THE PSYCHOLOGICAL REALITY OF PHONAESTHEMES

BENJAMIN K. BERGEN

Department of Linguistics

University of Hawai`i at Manoa

569 Moore Hall

1890 East-West Road

Honolulu, HI 96822

bergen@hawaii.edu

(808)956-3242
THE PSYCHOLOGICAL REALITY OF PHONAESTHEMES
The psychological reality of English phonaesthemes is demonstrated through a priming experiment with native speakers of American English. Phonaesthemes are well represented sound-meaning pairings in a language, such as English gl-, which occurs in numerous English words with meanings relating to light and vision. In the experiment, phonaesthemes, despite being non-compositional in nature, displayed priming effects much like those that have been reported for compositional morphemes. These effects could not be explained as the result of semantic or phonological priming, either alone or in combination. The results support a view of the lexicon in which shared form and meaning across words is a key factor in their relatedness, and in which morphological composition is not a prerequisite for internal word structure to play a role in language processing.*

* Thanks to Nancy Chang, Jerome Feldman, Larry Hyman, Sharon Inkelas, George Lakoff, and Madelaine Plauché for helpful comments on drafts of related work, and to Amy Schafer and Robert Blust, in addition to two anonymous reviewers for very useful suggestions on this paper.
INTRODUCTION. Phonaesthemes (Wallis 1699, Firth 1930) are frequent sound-meaning pairings, which cannot be defined entirely in terms of contrast, like the English onset gl-. As with other phonaesthemes, this particular sound sequence, gl-, is relatively infrequent in English, except among words with meanings related to ‘VISION’ and ‘LIGHT’, some of which are exemplified in 1a. Another well-documented phonaestheme is the English onset sn-, which occurs in a large number of words relating to ‘MOUTH’ and ‘NOSE’ (1b).

(1) a. gl- LIGHT, VISION glimmer, glisten, glitter, gleam, glow, glint, etc.
    b. sn- NOSE, MOUTH snore, snack, snout, snarl, snort, sniff, sneeze, etc.

When viewed simply as statistically aberrant distributions of sound-meaning pairings in the lexicon, phonaesthemes are found to be pervasive in human languages. They have been documented in such diverse languages as English (Wallis 1699, Firth 1930, Marchand 1959, Bolinger 1980), Indonesian (McCune 1983) and other Austronesian languages (Blust 1988), Japanese (Hamano 1998), Ojibwa (Rhodes 1981), and Swedish (Abelin 1999). While it is difficult to establish reliable estimates of how many of the world's languages display phonaesthemes, or how many words in a given language's lexicon have phonaesthemes in them, no systematic studies of particular languages have produced results suggesting that those languages have no phonaesthemes in them. In general, phonaesthemes seem to appear in content words over function words, and in more specific (or subordinate level) rather than more general (or basic level) words.

In addition to distributional evidence, phonaesthemes can be detected through their role in language change, in particular by the part they play in the generation of neologisms. Phonaesthemes have been implicated in both the production and perception of neologisms on the basis of experimental (Abelin 1999) and comparative studies (Blust 1988, 2003). The recurrent finding of
note in this area is that in a given set of related languages, phonaesthemes that appear in some languages will also appear in other languages, but in words that are not cognates. For example (from Blust 2003) the Austronesian phonaestheme ‹ŋ› occurs in initial position in words that have meanings related to 'NOSE' or 'MOUTH'. It appears in words meaning 'CHEW, MASTICATE, RUMINATE', such as Amis ‹ŋafŋaf›, Toba Batak ‹ŋaltok› and Trukese ‹ŋu›, among others, but none of these words are cognates.

While statistically significant representation in a lexicon and participation in the construction of neologisms are important convergent indications that phonaesthemes are a fact of language, these two types of evidence leave room for different interpretations of exactly what the status of phonaesthemes is. The extent to which phonaesthemes play a role in the synchronic mental organization of language remains an open question.

Indeed, their cognitive status remains controversial, in part because they do not fit well into linguistic theories that view compositionality as a central characteristic of morphological complexity, such as so-called Item and Arrangement models (Hockett 1954). On these compositional types of account, phonaesthemes belong outside the scope of regular morphological analysis. By contrast, non-compositional theories of morphology predict that recurrent sound-meaning pairings in a language, whether compositional or not, will rise to status as organizing principles of the lexicon, on the basis of their frequency. Examples of such non-compositional models are the Dynamic Model (Bybee 1985), the Usage-based Model (Langacker 1991), Seamless Morphology (Starosta To Appear), and most connectionist morphological models (e.g. Plaut and Gonnerman 2000).

This paper investigates the psychological reality of phonaesthemes, thereby providing an indication of the appropriate place for phonaesthemes in morphological theories. Looking ahead, it will be shown below that presenting language users with a sequential pair of words sharing a phonaestheme produces a processing advantage for the second word (a facilitory priming effect).
This phonaesthemic priming effect differs significantly from form priming, shown by words sharing only a phonological onset, and from semantic priming, which occurs between semantically related words. Rather, phonaesthemic priming very closely mirrors the priming effects that have been reported for morphologically related words. Moreover, phonaesthemic priming is not observed between simply any two words that by chance share both some phonological form and some meaning – it only surfaces when the form-meaning pairing is well attested in the lexicon. We will conclude from these results that while compositionality may play a role in morphological organization, statistical prevalence in the lexicon of form-meaning pairings is also a sufficient criterion for those pairings to display morpheme-like behavior in language processing tasks. As such, these results also lend support to the prediction made by non-compositional morphological models that the frequency, and not just the compositionality, of recurring form-meaning pairings is crucial to their mental representation.

1. PHONAESTHEMES AND COMPOSITIONALITY

1.1. COMPOSITIONALITY IN MORPHOLOGICAL THEORIES. Human language is frequently characterized as a finite system with infinite capacity (e.g. Chomsky 1995). On most views, the key property of linguistic systems that gives rise to this infinite capacity is the possibility of combining linguistic units with others in novel ways. Compositional morphological theories, such as Item and Arrangement models, view this property, PRODUCTIVITY, to be predicated upon the hierarchical nature of certain linguistic structures. Complex linguistic units such as complex words are believed to be composed of smaller units, that is, they are COMPOSITIONAL. Most linguistic theories view compositionality, along with the resulting productivity, as a defining feature of human language. Syntactic, phonological, and morphological compositionality have been particularly central in
shaping mainstream linguistics of the second half of the 20th century. The centrality of this capacity in theories of morphology is entirely unsurprising, since a large part of what language users know about the words of their language is indeed how to combine morphemes to construct novel words.

Many morphological theories view complex words as compiled in production or decomposed in perception. The processes can be seen as the simple concatenation of morphological units, by Item and Arrangement models (e.g. Lieber 1980, Di Sciullo and Williams 1987, and Halle and Marantz 1993). What these models share is the view that the morphological capacity is something akin to a metaphorical assembly line, which either compiles a complex word from its morphological parts in production, or takes apart a complex word to analyze the pieces during perception. Moreover, for many theories of the word-internal relation between sound and meaning, the ability to combine units in novel ways to produce larger ones even defines the domain of study. On one view for example, ‘processes of word-formation that are no longer productive [...] are of little or no interest to morphological theory’ (Baayen and Lieber 1991:802).

