Recognizing English Compound Words:  
The Role of Morphological Family Size

by

Rachel N. Berkowitz
Class of 2009

A thesis submitted to the faculty of Wesleyan University in partial fulfillment of the requirements for the Degree of Bachelor of Arts with Departmental Honors in Psychology
Abstract

The current study used lexical decision, naming, and eye-tracking tasks to examine the role of morphological family size in compound word recognition. Family size was manipulated using a factorial design. In the naming and lexical decision tasks, participants responded faster to compounds from large as opposed to small morphological families. In the eye-tracking task, target words were embedded into sentences; family size effects were found in gaze duration, number of fixations, and probability of refixation, but not in first fixation durations. These results suggest that compounds from large morphological families are easier to recognize than compounds from small morphological families, and that this is a semantic effect. Implications for the organization of the mental lexicon are discussed.
Recognizing English Compound Words: The Role of Morphological Family Size

When a person reads a word or a sentence, a number of processes occur. The person must recognize the word, identify its meaning, and, when reading a sentence, integrate the word into the meaning of the sentence. One question of particular interest to researchers is how our mental lexicon is organized; that is, how are the words we read represented in our minds, and how do we access those representations? A better understanding of the processes that occur in normal adults could ultimately have implications for teaching reading and helping individuals with reading disabilities.

According to the concept of the mental lexicon, we have a mental “dictionary” of words. When we read, we must match up the letters we are seeing with the word in our mental lexicon in order to recognize the word and determine its meaning (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). This takes place over several levels of processing. At the orthographic level, we recognize what the word looks like and what letters comprise it. The phonological level consists of information about how the word sounds, and is believed to be activated even during silent reading (Schilling, Rayner, & Chumbley, 1998). Together, the orthographic and phonological levels make up the lexical (or word form) level, which contains structural information about the word. The lexical level is believed to be activated relatively early, and contains no information about word meaning. Access to word meaning occurs at the semantic level of processing, a later stage. Word recognition can be facilitated or
inhibited at any stage of processing by a number of variables (Libben, 1998; Coltheart et al., 2001).

The way a word is represented in our mental lexicon is fairly straightforward when it comes to monomorphemic words, or words consisting of only one morpheme. A morpheme is the smallest unit of meaning; it can be a word or a smaller unit of a word. For example, the word *cat* is a morpheme, but so are the suffixes *-ing* and *-s*. Libben (2006) argues that the sounds, appearance, and meanings of monomorphemic words such as *cat* or *box* must simply be memorized and stored in the mental lexicon. With complex words, such as inflected, derived, or compound words, the picture becomes less clear. For example, are the complex words *walking* and *management* stored in their whole forms, or as roots and affixes (*walk+ing, manage+ment*) so that the reader computes the meaning of the whole word by combining the parts? If maximizing mental storage space is most important, it is likely that words are stored as their component parts. This requires additional mental energy to compute the meanings of words from their stored parts. On the other hand, if maximizing computational efficiency is most important, words should be stored in their whole-word forms; however this takes up more storage space in the lexicon. Ultimately, Libben argues, neither extreme takes place completely; rather it is some combination of both that allows for maximization of both storage and computation.

Compound words are particularly useful for investigating how the mental lexicon is organized because, unlike prefixes or suffixes, each of the two or more morphemes in a compound word is a lexeme, meaning that it can stand alone as a word. For example, the compound word *headband* consists of the lexemes *head* and
Compounding is a productive way of creating novel words, though this is more common in some languages than others (Hyönä & Pollatsek, 1998). Additionally, the spatial layout of compounds varies among languages: In English, compounds can be written as unspaced (wallpaper), hyphenated (single-minded), or spaced (front door). Unspaced compound words are relatively infrequent in English and rarely longer than two lexemes. In Finnish and German, however, spaced and hyphenated compounds are rare, while unspaced compounds are common and frequently consist of more than two lexemes. Hyönä and Pollatsek gave the words snowball, snowball fight, and snowball fight field as an example of productive compounding in Finnish leading to compounds with more than two lexemes—in Finnish, these compounds are unspaced. Additionally, the novel creation of unspaced compounds in Finnish and German is very common and guided by rules. In English, spaced compounds are created frequently, while novel unspaced compounds are uncommon, and no rules guide compounding (Juhasz, 2008).

A number of models have been proposed to explain how compounds are processed in all languages. One of the main questions facing researchers has been whether compound words are recognized via representation of the whole word (the whole word route) or whether they are broken down into their constituent lexemes (decomposition) (Andrews, Miller, & Rayner, 2004; Bertram & Hyönä, 2003; Hyönä & Pollatsek, 1998; Juhasz, Starr, Inhoff, & Placke, 2003; Pollatsek, Hyönä, & Bertram, 2000). One model of decomposition is the compositional model, in which lexemes are accessed in chronological order, and the meaning of the compound is subsequently computed from the meaning of the two lexemes (Andrews, 1986).
Somewhat similar to this is the “file drawer,” or *sequential search* model for all complex words (including compounds), in which the stem or root of a word serves as an access code to locate the rest of the word. The whole word is then looked up in the “file drawer” of words with the same access code (Taft & Forster, 1976). It is now widely accepted that decomposition does occur, but that the whole-word route plays an important role as well (Andrews et al., 2004; Bertram & Hyönä, 2003; Hyönä & Pollatsek, 1998; Juhasz et al., 2003; Pollatsek et al., 2000). One model, which takes into account both the whole-word and the decomposition routes, is called the *dual route* model, or *race* model, in which the whole-word route and the decomposition route compete. A variety of factors influence which route “wins” the race (Pollatsek et al., 2000).

**Decomposition**

It is well established that the frequency of a word affects the ease with which it is processed: more frequently occurring words are processed more quickly and easily than less frequent words (see Rayner, 1998, for a survey). By varying the frequency of the whole word or the frequency of the different lexemes in a compound word, it is therefore possible to index to what extent morphological decomposition is taking place. If a compound is recognized as a whole word, the frequencies of the individual lexemes should not matter; if decomposition occurs, however, the frequencies of the different lexemes will influence processing time.

Though it is commonly accepted that compounds are broken up into their constituent lexemes, the exact role that each lexeme plays is unclear. Hyönä and
Pollatsek (1998) manipulated the frequency of the first lexeme of Finnish compound words in an eye-tracking study. In this type of study, participants’ eye movements are recorded while they read sentences for meaning. A number of measures can be obtained, including the gaze duration (the time spent looking at the word before the eyes move to the next word) and the location and duration of the individual fixations made on a word. Hyönä and Pollatsek found that the frequency of the first lexeme did influence the pattern of fixations on the word: participants made longer first and second fixations on words with unique initial lexemes, and these words were also more likely to be fixated a third time. In a follow-up study, Pollatsek et al. (2000) separately manipulated the second lexeme frequency and the whole word frequency of Finnish compounds. They found both second lexeme effects and whole-word effects, and both effects surfaced at about the same time, later than the first-lexeme effects found by Hyönä and Pollatsek (1998). These findings would support a dual route model, in which the two routes occur in parallel and “race” each other.

Researchers were unable to replicate all of these effects in English, however. Juhasz, et al. (2003) orthogonally varied the frequency of both lexemes. They used a multi-methodological approach, with three tasks: lexical decision, naming, and eye-tracking. In the lexical decision task, participants are presented with letter-strings on a computer screen and must decide as quickly as possible whether or not the given letter-string is an acceptable word; in the naming task, participants are presented with words and asked to speak each word out loud as quickly as possible. In both tasks, shorter response latencies indicate that a word is easier to process. Juhasz et al. found a robust ending-lexeme effect in all three tasks, but were only able to find a
significant effect of first-lexeme frequency in one of the three tasks, the naming task. In contrast, Andrews et al. (2004) also manipulated first- and second-lexeme frequency in an eye-tracking study, and found marginally significant effects of both lexemes on gaze duration and on total time (the total time spent on a word, including regressions back to the word), but they did not find significant effects of first lexeme frequency on first-fixation duration (as Hyönä and Pollatsek (1998) did for Finnish lexemes). Taken together, these studies provide conflicting evidence for the role of the first lexeme in compound word recognition.

One difference between Finnish and English compounds that might explain the conflicting results is word length. Finnish compounds are, on average, much longer than English compounds (Juhasz, 2008), and are more likely to be viewed in two or more fixations. This was the hypothesis put forth by Bertram and Hyönä (2003). In a long Finnish compound word, only the first lexeme is visible during the first fixation, whereas in a shorter English compound, the entire word might be visible during the first fixation. Could these visual acuity constraints be causing first lexeme effects to be found in Finnish but not in English compounds? Bertram and Hyönä manipulated the first lexeme frequency and whole-word frequency of short and long compounds. They found that the first lexeme is initially involved in processing long but not short compounds, and that the whole-word frequency had a significant effect on both long and short compounds. These results support a race model, in which the decomposition route gets a “head start” for long compounds. When Juhasz (2008) tried to replicate this study in English, however, she found the opposite effect: stronger first lexeme effects for short compounds versus long
compounds, and larger whole word frequency effects on long compounds. This suggests that the inconsistency of first lexeme effects in research on English compounds is not due to word length, and that the role of the first lexeme is still unclear.

