Wesleyan University The Honors College

The Role of Hyphenation in English Compound Word Processing

by

Melanie Cherng
Class of 2008

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

Middletown, Connecticut April, 2008

Abstract

The present study examines the role of hyphenation in the processing of English compound words. English compound words that appear in the English language as both hyphenated (eg. grown-up) and unspaced (eg. grownup) were chosen as target stimuli. Half of the compound words chosen appeared more frequently in their hyphenated form while the other half of the compounds appeared more frequently in their unspaced form. The words were embedded in sentences in both their hyphenated and unspaced forms. Participants were instructed to read a randomly selected set of sentences while their eye movements were recorded.

First fixation duration showed a significant interaction between compound type and spatial layout presentation which indicated a significant processing gain for unspaced compound words shown as hyphenated. However, for compound words more familiar as unspaced, gaze durations significantly increased with the addition of the hyphen. There was no significant main effect for compound type or spatial layout presentation in gaze duration, indicating no significant benefit from the transformation of a hyphenated compound into an unspaced compound. The results indicate that while spatial segmentation benefits initial word processing by facilitating lexical decomposition, it has detrimental affects on later word processing when whole word retrieval occurs.

Background

Research on the complex processes involved when a person reads a word or sentence has played a crucial role in revealing the ways in which we process and store word meaning. Investigating this invaluable acquired skill many use to receive information in daily life not only furthers our understanding of word processing and recognition but also on a macro level, skill acquisition itself (Rayner, 1997). In turn, information gathered about the processes of this skill can help improve literacy education by potentially increasing reading efficiency for all and also give educators the knowledge necessary to help those who have difficulty with reading.

In reading research, researchers have approached the question "what processes are occurring during reading?" by investigating how words are represented in the mental lexicon and how these representations are retrieved as a person reads a sentence (Juhasz & Rayner, 2006). Though much research in the past has focused on the processing of monomorphemic words (words with only one unit of meaning), focus has turned in recent years on compound words as items for study. This shift in focus is largely due to the semantic complexity of compound words. A compound word by definition is composed of two joined and independent monomorphemic lexemes (eg. *blueberry*). Each lexeme brings with it its own independent meaning (*blue* and *berry*) in addition to the meaning that is created when the two lexemes are joined together (*blueberry*). Therefore, when a reader is presented with a compound word, there are several meanings that could be retrieved.

Sometimes, the lexemes' independent meanings are relatively the same as the meaning created when they are joined together. For example, transparent compound

words tend to have a whole word meaning nearly identical to the meaning of its lexemes (thus *blueberry* means a *berry* that is *blue*). In the case of opaque compound words, when two lexemes are joined together without a space in between, the meaning formed does not match the independent meanings of its lexemes (ie. *deadline* does not mean a *line* that is *dead*). In addition, the insertion or deletion of a space between the lexemes in a compound word can change the meaning of the word entirely as well (*blueprint* vs. *blue print*).

These layers of meaning present in a single compound open a floodgate of possibilities for lexical representation and processing: is the word *blueberry* represented in the mental lexicon as a whole, such that when we read it we directly access that whole representation, or is it broken down into its constituent lexemes *blue* and *berry* first, before the whole word meaning is ultimately accessed? Could the choice of decomposition or whole word access hinge on particular compound word combinations or specific sentence contexts or could both processes be occurring simultaneously? Researchers began manipulating compound word characteristics in their experiments in response to these questions with hopes of developing a model for word recognition and processing. Over the years, different models have developed, some emphasizing lexical decomposition in word recognition while others emphasizing immediate whole word recognition (Starr & Rayner, 2001). Most recently, researchers have come to a general agreement that both lexical decomposition and whole word retrieval occur simultaneously in what is called a 'dual route' model. Depending on the morphological and orthographic characteristics of a complex word, one route will occur faster than the other. The current controversy

is over whether the beginning or ending lexeme of a compound has greater influence in lexical decomposition (Juhasz, Starr, Inhoff, & Placke, 2003).

By manipulating the characteristics of the constituent lexemes in compound words, researchers have been able to determine whether lexical decomposition or whole word retrieval occurs in word recognition. Research on word recognition in both Finnish and English have found a robust connection between word frequency (the frequency a word appears in the English language) and word processing. The higher the frequency of a word, the shorter the fixation duration on the word and therefore, the quicker the recognition process for the reader (Bertram & Hyönä, 2003; Juhasz et al., 2003). Because word frequency has such a robust effect on reading time, researchers have used this to their benefit by manipulating the lexeme frequency in compound words in order to further understand and distinguish the different levels of word processing that may be occurring.