By itself, though, compositionality makes up only part of what language users know about how the words of their language pair form and meaning. They also have unconscious knowledge of groups of words that share meaning and are phonologically or prosodically similar (Sereno 1994, Kelly 1992). They know subtle correlations between a person’s gender and the phonology of their name (Cassidy et al. 1999) and between the number of syllables in a word and the complexity of the object it describes (Kelly et al. 1990). They also know when the pieces of a word bear meaning, whether or not those parts can be productively combined with others (Bybee and Scheibman 1999).

As these results indicate, and as we will see below, compositionality, as it is usually understood, does not suffice to capture what language users know about word-internal form-meaning relations. The next section describes the difficulties that assembly line models face, when confronted with phonaesthemes.
1.2. THE TROUBLE WITH PHONAESTHEMES. Phonaesthemes hold a prominent place among morphological phenomena that are troublesome to compositional accounts of morphology. Another of these difficult cases are cranberry morphs, like *cran-* or *boysen-* , which cannot stand alone and are permitted to attach to only one stem. Unlike cranberry morphs, phonaesthemes by definition appear in numerous words. They also attach to material that is no more like conventional morphemes than the phonaesthemes are themselves. These two properties of phonaesthemes – their frequency and the fact that they combine with material of a questionable morphological nature – are what make them particularly irksome.

Phonaesthemes are defined as form-meaning pairings that crucially are better attested in the lexicon of a language than would be predicted, all other things being equal. For an example of this type of hyper-attestation, consider the distribution of four phonaesthemes in the Brown corpus (Francis & Kucera 1967). A full 39% of the word types and a whopping 60% of word tokens starting with *gl-* have definitions that relate to ‘LIGHT’ or ‘VISION’ (Table 1) in an online version of Webster’s 7th Collegiate Dictionary.2 Similarly, 28% of word types and 19% of word tokens with a *sn-* onset have meanings related to ‘NOSE’ or ‘MOUTH’. Complete lists of each set of word type can be found in Appendix 1.

The overwhelming statistical pairings of forms like *gl-* and *sn-* with their associated meanings run contrary to what we would predict if the only correlations between the sounds of words and their meanings derive from shared component subparts, i.e. constituent morpheme. That is, on the basis of the assumption that a word’s phonological form is entirely arbitrary, given its semantics, there
should be the same portion of gl- words that have meanings related to ‘LIGHT’ or ‘VISION’ as there are sn- words that share these meanings. And yet, as Table 1 shows, this is clearly not the case. While more than one third of gl- word types relate to ‘LIGHT’ or ‘VISION’, only one percent of sn- words do. The reverse is true for sn- words and the meanings ‘NOSE’ and ‘MOUTH’ – while nearly one third of sn- word types relate to ‘NOSE’ or ‘MOUTH’, only four percent of gl- word types share this semantics. One can also compare the proportion of gl- or sn- words that bear these particular meanings with the distribution of fl- initial words in Table 1. While some fl- initial words have ‘LIGHT’ or ‘VISION’ meanings, like flicker and fluorescent, these words are less numerous than are gl- words with similar meanings.

Since morphological models are attempts to understand distributions of lexical and sub-lexical sound-meaning pairings in a language, phonaesthemes appear at first blush to be perfect fodder for them. But phonaesthemes are problematic for compositional morphologies because words that contain these phonaesthemes also contain a complement which is usually not itself a meaningful unit. For example, removing the onset phonaesthemes from glint and snarl yields -int and -arl. It would be problematic both semantically and structurally to say that -int and -arl were units which contributed to the words as a whole. Neither do -int and -arl have meanings on their own, nor can we mix and match the phonaesthemes and their complements: glarl and snint are mostly nonsensical (though as potential neologisms we can assign possible meanings to them, as discussed in Section 1.3 below).

It should be noted that although all the examples we have seen thus far are drawn from onsets, other parts of syllables like rimes in words like English smash, bash, crash, and mash (Rhodes and Lawler 1981 and Rhodes 1994) and even entire syllables in languages like Indonesian (Blust 1988) are appropriately classified as phonaesthemes. Thus in some cases, a word may actually appear to constructed compositionally from entirely phonaesthemic material. For example, sneer might be seen...
as composed from the onset phonaestheme *sn-* plus a rime phonaestheme *-eer*, also found in *leer* and *jeer*, with a meaning pertaining to 'EXPRESSION OF CONTEMPT'. Nevertheless, the overwhelming majority of words containing phonaesthemes include material that would be difficult to analyze as phonaesthemic, as a perusal of the examples in Appendix A will swiftly demonstrate. So while phonaesthemes display some potential compositionality, this is not their usual behavior.

It is the tension between compositional models’ apparent need to account for phonaesthemes and their inability to do so that drives us to look further into the details of the psychological status of phonaesthemes.

1.3. EXPERIMENTAL EVIDENCE ON PHONAESTHEMES. Phonaesthemes, as noted above, are pervasive in human languages. Traditionally, they have been documented primarily on the basis of distributional data. While there is a long tradition of psycholinguistic experiments focusing on how language users interact with novel words containing particular purported sound symbolic elements such as /i/ and /a/ (see the summary in Jakobson and Waugh (1979)), only recently has the psychological reality of large numbers of phonaesthemes been placed under systematic experimental scrutiny. As we will see below, the psychological status of phonaesthemes has been difficult to pin down for methodological reasons.

The standard linguistic test for evaluating the psychological reality of a proposed linguistic unit is based on compositionality and productivity. In her seminal study of regular English plural marking, Berko (1958) argued for the psychological reality of morphological knowledge on the basis of subjects’ generalizations of an existing pattern in the language to new forms, such as *wug*. When subjects accurately pluralized the novel form *wug* [wʌg] as *wugs* [wʌgz], this was taken as evidence that they had internalized a rule by for forming plural nouns. Such neologism tasks (now known as WUG-TESTS) are pervasive in psychological studies of morphological knowledge.
But a methodological predisposition towards compositionality and productivity leads to the following quandary. It is quite difficult to assess the psychological reality of phonaesthemes, which are generally not compositional or very productive, if the only metric we have is crucially based on productivity and compositionality themselves. A partial solution to this dilemma has arisen from the observation that phonaesthemes, while non-compositional, are in fact partially productive.

In the past three years, three theses (Hutchins 1998 and Magnus 2000 on English and Abelin 1999 on Swedish) have addressed the psychological reality of phonaesthemes from essentially the same perspective: that of neologisms. Magnus’ work is representative of all three, so her thesis can serve as a model to demonstrate their methodologies. Magnus uses two types of experiment to test the psychological status of phonaesthemes. Both of these are based upon subjects’ recognition and production of neologisms. The first methodology tests for knowledge of phonaesthemes by providing subjects with definitions for non-existent words and asking them to invent new words for those definitions. In Magnus’ study, subjects tended to invent novel words that made use of phonaesthemes of their language. For example, given the definition ‘to scrape the black stuff off overdone toast’, 27% of subjects invented a word that started with *sk*-. The second type of experiment tested the use of phonaesthemes in the perception of novel words. Subjects were presented a non-word and were asked to provide a definition for it. Again, they responded as if they were using phonaesthemes. For example, *glon* (bearing a *gl-* onset) evoked definitions relating to ‘LIGHT’ in 25% of the cases.