**Semantic Transparency**

Another variable that could influence compound recognition is the semantic transparency of the compound, or how related the meanings of its lexemes are to the meaning of the entire compound. For example, the word *paintbrush* is transparent: it is a *brush* used to *paint* with. A word like *deadline*, on the other hand, is semantically opaque because the meaning of neither of the lexemes contributes to the meaning of the entire word: a *deadline* is not a *line* that is *dead*, nor does its meaning have anything to do with the words *line* or *dead*. Some compounds, such as *strawberry*, are partially opaque; that is, one lexeme (*berry*) relates to the meaning of the whole compound and one lexeme (*straw*) does not.

For opaque compounds, decomposition should be unhelpful, because the meaning of the word is not related to the meaning of the lexemes. The question arises, then, whether opaque compounds are decomposed in the same way that transparent compounds are. Surprisingly, two studies found that transparency has little to no effect on compound word recognition (Frisson, Niswander-Klement, & Pollatsek, 2008; Pollatsek & Hyönä, 2005). In an eye-tracking study, Juhasz (2007) orthogonally varied lexeme frequencies of transparent and opaque compound words. She did find main effects of transparency, but found that transparent and opaque
words were decomposed similarly in the early stages of processing. The only interaction was in the go-past duration (the sum of all fixations on a word plus regressions back to words earlier in the sentence before moving the eyes past the word), which had a first lexeme effect for transparent compounds only. These findings together imply that decomposition occurs pre-semantically, i.e. without access to word meanings. This raises the question of how this pre-semantic decomposition occurs: how do we know where to separate the two lexemes without accessing the meaning of these lexemes?

Another factor influencing the effects of semantic transparency is the location of the semantic head of the compound. The semantic head is the lexeme that contributes most of the meaning of the compound. For example, the semantic head of *farmhouse* is the second lexeme (because a farmhouse is a *house*), while the semantic head of *mankind* is the first lexeme (because *man* contributes more to the meaning of the compound than *kind*). Two studies that examined whether lexical decomposition is mediated by the location of the semantic head found opposite results. Using a lexical decision task, Dunabeitia, Perea, & Carreiras (2007) found an ending lexeme effect for both Spanish and Basque, even though Basque commonly has an initial semantic head and Spanish (like English) often has a final semantic head. They concluded that the position of the semantic head is therefore irrelevant. This supports a compositional model in which the first and second lexemes are accessed sequentially and meaning is not accessed until the second lexeme is accessed. In contrast, Inhoff, Starr, Solomon, and Placke (2008) used a multi-methodology approach (lexical decision, naming, and eye-tracking) and found larger frequency
effects for the lexeme that was the semantic head; eye-tracking data revealed that these effects were early. Unlike the results of Dunabeita et al., these results suggest a dual route model.

Pollatsek and Hyönä (2005), who found no effect of transparency, suggested that their results were consistent either with a compositional model (where the lexemes are accessed in chronological order and then “put together” to determine meaning) or Taft and Forster’s (1976) “file drawer” model. If meaning is accessed as suggested in the “file drawer” model, transparency effects may have been confounded by two other variables: the number of competitors in the “drawer,” and the compound’s frequency relative to the frequencies of other words in the “drawer.”

This leads to the concept of morphological family size; in this case, the number of other compounds which begin with the same first lexeme as the compound of interest. In Taft and Forster’s model, the members of a compound’s morphological family would be competitors in the “file drawer.” Pollatsek and Hyönä (2005) did regression analyses to determine whether family size and relative position in the family could have confounded transparency effects. According to the authors, prior research had found a facilitative effect of large family size and a facilitative effect of high ranking within the family. Pollatsek and Hyönä found a facilitative effect of family size for both transparent and opaque compounds, with the exception of a small group of opaque compounds with low-frequency first lexemes. This suggested that in general, transparency does not affect word recognition, except possibly for this limited set of opaque compound words.
Taken together, these studies suggest that for the most part, semantic transparency does not influence early (orthographic) processing of compound words; however its effects on later processing (at the semantic level) remain to be determined. Pollatsek and Hyönä (2005) and Frisson et al. (2008) found no effect of transparency, while Juhasz (2007) did find effects of transparency on later processing measures. Additionally, the role of the semantic head of the compound remains unclear.

*Morphological Family Size*

The concept of family size was first experimentally investigated by Schreuder and Baayen in a 1997 study in which they varied different types of frequency for monomorphemic words in Dutch. Using lexical decision and subjective frequency ratings (participant ratings of how frequent they thought the words were), they found that in addition to the frequency of a word’s single form, the frequency of its plural also influenced processing. They also looked at two new variables: family frequency and family size. Since they were dealing with monomorphemic words, Schreuder and Baayen defined the concept of a morphological family slightly differently than did Pollatsek and Hyönä (2005). According to Schreuder and Baayen’s definition, a morphological family consisted of all the words “derived from a given stem by either compounding or derivation” (1997, p. 121). The family size was a type count of the number of words in the family, while the family frequency was the summed frequencies of all the words in the family. Schreuder and Baayen found a strong effect of family size, but no effect of family frequency after controlling for family
size. Using progressive demasking (a task designed to be sensitive to early stages of processing), they further determined that the effect of singular frequency was early, while the effect of family size was relatively late.

Baayen, Lieber, and Schreuder (1997) replicated the results of Schreuder and Baayen (1997) in English. They found that for monomorphemic words, it was the summed singular and plural frequencies and the family size that had an effect. The family frequency again had no effect above the effect of family size. Bertram, Baayen, and Schreuder (2000) expanded the results of these initial studies on monomorphemic words to complex words, and determined that the morphological family of a word is still activated when a complex word is presented, not only when the base word is presented. Furthermore, the transparency of the family members influenced family size effects: while there were strong effects of family size for words with inflectional suffixes, family size effects only emerged for words with derivational suffixes when opaque words were removed from the family count. Overall, the results of these three studies suggest that presentation of a word with a large morphological family leads to more activation within the lexicon, which facilitates recognition. This implies that a strict decomposed storage model is not accurate; in order for activation to spread to other family members, they must have some representation in their whole-word forms.

The above studies examined the family size of monomorphemic or affixed words, but very few studies have examined family size effects specifically in compound words. One such study was done by De Jong, Feldman, Schreuder, Pastizzo, and Baayen (2002), who examined compounds in both Dutch and English.
They defined the first and second lexeme family size as the number of other words containing the first or second lexeme of a compound, respectively; they examined first and second lexeme family frequency as well. They also examine what they call “position family size,” which they define as “a count of family members in which the constituent appears at the same position as it does in the target compound” (p. 558). They claim to find stronger effects of family frequency than of family size for both Dutch and English unspaced compounds, but stronger effects of family size for English spaced compounds. They conclude that, except for English spaced compounds (which are treated like monomorphemic words), it is the cumulative family frequency, not family size, which influences word recognition in compounds. This is in direct contradiction to the findings of Schreuder and Baayen (1997), Baayen et al. (1997), and Bertram et al. (2000), who found no effect of family frequency on monomorphemic and complex word recognition after taking into account family size. However, the De Jong et al. study has a number of methodological and statistical issues. First, they did not include filler words in their lexical decision experiment, so all of the words were compounds. This could have easily led to a decision strategy that does not mirror natural word recognition processes (see e.g. Andrews, 1986). Second, their claims of “stronger effects” for family frequency than for family size were based merely on comparing correlation coefficients, which is not a statistically valid comparison, especially since the two variables themselves are highly correlated. Therefore, although morphological family size has been shown to be an important variable influencing recognition of
monomorphemic, inflected, and derived words, more research into the effects of family size on compound word recognition is needed.

**Time Course**

In addition to introducing the variable of family size, Schreuder and Baayen (1997) brought up the question of the time course of family size effects. They used a progressive demasking study, which is believed to tap into early measures of processing. The effects of singular frequency that had been found in the lexical decision task were magnified in the progressive demasking task, while for family size no significant results were found in the progressive demasking task. For this reason, the authors suggested that singular frequency has an early effect, while family size has a later effect.

Other studies also provided evidence that family size is a semantic variable. Moscoso del Prado Martin, Bertram, and Häikiö (2004) and Moscoso del Prado Martin, Deutsch, Frost, Schreuder, De Jong, and Baayen (2005) successfully replicated the family size effect for Finnish and Hebrew. Because morphological families in Finnish are so large, Moscoso del Prado Martin et al. (2004) distinguished between “dominated family members” and “non-dominated family members.” Dominated family members are “direct descendents” of the target word which are more semantically related to the target word, while non-dominated family members are not direct descendents and are less semantically related to the target word (Moscoso del Prado Martin et al., 2004). They found that the dominated family size had a significant effect on reaction times, while non-dominated family size did not. In
Hebrew, some words have homonymic roots (roots with two different meanings). Moscoso del Prado Martin et al. (2005) found that processing was facilitated by having a large number of family members containing the root that shares the same meaning as the target word’s root (a large “related family size”) and was inhibited by having a large number of family members containing the root that was the homonym of the root in the target word (a large “unrelated family size”). Both of these findings support the hypothesis that family size is a semantic variable and hence influences later processing. It is important to note, however, that in Finnish the non-dominated family size had no significant effect, while in Hebrew the family size of the unrelated homonym had a significant inhibitory effect.