Researchers have also taken note that when manipulating word frequencies, the age at which a word is acquired (*age-of-acquisition*) (Juhasz et al., 2006) and word length also can influence fixation durations (Bertram et al., 2003). Early acquired words tend to have shorter fixations compared to later acquired words, and longer words have longer fixation durations and a greater number of fixations compared to shorter words. Because these characteristics can also greatly influence processing time, they have been either controlled or manipulated in studies that have also manipulated lexeme frequency.

In the effort to understand word processing and representation different methods have been employed in research. Researchers first began compound word

research by using two main tasks: lexical decisions and word naming. Lexical decisions involved the participants viewing a series of words and non-words and deciding whether the stimulus presented was a word or non-word. The word naming task involved the participant viewing and reading a series of words aloud as quickly and accurately as possible. Afterwards, reaction times and accuracy were used as basis for analyses.

However researchers have found that such tasks produce inconsistent results because they each individually capture only parts of the word recognition process rather than the process in its entirety. By using lexical decision making, naming and eye movement measurement in a three part study on lexical decomposition, Juhasz et al. (2003) were able to confirm that different task demands produced different results; each task produced results that captured only a portion of the word recognition process. Juhasz et al. (2003) found that all three tasks produced a strong ending lexeme effect in later processing. Not only do the results demonstrate that decomposition occurs but also, specifically, that the ending lexeme plays an important role in accessing whole word meaning. While they did find a beginning lexeme effect when the ending lexeme was low-frequency for lexical decision tasks, the role that the beginning lexeme played appeared to be small and the effect only occurred in early word processing. The beginning lexeme effect either became marginal or insignificant once experimental method was taken into account. There was no effect after considering lexical naming and eye movement measurements. This study in particular revealed that lexical decision and naming tasks alone often produces results that do not capture later processing effects. Drawing conclusions

from such results without eye movement research would therefore produce an inaccurate model of word processing.

Before further discussion on eye movement research related to word recognition, it is important to explain the terminology in more detail. Researchers are able to deduce word processing from eye movements by analyzing the number of *fixations* (when the eyes are still) and the time length of fixations (*fixation duration*) a participant makes on a word. The average fixation duration is 200-250 msec and it increases as the text becomes more difficult (Rayner & Pollatsek, 1991)**.** Researchers often measure *first fixation duration* (the first fixation time on the word of interest) and *gaze duration* (the sum total of all fixation time on a word not counting regressions back). It is typical for readers to also return to words they have already fixated in a sentence (*regression*), however, regressions back also increase as sentences become more difficult. When fixating on a word, we are able to see most clearly in the foveal region, the point on which we are fixating and 1 degree of visual angle to the left and right. Information is also collected from the parafoveal region, the region that extends out 5 degrees to right and left of the point of fixation (Rayner & Pollatsek, 1991).

Having found that lexical decision and word naming tasks produce inconsistent findings, researchers have turned to eye movement research as an additional and more precise measurement of word recognition. Lexical decision and word naming tasks place different processing demands on the participant in comparison to normal everyday word processing when reading (Juhasz et al., 2003). In eye movement research, the stimuli are presented in a form in which they are

usually found, in sentences. As a result, the task given most closely mirrors the kinds of reading tasks people do daily. Additionally, the context in which words are presented can also influence word processing and retrieval. Parafoveal information around the point of fixation can provide information of the word that follows before the eyes fixate onward. These integral parts of the word recognition process measurable by eye movement research are not identified by in research with lexical decisions and word naming. Many researchers have come to use either eye movement tracking in addition to lexical decision and word naming in order to capture the full scope of the word recognition process.