These studies each provide independent evidence for the role of phonaesthemes in the processing of neologisms. Magnus (2000) interprets her results as evidence that language users represent individual phones as bearing meaning (note that this is similar to Joseph's (1994) conclusion for the Greek phoneme */ts/*, made on the basis of distributional evidence). Abelin (1999) adds to this the insight that the fewer words there are that share the phonology of a
phonaestheme (excluding the phonasthemic words themselves), and the more common that phonaestheme is among the words sharing its phonology, the more likely subjects are to treat neologisms along the lines of that phonaestheme. Hutchins’ (1998) results confirm the role of phonaesthemes, even cross-linguistic ones, in the processing of neologisms. Interestingly, these findings are qualitatively similar to results of neologism experiments on purely phonological generalizations, such as the Obligatory Contour Principle in Arabic (Frisch and Zawaydeh 2001) and Labial Attraction in Turkish (Zimmer 1969) – frequent patterns, whether phonological or phonological-semantic, are used in neologism interpretation and production.

The results of these studies are intuitive, given what we know about lexical innovation in general. Probably the most famous use of English language literary neologisms are in The Jabberwocky, a poem that confounds young Alice in Lewis Carroll’s (1897) Through the Looking Glass. The first stanza of this poem appears in 2 below.

(2) ‘Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe.

Luckily for Alice, Humpty Dumpty is gracious enough to explain the poem: ‘Well, “slithy” means “lithe and slimy [..]”’ (Carroll 1897) and so on. As we will see in the remainder of this section, this literary example is particularly illustrative of the problem with neologism studies.

Although they provide valuable insight into how language users interact with new words, neologism studies do not by themselves constitute conclusive evidence for the psychological status of phonaesthemes. Even in light of the neologism results from the three phonaestheme works cited
above, one could still hold the position that phonaesthemes are only static, distributional facts about
the lexicon, which speakers of a language can access consciously. This is problematic since
essentially all normal morphological processing happens unconsciously. In other words, we know
that language-users are able to access all sorts of facts about their language upon reflection. People
can come up with a word of their language that is spelled with all five vowel letters and ‘γ’ in order,
or a word that has three sets of double letters in a row. (See below for the answers.3) These abilities
by themselves do not lead us to conclude, though, that orthographic order of vowel letters in a word
is a fundamental principle of implicit cognitive organization. For the same reason, subjects’ ability to
consciously access distributions of sound-meaning pairings in their language does not imply that
those pairings are meaningful for the subjects’ linguistic system.

In fact, different mechanisms could be evoked to account for the cognitive processing that the
neologism studies described above demonstrate.

a. Subjects could be taking a sample of the lexicon, and observing generalizations over a set of
   words that share structure with the neologisms.

b. They might be randomly selecting nearest neighbors.

c. They could be looking for cues in the stimulus itself; notice that the word ‘scrape’ appears in
   the example for sk- provided above, and the work ‘slimy’ appears in Humpty Dumpty’s
   explanation for ‘slithy’.

d. Finally, subjects just might be unconsciously acting in accordance with the structural
   configuration of their language system, which is structured in part on the basis of
   phonaesthemes.

Unfortunately, we cannot distinguish between any of these solutions on the basis of neologism tasks
like the ones described above.
What is needed to resolve this issue, then, is a way to tap into unconscious language processing of morphemes. Applying such a measure to phonaesthemes, we can evaluate the extent to which their behavior in language processing is like or unlike concatenative, productive morphological units. To reiterate, the stakes are high. If phonaesthemes have a demonstrable psychological reality, then the notion of morphemes as concatenative, productive units is insufficient for defining the human capacity to extract and use generalizations about the internal structure of words.

2. Method. From the perspective of their distribution, phonaesthemes are statistically well represented partial form-meaning pairings. In order to test whether phonaesthemes also have a psychological reality - whether they are internalized and used by speaker-hearers - we will have to answer two related questions. First, does the presence of a phonaestheme in a word affect the processing of that word? If phonaesthemes play a role in language processing, then words they occur in should be processed differently from words that do not. Second, if phonaesthemes do in fact affect lexical processing, is this effect significantly different for phonaesthemes than for subparts of words that correlate some form and some meaning, but not in a statistically significant way in the lexicon? In other words, is there a role for the frequency of a form-meaning pairings in determining processing effects?

The morphological priming methodology (first used by Kempley and Morton 1982, summarized by Drews 1996) provides an excellent starting point for developing a methodology for addressing such empirical questions. Morphological priming is the facilitation (speeding up) or inhibition (slowing down) of mental access to a TARGET word on the basis of some other PRIME word, which has been presented previously. A large number of morphological priming studies have found priming patterns that are unique to morphologically related PRIME and TARGET words, and are
not shared by morphologically unrelated words (see the summary in Feldman 1995). On the basis of this difference between morphological priming effects on the one hand and known phonological and semantic priming effects on the other, it has been argued that morphemes have a psychological reality. In the same way, we can hypothesize that if phonaesthemes have some cognitive status, then they should similarly display priming effects that are different from those exhibited by words that do not share phonaesthemes. The study described below aimed to test this possibility.

In this study, 20 native speakers of English, aged 18 to 36, were first presented with a PRIME stimulus, which was an orthographic representation of a word, on the screen of a digital computer. Following the setup of Feldman and Soltano (1999), it appeared slightly above the center of the screen for 150msec, just enough to be barely perceived by the subject. 300msec later, a second stimulus, the TARGET stimulus, also a typewritten word, was presented in the center of the screen for 1000msec or until the subject responded to it. This inter-stimulus latency of 300msec was chosen because that delay has been reported to most effectively separate morphological from non-morphological priming effects (Feldman and Soltano 1999). At this inter-stimulus interval, there is the least phonological or semantic priming relative to morphological priming. Subjects performed a lexical decision task; they were asked to decide as quickly as possible whether the TARGET was a word of English or not. They were to indicate their decision by pressing one of two keys on the computer keyboard - ‘z’ for non-words and ‘m’ for words.

The critical stimuli were 50 pairs of words, falling into five categories (Table 2). In the first condition, the PRIME and TARGET shared some phonological feature, always a complex onset, and some semantic feature, such as ‘NOSE’ or ‘LIGHT’. Additionally, a significant proportion of the words in English sharing that onset also shared the semantics, giving pairs like glitter and glow. In the second condition, the PRIME and TARGET shared only a phonological feature, always an onset, yielding stimuli such as druid and drip. PRIMEs and TARGETs in the third condition shared
only some semantics, like *cord* and *rope*. In the fourth condition, the pseudo-phonaestheme case, stimuli shared phonology and semantics but were part of a very small number of words occurring in the Brown corpus (usually only two) that did so. An example of such a pair is the nouns *crony* and *crook*. Finally, for comparison, stimuli in the fifth, baseline condition did not share any segments in their onset, or a meaning, giving pairs like *frill* and *barn*. The full list of stimuli presented can be seen in Appendix 2.