Evidence also supports morphological processing as an early effect. Pylkkanen, Feintuch, Hopkins, and Marantz (2003) used a magnetoencephalography (MEG) study, which measures brain activity every millisecond, to attempt to clarify the time course of family frequency and family size effects. They suggested that Schreuder and Baayen (1997) had not found effects of family frequency because these effects occur earlier than could be observed using a lexical decision paradigm. Pylkkanen et al. hypothesized that family size is a later, semantic effect, and that when looking at earlier processing measures like MEG, family frequency effects might appear. Though they did find an effect of family frequency on wave amplitude, they also found unexpected effects of family size in the early M30 measure. This contradicts their and Schreuder and Baayen’s hypothesis that family size effects occur later in processing, and suggests that morphological family size plays an earlier role than was previously expected. Pollatsek and Hyönä’s (2005) post hoc study of family
size also suggests that family size is an earlier variable. Since transparency is most likely to affect later semantic processing, and since Pollatsek and Hyönä found effects of family size on both transparent and opaque compounds (except for one specific group of opaque compounds), this is also evidence that family size acts at a pre-semantic level.

A recent study by Juhasz (2009) also informs our understanding of the time course of morphological processing, though not specifically of family size effects. Juhasz orthogonally manipulated sentence context (predictable or unpredictable) and first or second lexeme frequency in an eye-tracking study. She found that early processing measures such as first fixation durations were modulated by sentence context, but that later lexeme effects such as gaze durations were not affected by context. This implies that morphology acts at two different stages of processing: an early visual/orthographic stage (perhaps when segmentation occurs) that is influenced by sentence context, and a later semantic stage that is unaffected by context. Juhasz suggested that on the orthographic level, perhaps the first lexeme serves as an access code, which we use to locate the word in our mental lexicon. This idea is similar to Taft and Forster’s (1976) file drawer model, in which the first syllable of a complex word is used to “look up” the whole word in a “file drawer” of other words with that same first syllable. On the semantic level, Juhasz suggested that lexeme meaning is retrieved when it overlaps with the whole-word meaning (when the compound is fully or partially transparent), as also suggested by Libben (1998). Though these studies provide insight into the time course of morphological effects on word recognition, it is still unclear exactly how and when the family size variable takes effect, especially
because the majority of studies of morphological family size use tasks like lexical decision or subjective frequency ratings, which do not provide information about time course.

Current Study

The current study uses a multi-methodology approach comprised of lexical decision, word naming, and eye-tracking to examine the effect of family size on the recognition of compound words. The study aims to clarify the role of the first lexeme in compound recognition. Since family size is highly correlated with lexeme frequency, it could have acted as a confounding variable in prior studies that found conflicting evidence regarding the role of the first lexeme. Furthermore, by using eye-tracking, the current study intends to clarify the time course of the family size effect.

There are few prior studies of morphological family size which focus specifically on compound words and use a family count consisting only of other compound words. One such study was De Jong et al. (2002); however this study had a number of methodological and statistical problems. The current study will be the first to study how morphological family size specifically affects compound word recognition using a factorial design.

This study will also be the first to use a multi-methodology approach to study the role of morphological family size. The three tasks of lexical decision, naming, and eye-tracking impose different demands on the participant, and therefore reflect different parts of the reading process. In a multi-methodological study using these three tasks, Juhasz et al. (2003) found that the effect of beginning lexeme frequency
varied between the tasks; this confirms that the different task demands can produce different results. Additionally, while the lexical decision and naming tasks provide valuable information about word recognition processes, the task demands when a participant performs these tasks are different from the task demands when a person silently reads a word within a larger text. For these reasons, using the three tasks in combination is a valuable method of studying word recognition.

Eye-tracking more closely mirrors the natural reading process than lexical decision and naming tasks, and has not been used before to study morphological family size. It also provides more information on time course than other methods. A number of measures can be obtained from eye-tracking that reflect different stages of the word recognition process. The first-fixation duration measures the length of the first fixation on a word, while the gaze duration measures the amount of fixation time spent on a word before moving the eyes off of the word. Additionally, readers sometimes regress back to a word after reading it, and the number of regressions and the time spent on regressions can also be determined. The total time is a measure of the total time spent fixating a word, including regressions back to that word. In the current study, these measures will provide information about the time course of the family size effect. If the first lexeme acts as an access code at the orthographic level, family size effects should appear in early measures of processing, such as first-fixation duration. If family size effects are semantic, however, they should appear later, during gaze duration or regressions. It is also possible that morphological variables such as family size act at both early and later levels of processing, as suggested by Juhasz (2009).
Stimuli Selection and Normative Studies

*Compound words*

A master list was compiled of all compound words in the CELEX frequency database (Baayen, Piepenbrock, & Gulikers, 1995) and the Educator’s Word Frequency Guide (WFG) (Zeno, Ivens, Millard, & Duvvuri, 1995) with a morphological family size of fewer than 6 members or greater than 19 members. An attempt was made to exclude plurals and derivatives that were present in the databases. Family size was determined by counting the number of unspaced bilexemic compound words in both databases which shared the same first lexeme as the word of interest.

This master list of approximately 4,000 words was subjected to familiarity norming as follows: The list was divided into 27 questionnaires, with approximately 150 words per questionnaire. Each questionnaire was given to eight Wesleyan University undergraduates, who were asked to rate each word on a 7 point scale, with 1 indicating that a word was very unfamiliar and 7 indicating that a word was very familiar. Compounds which received mean familiarity ratings of less than 2 were removed from the master list, and family size was re-calculated to exclude those compounds.

A subset of potential stimuli was taken from the master list. This subset was normed for transparency as follows: the list was divided into four questionnaires, with approximately 150-160 words per questionnaire, and each questionnaire was given to eight undergraduate raters. Again, a 7 point scale was used, with 1 indicating that a
After transparency norming was complete, final target words were selected. Stimuli characteristics are presented in Table 1. The target words consisted of 27 compounds from large morphological families (with 20 or more members) and 27 compound words from small morphological families (with 5 or fewer members). All target words were 7-11 characters long. Target words were selected so that the two groups were matched on whole word frequency, first lexeme frequency, second lexeme frequency, familiarity, transparency, word length, the length of each lexeme, and the second lexeme’s morphological family size. All frequencies were obtained from the WFG (Zeno et al., 1995), using the U-frequency measure. This is a measure of the number of times a word appears in the database per million words. The second lexeme’s morphological family size was determined by counting the number of other bilexemic compounds in the WFG which shared the same second lexeme as the target word of interest. Independent samples t-tests revealed no significant differences on any of these variables (all $p > .25$, see Table 1 for means). Additionally, effort was made to select target words from different morphological families: each target word in the small family group had a unique first lexeme, and no more than two target words in the large family group shared the same first lexeme.

Another variable taken into consideration was the number of morphological family members which are of higher frequency than the target word. Because the target words from large morphological families by definition had many more family members than the target words from small morphological families, and because
compounds from large morphological families tended to be of higher overall word frequency than compounds from small morphological families, it was not possible to match the two groups on number of higher-frequency family members (hereafter referred to as HFFM) while also matching for overall word frequency. The large family group on average had more HFFM than the small family group \( t(52)=5.53, p<.001 \). However, target words were chosen which were the most frequent or close to the most frequent in their families, so that all target words had four or fewer HFFM.

**Monomorphemic words**

In addition to the compound words, 54 monomorphemic filler words were selected to mask the purpose of the experiments. These monomorphemic filler words were selected from the WFG (Zeno et al., 1995) and were matched to the target compound words on word length and overall word frequency (see Table 1).
Experiment 1: Lexical Decision

Method

Participants. Twenty-eight individuals from the Wesleyan University community were compensated $7 for their participation. All participants were native speakers of English and were not told about the purpose of the experiment prior to participating.

Apparatus. Stimuli were presented on a computer screen in black on a white background, using E-Prime software (Psychology Software Tools, Inc). Lexical decision times and error rates were recorded from the keyboard; the ‘z’ key (labeled ‘W’) signified word responses and the ‘m’ key (labeled ‘NW’) signified nonword responses. Participants sat a standard distance from the computer screen, as if for normal computer use.

Materials. In addition to the 27 compound words from large families, the 27 compound words from small families, and the 54 filler words described in the Stimuli Selection and Normative Studies section, 108 nonwords were created. All nonwords were created by altering one or two characters of an existing English word that was not one of the 108 words used in the experiment. Nonwords were orthographically legal, and were matched to target and filler words on word length. Three types of nonwords were created: 27 nonwords were created by changing the first lexeme of an existing compound word; 27 were created by changing the second lexeme of an existing compound word; and 54 were created by changing an existing monomorphemic word. This combination of compound and monomorphemic nonwords was intended to prevent decision making strategies.
The experiment therefore consisted of a total of 216 letter strings, 108 words and 108 nonwords: 27 compound words from large families (e.g. *airline*), 27 compound words from small families (e.g. *wallpaper*), 54 monomorphemic words (e.g. *scoundrel*), 27 first-lexeme-altered compound nonwords (e.g. *paspetball*), 27 second-lexeme-altered compound nonwords (e.g. *pocketgrife*), and 54 monomorphemic nonwords (e.g. *cabinet*).

