



The role of semantic transparency in visual word recognition of compound words: A megastudy approach

Say Young Kim^{1,2} · Melvin J. Yap¹ · Winston D. Goh¹

© Psychonomic Society, Inc. 2018

Abstract

Previous studies on visual word recognition of compound words have provided evidence for the influence of lexical properties (e.g., length, frequency) and semantic transparency (the degree of relatedness in meaning between a compound word and its constituents) in morphological processing (e.g., to what extent is *doorbell* influenced by *door* and *bell*?). However, a number of questions in this domain, which are difficult to address with the available methodological resources, are still unresolved. We collected semantic transparency scores for 2,861 compound words at the constituent level (i.e., how strongly the overall meaning of a compound word is related to that of each constituent) and analyzed their effects on speeded pronunciation and lexical decision performance for the compound words using the English Lexicon Project (<http://elexicon.wustl.edu>) data. The results from both tasks indicated that our human-judged semantic transparency ratings for both the first and second constituents play a significant role in compound word processing. Moreover, additional analyses indicated that the human-judged semantic transparency scores at the constituent level accounted for more variance in compound word recognition performance than did either whole-word semantic transparency scores or corpus-based semantic distance scores.

Keywords Visual word recognition · Compound word · Megastudy · Semantic transparency

Compounding is the most universal means of forming morphologically complex words across languages (Dressler, 2006), and typically two (or more) lexemes are involved (e.g., *door* and *bell* for *doorbell*). Thus, the meaning of a compound word may be related to the meaning of each lexeme for transparent compounds such as *doorbell*, but not for opaque compounds such as *honeymoon*. Previous studies on visual word recognition of compound words have provided evidence for the role of lexical properties such as frequency and length. For example, compound words with high-frequency lexemes are recognized faster than those with low-frequency lexemes (Andrews, 1986; Duñabeitia, Perea, & Carreiras, 2007; Juhasz, Starr, Inhoff, & Placke, 2003), and longer compound words elicited longer gaze durations in eye-tracking studies (e.g., Juhasz, 2008). In addition, morphological family size, or the number of morphologically

related words that share a given stem, is another facilitative factor that affects visual word recognition of compound words (Baayen, Lieber, & Schreuder, 1997). However, the findings regarding semantic transparency have been mixed (Fiorentino & Fund-Reznicek, 2009; Frisson, Niswander-Klement, & Pollatsek, 2008; Jarema, Busson, Nikolova, Tsapkini, & Libben, 1999; Juhasz, 2007; Libben, Gibson, Yoon, & Sandra, 2003; Monsell, 1985; Pollatsek & Hyönä, 2005; Sandra, 1990; Zwitserlood, 1994). In the present study, we focused on this aspect of compound word processing.

Semantic transparency in compound words

Semantic transparency refers to the extent to which two constituents in a compound word contribute to the meaning of the compound (Juhasz, 2018). Among morphologically complex words, compound words are optimal candidates for examining the role of semantic transparency, because a compound word typically consists of two independent lexical items that vary in semantic transparency. Since the meaning of the compound may or may not be determined by the meanings of the two constituents, the effect of semantic transparency can be used to reveal how each morpheme is processed and the extent

✉ Say Young Kim
sayyoungkim@hanyang.ac.kr

¹ Department of Psychology, National University of Singapore, Singapore, Singapore

² Department of English Language and Literature, Hanyang University, Seoul, South Korea

to which a compound word undergoes morphological decomposition during the word recognition process. For example, if semantically transparent compound words (e.g., *doorbell*) are recognized faster than opaque compound words (e.g., *deadline*) in lexical decision (i.e., classifying letter strings as either words or nonwords), one could infer that the semantic relatedness of the compound word to its constituents has an influence on its recognition.

Early studies using different methodologies and tasks have yielded mixed results regarding the role of semantic transparency in compound word recognition. In semantic-priming experiments, Sandra (1990) found that only fully transparent compounds (e.g., *milkman*), but not opaque compounds (e.g., *butterfly*), were significantly primed by words semantically related to either lexeme (e.g., *woman* for *milkman*, *bread* for *butterfly*). Zwitserlood (1994) further revealed that this priming effect could be extended to partially transparent compounds (such as *jailbird*), in which both constituents were primed by their related compound word.