Since the subjects were performing a lexical decision task, choosing whether a string of characters was a word of English or not, an equal number of non-words was required as well. An additional 50 nonce words were thus created, by manipulating one or two characters in each of the existing TARGET words (also seen in Appendix 2). For example, the real TARGET word *glow* became *glone* and *crook* became *croof*. Crucially, all of these non-words shared their onset (or in the case of the one vowel-initial word, the initial vowel) with the real TARGET word they were manufactured from. The nonce TARGET words were paired with the PRIME words that the words that were manipulated to form them were paired with. This ensured that subjects could not simply rely on a shared onset between PRIME and TARGET, or indeed a given onset on a TARGET word (such as sn- or gl-) as a cue that the TARGET was a word or not. Each subject saw half of the PRIME words in each condition followed by a real word TARGET and the other half followed by a manipulated non-word TARGET.

**INSERT TABLE 2 ABOUT HERE**

Translated into specific hypotheses about the result of this study, the questions at the top of this section appear as follows. First, if the presence of phonaesthemes affects processing, then responses to condition 1 (the phonaestheme case) should be significantly different from those to condition 5.
(the baseline, unrelated condition). To be certain that this effect does not derive from known priming effects resulting from phonological (Zwitserlood 1996) and semantic relatedness (Thompson-Schill et al. 1998) between PRIME and TARGET, the responses to class 1 must also be distinct from those to classes 2 and 3. Second, if this priming effect is restricted to statistically well represented form-meaning pairs, then responses to condition 1 (phonaesthemes) should be significantly different from those to condition 4 (pseudo-phonaesthemes).

Before assessing the results of the experiment, we need to briefly discuss some controls applied to it. The morphological priming methodology assumes that it takes subjects the same length of time to make lexical decisions for the TARGETs in each of the different conditions. If a set of TARGET words in one condition requires a longer lexical decision than those in another condition, then any differences in response time across conditions could be a result of this difference, rather than different priming effects. In fact, in most experiments using this methodology, the same TARGETs are used across conditions, with only the PRIMES varying. The current experimental design requires different TARGET words in the different conditions, as there are very few words (if any) that simultaneously instantiate phonaesthemes and pseudo-phonaesthemes. Two important factors that influence the rate of decision to a given TARGET word must therefore be controlled for - length and frequency.

The TARGET words in all conditions were matched as closely as possible for average token frequency and length, since these factors have demonstrable effects on morphological priming (Meunier and Segui 1999). As seen in Table 3, it was possible to find stimuli whose lengths were, on average, very similar - no conditions differ by as much as a phone or letter. Frequency is much more difficult to assess, since the type of corpus one selects for frequency measures radically influences the rates of occurrence of the words contained in it. Table 3 shows two standard written frequency measures, Thorndike-Lorge and Kucera-Francis, both taken from the MRC Psycholinguistic
database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). While there is good consistency across most of the conditions for each of these measures it's clear that according to Thorndike-Lorge, TARGET words in the baseline and pseudo-phonaestheme condition are the least frequent, while the Kucera-Francis numbers place the pseudo-phonaestheme words as least frequent.

INSERT TABLE 3 ABOUT HERE

These frequency differences should be a matter of concern, since our main interest is in determining whether different types of relation between a PRIME and TARGET yield different degrees of priming. It might take longer to make a lexical decision about less frequent TARGET words, such as those in the pseudo-phonaestheme condition, regardless of what precedes them, in which case it would be impossible to determine whether any differences among the conditions resulted from effects of the PRIME stimuli or simply from different default rates of response to the TARGET words.

Fortunately, there exist measures of lexical decision times for words in isolation, which will help us determine whether the frequency differences describes above should be of concern. Table 4 shows the mean response time in a lexical decision task to the TARGET words in each of the five conditions, according to the English Lexicon Project (Balota et al. 2002). As we can see, the phonaestheme, form, and meaning conditions are quite similar, while TARGET words in the pseudo-phonaestheme condition are responded to more quickly and those in the baseline condition more slowly. We will take these differences into consideration when analyzing the results from this experiment, in the next section.

INSERT TABLE 4 ABOUT HERE
3. RESULTS. The average reaction times per condition fell out as shown in the first row of Table 5. By far, the fastest average reaction time was to the phonaestheme condition, which was 59 msec faster than the baseline (unrelated) condition. That is to say, subjects responded to a TARGET stimulus 59 msec faster when the PRIME and TARGET shared a phonaestheme, relative to the condition where PRIME and TARGET were semantically and phonologically unrelated. By comparison, PRIME and TARGET pairs that shared meaning but no form were identified only 23 msec faster than the baseline condition. By contrast, words sharing an onset with their PRIME, those in the form condition, took slightly longer to decide on than did those in the baseline condition. In other words, phonaesthemically related words led to much faster recognition of the TARGET than did semantically or phonologically related words.

Moreover, shared form and meaning together were not sufficient to explain the phonaesthemic priming. Recall that the phonaestheme condition elicited reaction times 59 msec faster than the baseline. By comparison, the pseudo-phonaestheme condition, containing words that shared form and meaning but were among very few words in the lexicon sharing both, yielded reactions only 7 msec faster than the baseline. This difference indicates that the extent of the form-meaning pairing’s distribution in the language is crucial to its role in processing. Statistically prevalent form-meaning pairings yield faster responses than do similar words that are statistical loners. The significance of the effect of condition on reaction time was evaluated using a one-way ANOVA subject analysis. The effect was significant, with p < 0.05. The differences between the phonaestheme condition and
each other condition were also evaluated using Fisher's PLSD and Sheffe's, which both confirmed that the phonaestheme condition was significantly different (p < 0.5) from all other conditions.

While these results very clearly indicate that TARGETs are responded to much more quickly when they share a phonaestheme with their PRIME than when they share form, meaning, both, or nothing with it. But as mentioned in the previous section, there remains the possible confound of how long it takes to make a lexical decision on the TARGETs in the different conditions in isolation. If the reaction time differences we observed in the first row of Table 5 are due to solely to different default rates to decision for the TARGETs in the different conditions, then there should be a strong correlation between the default response time and the primed response time. As might be hypothesized from a scan of the data in Table 5, there is no such statistically significant correlation. (The correlation coefficient of the two rows is 0.2, which falls well below the requisite 0.9 required for significance with five observations.)