*Procedure.* Participants were tested individually. They were told that a 7-11 character letter string would appear on the screen, and that they should press the ‘m’ key (labeled ‘W’) if they thought it was an acceptable English word, and the ‘z’ key (labeled ‘NW’) if it was not an acceptable English word. Speed of response was stressed. The session began with 12 practice trials, and then segued into the experimental section consisting of the 216 stimuli described above. Stimuli were presented in a randomized order. Each trial began with a fixation cross for 1000 ms, followed by the stimulus (word or nonword). Stimuli were presented in the center of the screen in Courier New 18pt font, bold, in lower-case. Upon pressing a key (indicating that a decision had been made), the stimulus disappeared and the fixation cross for the next letter-string appeared. Lexical decision time was measured as the time between the presentation of the stimulus and the pressing of the key and was recorded using E-Prime software.

*Design and Data Analysis*

ANOVA analyses were used to compare reaction times and error rates for the three word conditions (compounds from large families, compounds from small families,
and monomorphemes), followed by t-tests to compare each group to one another. Error variance was computed by participants (F1) and items (F2). Paired samples t-tests were used in the by participants analysis (t1), and independent samples t-tests were used in the by items analysis (t2). Additionally, an ANCOVA was used to determine the influence of the number of higher frequency family members (HFFM). Similar statistics were computed for the nonword stimuli.

Results

Mean response latencies and error rates for the three word conditions (compounds from large families, compounds from small families, and monomorphemes) are shown in Table 2. Incorrect responses were excluded from the analysis of response latencies.

An ANOVA revealed a significant difference among the three conditions [F1(2, 54)=18.02, p<.001; F2(2,105)=4.70, p<.025]. In the items analysis, it was noticed that one item (streambed, from the small family group) led to extremely long response latencies; when this item was deleted, the results remained significant [F2 (2, 104)=8.31, p<.001]. Follow up t-tests revealed a significant 40ms effect of family size, with participants responding faster to compounds in the large family condition than the small family condition [t1(27)=3.92, p<.005; t2(52)=2.07, p<.05]. Participants also responded significantly faster to compounds in the large family condition than to the monomorphemic controls [t1(27)=5.99, p<.001, t2(79)=4.06, p<.001]; response times for the small family condition were not significantly different than for the monomorphemic controls (both ps>.15).
Analysis of error rates also revealed a significant difference among the three word conditions \([F_1(2, 54)=17.98, p<.001; F_2(2,105)=3.12, p<.05]\). Again, when the item *streambed*, which had an extremely high error rate, was deleted in the items analysis, the results remained significant \([F_2(2, 104)=3.21, p<.05]\). Follow up t-tests revealed that participants made significantly fewer errors in the large family condition compared to the small family condition \([t_1(27)=5.60, p<.001; t_2(52)=2.15, p<.05]\) and the monomorphemic controls \([t_1(27)=5.42, p<.001; t_2(79)=4.06, p<.001]\). Differences in error rates in the monomorpheme and small family conditions approached significance by participants \([t_1(27)=1.85, p<.08]\) but not by items \((p>.5)\).

As discussed in the *Stimuli Selection and Normative Studies* section, it was not possible to match the large family and small family conditions on number of higher frequency family members; the large family group had more HFFM than the small family group. For this reason, an ANCOVA was conducted to determine the influence of the number of HFFM on response latencies. Morphological family size remained significant after co-varying out the number of HFFM \([F_2(1,51)=6.35, p<.025]\); the covariate (number of HFFM) was not significant \((p>.15)\).

Analysis of nonwords revealed significant differences among the three conditions (first-lexeme altered compounds, second-lexeme altered compounds, and monomorphemic nonwords), both in response latencies and in the amount of errors \([F_1(2, 54)=5.53, p<.01; F_2(2,105)=5.48, p<.01]\) for reaction times; \([F_1(2, 54)=4.83, p<.025; F_2(2,105)=3.87, p<.025]\) for error rates. For reaction times, follow up t-tests revealed that participants were slower in the second-lexeme altered condition than in both the first-lexeme altered condition \([t_1(27)=2.84, p<.01; t_2(52)=2.98, p<.005]\) and...
the monomorphemic nonword condition \([t_1(27)=1.75, p=.09; t_2(79)=2.88, p<.01]\); participants were significantly faster in the first-lexeme altered condition than in the monomorphemic nonword condition by participants only \([t_1(27)=2.86, p<.01; t_2(79)=.61, p>.5]\). For error rates, follow up t-tests revealed that participants made fewer errors in the first-lexeme altered condition than in both the second-lexeme altered condition \([t_1(27)=2.47, p<.025; t_2(52)=1.77, p=.08]\) and the monomorphemic nonword condition \([t_1(27)= 2.79, p<.025; t_2(79)=2.65, p<.01]\); the 2nd-lexeme altered condition and the monomorphemic nonword conditions were not significantly different from each other (both \(ps>.2\)).

**Discussion**

Morphological family size exerted a significant effect on both lexical decision times and error rates, with participants responding 40ms faster and with fewer errors to compounds from large morphological families than to compounds from small families. These effects were not due to the presence of a greater number of higher frequency family members (HFFM) in the large family group than in the small family group.

These results indicate that compounds from large morphological families are easier to recognize than compounds from small morphological families. There are two important implications of this new evidence that morphological family size influences compound word recognition: first, it confirms that lexical decomposition is taking place, as was previously determined (Andrews et al., 2004; Bertram & Hyönä, 2003; Hyönä & Pollatsek, 1998; Juhasz, et al., 2003; Pollatsek et al., 2000); and
second, it suggests that the first lexeme does play an important role in compound processing. Prior studies, which manipulated the first lexeme frequency, found conflicting results for the role of the first lexeme in compound word processing; perhaps this is because these studies did not take morphological family size into account.

Nonwords were included merely to create the decision task, and analysis of responses to nonwords is not particularly useful to the current study. However, one item of some relevance bears noting: of the nonword conditions, first-lexeme altered compound nonwords had the fastest response times and the fewest errors. This could suggest the importance of the first lexeme in compound word recognition, such as in the sequential search, or “file drawer,” model in which the first lexeme acts as an access code to locate the meaning of the whole word. On the other hand, these results could also be explained by the compositional model, in which access to the first lexeme precedes access to the second lexeme. It is unclear; therefore, whether the faster response times to first-lexeme altered nonwords reflect a special role of the first lexeme, or merely that the first lexeme is accessed before the second lexeme.

While the results of the current study do indicate a significant role of morphological family size in compound word recognition, the lexical decision task alone gives no indication of the time course at which these effects are occurring. Furthermore, the task demands of lexical decision require higher levels of processing than are required for ordinary reading and word recognition: in addition to recognizing the word, the participant must make a word/nonword decision and then execute a motor response. Because of this, lexical decision can magnify effects that
might be difficult to catch using other measures. Now that a significant effect of morphological family size has been found, however, other methods are necessary to better determine at what stage of processing these events occur.

The purpose of the second experiment was to use the naming task to examine the role of morphological family size. The naming task is believed to be sensitive to earlier measures of processing and does not require participants to make a word-nonword decision.
Experiment 2: Naming

Method

Participants. Twenty-two individuals from the Wesleyan University community were compensated $7 or given credit in Psych 105 for their participation. All participants were native speakers of English and were not told about the purpose of the experiment prior to participating.

Apparatus. Apparatus was identical to the apparatus used in Experiment 1, except that a microphone was used to record response latency instead of a keyboard.

Materials. Stimuli consisted of the 27 compound words from large families, the 27 compound words from small families, and the 54 monomorphemic words described in the Stimuli Selection and Normative Studies section and used in Experiment 1. No nonword stimuli were used.

Procedure. Participants were tested individually. They were told that words would appear one at a time on the screen, and that they should read each word out loud into the microphone as soon as they could after it appeared. The session began with 20 practice trials, and then segued into the experimental section consisting of the 108 stimuli described above. Stimuli were presented in a randomized order. Each trial began with a fixation cross for 1000 ms, followed by the stimulus word. Stimuli were presented in the center of the screen in Courier New font, size 18, bold, in lower-case. Upon voice key activation, the stimulus disappeared and was followed by a 1500ms wait period before the next fixation cross. Response latency was measured as the time between the presentation of the stimulus and voice key activation and was recorded.
using E-Prime software. The experimenter was present in the room to record failures to activate the voice key as well as any stuttering or speech errors.

**Design and Data Analysis**

Data analysis for response latencies was identical to that of Experiment 1. There were no nonword analyses and no analyses of error rates. Any trials on which the participants stuttered, mis-triggered the voice key, or misread a word were excluded from analysis.

**Results**

Mean response latencies for the three conditions are shown in Table 3. The data from two participants was excluded from analysis: one participant who mis-triggered the voice key on over 50% of trials, and one who had previously participated in stimuli norming. Of the data from the remaining twenty subjects, trials on which the participant stuttered, mispronounced the word, or failed to activate the voice key were also excluded from analysis, leading to exclusion of approximately 5% of the data.