Other studies have examined the role of semantic transparency in compound processing using either primed (Fiorentino & Fund-Reznicek, 2009; Jarema et al., 1999; Shoolman & Andrews, 2003) or unprimed (e.g., Libben et al., 2003) lexical decision tasks. Libben et al. demonstrated that compound words with a transparent second constituent (i.e., transparent-transparent compounds such as *bedroom* or opaque-transparent compounds such as *strawberry*) elicited faster reaction times in lexical decision than did those with an opaque second constituent (i.e., transparent-opaque compounds such as *jailbird* or opaque-opaque compounds such as *honeymoon*). These findings were interpreted as a headedness effect: English is a right-headed language (see Jarema et al., 1999, for a discussion). However, this headedness effect was not observed in masked-priming experiments. Both transparent (e.g., *teacup*) and opaque (e.g., *honeymoon*) compound words as primes elicited faster latencies for their constituent targets, regardless of whether the target was the head of the compound (*cup* or *moon*) or not (*tea* or *honey*), than did unrelated primes (Fiorentino & Fund-Reznicek, 2009), suggesting that morphological decomposition is robust regardless of constituent word position. In addition, both the first and second constituents also primed target compounds regardless of semantic transparency (Shoolman & Andrews, 2003). In summary, the role of each constituent in the recognition of compound words is not yet fully understood.

Several eye-tracking studies have also provided inconsistent results regarding the role of semantic transparency. Pollatsek and Hyönä (2005) did not find an effect of semantic transparency on gaze durations for Finnish compound words. Instead, they found only a first-constituent frequency effect for both transparent and opaque compounds. This null effect of semantic transparency was reexamined using English compound words by Juhasz (2007, 2018), who found a semantic

effect on gaze durations for English compound words in a by-subject analysis, but not in a by-item analysis (Juhasz, 2007), and no significant effect of semantic transparency on fixation duration measures (Juhasz, 2018). Frisson et al. (2008) found a significant effect of semantic transparency only when a space was inserted between the constituents, suggesting that compound words are not automatically decomposed into their constituents during visual analysis (see also Ji, Gagné, & Spalding, 2011).

In sum, semantic transparency is a critical factor for understanding whether morpheme-based processing in compound words occurs, and if so, how each constituent's semantic transparency contributes to the recognition of the full compound word. However, our understanding of the role of semantic transparency is based largely on evidence across studies with small numbers of items and on factorial designs in which semantic transparency is typically treated as a categorical variable. Additionally, the extent to which the position of a constituent in a compound word influences morphological decomposition remains an open question. In the present study, we examined the role of the semantic transparency of compound words as a continuous variable by focusing on each constituent's contribution, using a large dataset.

Compound words with a megastudy approach

Most previous studies examining the role of semantic transparency have employed a factorial design in which multiple variables, each with discrete possible levels, are crossed with one another so there are observations at every combination of levels across the different variables. However, a large number of variables have been shown to influence word recognition, and it is challenging to choose and manipulate a subset of variables while holding many other correlated variables constant. Many psycholinguistic variables are continuous in nature, so converting the continuous variables (e.g., word frequency) into categorical variables (high vs. low frequency) may magnify or diminish the influence of the variables (e.g., Balota, Yap, Hutchison, & Cortese, 2012). In contrast, the megastudy approach collects data for as many stimuli as possible without any constraints, allowing researchers to explore the interrelationships among the variables of interest and evaluate the generalizability of theoretical models. Indeed, several megastudies have advanced our understanding regarding the role of word length in visual word recognition (e.g., New, Ferrand, Pallier, & Brysbaert, 2006; Yap & Balota, 2009). The megastudy approach, which has rarely been employed to study morphologically complex words, can help provide complementary evidence and insights into the properties that influence the visual word recognition of complex words.