Giving these data a more detailed analysis, we might be interested in investigating whether differences in the reaction time between the phonaestheme condition and any others are in fact due to absolute differences in response time. One way to assess this is to compare the average difference between reaction time in isolation and in the priming experiment for each condition. If the phonaestheme condition is responded to faster under priming than in isolation, while the other conditions are not, then this will be yet another indication that it is the phonaesthemic relation, rather than independent response time, that is responsible for the fast decisions in the phonaesthemic condition reported above. This is almost precisely what we find - as seen in the last row of Table 5, TARGETs in the phonaestheme are responded to much faster (38.3 msec faster) in the current priming experiment than in isolation, while the form, meaning, and pseudo-phonaestheme TARGETs require slower lexical decision responses in the priming experiment than in isolation. The baseline condition showed a slightly faster response in the priming experiment than
in isolation. Since the actual conditions of the current experiment and those that yielded the mean isolated lexical decision times differ, we cannot quantitatively compare their results. Nevertheless, they support the conclusion that even when the response times to these TARGET words in isolation are taken into account, the faster response times in the phonaestheme condition do not result from shared form or meaning, or both.

One final confound should be also addressed - the possibility that the stimuli in the phonaestheme condition are more closely related in terms of meaning than those in the other conditions, and that this is the cause for faster response time to words in this condition. While it is difficult to establish reliable and useful semantic similarity ratings, it is clearly useful to try to apply some quantitative analysis to the problem. One useful semantic similarity metric is a similarity rating produced by Latent Semantic Analysis (LSA - Landauer et al. 1998). LSA is, among other things, a statistical method for extracting and representing the similarity between words or texts on the basis of the contexts they do and do not appear in. Two words or texts will be rated as more similar the more alike their distributions are. LSA has been shown to perform quite like humans in a range of behaviors, including synonym and multiple-choice tasks. Of relevance to the current discussion is the pairwise comparison function, which produces a similarity rating from -1 to 1 for any pair of texts. Identical texts have a rating of 1, while completely dissimilar ones would have a rating of -1. To give examples from the pairs used in the phonaestheme experiment, *glitter* and *glow* have an LSA value of 0.27 (somewhat similar) while *barn* and *frill* have an LSA of 0 (no particular similarity or dissimilarity).

The mean LSA scores for each condition were as seen in Table 6. While the LSA values for word pairs in the form and baseline conditions were close to 0, those of the pseudo-phonaestheme, phonaestheme, and meaning conditions were higher. The significance of differences among these conditions were evaluated using an ANOVA, which took the semantic relatedness of each PRIME-
TARGET pair, measured as an LSA value, as the dependent variable and condition (phonaestheme, form, meaning, etc.) as the independent variable. According to this analysis, the phonaestheme condition was not significantly different from the pseudo-phonaestheme condition (p=0.16), or from the meaning condition (p=0.65) but was significantly different from the form and baseline conditions (p=0.01 for each). The pseudo-phonaestheme was however distinguishable from the meaning condition (p<0.05) but not from the baseline (p=0.18) or the form condition (p=0.16). In other words, the pairs of words in the phonaestheme condition were just as similar in terms of meaning as were the words in the meaning-only and pseudo-phonaestheme conditions, while word pairs in the form-only and baseline conditions were significantly different.

The current experiment has demonstrated phonaesthемically related facilitory priming by PRIMEs that are both semantically and phonaesthемically related to a TARGET that cannot be accounted for simply on the basis of other sorts of priming. The meaning priming and form priming both yield quantifiably different priming effects from phonaesthemic priming. As a potential complication, it has been proposed in the morphological priming literature that since stimuli that share a morpheme share both form and meaning, the sum of the priming effects from each of these domains might give rise to the actual priming observed with phonesthemes (see Feldman and Soltano 1999 for discussion). However, taken together, form and meaning priming are still distinct from phonaestheme priming no matter how they are measured. If the difference between the baseline reaction time and the reaction time in each other condition is taken as the priming effect for that condition, then the sum of the meaning and form priming effects (20mesc) differs a great deal from the phonaesthemic priming effect (59mesc). If, however, it is the difference between the
reaction time to a condition in isolation and the time to respond in the priming context that is used, then the difference is even greater - the form and meaning priming effects would sum to a delay of -25 msec, while the phonaesthetic priming effect would be an increase in speed at 38 msec. No other metrics aside from summation have been proposed in the literature to deal with combined phonological and semantic effects. Phonaesthetic priming is an entity unto itself, resulting from the statistical over-representation of a particular pairing between form and meaning in the lexicon.

These results disconfirm the possibility that language-users access phonaesthetic knowledge only during tasks allowing reflection, such as the neologism tasks described in Section 1.3., above. Instead, phonaestheme effects emerge even while individuals are performing a task that is tightly constrained by time pressure, and is therefore processed unconsciously, like natural language is.

4. DISCUSSION. The experimental evidence described above argues for the psychological reality of phonaesthetic priming effects. These effects bear implications for questions as fundamental as the definition of the morpheme and the role of frequency in language learning, representation, and change.

4.1. PHONAESTHHEME AND MORPHEME PRIMING. Despite the fact that they are mostly non-compositional, phonaesthemes behave just like central morphological units in terms of two definitional criteria. First, unlike the most frequent and productive affixes, but like other semi-productive units like English en- (Baayen and Lieber 1991), phonaesthemes are partially productive. Evidence of this productivity comes from the use of phonaesthemes in interactions with neologisms, as shown in Hutchins (1998) and Magnus (2000). Second, as we will now see, phonaesthemes mirror the priming behavior of canonical morphemes.
In general, morphologically related forms demonstrate faciliatory priming that differs from both semantic priming and phonological priming in terms of its degree and time course. Consider the results of an experiment described by Feldman and Soltano (1999), which was nearly identical in format to the one described above. The key difference between the two experiments was that in Feldman and Soltano, the test stimuli were morphologically related, instead of being phonaesthetically related. Unlike the phonaestheme stimuli, the morphologically related forms in Feldman and Soltano’s experiment shared a productive suffix, like the regular past tense. Table 7 lays out the priming effects from Feldman and Soltano’s experiment in comparison with those from the phonaestheme experiment (measured as the difference between the mean reaction time of the baseline condition and that of each of the other conditions).

INSERT TABLE 7 ABOUT HERE

We cannot compare the measures of priming in these two experiments quantitatively, because the test conditions differed slightly. We can nevertheless draw a valuable conclusion from their juxtaposition. The phonaesthemic priming detected in the experiment described above is qualitatively similar to evidence for morphological priming, like the morphological priming reported by Feldman and Soltano. In both cases, at the same inter-stimulus interval, the morphologically or phonaesthetically related condition yielded significantly greater faciliatory priming than form-only priming, meaning-only priming, or the sum of form and meaning priming.

This similarity between phonaesthemic priming and morphological priming, along with the partial productivity of phonaesthemes, suggests that phonaesthemes have a status in the mental language processing system that is similar to that of canonical morphemes. Observations of the importance of phonaesthemes as organizing units has led some, such as Blust (1988), to suggest that
Phonaesthemes may be represented as units, below the level of the morpheme, but above the level of the phoneme. While such a solution is enticing since it could preserve the compositionality of morphemes, it nevertheless spawns as many important difficulties. For one, phonaesthemes are not compositional, which leaves open the question of how the remainder of a morpheme is represented at the phonaesthemic level. Second, instantiating a phonaesthemic level provides no a priori explanation for why frequent sound-meaning pairings would give rise to morpheme-like and not phoneme-like priming effects. Third, and perhaps most importantly, such a solution does not provide any explanation for the importance of frequency effects at the phonaesthemic level. Phonaesthemes, as we will see below, are just the tip of the frequency iceberg.