An ANOVA revealed a significant difference among the three conditions \[F_1(2, 38)=25.3, p<.001; F_2(2, 105)=7.96, p<.025\]. Follow up t-tests revealed a significant 30ms difference between the large and small family conditions, again with participants responding faster to compounds in the large family condition than in the small family condition \[t_1(19)=4.88, p<.001; t_2(52)=3.80, p<.001\]. Participants also responded significantly faster to compounds from the large family group than to
monomorphemic words \([t_1(19)=7.22, p<.001; t_2(79)=3.81, p<.001]\). There was no significant difference in response times to compounds from the small and monomorphemic groups (both \(ps>.10\)).

As in Experiment 1, an ANCOVA was done to determine whether the presence of HFFM could have influenced the results. Again, morphological family size remained significant after co-varying out the number of HFFM \([F_2(1,51)=9.50, p<.005]\); the covariate (number of HFFM) was not significant (\(p>.75\)).

**Discussion**

There was a significant, 30ms effect of morphological family size on naming latencies, with compounds from large families being recognized more quickly than compounds from small morphological families. This replicates the findings of Experiment 1, and suggests that compounds from large families are easier to recognize than compounds from small families.

Unlike Experiment 1, however, the naming task does not include nonwords, and does not require categorizing words and nonwords. It is therefore thought to be more sensitive to early stages of processing than the lexical decision task (Schilling et al., 1998). Additionally, the naming task requires information at the phonological level of processing, which is activated prior to the semantic level. For this reason, meaning may not be accessed during the naming task (Schilling et al., 1998). The presence of an effect in the naming task, therefore, could be an indication that morphological family size influences pre-semantic levels of processing.
Like lexical decision, however, naming is subject to task demands not present in normal reading. The subject must articulate a response, which requires positioning the mouth and generating speech. It could be that the morphological family size has an effect on the time taken to articulate a response, rather than during the phonological or other early processing stages. Because of these task demands, and because the semantic level may not be activated during naming, this task does not very well imitate the processes that occur during normal silent reading, in which the semantic level is activated and no articulated response is required.

The third experiment uses eye-tracking to measure the effects of morphological family size. Eye-tracking more closely imitates the natural reading process, and does not require the participant to make word-nonword decisions or to articulate responses. Furthermore, eye-tracking provides measures of both early and late processing, and therefore provides more information about time course than the lexical decision and naming tasks.
Experiment 3: Sentence Reading

Method

Participants. Forty-six Wesleyan University students were given credit in Psych 105 for their participation. All participants had normal or corrected-to-normal vision, were native speakers of English, and were not told about the purpose of the experiment prior to participating.

Apparatus. Eye position was recorded every millisecond by an Eyelink 1000 (SR Research, LTD) eye-tracker. Head movement was restrained by a chin and headrest. The distance between the eye and the computer screen (a 20-inch ViewSonic CRT monitor) was 30cm, at which distance 5 characters of text subtended 1 degree of visual angle. Viewing was binocular, but data was only recorded from the right eye. Sentences were presented on the screen in a single line of text near the center of the screen in black on a white background in Courier New 14pt font.

Materials. The 27 compounds from large families and 27 compounds from small families used in Experiments 1 and 2 were embedded in sentences. All sentences were less than 80 characters long. Each compound from a large family was paired with a compound from a small family, and matched sentences were written for each pair so that, aside from the target word, the sentence frames were identical up to and including the post-target word. Two sentence frames were written for each pair, so that each participant could see all 54 target words without seeing the same sentence frame twice (see Appendix).

The sentences were written so that the context preceding the target word was neutral, and predictability was assessed as follows: eight participants who did not
participate in the eye-tracking portion of the experiment completed a cloze norming task, in which they were given the sentences up to but not including the target word and were asked to generate a word which could serve as the next word in the sentence. No participants generated any of the entire target compounds, and no more than one participant per compound ever generated one of the lexemes of that compound. How well the target words fit into the sentence frames was assessed as follows: for each sentence, eight participants who did not participate in the eye-tracking experiment or the cloze-norming task were given “goodness-of-fit” questionnaires which had the sentence up to and including the target word with the target word in bold. They were asked to rate how well the bolded word fit into the preceding sentence fragment on a scale of 1 to 7, with 1 indicating that the word does not fit at all, and 7 indicating that the word fits very well. With the exception of one sentence, all sentences received average ratings of 4 or better, and a t-test revealed no significant differences between the large family group and the small family group in goodness-of-fit ($p<.75$).²

In addition to the target sentences, 96 filler sentences and 5 practice sentences were created, so that each participant saw a total of 155 sentences. Thirty-eight yes-or-no comprehension questions were also created to follow filler sentences.

Procedure. Participants were tested individually. They were told that they would be asked to read sentences on a computer screen while an eye-tracker recorded their eye movements. After adjusting the chin and headrest, a calibration was performed. If average error was less than .4 degrees of visual angle and maximum error was less than .5 degrees of visual angle, the experimenter continued on to the
experiment. After calibration, the participant was asked to fixate a black box on the left side of the screen. Fixation of this box caused the sentence to appear. Participants were instructed to read the sentence normally, and when finished reading, to look off the screen to the right and press a button. Pressing the button caused the sentence to disappear and the black fixation box on the left side of the screen to reappear. This fixation box also served as a calibration check: if at any point fixation of the box failed to trigger the sentence to appear, the experimenter performed a re-calibration. After 25% of the sentences, participants were presented with a yes-or-no comprehension question about the sentence they had just read. One button on the gamepad indicated ‘yes’ responses and one indicated ‘no’ responses. After the five practice sentences, the target and filler sentences were presented in a randomized order. The target word of the pair that was shown in each sentence frame was counter-balanced, with the constraint that all participants saw all 54 target words once without repeating sentence frames.

Design and Data Analysis

The data from any participant who had a track loss in the target region (the pre-target word, target word, and post-target word) on more than 20% of trials was excluded from analysis. Of the data from the remaining subjects, all trials on which there was a track loss in the target region were also excluded from analysis.

Six measures were recorded and analyzed: the first fixation duration, second fixation duration, gaze duration, number of fixations, probability of refixation, and total time. First fixation duration is length of the initial fixation on the word during
the first reading of the word. Similarly, second fixation duration is the length of the second fixation on a word, if one occurs during the first time the word is read. Gaze duration is the cumulative time spent fixating the word before moving the eyes forward or backward to another word. Number of fixations is the number of fixations made on a word before moving the eyes to another word. Probability of refixation is very similar to number of fixations and is the percentage of times a word was refixated before the eyes were moved to another word. Total time is the cumulative time spent fixating the word, including re-readings of the word. In all of these analyses, the space preceding a target word was considered part of the target word.

Paired sample t-tests were used to compare responses to compounds from the large and small family groups. Again, analyses were computed by subjects ($t_1$) and by items ($t_2$). Repeated measures ANCOVAs were used to determine the influence of the number of HFFM.

Results

Of the 46 participants, the data from twelve participants was discarded because there were track losses on the target region on over 20% of the trials. Data from an additional 4 participants was excluded for a variety of reasons, leading to the analysis of data from 30 subjects. Of the data from these subjects, any trial on which there was a track loss on the target region (the pre-target, target, and post-target words) was discarded. Mean responses to words from small and large families for first fixation duration, second fixation duration, gaze duration, number of fixations, probability of refixation, and total time are shown in Table 4.
Main findings. Paired sample t-tests did not reveal any significant effects of morphological family size on first or second fixation durations (all $p$s > .3). There was a significant 17ms effect on gaze duration, with longer gaze durations for words from small families than for large families [$t_1(29) = 3.80, p < .005; t_2(53) = 2.03, p < .05$]. There was also an effect on the number of fixations, with compounds from small families receiving more fixations than compounds from large families. This effect was significant by participants but not by items [$t_1(29) = 2.21, p < .05; t_2(53) = 1.47, p = .15$]. Additionally, compounds from small families had a significantly higher probability of refixation than compounds from large families [$t_1(29) = 2.78, p < .01; t_2(53) = 1.99, p = .05$]. There was no significant effect of morphological family size on total time (both $p$s > .15).

ANCOVA results. The effect of family size on gaze duration and on the probability of refixation remained significant after co-varying out the number of HFFM using a repeated measures ANCOVA [$F_2(1,52) = 9.90, p < .005$ for gaze duration; $F(1,52) = 7.89, p < .01$ for probability of refixation]. A significant effect on the number of fixations emerged by items [$F_2(1,52) = 4.19, p < .05$], though there was a marginally significant effect of the covariate (HFFM) [$F_2(1,52) = 2.90, p < .10$]. A significant effect on total time also emerged after co-varying out HFFM [$F_2(1,52) = 6.33, p < .025$] (see Footnote 4). There remained no significant effect of morphological family size on either first or second fixation durations (both $p$s > .20), though there was a marginally significant effect of the covariate (the number of HFFM) for first fixation duration [$F_2(1,52) = 3.30, p = .075$].
Discussion

Experiment 3 again revealed a significant effect of family size, with compounds from large morphological families receiving shorter gaze durations, a smaller likelihood of being refixated, and fewer fixations than compounds from small morphological families. When the influence of higher frequency family members (HFFM) was co-varied out, words from large families also received significantly shorter total viewing times than compounds from small families. This replicates the results obtained in Experiments 1 and 2 that compounds from large families are processed more easily than are compounds from small families. It also confirms that the results of Experiments 1 and 2 were not merely due to task demands such as the requirement to make a decision or to enact a verbal or motor response: Compounds from large families are easier to recognize than compounds from small families, even in a more natural reading context. However, the absence of a family size effect in the first fixation durations indicates that the family size variable is likely to influence later, rather than earlier, stages of processing.

