for the highest-frequency bin.

Table 1. Error rate in up detection depending on the frequency of the verb + up collocation

frequency bin	1 lowest	2	3	4	5	6	7 highest
error rate	20%	5%	6%	5%	3%	2%	6%

Figure 3 presents reaction time (RT) data (correct trials only). As predicted by the hypothesis of between-level competition between prefabs and their component words, detection of *up* is more difficult in ultra-high-frequency verb + *up* collocations than in medium-frequency collocations. The difference in reaction time between frequency bin 7 (the highest-frequency bin containing the collocations *get up*, *sign up*, *go up*, and *set up*) and bin 6 (containing slightly less frequent collocations, including *keep up*, *line up*, *stand up*, *catch up*) is statistically significant according to a one-way ANOVA (for reaction time relative to particle onset, $p = .005$, for reaction time relative to particle offset, $p = .002$). Interaction with subject identity is not significant ($p > .1$). The significance of this effect is further confirmed by the fact that a quadratic function, which is U-shaped, provides a much better fit to the data than a monotonic, logarithmic one (the quadratic function explains 96% of the variance in reaction time as a function of phrase frequency while the logarithmic function explains 57% of the variance in reaction time measured relative to the onset and 46% of the variance in reaction time relative to the offset). The effect is observed regardless of whether we estimate phrase frequency via base frequency or surface frequency (for surface-frequency estimates, the difference between groups 7 and 6 is significant at $p<.05$, while the difference between groups 7 and 5 is significant at $p = .002$, interactions with subject identity are not significant, $p > .2$).

The difference in fit almost disappears if frequency bin 7 is removed (the fit of the logarithmic function increases to 94–95% of the variance) indicating that

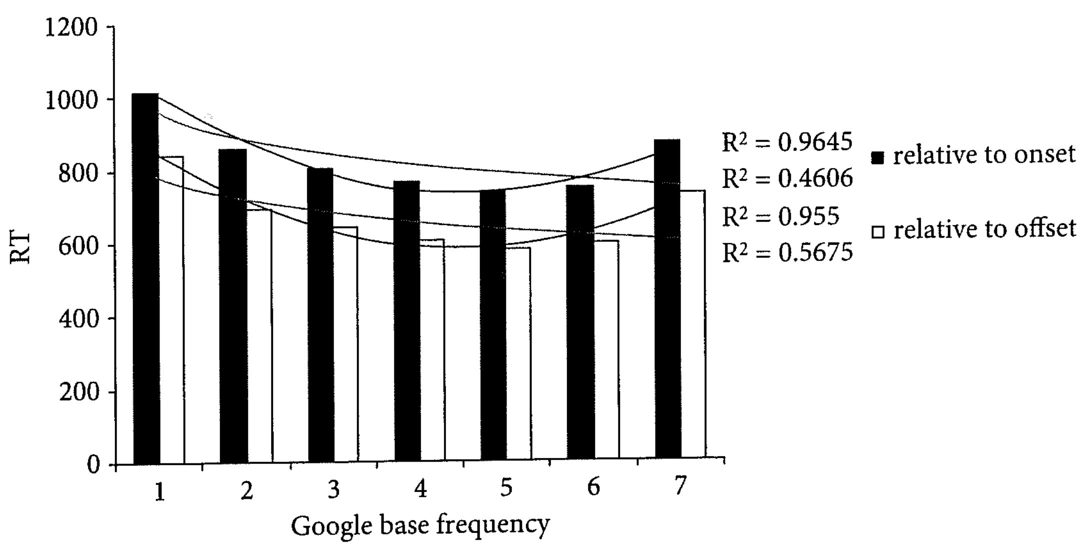

Figure 3. The U-shaped effect of the frequency of verb + up collocations on the speed with which *up* is detected. For both RT measured from the beginning of the word and RT measured from the end to the word, the top R^2 value indicates the amount of variance accounted for by the U-shaped function while the bottom R^2 value indicates the amount of variance accounted for by the monotonic function.

throughout most of the frequency range, increased predictability helps to detect the particle. Just like in Sosa & MacFarlane (2002) and consistent with the accuracy results above, effects of phrase-word competition are only observed with extremely high-frequency phrases. Throughout most of the frequency continuum, *up* detection is easier in higher-frequency phrases than in lower-frequency ones, supporting the hypothesis that, other things being equal, predictability of the to-be-detected unit speeds up detection (Morton & Long 1976; Dell & Newman 1980).