In Finnish eye movement studies, researchers manipulated the lexeme frequency of nonspaced compound words and found that the beginning lexeme is accessed first and then followed by whole compound and ending lexeme retrieval. This finding suggested a dual route model for compound word recognition and processing in Finnish (Hyönä & Pollatsek, 1998; Pollatsek, Hyönä, & Bertram, 2000). Researchers have confirmed also through lexeme frequency manipulation that decomposition processing exists for nonspaced compound words in the English language as well (Andrews, Miller, & Rayner, 2004; Juhasz et al., 2003) but have differed on whether the beginning or ending lexeme had a greater frequency effect. Juhasz et al. (2003**)** found a significant ending lexeme frequency effect that may indicate that the ending lexeme plays a greater role in compound word processing, speculating that perhaps this is because in English the meaning of a compound word is often close to the meaning of the ending lexeme, particularly in adjective-noun compound combinations. However, Andrews et al. (2004) countered these results by

manipulating lexeme frequency and finding a significant frequency effect in the beginning lexeme.

This dilemma presents itself not only in Finnish and English but in Spanish and Basque as well (Dunabeitia, Perea, & Carreiras, 2007). Research done by manipulating lexeme frequency for compound words in Spanish and Basque confirmed the decomposition model as did the studies previously mentioned, but found significant ending lexeme frequency effects for both Spanish and Basque. However, while in Spanish, compound word meaning tends to be determined by the ending lexeme, in Basque, compound word meaning tends to be determined by the beginning lexeme. Thus, regardless of where meaning was centered, an ending lexeme effect prevailed. Dunabeitia et al. (2007) suggested that these results indicated that lexical decomposition is perhaps a blind to semantics process. In spite of the controversy over whether the beginning or ending lexeme has a greater role in compound word processing, all of these studies have confirmed through word frequency manipulation that decomposition does in fact occur in word processing.

Not only have researchers manipulated frequency in order to further understand the process of lexical decomposition, researchers have also additionally manipulated the degree of transparency a compound word has as well as the age at which the word was acquired. Manipulating both lexeme frequencies and the transparency of compound Finnish words, researchers found little to no effect of transparency in the processing of compound words (Pollatsek & Hyönä, 2005). However, Juhasz (2007) in a similar study conducted with English compound words did find a main effect of transparency. Furthermore, the study revealed that lexical

decomposition occurred for both transparent and opaque compound words, suggesting that decomposition in early processing has less to do with semantic processing (Juhasz, 2007).

Having found the manipulation of word frequency, age-of-acquisition, word length, and transparency useful in order to determine the pathways of word processing and recognition, researchers have turned to spatial layout in recent years as the next variable of study. As mentioned in a review of literature regarding spatial segmentation for German compound words (Inhoff & Radach, 2002) spatial segmentation in the form of interword spaces can help early word processing by indicating to reader meaning units visually. However, when spaces are inserted in unspaced compound words, readers have longer gaze durations as a result of making misassumptions about where the meaning unit ends. Spatial unity on the other hand, greatly benefits the later stage of word processing by indicating that the lexemes form one meaning unit.

Particularly for the English language, spatial layout is of interest because many English compound words exist in unspaced, spaced and hyphenated forms. In Finnish, a language in which much word recognition and processing research has been done, most compound words are unspaced. Interested in whether adding or deleting a space in a compound word increases or decreases the speed of word processing, Juhasz et al. (2005) manipulated the spacing of spaced and unspaced compound words and found that spatial unification of spaced compound words facilitated reading and did indeed speed up compound word processing (Juhasz, Inhoff, & Rayner, 2005). However, it was unclear in the results whether this

advantage in word processing was because spatial unification helped the reader gain more parafoveal information about the second lexeme or because spatial unification was related to how compound words are represented in the mental lexicon.

Though research has been done on spacing in compound word processing, little to no research has been done on the effect of hyphenation and whether it facilitates word processing. A recent move by dictionaries such as the Oxford English Dictionary towards converting many hyphenated words into either spaced or unspaced compound words presents the question of whether or not hyphenation serves an important purpose in word recognition and whether the move to delete them facilitates or impedes word processing. In this study, hyphenation of two groups of compound words was manipulated: those that generally appear hyphenated and those that generally appear unhyphenated. The eye-movements of participants were recorded in order to see if the deletion or insertion of a hyphen increases or decreases word processing speed.

Methods

Participants. Forty-five undergraduate students at Wesleyan University participated and received introductory psychology course credit or monetary compensation. Of the 45 participants, the data of 13 were discarded because track loss was more than 25%. All participants had normal or corrected-to-normal vision and were native English speakers unaware of the purpose of the study. Informed consent was obtained from participants.