There were no significant differences in first or second fixation duration. However, there were significant differences in the number of fixations and probability of refixation. This provides an explanation for the difference in gaze durations and total viewing times for compounds from large and small families: Compounds from small families are more likely to require multiple fixations in order to process, before the reader continues on to the next word.

The results of the current experiment also indicate that HFFM could be playing a role in compound recognition. In several of the measures, there were
significant interactions between the covariate (HFFM) and family size, and in two of the measures there was a marginally significant effect of HFFM. Because of the small range of HFFM in the stimuli for the present experiment (all compounds had fewer than four HFFM), it was not possible to investigate either the interaction between HFFM and family size or the main effect of HFFM further. However, an additional study should be conducted to investigate the role of the number of HFFM on compound recognition and the potential interaction between number of HFFM and family size.
General Discussion

The current study used lexical decision, naming and eye-tracking to determine the effect of having a large or small morphological family size on compound word recognition. The goals were to determine whether the effect of morphological family size found in prior studies of complex words can be extended to compound words specifically, and to use eye-tracking to determine at what stage of processing this effect occurs.

All three experiments revealed a strong effect of family size, with compounds from large morphological families receiving shorter lexical decision times, fewer lexical decision errors, shorter naming times, and shorter gaze durations than compounds from small morphological families. This suggests that compound words from large families are easier to recognize than compounds from small families. In the eye-tracking study, a family size effect was found in the gaze durations, in the number of fixations, and in the likelihood of refixation, but not in the first or second fixation durations. This implies that the difference in gaze durations between compounds from small and large families is due to readers’ greater likelihood of making refixations on words from small morphological families.

These results extend the findings of Bertram et al. (2000) and de Jong, Schreuder, and Baayen (2000), who found morphological family size effects in Dutch complex words, to English words as well. Furthermore, the current results demonstrate a morphological family size effect specifically for compound words. This confirms that morphological decomposition is taking place, and that the first lexeme plays an important role in compound recognition.
Prior studies manipulated first lexeme frequency and found conflicting results regarding the presence and time course of first lexeme effects (Andrews et al. 2004; Bertram and Hyönä 2003; Hyönä and Pollatsek 1998; Juhasz et al. 2003; Juhasz et al. 2008). Since morphological family size and first-lexeme frequency are highly correlated, these conflicting results might be explained by taking into account morphological family size. According to the current study, family size influences later processing measures, while frequency is known to influence early processing measures. Perhaps the studies that found early effects of first lexeme frequency (Bertram and Hyönä, 2002; Hyönä and Pollatsek, 1998; Juhasz, 2008) were indeed observing first lexeme frequency effects. In the studies that found no effects or later effects of first lexeme frequency, it is possible that early first lexeme frequency effects were masked by effects of morphological family size (Andrews et al., 2004; Juhasz et al., 2003).

The existence of a family size effect suggests that a word’s family members influence the recognition of that word. This supports a combination of full-form and decomposed storage in the mental lexicon. A strict decomposed storage model is not accurate, because family size could not influence word recognition unless the family members have some representation in their whole word forms. At the same time, however, the word must be decomposed in order for the reader to separate the first lexeme from the rest of the word and therefore access the word’s family at all.

In Experiment 3, effects of morphological family size were observed for gaze duration and the probability of refixation, which are measures of later processing. No significant effects were found on first fixation duration, an earlier measure. This
suggests that morphological family size influences the semantic stage of processing, the later stage at which meaning is accessed, and does not act at the earlier, lexical level at which structural information about the word is accessed. This is consistent with earlier studies that found morphological family size to be a semantic variable (Baayen et al., 1997; Bertram et al., 2000; de Jong et al., 2000; de Jong, Schreuder, & Baayen, 2003; Moscoso del Prado Martin et al., 2004; Moscoso del Prado Martin et al., 2005; Schreuder and Baayen, 1997). Since there were no significant findings in the first fixation durations, the current results do not provide evidence that the first lexeme acts as an access code through which we locate the rest of the compound. If the first lexeme does act as an access code, this process is not influenced by morphological family size.

The conclusion that morphological family size is a semantic variable, however, is somewhat inconsistent with the results from Experiment 2. The naming task is thought to be sensitive to early stages of processing, since it is not necessary to access word meaning in order to access the pronunciation of that word. Therefore, the presence of a strong effect of morphological family size in the naming task suggests that it affects the lexical stage of processing. However, as mentioned in the Discussion section of Experiment 3, the naming task is subject to further task demands beyond ordinary reading, such as the need to coordinate the muscles to articulate a response. It is possible, therefore, that these additional task demands caused an effect of morphological family size to occur in the naming task even though morphological family size is, in fact, a semantic variable.
Most models that have attempted to explain the morphological family size effect have suggested that it arises due to spreading activation in the mental lexicon (Baayen et al., 1997; Bertram et al., 2000; de Jong et al., 2000; de Jong et al., 2003; Schreuder and Baayen, 1997). Reading a word activates the ‘access representation’ of that word, which is based on its structural properties. Activation then spreads from the access representation to its semantic and syntactic representations (Schreuder and Baayen, 1997). For example, when the access representation for the word *greatness* is activated, it activates the semantic representation for *great*, as well as the syntactic representation *noun*. Semantic representations are shared by semantically related words; for example, the semantic representation *great* is shared by the words *great, greatness,* and *greater* (de Jong et al., 2003). Activation then spreads back from the semantic representation (*great*) to the words that share this semantic representation (*greatness, greater, great,* etc). In this way, a word’s morphological family members become activated during recognition of that word. When a word with a large morphological family is read, a large number of family members become co-activated; this increased activation in the mental lexicon speeds recognition of the target word. Words from small families co-activate fewer family members and therefore take longer to recognize.

This explanation is in line with the current findings that morphological family size influences the semantic level of processing. However, the idea that activation spreads to semantically related family members may be less applicable when specifically studying compound words. Unlike family members that are derived or inflected from a shared base word (e.g. *greatness, greater, greatly, greatest*),
compounds are often quite semantically different from their family members, and thus might not share semantic representations. The compounds *footnote* and *foothold*, for example, are members of the same morphological family that have very different meanings. In English, compounds are especially likely to be semantically unrelated when the morphological family is defined by the first lexeme, as was done in the current study, because the semantic head of English compounds is usually the second lexeme.

How can the theory that activation spreads from a word to its semantically related family members be reconciled with the fact that compound words in the same family are often not semantically related? One potential explanation is that activation only spreads to transparently related family members. Compounds from large morphological families inherently have more transparent family members than compounds from small morphological families. Perhaps it is only the transparent family members that are driving the family size effect found in the current study. Schreuder and Baayen (1997) and Bertram et al. (2000) found that removing opaque words from the family size count increased the effect of morphological family size on lexical decision latencies; this supports the idea that transparent family members drive the family size effect. Similarly, Moscoso del Prado Martin et al. (2004), and Moscoso del Prado Martin et al. (2005) determined that the morphological family size effect is driven only by the transparent family members in Finnish and Hebrew, and that the semantically unrelated family members had either no effect or had an inhibitory effect on reaction times.
If only the transparent family members drive the family size effect in compounds, then no effect of family size should be found for opaque compounds. Due to the number of target words in the current study, it was not possible to conduct a post-hoc test to determine whether this was the case: examining the transparent and opaque family members separately led to sample sizes that were too small to obtain significant results. Future research is needed, therefore; both to determine whether family size effects exist for opaque compounds, and also to determine whether semantically opaque compound family members contribute to the family size effect.

It is also possible that the meanings of the individual lexemes are activated at some point during compound recognition, even in opaque compounds. If this is the case, it could explain the activation of semantically unrelated family members. For example, if the meaning of the lexeme *foot* is activated when the partially opaque compound *footnote* is read, then the word *foothold* and other semantically unrelated family members would become activated despite the differences in their overall meanings.

If the meanings of the individual lexemes are activated during recognition, opaque compounds should take longer to recognize than transparent compounds. Prior studies of transparency revealed conflicting results as to whether this occurs. Pollatsek and Hyönä (2005) and Frisson et al. (2008) found no effects of transparency, and concluded that lexeme meaning is not accessed during compound recognition. Juhasz (2007), however, found that opaque compounds received longer gaze durations than transparent compounds, which suggests that lexeme meanings are accessed at the semantic level of processing. Furthermore, Inhoff et al. (2008)
manipulated lexeme frequency and found greater effects of lexeme frequency when the meaning-dominant lexeme (semantic head) was manipulated. Like Juhasz’s results, the findings of Inhoff et al. suggest that the meanings of individual lexemes are activated. Unlike Juhasz (2007), however, Inhoff et al. (2008) found effects in the first fixation durations. Again, further research is needed to clarify whether or not the meanings of the independent lexemes are activated during recognition of compound words.