Juhasz, Lai, and Woodcock (2015) developed a database of 629 compound words with ratings on compound familiarity, age of acquisition, semantic transparency, lexeme meaning dominance (LMD; the degree to which the meaning of a compound word is contained in its first and second constituents), imageability (the degree to which a word evokes a mental image), and sensory experience (the degree to which a word evokes a sensory and/or perceptual experience in the reader's mind). They found that certain types of semantic variables, such as semantic transparency or LMD, had no or minimal effect on the visual recognition of compound words. Although this was one of the first reported megastudies on compound word processing, it had one limitation that the present study aimed to address: Namely, Juhasz et al. (2015) measured semantic transparency scores only at a whole-word, not at a constituent, level, where a 1-to-7 scale was used to judge how strongly the lexemes were related to the meaning of the compound word (i.e., higher numbers indicated greater transparency).

It is apparent that the two constituents in a compound can contribute to the meaning of the compound in a balanced (e.g., *doorbell*) or an unbalanced (*strawberry*) manner. Indeed, Libben et al. (2003) categorized compound words into four types: (1) both constituents are transparent, as in *bedroom*; (2) only the first constituent is transparent, as in *jailbird*; (3) only the second constituent is transparent, as in *strawberry*; and (4) neither constituent is transparent, as in *hogwash*. Thus, even if the semantic transparency ratings of *jailbird* and *strawberry* are similar, they do not specify which constituent contributes to the degree of semantic transparency. Previous studies have provided support for position effects (e.g., Jarema et al., 1999; Libben et al., 2003), so a whole-word-based transparency norm imposes limitations on the study of semantic transparency effects in the visual word recognition of compound words. Therefore, collecting constituent-based semantic transparency ratings would allow one to explore a wider range of questions, such as how individual constituents play a role in compound word recognition.

A recent eye-tracking study using 445 noun–noun compounds within a sentence context (Schmidtke, Van Dyke, & Kuperman, 2018) demonstrated significant interactions between semantic transparency and individual reading experience (i.e., exposure to printed material and vocabulary sizes) on compound processing, in which less experienced readers processed transparent compound words more slowly than opaque compounds, but more experienced readers processed transparent words more rapidly than opaque compounds. The results of this study suggest that how semantic transparency is estimated—the authors used human-judged similarity ratings between each component and the whole compound (i.e., either head–compound or modifier–compound)—might be partly responsible for the inconsistent findings regarding the role of semantic transparency in earlier work.

In addition, we also attempted to control for more variables, in order to identify the unique influence of semantic transparency. Specifically, beyond the lexical variables commonly controlled, two new ones were taken into account: phonological onset, to which naming performance is particularly sensitive (Yap & Balota, 2009), and morphological family size (i.e., the number of morphologically related words that share a given stem) for each constituent, which has been found to be critical for morphological processing (Juhasz & Berkowitz, 2011; Kuperman, Bertram, & Baayen, 2008; Pykkänen, Feintuch, Hopkins, & Marantz, 2004).

Corpus-based semantic transparency

In addition to human judgments of semantic transparency, one could consider corpus-based calculations for estimating the semantic transparency of compound words (e.g., Wang, Hsu, Tien, & Pomplun, 2014). Latent semantic analysis (LSA: <http://lsa.colorado.edu>; Landauer & Dumais, 1997) has been used to calculate the semantic similarity between two words (or phrases), through the analysis of a large text corpus. The technique used by LSA is based on generating vectors in semantic space that represent the co-occurrence of words in passages of text, so-called *count models* (Mandera, Keuleers, & Brysbaert, 2017). One can compute the semantic similarity values for any two terms using the LSA cosine value, which ranges from -1 to 1 . Therefore, this calculation allows researchers to estimate the semantic transparency of two morphemes in complex words, such as derived words (Diependaele, Duñabeitia, Morris, & Keuleers, 2011; Feldman, O'Connor, & Moscoso del Prado Martin, 2009; Rastle, Davis, Marslen-Wilson, & Tyler, 2000) or compound words (El-Bialy, Gagné, & Spalding, 2013).