Apparatus. A SRI EyeLink 1000 eye tracker was used to record the participants' eye movements. The eye tracker records eye position every millisecond and was interfaced with a Dell compatible computer. Only the eye movements of the right eye were measured although participants viewed the stimuli with binocular vision. Sentences were presented in Courier New 14 pt font on one line against a white background near the horizontal midline of the screen.

Procedure: Participants were given instructions and asked to sign informed consent forms upon arrival. The chin-rest and mirror of the eye-tracker tower mount were adjusted in accordance to the needs of the participant. Calibration was then carried out and was accepted if the average error was less than .4 degrees of visual angle and the maximum error was less than .5 degrees of visual angle. After calibration, the participant was asked to fixate on a rectangular marker on the left hand side of the screen. A successful fixation triggered the appearance of sentence stimulus. Participants were instructed to read normally for comprehension. After reading the sentence, the participant was asked to fixate to the right off-screen and press a button which replaced the sentence with another left-side marker which

ensured the accuracy of the calibration for each stimulus presented. If at any point the participant could not initiate the appearance of the stimulus by fixating on the left-side marker the participant was instructed to notify the experimenter to conduct recalibration. Sentence comprehension was checked by the random insertion of yes or no questions that referred to the sentence meaning of the previous stimulus. Participants were instructed to press the button that matched the correct answer.

Stimuli. Forty bimorphemic compound words that appear in both hyphenated and unspaced compound word form were selected from the Educator's Word Frequency Guide (Zeno, Ivens, Millard, & Duvvuri, 1995) and matched on first and second lexeme frequency and first and second lexeme length. Table 1 shows average first and second lexeme frequencies and lengths and high and low familiar frequencies for both compounds more familiar as hyphenated and compounds more familiar as unspaced. Half of the target words selected appeared more frequently in the unspaced form in comparison to the hyphenated form and the other half selected appeared more frequently in the hyphenated form in comparison to the unspaced form. A 2 x 2 factorial design was used to determine whether spatial layout type interacted with spatial layout familiarity. There were four conditions in this study: hyphenated type-familiar spatial layout condition, hyphenated type-unfamiliar spatial layout condition, unspaced type–familiar spatial layout condition and unspaced type-unfamiliar spatial layout condition. The target words were embedded in sentence contexts in both their hyphenated and unspaced forms. Each participant was shown only 10 sentences per condition. These sentences were presented in a set with filler

sentences and sentences of another study being conducted simultaneously to ensure that participants were unaware of the study's purpose.

Normative Data. Before the eye tracking experiment was conducted, norming was conducted for goodness of fit and predictability. Sixteen participants were given a goodness of fit questionnaire and asked to assess on a scale of 1 to 7 whether the sentence context suited the target word, 1 being very unsuitable and 7 being very suitable. For all sentence types there was an average rating of 5.7 indicating that all target words were compatible with their respective sentence contexts. Goodness of fit did not significantly vary across conditions ($Fs<.1$). Eleven participants were given cloze questionnaires that presented them with each sentence beginning right up to but not including the target word and were asked to predict the word that would follow immediately after. The percentage of items predicted did not differ significantly across conditions ($ps>.1$).

Results

A 2 (compound type: hyphenated or nonspaced) x 2 (compound presentation: hyphenated or unspaced) ANOVA was conducted. An average of 9.45% of the data was deleted as a result of track loss. Four dependent measures were analyzed: first fixation, gaze duration, landing position and probability of a single fixation. First fixation duration is the amount of time (ms) a participant fixates on the target word for the very first fixation. Gaze duration is total sum of all fixations on the target word before the eyes leave the word. Also measured were probability of a single fixation and initial landing position of first fixation. Probability of a single fixation measures the probability of only one fixation on the target region and initial landing position measures how far the first fixation lands into the target region. Table 2 presents the means for these measures as a function of the four conditions: hyphenated-familiar, hyphenated-unfamiliar, unspaced-familiar, unspaced-unfamiliar.

For first fixation duration, there was no significant main effect for the familiarity of spatial layout presentation (all *p*s>0.05) but a significant main effect for compound type, $F(1,31) = 10.46$, $p < 0.05$. Compounds that are more familiar as hyphenated compounds received 14 ms shorter first fixation durations. The interaction between compound type (hyphenated or unspaced) and spatial layout presentation (familiar or unfamiliar) was significant, $F(1,31) = 7.60$, $p<0.05$. First fixations on compounds that are more familiar as unspaced were 18 ms shorter if they were shown in an unfamiliar spatial layout (hyphenated), $t(31) = 2.40$, $p <0.05$. However, first fixations on compounds that are more familiar as hyphenated were

only 9 ms longer if they were shown in an unfamiliar spatial layout (unspaced), $t(31) = -1.47$, $p > 0.05$.