**4.2. OTHER STUDIES OF STATISTICAL SOUND-MEANING PAIRINGS.** It is not only phonaesthemes, but virtually any form-meaning pairing with sufficient statistical reliability that can play a demonstrable role in the mental processing of language. The range of statistical form-meaning pairings that have been shown to play a role in processing (of which only a selection are surveyed below) testifies to the industriousness of the human capacity to encode statistical form-meaning correlations within words. Research on a number of seemingly unrelated topics has indicated that language users integrate and make use of statistical correlations between sound and meaning, even when these relations do not play a productive role in the linguistic system.

Language users internalize subtle correlations between a person’s gender and the phonology of their name. Cassidy et al. (1999) documented these correlations, finding that while male names like Richard and Arthur tend to have trochaic stress, female names like Irene and Michelle tend to have iambic stress. Additionally, male names like Bob and Ted tend to end in consonants, while female names like Sue and Mary tend to end in vowels. The researchers then presented adults and four-year-old children with words that bore statistically prevalent characteristics of female or male first names,
asking them to perform neologism and sentence completion tasks. Cassidy and her colleagues found that subjects tended to process the names along the lines predicted by the phonological cues to gender.

In another study, Kelly et al. (1990) found that adults and children have internalized and make use of the correlation in English between the number of syllables in a word and the complexity of the object it describes. Geometric shapes are particularly exemplary of this phenomenon; consider the tendency expressed by the visual complexity and syllabicity of line, square, and point with those of trapezoid, hexagon, and parabola. When presented with novel shapes, subjects were more likely to pair complex shapes with polysyllabic words than they were simpler shapes.

Finally, a segment’s morphological expressiveness correlates probabilistically with the extent to which speakers are willing to reduce that segment. English has a word-final deletion process, which elides the final coronal stops of many words in most dialects. It has consistently been shown that individuals reduce the final \( t \) or \( d \) of a word more, the more frequent the word in question is (Labov 1972, Guy 1980, Bybee 2000). But in addition to this effect, if a final coronal stop has some morphological status – if, for example, it indicates the past tense such as in the words tossed and buzzed – then it is less likely to be reduced. Just like probabilistic effects from gender and visual complexity, this meaning-bearingness effect provides external support that language users internalize statistical form-meaning correlations and make use of them during language processing.

4.3. USAGE-BASED MODELS. The key property that phonaesthemes and the three other cases summarized in the previous sub-section share is that their only manifestation is as recurrent pairings of formal and semantic properties of words. For phonaesthemes, this recurrence is sufficient to allow them to play a role in unconscious language processing, as well as, possibly, in the invention and interpretation of neologisms. While compositional morphologies have very little to say about the
mechanisms by which phonaesthemes might be learned and represented, usage-based views of morphology do. Network models (e.g. Bybee 1985, Langacker 1991), and connectionist models (e.g. Plaut and Gonnerman 2000) both predict that statistical recurrences across words, like phonaesthemes, will automatically rise to the status of organizing structures in a language.

Acquiring phonaesthemes, like all the other form-meaning pairings described above, is accounted for in network models as the emergent product of implicit abstraction over specific lexical representations. For example, in Bybee’s (1985) network morphology, phonological and semantic similarities across words give rise to connections between them. This might be represented as in Figure 1, where we can see that connections between words (depicted as solid lines) emerge where form elements or meaning elements are shared across words. Where form and meaning are shared across a set of words, a morphological relation emerges; that is, a generalization appears where previously there were only memorized word forms. This emergent generalization is usually referred to as a schema (Bybee and Slobin 1982, etc.). Morphological schemata, as simple generalizations over individual tokens, are based not on compositionality or productivity, but rather on the simple recurrence in the language of a form meaning association. Langacker’s (1991) model works similarly, generalizing over individual examples such that more abstract schemas emerge. In both, the only difference between compositional and non-compositional morphological units is whether or not the remainder of the words in which they occur also instantiates a more general schema.

INSERT FIGURE 1 ABOUT HERE

Distributed connectionist models of morphology share the premise that morphological relations are emergent from semantic and phonological generalizations over stored lexical forms. They differ
from network models like Bybee’s, however, in two main ways. First, they are quantitative models, which, when implemented in a computational environment, facilitates their evaluation. Second, rather than a unique representation for each word, they posit a network of phonological and semantic units, where a pattern of activation over those units represents a word. In the representative miniature connectionist model in Figure 2, a tendency for form units /g/ and /l/ to co-occur with the meaning unit LIGHT is quantitatively encoded in the strength of connections among the three units. Activation of /g/ and /l/ will increase the likelihood that LIGHT will also become active, due to emergent strengthening of connections between these nodes and the network’s internal representation in its hidden layer.

In all usage-based models, the strength of an emergent schema, which allows it to play a role in unconscious and conscious language tasks, will depend at least on its frequency of occurrence in the words of a language and the frequency of those words that exemplify it. Schema strength does not depend on the compositionality of the form-meaning pairing. Thus, usage-based morphologies provide a natural explanation for the behavior of phonaesthemes we have seen above.

It is important to note that a certain class of morphological theories are constructed from two independent morphologies – an Assembly Line model and a usage-based model. These "Dual Route" models (introduced by Pinker and Prince 1988) recognize that not all morphology is compositional, and view the morphological capacity as composed of a regular, compositional component, and an irregular, lexical component. Phonaesthemes would in such a model be placed in this second group. But separating morphological knowledge into two separate modules raises a host of problems, particularly when phonaesthemes are on the table. From the discussion above, we can
see that the only demonstrated difference between phonaesthemes and other morphological units is their compositionality. They can all be more or less productive, can be used to invent and interpret neologisms, and demonstrate priming effects that are distinct from phonological and semantic priming effects. Thus, a Dual Route model would need to explain why all these properties are shared by compositional and non-compositional morphological units, which are processed in two separate routes. This is a less parsimonious solution than those, like usage-based models, which use a single mechanism for both types of morphological unit.

4.4. FUTURE DIRECTIONS. The results of the present work raise the question of the precise relationship between morphological and phonaesthetic priming. Only by combining these two types of test stimuli in a single experiment can answer this question. Also of future interest is the specific role that statistical frequency plays in phonaesthetic priming. Meunier and Segui (1999) have shown that frequency plays a role in morphological priming, but we don’t yet know what the exact character of the effect of frequency in phonaesthetic priming is. Is being a phonaestheme an all-or-none enterprise, or does frequency correlate gradedly with degree of phonaesthetic priming? The present study did not include frequency as a graded parameter, but rather as a discrete one; statistically significant phonaesthemes were assigned to the same condition, regardless of how significant they were. Another factor that might play a role in phonaesthetic priming is the degree of semantic overlap between PRIME and TARGET. This factor has been shown to play a role in the degree of morphological priming between stimuli, and might also do so for phonaesthemes (Laura Gonnerman, p.c.).