In order to examine how consistent our results are with the results of Sosa & MacFarlane (2002), we examined where the collocations used in the previous study fit onto the frequency continuum derived from Google. We obtained a mean log frequency of 8.15 for their lowest-frequency group, 8.36 for the medium-low-frequency group, 8.77 for the medium-high-frequency group and 8.92 for the ultra-high-frequency group. Thus, their lowest-frequency bin is similar in frequency to our bin 6 (mean log frequency = 8.22) while our group 7 is similar to their medium-high-frequency group (mean log frequency = 8.72). Thus, we find the inhibitory frequency effect at a similar (slightly lower) frequency level than Sosa & MacFarlane. The absence of facilitatory predictability effects in Sosa and MacFarlane's data is consistent with our findings: such effects are found much lower on the frequency continuum (between bin 1 with mean frequency of 3.74 and bin 5 with mean frequency of 7.72) than the range of frequencies used by Sosa & MacFarlane.

Importantly, the duration of the particle does not depend on phrase frequency. As can be seen in Figure 3, the difference between reaction time relative to particle onset and reaction time relative to particle offset is constant throughout the frequency range. Thus, the slow-down in detection observed in ultra-high-frequency phrases is not due to the presence of phonological reduction in those phrases. Thus, the findings of the present study support the hypothesis that phonological reduction is not a precondition for storage (Bybee 2001).

3.2 Word-internal /ʌp/

An alternative interpretation of the results in the previous section is a parseability-salience tradeoff: at some point on the phrase frequency continuum, *up* becomes so predictable that it is always parsed out of the signal. Above that point, further increases in phrase frequency can only decrease how surprising the occurrence of *up* is without increasing the likelihood of *up* being parsed out. To test this hypothesis, we turn to data from trials in which /ʌp/ occurs inside another word. In such cases, parseability of /ʌp/ should be decreased, thus /ʌp/ may be easier to detect in high-frequency words than in low-frequency words. On the other hand, since words are stored in the lexicon, the hypothesis of between-level competition predicts that /ʌp/ should be harder to detect in high-frequency words because such words are stronger competitors. A U-shaped function is not predicted because even the lowest-frequency words are expected to be stored in the lexicon.

Since word-internal occurrences of /ʌp/ are not all equal in terms of location within the word, length of the bearing word, morphological and syllabic constituency, stress, and, as it turns out, duration, we tested for effects of each of these variables. While stress and within-word location did not have a significant main effect, morphological and syllabic constituency, word length, and duration did.

Table 2. shows that /ʌp/ is easier to detect when it is a morpheme than when it is not ($p < .0005$ for both accuracy and reaction time). This result is consistent with Zwitserlood et al.'s (1993) findings for syllable monitoring in Dutch.

Table 2. /ʌp/ is easier to detect when it is a morpheme than when it is not[3]

	Morpheme	Not morpheme
Accuracy	90%	72%
Reaction time	813	1023

As shown in Table 3, accuracy of /ʌp/ detection is also affected by the length of the word in which /ʌp/ occurs: /ʌp/ is more likely to be missed in longer words than in shorter ones ($p = .002$ in a multinomial logistic regression that also included morphological constituency, syllabic constituency, and presence/absence of stress) especially if /ʌp/ is not a morpheme (the interaction is significant at $p = .026$). Table 3 shows that this is not a side effect of differences in duration of /ʌp/ within long and short words: while in general, longer instances of /ʌp/ are easier to detect (Table 6), instances of /ʌp/ that occur in longer words do not tend to be shorter than those occurring in short words (in fact, instances of /ʌp/ tend to be somewhat longer in longer words).