However, some limitations are associated with this approach. For instance, count models (e.g., LSA) heavily focus on mathematical functions to estimate semantic similarity, which may not reflect the human cognitive system (Mandera et al., 2017). Critically, count models have been less than successful in terms of predicting human performance in speeded pronunciation and lexical decision. Recently, Mandera et al. addressed this issue by comparing the performance of count models with that of *prediction-based models* (or skip-gram model), which are based on neural networks (Mikolov, Chen, Corrado, & Dean, 2013), to predict a target word from the neighboring words (i.e., context words) that tend to co-occur with it (e.g., the continuous bag-of-words model; CBOW) or to predict the context words from a target word (e.g., the skip-gram model). In short, the target word, $Word_i$, in the CBOW can be predicted from the preceding words, such as $Word_{i-2}$, $Word_{i-1}$, and the words following it, such as $Word_{i+1}$ $Word_{i+2}$. When Mandera et al. compared the

performance of the prediction-based models (CBOW and skip-gram) with the count model using a large dataset of semantic priming (Hutchison et al., 2013) and other behavioral tests, they found that the prediction-based models (e.g., CBOW) explained the most variance in behavioral data (e.g., RTs). In light of this, we utilized Mandera et al.'s semantic vectors, based on distributional semantics, for our compound words (available at <http://crr.ugent.be/snaut/>).

The present study

The purposes of the present study were twofold. First, we wished to provide a database of semantic transparency ratings for 2,861 bilinear English compound words at the constituent level. The ratings were then analyzed using an existing behavioral database of compound word performance on two visual word recognition tasks, lexical decision and speeded pronunciation, using data obtained from the English Lexicon Project (ELP; Balota et al., 2007).¹ Second, the results of our analysis were compared to those using existing studies, based on either human-judged semantic transparency at the whole-word level (Juhász et al., 2015) or distance scores derived from the prediction-based distributional semantics algorithm at the constituent level (Mandera et al., 2017). This comparison allowed for an evaluation of the relative efficacy of different semantic transparency measures (human rating scores vs. corpus-based calculation) in accounting for the role of semantic transparency in compound word recognition, based on performance data from the ELP.

Method

Participants

A total of 254 participants from the National University of Singapore, whose first and dominant language was English, participated for course credit or were paid SGD 5 for a 30-min session. The total length of the task varied between 30 and 45 min across participants.

¹ The ELP provides a behavioral dataset of lexical decision and speeded pronunciation latencies for 40,481 words and nonwords (available at <http://lexicon.wustl.edu>). Across six universities in the United States, lexical decision data were collected from 816 participants, with each participant performing approximately 3,400 trials, and speeded pronunciation data were collected from 444 participants, with each participant performing approximately 2,500 trials (see Balota et al., 2007, for more about the dataset).

Materials

We took 629 English compound words from Juhász et al. (2015) and added another 2,271 words from the ELP (Balota et al., 2007). The final number of compound words analyzed was 2,861, after excluding items that were incorrectly coded as compound words (e.g., derived words such as *familiarness* or *antithyroid* were coded as compound words) in the ELP.

Procedure

On each trial, participants were presented a compound word and either its first or second constituent (e.g., *doorbell-door* or *doorbell-bell*) simultaneously on the computer monitor and were asked to judge how strongly the meanings of two words were related, using a 1-to-7 scale (1 = *not related*, 7 = *highly related*). Due to the large number of compound words, ten sets of word pairs (compound words and one of their constituents) were created, and each participant was assigned to a set of either 590 or 540 word pairs. No participant rated semantic transparency for both constituents of any compound word. In the final database, each compound-constituent pair was rated by 28–32 different participants. The judgment task was presented using E-Prime 1.2 (Schneider, Eschman, & Zuccolotto, 2002), and responses were made via the computer keyboard. The order of presentation of the words within the set assigned to each participant was randomized.

Results and discussion

Reliability

To ensure the reliability of the transparency ratings, split-half reliabilities (e.g., Cunningham, Perry, & Stanovich, 2001; McBride-Chang & Manis, 1996) were calculated separately for both constituents' semantic transparency scores and were found to be fairly high: $r = .783$, $p < .001$, for the first constituent, and $r = .773$, $p < .001$, for the second constituent (DeDe, Ricca, Knilans, & Trubl, 2014; Nunnally, 1978). In addition, we examined the correlation with the previous database of 629 compound words (Juhász et al., 2015). Because the Juhász et al. (2015) database reflects only whole-word-based transparency, we calculated the composite transparency scores for our data by averaging the transparency scores of each constituent for these 629 compound words. The correlation between the ratings from Juhász et al.'s (2015) database and our composite transparency scores was found to be fairly high ($r = .757$, $p < .001$).