In addressing the psychological reality of phonaesthemes, we have addressed only the degree to which phonaesthetic knowledge has been internalized by language users. Other questions about phonaesthemes deserve attention, though. For example, the study described above assumes that the
potential universality of or motivation for phonaesthemes is orthogonal to their representation. In contrast with this approach, phonaesthemes have predominantly been investigated not in terms of their representation, but in terms of their motivation – that is, with respect to their sound-symbolic properties. Various studies have attempted to document cross-linguistic phonaesthemes for the purpose of demonstrating that there is some correlation between the sound structure of words that have phonaesthemes and the meanings they bear. Famous cases include the often reported use of high vowels to express small size and low vowels to indicate largeness, as in English *teeny* and *tiny* versus *humungous* and *large* (e.g. Ohala 1984). The results of these studies are frequently interpreted as disproving the arbitrariness of the sign, e.g. Bolinger (1949). Both the results and this conclusion, though, are hotly contested. While the motivation describes ultimate historical or acquisitional causes for the phonaestheme’s existence, there may in fact be some relationship between the semantic motivation for a phonaestheme and its distribution. More semantically unified or more semantically ‘basic’ meanings, however basicness is measured, might be more prevalent. One way to test these hypotheses would be to repeat the experiment presented above with phonaestheme stimuli, which varied in terms of their proposed universality.

And finally, if statistical prevalence of a sound-meaning pairing in a language results in a demonstrable representation of that pairing in the cognitive system, which itself may give rise to the ability of individuals to invent new words based on that pairing, and to interpret others’ neologisms making use of that pairing, then, we would expect statistical distribution of a phonaestheme to affect its extension in a language over time. Phonaesthemes that are well represented should yield new forms over time for which there is no cognate in closely related languages. One such case, discussed by Hock and Joseph (1996) (from Samuels 1972), concerns the rime of words like *drag, flag, lag, and sag*, which all share the meaning of ‘slow, tiring, tedious motion’. According to Hock and Joseph, this cluster gained a new member in the 16th century when English *sacke* became *sag* through an
irregular sound change. Elsewhere, Blust (1988) reports on the historical developments of Austronesian roots, which are like English phonaesthemes, except that they occur word-finally and can encompass as much as an entire syllable. Blust has found numerous cases of cognateless forms in various Austronesian languages which it seems is best explained as a case of the richer roots getting richer, or as Blust (2003) calls it, a "snowballing effect". In the future, we hope to investigate how much of a role the precise distributional statistics of phonaesthemes influences their extension over time.

5. CONCLUSION. Like other non-categorical pairings between phonology and semantics, phonaesthemes have a significant psychological status. Specifically, when a form-meaning pairing recurs with sufficient statistical weight, it comes to take on priming behavior that cannot be explained as the result of form or meaning priming, alone or in combination. Assembly line and dual route models of morphology are unable to account for the similarity between phonaesthetic priming and morphological priming, and the experimental results have been interpreted as supporting usage-based models. Phonaesthemes are a testament to the diligence of the human ability to encode and use subtle statistical associations in the linguistic environment.
REFERENCES


   Dissertation..


   Linguistics 29.801-43.


   (2002). The English Lexicon Project: A web-based repository of descriptive and behavioral
   University.

Barlow, Michael and Suzanne Kemmer (eds.). 2000. Usage-Based Models of Language. Stanford:

   CSLI.


   Benjamins.


   New York: Longman.

Bybee, Joan. 1985. Morphology: A study of the relation between meaning and form. Amsterdam:

   John Benjamins.

Bybee, Joan. 2000. The Phonology of the Lexicon: Evidence from Lexical Diffusion. Usage-Based

Bybee, Joan and Joanne Scheibman. 1999. The effect of usage on degrees of constituency the reduction of don't in English. Linguistics 37-4.575-596.


Hockett, Charles F. 1954. Two Models of Grammatical Description. Word 10.210-231


Marchand, Hans. 1959. Phonetic symbolism in English word formations. Indogermanische Forschungen 64: 146-168.


<table>
<thead>
<tr>
<th></th>
<th>Types</th>
<th>Tokens</th>
<th>Types</th>
<th>Tokens</th>
<th>Types</th>
<th>Tokens</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIGHT / VISION</td>
<td>38.7%</td>
<td>59.8%</td>
<td>1.1%</td>
<td>0.3%</td>
<td>0%</td>
<td>0%</td>
<td>10.9%</td>
<td>10.1%</td>
</tr>
<tr>
<td>NOSE / MOUTH</td>
<td>4%</td>
<td>1.4%</td>
<td>28.4%</td>
<td>19.0%</td>
<td>25.3%</td>
<td>27.3%</td>
<td>5.6%</td>
<td>2.7%</td>
</tr>
</tbody>
</table>

Table 1.

Phonaesthemes gl-, sn-, sm-, and fl- in the Brown Corpus
<table>
<thead>
<tr>
<th>Condition</th>
<th>Characteristics</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonaestheme</td>
<td>PRIME and TARGET shared a semantic feature and a phonological onset and were a statistically significant subclass of the lexicon.</td>
<td>glitter:glow</td>
</tr>
<tr>
<td>Form</td>
<td>PRIME and TARGET shared an onset.</td>
<td>druid:drip</td>
</tr>
<tr>
<td>Meaning</td>
<td>PRIME and TARGET shared a semantic feature.</td>
<td>cord:rope</td>
</tr>
<tr>
<td>Pseudo-Phonaestheme</td>
<td>PRIME and TARGET shared a semantic feature and an onset but were not a statistically significant subclass of the lexicon</td>
<td>crony:crook</td>
</tr>
<tr>
<td>Baseline</td>
<td>PRIME and TARGET were unrelated in form and meaning</td>
<td>frilk:barn</td>
</tr>
</tbody>
</table>

**TABLE 2.**

Five test conditions with examples
<table>
<thead>
<tr>
<th></th>
<th>Thorndike-Lorge</th>
<th>Kucera-Francis</th>
<th>Letters</th>
<th>Phones</th>
<th>Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonaestheme</td>
<td>181.4</td>
<td>15.75</td>
<td>5.2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Form</td>
<td>136.56</td>
<td>12.7</td>
<td>5</td>
<td>4.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Meaning</td>
<td>160.5</td>
<td>13.1</td>
<td>4.7</td>
<td>4.1</td>
<td>1.4</td>
</tr>
<tr>
<td>Pseudophonaestheme</td>
<td>111.5</td>
<td>5.5</td>
<td>5.1</td>
<td>4.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Baseline</td>
<td>96.3</td>
<td>12.22</td>
<td>4.8</td>
<td>3.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Table 3.

For each condition, mean frequency (written frequency in Thorndike-Lorge and Kucera and Francis), and length (in number of letter, number of phones, and number of syllables)
<table>
<thead>
<tr>
<th></th>
<th>Mean RT in isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonaestheme</td>
<td>645</td>
</tr>
<tr>
<td>Meaning</td>
<td>631</td>
</tr>
<tr>
<td>Form</td>
<td>655</td>
</tr>
<tr>
<td>Pseudo-phonaestheme</td>
<td>610</td>
</tr>
<tr>
<td>Baseline</td>
<td>681</td>
</tr>
</tbody>
</table>

**Table 4.**

For each condition, mean reaction time to the stimuli in isolation (according to the English Lexicon project)
### TABLE 5.