Table 3. The effect of word length on accuracy of /ʌp/-detection (number of segments by percent correct)

Length (segments)		3	4	5	6	7	8	10
% correct	Morpheme	N/A	95%	92%	90%	87%	86%	N/A
	Not morpheme	88%	76%	73%	58%	55%	N/A	55%
duration of /ʌp/ (ms)	Morpheme	N/A	93	94	99	102	116	N/A
	Not morpheme	74	64	84	134	112	N/A	47

3. Reaction time for word-internal occurrences of /ʌp/ is relative to the onset of /ʌp/.

The effect of word length is consistent with the hypothesis of between-level competition. There is a greater chance that not all parts of a word will be fully perceived prior to word identification in a long word than in a short word. Thus, processing of a part is more likely to be interrupted prior to completion in a long word than in a short word. If this hypothesis is correct, then, given that words are processed mostly left-to-right, the effect of word length should be most apparent in the word-final position, less apparent in the word-medial position and least apparent in the word-initial position. This is indeed the case in the data: the effect of word length is highly significant in the word-final position according to a one-way ANOVA ($p < .0005$ for non-morphemic and $p = .008$ for morphemic /ʌp/'s), marginally significant in the word-medial position ($p = .087$ for non-morphemic and $p = .063$ for morphemic /ʌp/'s), and not significant in the word-initial position ($p = .172$ for non-morphemic and $p = .186$ for morphemic /ʌp/'s).

Table 4 shows that detection of /ʌp/ is slower when /ʌp/ straddles a syllable boundary than when it does not ($p < .0005$). There was no difference between cases in which /ʌp/ is a syllable and when it is the rime (whether or not the rime was followed by an appendix). Syllabic constituency does not have a significant effect on accuracy, although the numerical trend is in the same direction as the effect on reaction times (87% correct when /ʌp/ is a syllabic constituent vs. 85% when it straddles a syllable boundary).

Table 4. The effects of morphological and syllabic constituency on the speed of /ʌp/ detection (ms)

	Morpheme	Not a morpheme
Syllabic constituent	796	960
Not a syllabic constituent	964	1187

The effect of syllabic constituency on sequence monitoring has been previously obtained by Mehler et al. (1981) for French, Bradley et al. (1993) for Spanish, and Zwitserlood et al. (1993) for Dutch. It has not previously been found in English (Cutler et al. 1986; Bradley et al. 1993). A possible reason for why previous studies have not found a syllabic constituency effect is that both Cutler et al. (1986) and Bradley et al. (1993) had subjects monitor for sonorant-final targets[4] whereas we used a stop-final target. A post-vocalic sonorant in English is more closely associated with the preceding vowel than an intervocalic stop is (Treiman & Danis 1988; Derwing 1992). Thus, previous syllable monitoring studies in English may not have included (many)

4. Cutler et al. (1986) used /l/, Bradley et al. (1993) used mostly /l/ and nasals except for two stimuli containing /s/.

targets that crossed a syllable boundary. This hypothesis is supported by the results of Ferrand et al. (1997) who failed to observe an effect of prime-target syllable structure consistency in masked priming in English when using Bradley et al.'s (1993) stimuli but were able to obtain it when stimuli with clear syllable boundaries were used.

The findings in Tables 2–4 indicate that /ʌp/ is more detectable when it is a constituent (whether morphological or phonological) than when it is not. These findings support a view of constituency as unithood: constituents are more likely to be parsed out of the signal than phoneme strings that straddle a constituent boundary. Especially in longer words, not all parts of the word are parsed out of the signal. Being a constituent makes a phoneme string more likely to be detected.

There is no interaction between morphological and syllabic constituency for either accuracy or reaction time ($p > .3$), indicating that being a syllabic constituent increases detectability even when /ʌp/ is a morphological constituent. Similarly, being a morpheme increases detectability of units that are syllables or rimes. This suggests that a morphological or syllabic constituent is not always parsed out of the signal. Rather, the fewer the constituent boundaries that lie within a phoneme string, the more likely the string is to be parsed out.

However, before we conclude that constituency affects detectability, we need to address the fact that constituency of the particle correlates with particle duration in the stimuli, as shown in Table 5. Main effects of morphological and syllabic constituency are significant ($p < .0005$ in an ANOVA that included morphological constituency, syllabic constituency and word length as fixed factors and subject as random factor). There is no significant interaction.