For gaze duration, neither main effects were significant (all *p*s > 0.05). However there was a significant interaction between compound type and spatial layout presentation, $F(1, 31) = 5.17$, $p<0.05$. If a compound was more familiar as unspaced, gaze durations were 30 ms longer if it was presented in its unfamiliar spatial layout (hyphenated), $t(31) = -2.20$, $p < 0.05$. This pattern was opposite to what was found for first fixation durations. If a compound was more familiar as hyphenated, gaze durations were only 19 ms shorter if it was presented in its unfamiliar spatial layout (unspaced), $t(31) = 1.48$, $p>0.05$, and therefore this difference did not reach significance.

For initial landing position, neither main effects were significant (all *p*s >0.05). Furthermore, there was no significant interaction ($p>0.05$). Whether the compounds, hyphenated or unspaced, were shown in familiar or unfamiliar spatial layouts did not significantly influence where first fixations would land.

Finally, for the probability of a single fixation, neither main effects were significant (all *p*s >0.05). However, there was a significant cross over interaction, $F(1,31) = 23.54$, $p<0.05$. If a compound was more familiar as unspaced a single fixation was 20% more likely if it was shown in its familiar spatial layout (unspaced), $t(31) = 4.13$, $p<0.05$. If a compound was more familiar as hyphenated, a single fixation was 10% more likely if it was shown in its unfamiliar spatial layout (unspaced), $t(31) = -2.77$, $p<0.05$.

Discussion

When analyzing first fixation durations, the most important effect is the significant interaction between compound type and spatial layout presentation. This indicated a significant processing gain for unspaced compound words shown as hyphenated. These findings are similar to previous findings in compound word spacing research. Regardless of correct or incorrect presentation, compound words presented with a space had shorter first fixation durations (Juhasz et al., 2005). This similarity indicates that in early processing measures, spaces and hyphens both provide spatial segmentation information that is helpful in lexical decomposition. Previous research has indicated that lexical decomposition occurs in early word processing (Juhasz et al., 2003). Thus, spatial segmentation of compound words by hyphens benefits the word parsing that must occur in lexical decomposition.

Encompassing all fixations before the eyes fixate past the target region, including refixations, the dependent measure of gaze duration captures the effect of spatial layout on word processing over time. Past spatial layout research has indicated that by including refixations on the target word, gaze durations reveal later word processing (Juhasz et al., 2005). Though spatial segmentation benefits lexical decomposition, an early word processing task, it detracts from whole word recognition, a later word processing task. Spatial unification has previously been found to benefit later whole word recognition. When a compound word is unspaced, the reader is given visual and semantic information that the two lexemes compose a whole meaning unit.

Unlike past spatial layout research in compound word spacing, in this study there was no significant main effect for compound type or spatial layout presentation in gaze duration. Past research with spacing has indicated a significant main effect of compound type (Juhasz et al., 2005). In that particular study, compound words presented as unspaced regardless of correctness had shorter fixation durations. Spatial unification produced eye fixations closer to the center of word, and therefore more information could be collected from fewer fixations. In the present study, the hyphenated and unspaced word type, regardless of presentation, had very small effects in gaze durations. There were also very small effects in gaze durations between familiar and unfamiliar spatial layout presentations.

The absence of significant main effects in gaze duration is unsurprising however in comparison to past research in compound word spacing. Sometimes the insertion or deletion of a space in compound words creates compound words that do not orthographically exist in the English language. However, the hyphenated and unspaced compounds in this present study are used in both their hyphenated or unspaced form, differing only with which format are readers generally more familiar. Furthermore, though the hyphen represents spatial segmentation in some aspects similar to a space, it still functions as a symbolic connector between lexemes both orthographically and semantically. Deleting a space significantly benefited word processing for spaced compound words in Juhasz et al. (2005) because it visually unified two related lexemes and thus indicated a unit of meaning to the reader. Deleting a hyphen was not significantly beneficial in the present study perhaps because a hyphen already signified to the reader the connection between the two

lexemes. Furthermore, a hyphenated compound word is more unified than a space compound word.