There is a third potential explanation for how activation could spread among semantically unrelated compound family members. Perhaps the morphological context of a compound word enables other compounds to become activated as well, even if they are not semantically related to the target word. Studies by Bertram et al. (2000), de Jong et al. (2000), and de Jong et al. (2003) suggest that the family size effect is mediated by a word’s morphological context. When a Dutch adjective with the suffix -heid (‘-ness’) is presented, adjectives in the morphological family that are incompatible with -heid, such as zeegroen (‘sea green’) and steenkoud (‘ice cold’), are not activated in the lexicon (Bertram et al., 2000). Similarly, de Jong et al. (2003) found that when a Dutch adjective is either followed by the suffix -er (‘-er’) or preceded by the word niet (‘not’), only the adjectival family members contribute to the family size effect. This is because -er and niet provide contextual information that the target word is an adjective, which limits the spread of activation to other adjectives.

Context can increase the amount of activation in the lexicon, as well. De Jong et al. (2000) found that when a Dutch verb is presented without morphological
context (e.g. *kwets*, ‘hurt’), only the nouns in the verb’s morphological family become activated. However, when the same verb is presented with the third person singular suffix -\(t\) (e.g. *kwetst*, ‘hurts’), the verbs in the family also become activated. This suggests that the presence of an obvious verb marker (in this case, the suffix -\(t\)) is necessary in order for activation to spread to the verbs in the family. These three studies (Bertram et al., 2000; de Jong et al., 2000; and de Jong et al., 2003) provide compelling evidence that morphological context, such as the presence or absence of affixes, can influence which family members become activated in the mental lexicon.

This idea helps to explain how activation could spread among semantically unrelated compound family members. When a compound is read, perhaps the presence of two independent lexemes provides the contextual information needed to spread activation to other compound family members. Although compound family members that do not share semantic representations with the target word might not be activated by a monomorphemic target word, contextual cues might enable those family members to become activated by a compound word. For example, the target word *foot* might not activate *footnote*, because the two words do not share a semantic representation. However, when the word *foothold* is presented, perhaps the contextual information that the target is a compound enables activation to spread to *footnote* and other compound members of the *foot* family. This is similar to the findings of de Jong et al. (2000) that a verb without -\(t\) only activates nouns, whereas a verb with -\(t\) activates verbs as well. Clearly, further research is needed to test this hypothesis.

One variable that was not controlled for in the current study was family frequency, which is the summed frequency of all of a word’s family members. Given
the large number of other variables that were controlled for in the current study, it was not possible to control for this variable. Because family frequency is highly correlated with family size, it was also not statistically valid to attempt to co-vary out the effect of family frequency in a post-hoc analysis. Since most prior studies of morphological family size found no effect of family frequency above the effect of family size, controlling for family frequency is probably unnecessary (Baayen et al., 1997; Bertram et al., 2000; de Jong et al., 2000; Schreuder and Baayen, 1997). One study, however, did claim to find stronger effects of family frequency than family size, specifically in compound words (De Jong et al., 2002). For this reason, future studies are needed to confirm that the family size effect for compound words found in the present study is in fact due to family size and not to family frequency.

As mentioned in the *Stimuli Selection and Normative Studies* section, it was not possible to control for the number of higher frequency family members (HFFM) in the current study. An attempt was made, however, to select target words that were either the most frequent or close to the most frequent words in their families (all had fewer than four HFFM). In addition, ANCOVAs were conducted to determine the effect of the number of HFFM in all three experiments. These ANCOVAs revealed no observable effect of HFFM on lexical decision or naming times. However, as mentioned in the *Discussion* section of Experiment 3, several of the ANCOVAs in Experiment 3 revealed either a marginally significant effect of HFFM or a significant interaction of HFFM and family size. This suggests that the number of HFFM could play an important role in compound word recognition.
This potential role of HFFM in compound word recognition is consistent with the idea that a large family size leads to increased activation in the mental lexicon. When the target word is one of the most frequent members of its family, this increased activation speeds the recognition process. If a target word has a number of HFFM, however, it is possible that the activation of a large number of family members that are more frequent than the target word could slow, rather than speed, recognition. It is even possible that having a large number of HFFM could eliminate, or even potentially reverse, the family size effect found in the current study. For this reason, future studies are needed to determine the role that HFFM plays in compound recognition.

The current study is the first to use a multi-methodology approach to study the role of morphological family size, and is the first to use eye-tracking to study this variable. It is also one of the first studies to examine family size specifically in compound recognition. This study reveals a significant effect of morphological family size on compound word recognition, and suggests that this effect occurs either at the semantic stage of processing or at a separate morphological stage. These findings have significant implications for our understanding of the mental lexicon. Namely, they suggest that activation spreads from a target word to its compound family members during reading. Furthermore, this study indicates that further research is needed on the role of opaque family members, the role of morphological context, and the role of HFFM on the family size effect.
References


Appendix

*Target sentences used in this study.* Participants saw one sentence of each pair in a randomized order, with the condition that each participant saw every target word once. (For example, a participant who saw the first sentence of the first pair would see the second sentence of the second pair.) Target words are bolded here for emphasis, but were not bolded in the actual study. An ‘l’ indicates that the target word is from a large morphological family; an ‘s’ indicates that the target word is from a small morphological family; these markings were not present in the actual study.

Yesterday evening I went to the **supermarket** to buy snacks for the party. (l)
Yesterday evening I went to the **boardwalk** to watch the sun set over the ocean. (s)

There’s a proposal to build a **supermarket** next to Barnes and Noble. (l)
There’s a proposal to build a **boardwalk** next to Barnes and Noble. (s)

There are always a lot of people in the **crosswalk** just after classes get out. (l)
There are always a lot of people in the **hallway** just after classes get out. (s)

I saw Justin in the **crosswalk** outside the science center. (l)
I saw Justin in the **hallway** outside the principal’s office. (s)

Yesterday we moved all the **firewood** to make more room in the driveway. (l)
Yesterday we moved all the **livestock** to the other barn. (s)

The boy’s job was to tend to the **firewood** and make sure it was neatly stacked. (l)
The boy’s job was to tend to the **livestock** and make sure they were well fed. (s)

Steph was disappointed that the **airline** didn’t give out snacks on the plane. (l)
Steph was disappointed that the **campsite** didn’t have an outhouse. (s)

I thought that the **airline** would give out snacks on the plane. (l)
I thought that the **campsite** would have an outhouse. (s)

Mark thought that his **headmaster** gave a very moving speech at graduation. (l)
Mark thought that his **classmate** gave a very moving speech at graduation. (s)
I don’t think that my headmaster knows who is responsible for the vandalism. (l)
I don’t think that my classmate knows what the math assignment is. (s)

Jessie has a handmade miniature elephant that was crafted in India. (l)
Jessie has a purebred miniature poodle that she takes to dog shows. (s)

Becca asked for a handmade miniature teapot to add to her china collection. (l)
Becca asked for a purebred miniature poodle so she could take it to dog shows. (s)

Ted started the underground movement to smuggle weapons over the border. (l)
Ted started the nationwide movement to eradicate injustices in the workplace. (s)

Bill pledged to support the underground network of drug dealers. (l)
Bill pledged to support the nationwide network of gay-rights activists. (s)

After it rained on Saturday, the seashore looked particularly beautiful. (l)
After it rained on Saturday, the coastline looked particularly beautiful. (s)

Tidal lagoons help to protect the seashore from further erosion. (l)
Tidal lagoons help to protect the coastline from further erosion. (s)

The man kept the handbook under a pile of papers on the table. (l)
The man kept the shotgun under his coat, hidden from view. (s)

I looked for the handbook in the cabinet but I wasn’t able to find it. (l)
I looked for the shotgun in the cabinet but I wasn’t able to find it. (s)

Benny thinks that sunrise is the most beautiful time of the day. (l)
Benny thinks that nowhere is more beautiful than the Grand Canyon. (s)

It has been said that sunrise is the most beautiful time of day. (l)
It has been said that nowhere is more beautiful than New Zealand. (s)

I thought that the crossword in the New York Times was hard today. (l)
I thought that the wallpaper in the dining room needed to be replaced. (s)

Peter said that the crossword in the New York Times was hard today. (l)
Peter said that the wallpaper in our living room is really ugly. (s)

During the hot summer, the waterfall almost completely dried up. (l)
During the hot summer, the streambed almost completely dried up. (s)

In the spring, the waterfall becomes filled with rushing water. (l)
In the spring, the streambed becomes filled with rushing water. (s)
It seems as if her **foothold** on the slippery rock is pretty shaky. (l)
It seems as if her **viewpoint** on the situation is pretty radical. (s)

We didn’t think that her **foothold** on the rock-climbing wall looked very secure. (l)
We didn’t think that her **viewpoint** would be so radical. (s)

I couldn’t find the **footnote** that would explain the obscure reference. (l)
I couldn’t find the **paintbrush** that I had been using yesterday. (s)