Table 5. The effect of constituency on duration of /ʌp/ (ms)

	Morpheme	Not a morpheme
Syllabic constituent	100	86
Not a syllabic constituent	84	67

There is a significant correlation between /ʌp/ duration and how easy it is to detect. Shorter, more reduced, instances of /ʌp/ are detected more slowly (Pearson $r = -.27$, $p < .0005$).[5] Therefore, we conducted a linear regression analysis with logarithmically scaled reaction time as a dependent variable and syllabic constituency (1 vs. 0), morphological constituency (1 vs. 0), presence of stress on /ʌp/, /ʌp/ duration, word length (in segments), distance from sentence onset to /ʌp/ onset, log word frequency, and location of the stimulus in the list of sentences as independent variables. Both of the constituency variables were significant ($t = -4.123$, $p = .001$ for syllabic constituency, $t = -3.227$, $p < .0005$ for morphological

5. We used log10(reaction time) for correlation analyses.

constituency) as was duration of /ʌp/ ($t = -4.206$, $p < .0005$). These results suggest that constituency has an effect on detectability above and beyond duration.

In this analysis, the effect of word frequency only approached significance ($p = .089$, $t = 1.702$). The direction of the trend was as predicted by the hypothesis of between-level competition: /ʌp/ was more difficult to detect in high-frequency words than in low-frequency words. However, we reasoned that the word frequency effect may not manifest itself when /ʌp/ occurs in the word-initial position but only when /ʌp/ occurs word-medially or word-finally. For instance, Lively & Pisoni (1990) observe a much stronger word frequency effect in phoneme categorization when the phoneme was in the final position than when it was in the initial position of a CVC word. In addition, we have observed earlier that the effect of word length on detectability of the word's parts is stronger for non-initial parts.

Thus, we broke the data down by where in the word /ʌp/ was located. Table 6 shows correlations between /ʌp/ duration, log frequency and logarithmically scaled reaction time depending on where in the word /ʌp/ is located. All correlations are significant ($p < .001$) except the one between word frequency and reaction time in the word-initial position, indicating that while word frequency does not appear to affect detection of word-initial targets, this is not simply because word-initial data is messier. The correlations between word frequency and speed of /ʌp/ detection are in the direction predicted by the between-level competition hypothesis: the higher the frequency of the word, the harder /ʌp/ is to detect when it occurs inside it.

Table 6. Correlations (r) between independent variables and reaction time to /ʌp/ depending on the location of /ʌp/ within the word

	Initial	Medial	Final
Word frequency	.052	.285	.221
/ʌp/ duration	−.264	−.231	−.282

When word-initial instances of /ʌp/ are excluded from the regression analysis, word frequency is a significant predictor of reaction time ($t = 2.999$, $p = .003$). Figure 4 shows that when a variety of functions is fit to the data, all of them display a monotonic relationship between word frequency and reaction time. Thus as word frequency increases, time taken to detect /ʌp/ inside the word rises throughout the frequency range. Unlike the effect of phrase frequency, the effect of word frequency is not U-shaped, as expected if (1) all words we presented to subjects are stored in the lexicon, (2) lexical units compete with their parts during recognition, and (3) high-frequency lexical units are stronger competitors.