While there was no significant benefit of spatial unification, indicated by neither main effects being significant, this was qualified by a significant interaction between compound word type and spatial layout presentation in gaze duration. Again, there was no significant benefit of spatial unification; readers did not significantly benefit from the deletion of the hyphen for compound words more familiar as hyphenated. However, for compound words more familiar as unspaced, gaze durations significantly increased with the addition of the hyphen. Readers incurred significant processing costs when they were more familiar with a compound as more spatially unified (unspaced). As noted in a spatial segmentation literature review (Inhoff et al., 2002) the spatial segmentation in compound words that are more familiar to readers as spatially unified introduces parsing errors to the word recognition process. Readers must correct misassumptions about where units of meaning end and begin through refixations. Though readers may have benefited from spatial segmentation in first fixation durations during early processing, spatial segmentation hinders the whole word retrieval process.

The measure of initial landing position also indicates that the hyphen functions as a spatial unifier. There were no significant main effects or interactions in landing position. When analyzing the probability of a single fixation, the results do show an interaction between compound type and spatial layout presentation. If a compound word was more familiar as unspaced, there was higher probability of a single fixation when it was presented as unspaced. If a compound word was more

familiar as hyphenated, there was a higher probability of a single fixation also when it was it was presented as unspaced. Only these findings support that spatial unification (the deletion of the hyphen) does benefit word the recognition process. If there is a high probability of a single fixation for a target region, it indicates that the reader was more likely able to correctly collect enough information to retrieve the whole word meaning from one fixation. A low probability would indicate multiple fixations were required in order to retrieve the word meaning.

Though evidence has suggested that spatial unification is beneficial to word processing speed in previous compound word spacing studies, this study reveals that spatial unification by deleting the hyphen, does not significantly benefit word processing speed. The study was able to replicate past findings that spatial segmentation helps facilitate lexical decomposition in early word processing. However, it appears that because hyphens function also as symbolic connections and because the compounds used in this study appear in the English language as both hyphenated and unspaced, the deleting the hyphen did not increase reading speed significantly. It should be noted that the insertion of a hyphen in compound words more familiar as unspaced incurred significant processing costs. The findings become particularly important when noting the recent move by the Oxford English Dictionary in deleting the hyphen in many hyphen words and transforming them into unspaced and spaced words. This study has shown that the transformation of a hyphenated compound word into an unspaced compound posed no threat to word processing, though it also did not have significant improvement upon word processing.

Additional research on the impact replacing a hyphen with a space should be done particularly as past research indicates moving in the direction of further spatial segmentation has proven to decrease word processing speed (Juhasz et al., 2005). If complete compound spatial unification is an unspaced spatial layout and complete compound spatial segmentation is spaced spatial layout, then where does hyphenated spatial layout fall on the continuum of spatial unification to segmentation? The fact that there was no significant processing gain from transforming a hyphenated compound to an unspaced compound seems to indicate that a hyphen can spatially unify two lexemes as efficiently as a space deletion. If past compound word spacing studies have found that word processing is slower when a spatially unified compound is converted to a spatial segmented compound with the addition of a space (Juhasz et al., 2005), then it is likely that inserting a space to replace the hyphen will be detrimental to word processing speed. More interestingly though is the question of whether inserting a space in the place of a hyphen in a hyphenated compound word has as much as an effect on word processing speed as inserting a space in an unspaced compound word. Such a comparison could reveal the degree to which a hyphen functions as a spatial unifier and offer greater insight on the effects of spatial layout on word processing.

References

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology, 16*(1/2), 285-311.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morpholofical structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*(48), 615-634.

Dunabeitia, J. A., Perea, M., & Carreiras, M. (2007). The role of the frequency of constituents in compound words: evidence from Basque and Spanish. *Psychonomic Bulletin Review, 14*(6), 1171-1176.

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1612-1627.

Inhoff, A. W., & Radach, R. (2002). The biology of reading: the use of spatial segmentation in the reading of complex words. In *Comments on Modern Biology: Part C. Comments on Theoretical Biology. 7,* (pp. 121-138).

Juhasz, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. P. Van Gompel, M. H. Fischer & W. S. Murray (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 373-390): Elsevier.

Juhasz, B. J., Inhoff, A. W., & Rayner, K. (2005). The role of interword spaces in the processing of English compound words. *Language and Cognitive Processes. Vol, 20*(1), 291-316.