<table>
<thead>
<tr>
<th></th>
<th>Phonaestheme</th>
<th>Form</th>
<th>Meaning</th>
<th>Pseudo-P</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average RT under priming (msec)</strong></td>
<td>606.7</td>
<td>668.2</td>
<td>642.7</td>
<td>658.7</td>
<td>665.3</td>
</tr>
<tr>
<td><strong>Average RT in isolation (msec)</strong></td>
<td>645</td>
<td>655</td>
<td>631</td>
<td>610</td>
<td>681</td>
</tr>
<tr>
<td><strong>Difference (msec)</strong></td>
<td>-38.3</td>
<td>13.2</td>
<td>11.7</td>
<td>48.7</td>
<td>-15.7</td>
</tr>
</tbody>
</table>

Average TARGET reaction times by condition, along with mean reaction time to the stimuli in isolation (according to the English Lexicon project) and the difference between the two measures.
### Table 6.

For each condition, mean LSA semantic similarity rating for word pairs

<table>
<thead>
<tr>
<th></th>
<th>Mean LSA rating for PIME-TARGET pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonaestheme</td>
<td>0.23</td>
</tr>
<tr>
<td>Meaning</td>
<td>0.26</td>
</tr>
<tr>
<td>Form</td>
<td>0.06</td>
</tr>
<tr>
<td>Pseudo-phonaestheme</td>
<td>0.12</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Morphological priming (Feldman and Soltano 1999) and phonaesthetic priming (the current work)
in milliseconds

<table>
<thead>
<tr>
<th></th>
<th>Morphological</th>
<th>Phonaesthetic</th>
<th>Formal</th>
<th>Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feldman &amp; Soltano</td>
<td>49</td>
<td>--</td>
<td>-18</td>
<td>34</td>
</tr>
<tr>
<td>The present study</td>
<td>--</td>
<td>59</td>
<td>-3</td>
<td>23</td>
</tr>
</tbody>
</table>

**TABLE 7.**
FIGURE 1.

A piece of a network model of generalization over phonaesthemes
Figure 2.

A miniature connectionist model for learning and representing phonaesthemes.
These words were among those extracted from the Brown corpus. They were selected as containing the phonaesthetic meanings if one of their senses in Webster's 7th dictionary referred to some aspect of 'LIGHT' or 'VISION' for the case of gl- and 'NOSE' or 'MOUTH' for sn-.

<table>
<thead>
<tr>
<th>gl-</th>
<th>sn-</th>
</tr>
</thead>
<tbody>
<tr>
<td>glance</td>
<td>glimpse</td>
</tr>
<tr>
<td>glanced</td>
<td>glimpsed</td>
</tr>
<tr>
<td>glances</td>
<td>glimpses</td>
</tr>
<tr>
<td>glancing</td>
<td>glint</td>
</tr>
<tr>
<td>glare</td>
<td>glinted</td>
</tr>
<tr>
<td>glared</td>
<td>glinting</td>
</tr>
<tr>
<td>glaring</td>
<td>listen</td>
</tr>
<tr>
<td>glaringly</td>
<td>glistened</td>
</tr>
<tr>
<td>glass</td>
<td>glistering</td>
</tr>
<tr>
<td>glass-bottom</td>
<td>glitter</td>
</tr>
<tr>
<td>glasses</td>
<td>glittered</td>
</tr>
<tr>
<td>glass-fiber</td>
<td>glittering</td>
</tr>
<tr>
<td>glassless</td>
<td>gloom</td>
</tr>
<tr>
<td>glass-like</td>
<td>gloomily</td>
</tr>
<tr>
<td>glassy</td>
<td>gloomy</td>
</tr>
<tr>
<td>glaucoma</td>
<td>gloss</td>
</tr>
<tr>
<td>glaze</td>
<td>glossy</td>
</tr>
<tr>
<td>glazed</td>
<td>glow</td>
</tr>
<tr>
<td>glazes</td>
<td>glowed</td>
</tr>
<tr>
<td>glazing</td>
<td>gloered</td>
</tr>
<tr>
<td>gleam</td>
<td>flowering</td>
</tr>
<tr>
<td>gleamed</td>
<td>glowing</td>
</tr>
<tr>
<td>gleaming</td>
<td>glows</td>
</tr>
<tr>
<td>glimmer</td>
<td></td>
</tr>
<tr>
<td>glimmering</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix 2

<table>
<thead>
<tr>
<th>Phonaestheme</th>
<th>Meaning</th>
<th>Pseudo-Phon</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prime</strong></td>
<td><strong>Word</strong></td>
<td><strong>Nonce</strong></td>
</tr>
<tr>
<td>glimmer</td>
<td>gleam</td>
<td>gleat</td>
</tr>
<tr>
<td>glint</td>
<td>glimpse</td>
<td>glim</td>
</tr>
<tr>
<td>glisten</td>
<td>glare</td>
<td>glame</td>
</tr>
<tr>
<td>glitter</td>
<td>glow</td>
<td>glone</td>
</tr>
<tr>
<td>smirk</td>
<td>smell</td>
<td>smett</td>
</tr>
<tr>
<td>smooch</td>
<td>smoke</td>
<td>smoofe</td>
</tr>
<tr>
<td>sniffle</td>
<td>sneeze</td>
<td>sneeg</td>
</tr>
<tr>
<td>snore</td>
<td>snack</td>
<td>snab</td>
</tr>
<tr>
<td>snort</td>
<td>sniff</td>
<td>snick</td>
</tr>
<tr>
<td>snout</td>
<td>snarl</td>
<td>snart</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prime</strong></td>
<td><strong>Word</strong></td>
</tr>
<tr>
<td>dial</td>
<td>ugly</td>
</tr>
<tr>
<td>frill</td>
<td>barn</td>
</tr>
<tr>
<td>jest</td>
<td>quill</td>
</tr>
<tr>
<td>lure</td>
<td>twin</td>
</tr>
<tr>
<td>muck</td>
<td>wrangle</td>
</tr>
<tr>
<td>poppy</td>
<td>shiver</td>
</tr>
<tr>
<td>sedan</td>
<td>roast</td>
</tr>
<tr>
<td>thirst</td>
<td>hymn</td>
</tr>
<tr>
<td>vigil</td>
<td>yearn</td>
</tr>
<tr>
<td>wry</td>
<td>nest</td>
</tr>
</tbody>
</table>
Notes

\footnote{Word-and-Paradigm and Item-and-Process models \citep{Hockett1954} do not make strong use of compositionality, though they do depend on productivity.}

\footnote{The distinction between word types, the list of word entries that appear in the corpus, and word tokens, the number of actual instances of each of those words, is a common and necessary one in studies of corpus frequency.}

\footnote{Facetiously and bookkeeper, respectively, although Brian Joseph \citep{Joseph} notes that if one is willing to adopt a flexible definition of word, then a word with five double sets of letters in sequence is boobbookkeeper (a bookkeeper who is a boob).}