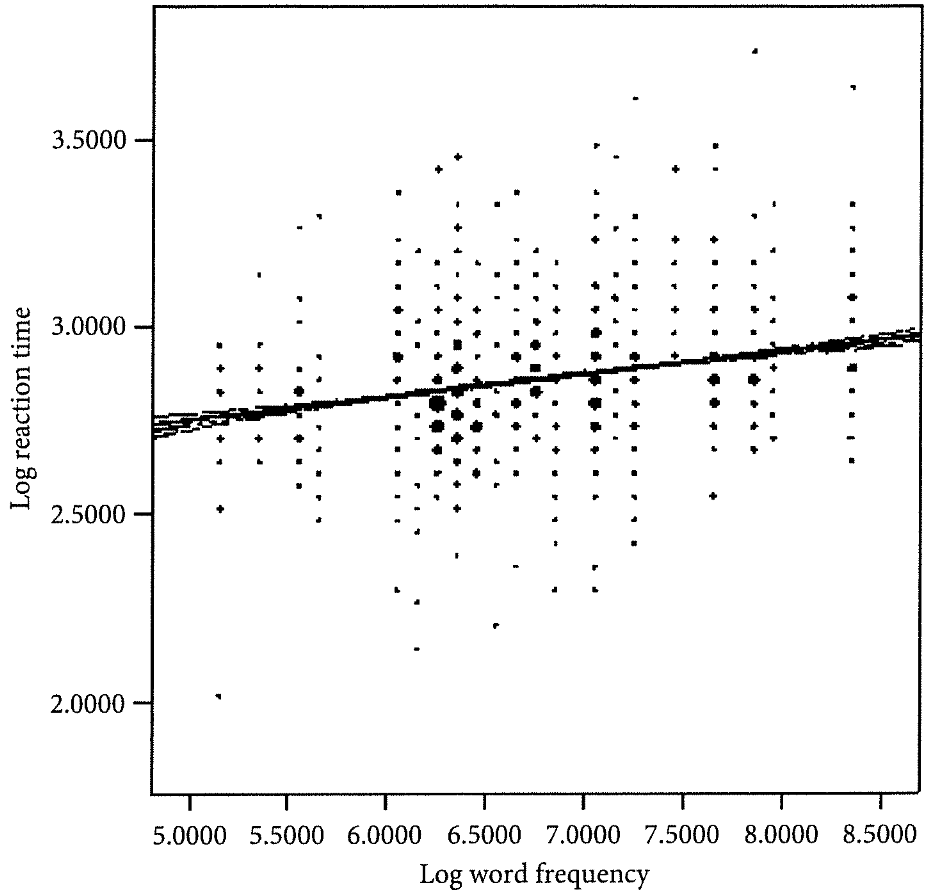

Figure 4. The monotonic relationship between word frequency and detectability of /ʌp/ within the word.[6]

3.3 Summary of the results

When *up* is a particle:

1. The higher the frequency of the verb-particle collocation, the easier the particle is to detect, except for the highest-frequency collocations.
2. Detection of the particle is harder in the highest-frequency verb-particle collocations than in less frequent collocations.

When /ʌp/ is inside another word and is not word-initial:

3. The higher the frequency of the word, the harder it is to detect /ʌp/ inside it.
4. The longer the word, the harder it is to detect /ʌp/ inside it.

6. Circle size indicates number of data points. The trendlines shown are linear, quadratic, cubic and sigmoid.

Regardless of whether /ʌp/ is word-initial:

5. /ʌp/ is harder to detect when it crosses a morphological or syllabic constituent boundary than when it is a morphological or syllabic constituent.
6. Short instances of /ʌp/ are harder to detect than longer instances.

4. Discussion

4.1 Theoretical interpretation

The phoneme sequence /ʌp/ is more difficult to detect inside a high-frequency word than inside a low-frequency word. Thus, parts of frequent lexical units are less accessible to detection than parts of rare lexical units. Given this finding, we would predict that, if prefabs are lexical units, parts of frequent prefabs should be harder to detect than parts of rare prefabs. Finding an inverse relationship between frequency of a whole and detectability of its parts should indicate that at least the high-frequency wholes are stored in the lexicon. Such an inverse relationship is found for verb-particle phrases containing *up* but only at the very top of the phrase frequency continuum. These results are consistent with Sosa & MacFarlane's (2002) findings on word+*of* collocations. They indicate that the highest-frequency phrases are stored in memory as lexical units but they also **suggest** that a phrase needs to be extremely frequent to be stored in the lexicon.[7]

Why are parts of high-frequency lexical units harder to detect than parts of less frequent lexical units? There must be some mechanism that would make activating the prefab interfere with bottom-up activation of the component words and activating a word interfere with bottom-up activation of the component morphemes, syllables, and bigrams. In other words, the results can only be explained if linguistic units in a part-whole relationship compete for activation during the perception process. This hypothesis is also supported by our finding that /ʌp/ is more likely to be missed in a long word, where recognition of /ʌp/ is less likely to be necessary for lexical access.

7. However, as Figure 1 shows, it is also possible that the activation level of the phrase begins to rise slowly as phrase frequency increases, and that until a certain point these frequency-dependent increases in the amount of competition the phrase generates are not enough to offset increases in word predictability that are also caused by increases in phrase frequency. If that is the case, a more prudent conclusion is that the phrase representation does not participate in the lexical access process to a significant degree unless the phrase is extremely frequent.

This idea can be implemented in several non-mutually-exclusive ways. Some possibilities include (1) competition for a limited supply of activation coming from either the acoustic signal or previously perceived context, (2) top-down inhibition, where wholes inhibit their parts when activated beyond a particular threshold (Libben 2005: 276), or (3) removal of the activation source at the completion of lexical access by ceasing to process the acoustic signal that has been parsed into lexical units (Healy 1994).