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition* (Vol. 13, pp. 846-863): Psychology Press.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*(94), 223-244.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*(20), 261-290.

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 820-833.

Rayner, K. (1997). Understanding eye movements in reading. *Scientific Studies of Reading Vol 1(4) 1997, 317-339.*

Rayner, K., & Pollatsek, A. (1991). Reading Processes and Comprehension. In Encyclopedia of Human Biology In (Vol. 6, pp. 481-492): Academic Press, Inc. .

Starr, M. S., & Rayner, K. (2001). Eye movements during reading: some current controversies. *TRENDS in Cognitive Sciences, 5*(4), 156-163.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

Acknowledgements

This research was carried out by Melanie Cherng with the mentorship of Barbara Juhasz at Wesleyan University. I appreciate deeply the guidance and encouragement Professor Juhasz has given me in the two years I have done research in her lab. This process has been an incredible learning experience. I would also like to thank Margaret Gullick and Rachel Berkowitz for their aid and support at various stages of the research process. Without their help, this study could not have been successfully completed.

I send my love and deep appreciation to my parents for the blessing of my education here at Wesleyan and their continual spiritual, emotional and intellectual support in all my endeavors. Finally, without the strong, loving support network of the dear friends I have discovered here, I would not have survived. My loves: Maya, Diana and Katrina; my dance community and the other kindred spirits in my world: thank you for being the gems in my life.

You remind me to be exuberant always. I am blessed.

TABLE 1: Average first lexeme frequency (1st LF), first lexeme length (1st LL), second lexeme frequency (2nd LF), second lexeme length (2nd LL), high frequency (HF), and low frequency (LF) as a function of compound type.

Type	*Example*	*1st LF*	*1st LL*	*2nd LF*	*2nd LL*	*HF*	*LF*
Hyphenated	grown-up	413 (403)	4.20 (.95)	868 (982)	3.65 (1.18)	.91 (1.55)	.21 (.66)
Unspaced	grownup	478 (772)	4.20 (.77)	901 (986)	3.65 (1.18)	1.29 (1.73)	.12 (.23)

Note: Freq = average written frequency per million (Zeno et al., 1995).

TABLE 2: First fixation duration (FF), gaze duration (GD), the probability of a single fixation (prob.) and the landing position of the first fixation (FFP) as a function of compound type and spatial layout presentation.

Type	*Spatial Layout Presentation*	*Example*	*FF*	*GD*	*Prob*	*FFP.*
Hyphenated	*Familiar*	grown-up	228 (39)	317 (74)	52 (18)	3.18 (.94)
Hyphenated	*Unfamiliar*	grownup	237 (50)	298 (77)	62 (16)	3.03 (.94)
Unspaced	*Familiar*	airlift	256 (48)	305 (59)	66 (19)	2.99 (.82)
Unspaced	*Unfamiliar*	air-lift	238 (43)	335 (78)	46 (23)	2.95 (.87)

Note: Fixation durations are in milliseconds and numbers in parentheses are standard deviations.

Appendix

Sentences used in this study. The first sentence in each pair contains the compound in hyphenated form and the second sentence contains the same compound in unspaced form. The first set of sentences contain target compound words that are more familiar as unspaced and the second set of sentences contain target compound words that are more familiar as hyphenated. Target words are presented in bold.

Sentences Containing Compounds more Familiar as Unspaced

My birthday cake was almost an **after-thought**, since I focused on dinner first.

My birthday cake was almost an **afterthought**, since I focused on dinner first.

They had to call a helicopter to **air-lift** her to the nearest hospital.

They had to call a helicopter to **airlift** her to the nearest hospital.

Through the comic series, Superman's **arch-enemy** is Lex Luthor.

Through the comic series, Superman's **archenemy** is Lex Luthor.

He threaded the flower through the **button-hole** on his jacket.

He threaded the flower through the **button-hole** on his jacket.

Anna headed to the dentist for her **check-up** and regular tooth-cleaning.

Anna headed to the dentist for her **checkup** and regular tooth-cleaning.

My friend's hands are almost **child-like** because his fingers are so short.

My friend's hands are almost **childlike** because his fingers are so short.

My mom showed me her old 80's **get-up** and clothes, which look rather silly now.