Brittany checked to see that her **footnote** had the proper citation. (l)
Brittany checked to see that her **paintbrush** had the right kind of bristles. (s)

The film company hired a **superstar** for the leading role. (l)
The film company hired a **bodyguard** for the famous actress. (s)

I was able to see the **superstar** through the crowd, but couldn’t get close. (l)
I was able to see the **bodyguard** through the crowd, but couldn’t see Madonna. (s)

It seems that **seafood** is the most expensive item on the menu. (l)
It seems that **whoever** is the last to leave should turn the lights off. (s)

According to Gabe, **seafood** was our best option at this restaurant. (l)
According to Gabe, **whoever** was in the kitchen last night left a mess. (s)

This report describes the **background** and mission of the organization. (l)
This report describes the **wholesale** and retail prices over the past six months. (s)

I wanted more information on the **background** and mission of the organization. (l)
I wanted more information on the **wholesale** and retail prices before deciding. (s)

I didn’t think that the **firefly** would be so difficult to catch. (l)
I didn’t think that the **yardstick** would work so I asked for a measuring tape. (s)

Ken wanted to play with the **firefly** but Betty had already let it fly away. (l)
Ken wanted to play with the **yardstick** but his mother wouldn’t let him. (s)

It is difficult to see **underwater** in the murky pond. (l)
It is difficult to see **oneself** in a non-judgmental manner. (s)

It is hard to see **underwater** in the murky pond. (l)
It is hard to see **oneself** in a non-judgmental manner. (s)

Somehow my brother managed to **overcome** his fear of the dark. (l)
Somehow my brother managed to **sleepwalk** his way downstairs. (s)
I didn’t think she’d be able to **overcome** her fear of snakes without help. (l)
I didn’t think she’d be able to **sleepwalk** her way upstairs and into our bedroom. (s)

The man had very weak **eyesight** and needed a pair of glasses. (l)
The man had very weak **willpower** and rarely finished what he set out to do. (s)

She had very strong **eyesight** and was able to read even the tiniest print. (l)
She had very strong **willpower** and was determined to finish the project. (s)

I heard that the new **airport** they are building will be finished by next August. (l)
I heard that the new **railroad** they are building will avoid the center of town. (s)

Unfortunately, the **airport** is the only way to leave this isolated town. (l)
Unfortunately, the **railroad** is the only way to leave this isolated town. (s)

We didn’t think that his **overnight** visit would stretch out into an entire week. (l)
We didn’t think that his **weekend** visit would stretch out into an entire week. (s)

She packed a small bag for her **overnight** visit to her grandparents’ house. (l)
She packed a small bag for her **weekend** visit to her grandparents’ house. (s)

I was sick of **sunshine** and just wanted it to rain so the crops wouldn’t dry up. (l)
I was sick of **wartime** and just wanted my husband to come home safely. (s)

I read a paper about **sunshine** and UV-B radiation, a good source of vitamin D. (l)
I read a paper about **wartime** and PTSD rates among the children of Uganda. (s)

The man never thought that his **eyelid** would become painfully infected. (l)
The man never thought that his **wrongdoing** would land him in jail. (s)

In my opinion, this **eyelid** could be infected and should be seen by a doctor. (l)
In my opinion, this **wrongdoing** could be harmful to the committee’s reputation. (s)

John said that the **backyard** weeds were very difficult to remove. (l)
John said that the **widespread** weeds are worrying gardeners across North America. (s)

We joined in the **backyard** celebration after the ceremony was over. (l)
We joined in the **widespread** celebration after the Red Sox won the World Series. (s)

Samantha had a **headache** and left work early today. (l)
Samantha had a **carefree** and easygoing attitude about the whole ordeal. (s)

Kim thought that her **headache** and sore throat meant she was getting the flu. (l)
Kim thought that her **carefree** and easygoing attitude helped her to make friends. (s)
Footnotes

1There was one exception to this: the sentences “We didn’t think that her foothold on the rock-climbing wall looked very secure.” and “We didn’t think that her viewpoint would be so radical.” did not have identical post-target words. (Target words bolded for emphasis.)

2The fragment, “Benny thinks that nowhere” received a mean rating of 3.875.

3One participant did not fixate on the last several words for a majority of the sentences, one participant requested to terminate the experiment early, one participant had a maximum error that was larger than .5 degrees of arc, and one participant had participated in the naming task.

4In the repeated measures ANCOVAs, there were significant interactions between number of HFFM and family size for gaze duration \([F_2(1,52)=5.43, p<.025]\), probability of refixation \([F(1,52)=3.77, p=.058]\), and total time \([F_2(1,52)=7.93, p<.01]\). There were no significant effects of the covariate (HFFM) for any of these measures.
Table 1.

*Stimuli Characteristics*

<table>
<thead>
<tr>
<th></th>
<th>Large family</th>
<th>Small family</th>
<th>Mono-morphemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family size</td>
<td>66.74*</td>
<td>3.19*</td>
<td></td>
</tr>
<tr>
<td>Whole word frequency</td>
<td>6.86</td>
<td>5.26</td>
<td>5.81</td>
</tr>
<tr>
<td>Number of HFFM</td>
<td>1.63*</td>
<td>.11*</td>
<td></td>
</tr>
<tr>
<td>1st Lexeme frequency</td>
<td>457.22</td>
<td>471.81</td>
<td></td>
</tr>
<tr>
<td>2nd Lexeme frequency</td>
<td>260.26</td>
<td>284.26</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>6.53</td>
<td>6.54</td>
<td></td>
</tr>
<tr>
<td>Transparency</td>
<td>5.33</td>
<td>5.36</td>
<td></td>
</tr>
<tr>
<td>Whole word length</td>
<td>8.37</td>
<td>8.48</td>
<td>8.41</td>
</tr>
<tr>
<td>1st Lexeme length</td>
<td>3.96</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td>2nd Lexeme length</td>
<td>4.41</td>
<td>4.26</td>
<td></td>
</tr>
<tr>
<td>2nd Lexeme family size</td>
<td>14.26</td>
<td>16.85</td>
<td></td>
</tr>
</tbody>
</table>

Note: Frequencies (WFG) are in words per million; familiarity and transparency are mean rankings on a 7-point scale.

* significant difference ($p<.001$); there were no significant differences for any other variables (all $ps>.25$)
Table 2.

*Participant mean response latencies and error rates for Experiment 1.*

<table>
<thead>
<tr>
<th>Word type</th>
<th>Response time (ms)</th>
<th>% Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large family</td>
<td>623 (88)</td>
<td>1.1 (1.7)</td>
</tr>
<tr>
<td>Small family</td>
<td>664 (112)</td>
<td>6.1 (5.4)</td>
</tr>
<tr>
<td>Monomorpheme</td>
<td>675 (105)</td>
<td>4.2 (3.4)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses represent the standard deviation.
Table 3.

*Participant mean response latencies for Experiment 2.*

<table>
<thead>
<tr>
<th>Word type</th>
<th>Response time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large family</td>
<td>510 (15)</td>
</tr>
<tr>
<td>Small family</td>
<td>540 (17)</td>
</tr>
<tr>
<td>Monomorpheme</td>
<td>549 (16)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses represent the standard deviation.
Table 4.

Participant first fixation durations, second fixation durations, number of fixations, probability of refixation, gaze durations, and total times for Experiment 3.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Large family</th>
<th>Small family</th>
</tr>
</thead>
<tbody>
<tr>
<td>First fixation (ms)</td>
<td>225 (26)</td>
<td>227 (34)</td>
</tr>
<tr>
<td>Second fixation (ms)</td>
<td>186 (36)</td>
<td>187 (47)</td>
</tr>
<tr>
<td>Number of fixations</td>
<td>1.10 (.192)†‡</td>
<td>1.15 (.235)†‡</td>
</tr>
<tr>
<td>Probability of refixating</td>
<td>16.7% (14.4)*</td>
<td>21.1% (11.9)*</td>
</tr>
<tr>
<td>Gaze duration (ms)</td>
<td>263 (41)*</td>
<td>280 (58)*</td>
</tr>
<tr>
<td>Total time (ms)</td>
<td>344 (104)‡</td>
<td>360 (86)‡</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses represent the standard deviation.

* significant difference by participants and items ($p<.05$)

† significant difference by participants ($p<.05$) but not by items

‡ significant difference by items after co-varying out HFFM ($p<.05$)
Acknowledgements

First and foremost, I could never have completed this project without the guidance and support of my advisor, Dr. Barbara Juhasz, who started me on this project when I first joined her lab two-and-a-half years ago and who has helped me along every step of the research process since then. Thank you for reading draft after draft and for patiently answering my innumerable questions. I’ve had a great experience working in your lab, and I’ll miss it.

To all the members of the Eye-Movement and Reading lab, thanks for your company and for being such wonderful people to work with. A special thanks to Joanna Dicke for helping to write sentences and run participants, to Liza Litvina for running participants, and to Kacey Wochna for helping with APA formatting. To my friends and housemates, thanks for supporting me throughout this process. Whether you listened to me kvetch, gave me late-night words of encouragement, or helped me edit, I couldn’t have done it without you.

To my parents: I would not be the person I am without the many blessings you have given me. Thank you for your encouragement and love, always.