Finally, we observe that /ʌp/ is easier to detect when it is a constituent than when it is not a constituent. This finding suggests that the acoustic signal is parsed into morphemes and syllables during speech perception making /ʌp/ easier to detect when it matches one of the units automatically extracted from the signal and more difficult to detect when the component segments of /ʌp/ need to be matched to segments that occur in different, though adjacent, units.

4.2 The facilitatory effect of word frequency on phoneme monitoring in word lists

In the present study, we observed that sequence detection is easier in low-frequency words than in high-frequency words. This is consistent with letter-detection results observed by Healy (1976) and Minkoff & Raney (2000). However, a word frequency effect in the opposite direction is often observed in phoneme monitoring (Rubin et al. 1976; Cutler et al. 1987; Eimas et al. 1990; Lively & Pisoni 1990) and letter monitoring (Howes & Solomon 1951; Johnston 1978) where phonemes and letters in high-frequency words are easier to detect than those in low-frequency words.

There is a systematic difference between experiments that find a word-frequency advantage in letter or phoneme detection and those that find a disadvantage: the word-frequency advantage is found with single-word presentation while multi-word presentation yields a word-frequency disadvantage (Healy et al. 1987; Hadley & Healy 1991).[8]

Healy et al. (1987) explain the difference between single-word and multi-word presentation using the Unitization Hypothesis. According to the hypothesis, readers move on to the next word in text as soon as they have identified the current word, terminating processing of smaller units within the current word. When only a single word is visible, there is no subsequent word, hence the subjects will continue processing the word they have already identified, at which point determining the identity of individual letters will be facilitated by having identified the word

8. Eimas et al. (1990) presented target words in a sentence context but the context was constant (the next word is…) and the target word was always the last word in the sentence.

because the reader will be able to use his/her knowledge of what the word is to infer whether the target letter has been presented.

This explanation predicts that the word-frequency disadvantage should not be observed when the target word is in the sentence-final position. Our data are consistent with this prediction: there is no significant correlation between log word frequency and log reaction time for words in the sentence-final position even if only words in which /ʌp/ is not word-initial are included ($r = .047$, $p = .569$). However, this subset of words is small (12 words), so the reliability of this result is questionable.

5. Conclusion

Listeners find it more difficult to detect /ʌp/ in a high-frequency lexical unit than in a low-frequency one or, more concisely, **the stronger the whole the weaker the parts** (Bybee & Brewer 1980; Hay 2003; Healy 1976; Sosa & MacFarlane 2002). While all words are lexical units, leading to a monotonic relationship between word frequency and difficulty of /ʌp/ detection, our results suggest that only high-frequency phrases are stored in the lexicon. Since, other things being equal, predictable units are easier to detect, there is a U-shaped relationship between the frequency of the verb-particle collocation and detectability of the particle. For collocations that are not stored in the lexicon as units, the more probable the particle, the easier it is to detect due to a strong association between the particle and the co-occurring verb. For phrases that are stored in the lexicon, the more frequent the phrase, the more it interferes with the detection of the particle. Finally, /ʌp/ is easier to detect when it matches a morphological or syllabic constituent than when the segments of /ʌp/ are separated by a morpheme or syllable boundary, providing evidence for the hypothesis that syllables and morphemes are extracted from the acoustic signal and take part in the part-whole competition operating during lexical access.

References

Alegre, Maria & Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41–61.

Bod, Rens. 2001. Sentence memory: The storage vs. computation of frequent sentences. Paper presented at the CUNY Sentence Processing Conference. Philadelphia PA.

Bradley, Dianne C., Rosa M. Sánchez-Casas & Juan E. García-Albea. 1993. The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes* 8: 197–233.

Bybee, Joan. 2002. Sequentiality as the basis of constituent structure. In *The evolution of language out of pre-language,* T. Givón & B.F. Malle (Eds), 109–32. Amsterdam: John Benjamins.

Bybee, Joan. 2001. *Phonology and language use.* Camdridge: CUP.

Bybee, Joan & Mary A. Brewer. 1980. Explanation in morphophonemics: Changes in Provençal and Spanish preterite forms. *Lingua* 52: 201–42.