My mom showed me her old 80's **getup** and clothes, which look rather silly now.

The airplane ride left the **heavy-set** man feeling squished and claustrophobic.

The airplane ride left the **heavyset** man feeling squished and claustrophobic.

We couldn't figure out what the **hold-up** on the highway was until we were past.

We couldn't figure out what the **holdup** on the highway was until we were past.

The children played **keep-away** in the yard during recess.

The children played **keepaway** in the yard during recess.

A football game's **kick-off** is one of the few parts I understand.

A football game's **kickoff** is one of the few parts I understand.

The painting was incredibly **life-like** and captured the subject's personality.

The painting was incredibly **lifelike** and captured the subject's personality.

I rarely buy any kind of **make-up** since I only use it on special occasions.

I rarely buy any kind of **makeup** since I only use it on special occasions.

The union workers had a **nation-wide** strike until their demands were met.

The union workers had a **nationwide** strike until their demands were met.

We had to count carefully to play the **off-beat** rhythms correctly.

We had to count carefully to play the **offbeat** rhythms correctly.

Liz was insulted by my **off-hand** comment about her outfit.

Liz was insulted by my **offhand** comment about her outfit.

The legislature started instigating **state-wide** changes in property taxes.

The legislature started instigating **statewide** changes in property taxes.

Though the cat and dog played dumb, the **tell-tale** signs of a fight were there.

Though the cat and dog played dumb, the **telltale** signs of a fight were there.

When I get home late, I **tip-toe** through my apartment so I don't wake anyone.

When I get home late, I **tiptoe** through my apartment so I don't wake anyone.

The shape of the sine curve **wave-form** is my favorite of the trig functions.

The shape of the sine curve **waveform** is my favorite of the trig functions.

Sentences Containing Compounds more Familiar as Hyphenated

Due to the drought, there was a **city-wide** ban on lawn sprinklers and hose use.

Due to the drought, there was a **citywide** ban on lawn sprinklers and hose use.

The deal seemed straightforward and **clear-cut**, so I was surprised by the catch.

The deal seemed straightforward and **clearcut**, so I was surprised by the catch.

I jumped and had a few moments of **free-fall** before opening the parachute.

I jumped and had a few moments of **freefall** before opening the parachute.

When I was little, I didn't want to be a **grown-up** and have responsibilities.

When I was little, I didn't want to be a **grownup** and have responsibilities.

I only took a quick **half-hour** lunch break because I came in late to work.

I only took a quick **halfhour** lunch break because I came in late to work.

The new gaming system was **hand-held** and came with a stylus pointer.

The new gaming system was **handheld** and came with a stylus pointer.

The new set of **high-rise** apartments was the tallest thing on campus.

The new set of **highrise** apartments was the tallest thing on campus.

Even though the senator is **left-wing** now he once was a moderate conservative.

Even though the senator is **leftwing** now he once was a moderate conservative.

Though his family was in the upper **middle-class**, he was on financial aid.

Though his family was in the upper **middleclass**, he was on financial aid.

Because of a computer error, there was a **mix-up** between two tables' orders.

Because of a computer error, there was a **mixup** between two tables' orders.

The actress went from **one-time** extra to a major TV star in only a few years.

The actress went from **onetime** extra to a major TV star in only a few years.

To help pay for school, I found a **part-time** job during the semester.

To help pay for school, I found a **parttime** job during the semester.

Hilary Clinton's **right-wing** conspiracy theory hasn't resurfaced in awhile.

Hilary Clinton's **rightwing** conspiracy theory hasn't resurfaced in awhile.

The brick building was **run-down** and looked abandoned.

The brick building was **rundown** and looked abandoned.

It is only by having **self-respect** that others will also respect you.

It is only by having **selfrespect** that others will also respect you.

Billy was always a **show-off** and wanted all the attention to himself.

Billy was always a **showoff** and wanted all the attention to himself.

The show was so popular that a **spin-off** was made about the lead actress.

The show was so popular that a **spinoff** was made about the lead actress.

Almost all of the internet **start-up** companies seem to be running out of steam.

Almost all of the internet **startup** companies seem to be running out of steam.

Though the photo was good, it needed a **touch-up** to digitally fix the spots.

Though the photo was good, it needed a **touchup** to digitally fix the spots.

The new professor was one of the most **well-read** people in the Math department.

The new professor was one of the most **wellread** people in the Math department.