Coltheart, Max. 1981. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology* 33A: 497–505

Corcoran, Derek. W.J. 1966. An acoustic factor in letter cancellation. *Nature* 210: 658.

Cutler, Anne, Jacques Mehler, Dennis G. Norris & Juan Segui. 1987. Phoneme identification and the lexicon. *Cognitive Psychology* 19: 141–77.

Cutler, Anne, Jacques Mehler, Dennis G. Norris & Juan Segui. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25: 385–400.

Dell, Gary S. & Jean E. Newman. 1980. Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior* 19: 607–23.

Derwing, Bruce L. 1992. A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: Preliminary results for five diverse languages. *Language and Speech* 35: 219–35.

Eimas, Peter D., Susan B. Marcovitz-Hornstein & Paula Payton. 1990. Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language* 29: 160–80.

Ferrand, Ludovic, Juan Segui & Glyn W. Humphreys. 1997. The syllable's role in word naming. *Memory and Cognition* 25: 458–70.

Hay, Jennifer. 2003. *Causes and consequences of word structure.* London: Routledge.

Hadley, Jeffrey A. & Alice F. Healy. 1991. When are reading units larger than the letter? Refinement of the unitization reading model. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17: 1062–73.

Healy, Alice F. 1994. Letter detection: A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin and Review* 1: 333–44.

Healy, Alice F. 1976. Detection errors on the word *the*: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance* 2: 235–42.

Healy, Alice F., William L. Oliver & Timothy P. MacNamara. 1987. Detecting letters in continuous text: Effects of display size. *Journal of Experimental Psychology: Human Perception and Performance* 9: 413–26.

Howes, Davis H. & Richard L. Solomon. 1951. Visual duration threshold as a function of word probability. *Journal of Experimental Psychology* 41: 401–10.

Inhoff, Albrecht W. & Keith Rayner. 1986. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics* 40: 431–9.

Johnston, James C. 1978. A test of the sophisticated guessing theory of word perception. *Cognitive Psychology* 10: 123–53.

Keller, Frank & Mirella Lapata. 2003. Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics* 29: 459–84.

Libben, Gary. 2005. Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics* 50: 267–83.

Lieberman, Philip. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6, 172–87.*

Lively, Scott E. & David B. Pisoni. 1990. Some lexical effects in phoneme categorization: A first report. *Research on Speech Perception Progress Report* 16: 327–59. Bloomington IN: Indiana University Speech Research Lab.

McClelland, James L. & David E. Rumelhart. 1981. An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review* 88: 375–407.

McDonald, Scott A. & Richard C. Shillcock. 2004. Eye-movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science* 14: 648–52.

Mehler, Jacques, Jean-Yves Dommergues, Uli Frauenfelder & Juan Segui. 1981. The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20: 298–305.

Minkoff, Scott R.B. & Gary E. Raney. 2000. Letter-detection errors in the word *the*: Word frequency versus syntactic structure. *Scientific Studies of Reading* 4: 55–76.

Morton, John & John Long. 1976. Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior* 15: 43–51.

Reali, Florencia & Morten H. Christiansen. 2007. Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology* 60: 161–70.

Rubin, Philip, Michael T. Turvey & Peter Van Gelder. 1976. Initial phonemes are detected faster in words than in non-words. *Perception and Psychophysics* 19: 394–8.

Solan, Zach, David Horn, Eyton Ruppin & Shimon Edelman. 2005. Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences* 102: 11629–34.

Sosa, Anna V. & James MacFarlane. 2002. Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language* 83: 227–36.

Treiman, Rebecca & Catalina Danis. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language* 27: 87–104.

Underwood, Geoffrey, Norbert Schmitt & Adam Galpin. 2004. The eyes have it: An eye-movement study into the processing of formulaic sequences. In *Formulaic sequences.* Acquisition, processing and use [Language Learning & Language Teaching 9], N. Schmitt (Ed.), 153–72. Amsterdam: John Benjamins.

Wray, Alison. 2002. *Formulaic language and the lexicon.* Cambridge: CUP.

Zwitserlood, Pienie, Herbert Schriefers, Aditi Lahiri & Wilma van Donselaar. 1993. The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19: 1–12.