Item-level regression analyses for the effect of semantic transparency

Following the ELP and other megastudies (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Goh, Yap, Lau, Ng, & Tan, 2016; Sze, Yap, & Rickard Liow, 2015), a hierarchical multiple regression analysis was used. In this approach, the dependent variable (response latency from the ELP, in this case) was first regressed on control variables that are known to influence word recognition. These first steps were done to account for variance in the latencies that can be explained by the control variables. In the final steps, the variables of interest—semantic transparency scores, in this case—were added to the regression, to determine the amount of additional variance in the latencies that could be explained, *over and above* the variance already accounted for in the earlier steps. This would allow for a determination of the extent to which the variables of interest could account for unique variance relative to that already accounted for by the control variables. The values of the criterion and semantic transparency variables in this approach were estimated at the item level—that is, averaged across responses from individual participants.

Control variables

To identify the unique role of constituent-based semantic transparency in compound processing, we controlled ten variables at the surface and lexical levels. To capture the variance associated with voice key biases, phonological onsets as dichotomous variables were used to code the initial phoneme of each word (1 for *presence of feature*, 0 for *absence of feature*) on 13 features (affricate, alveolar, nasal, etc.). These were entered in the model first (see Spieler & Balota, 1997, for details) as the *surface* level.

For the *lexical* level, we controlled word length variables for both the compound and constituent words. These included the numbers of syllables and letters in the compound and the number of letters in each constituent.² The logarithm-transformed SUBTLEX values for contextual diversity (i.e., lgSUBTLCD) for both the compound and each constituent were used as indices of word frequency.³ Finally, the morphological size for

² Despite the fact that the length variables at compound level and constituent level were highly correlated, we did not find any multicollinearity issue among them (all VIFs < 8).

³ This index substantially outperformed lgSUBTLWF (Brysaert & New, 2009), consistent with some earlier studies (e.g., Adelman, Brown, & Quesada, 2006).

Table 1 Means and standard deviations for the full set of predictors and dependent variables

	Mean	Std. Deviation	<i>N</i>
1. LDT-ZRT	.045	.411	2,849
2. LDT-Acc	.864	.176	2,849
3. Pronunciation-ZRT	-.101	.376	2,849
4. Pronunciation-Acc	.969	.061	2,849
5. ST-c1 (present study)	4.048	.786	2,861
6. ST-c2 (present study)	3.729	.796	2,861
7. ST-(from Juhasz et al., 2015)	4.532	1.203	629
8. Dist-c1 (from Mandera et al., 2017)	.746	.132	2,609
9. Dist-c1 (from Mandera et al., 2017)	.744	.133	2,616
10. Number of syllables	2.266	.487	2,850
11. Length (compound)	8.480	1.444	2,850
12. Length (c1)	4.250	1.073	2,861
13. Length (c2)	4.244	.947	2,860
14. Freq-LgSUBTLCD	1.276	.663	2,443
15. Freq-LgSUBTLCD1	3.171	.629	2,861
16. Freq-LgSUBTLCD2	3.188	.630	2,861
17. Morph size (c1)	14.633	20.498	2,861
18. Morph size (c2)	17.865	21.177	2,861

LDT: lexical decision task; ZRT: z-transformed reaction times; Acc: accuracy; Pronunciation: speeded pronunciation task; ST: semantic transparency score; c1: first constituent; c2: second constituent; Dist: semantic distance score

each constituent—the count of the number of times each constituent appeared, regardless of constituent position, across the compound words in the present database—were also entered into the models.

Table 1 shows descriptive statistics for the semantic transparency scores, control variables, and dependent measures. There are three sets of semantic transparency scores: those from the present study, those from Juhasz et al.'s (2015) database of 629 compound words, and the semantic distance scores from Mandera et al. (2017). Table 2 presents the intercorrelations between the predictors and dependent measures.

Item-level regression analyses

Three-step hierarchical item-level regression analyses were conducted for both lexical decision and speeded pronunciation performance, with reaction time (RT) and accuracy as dependent measures. Phonological onsets were entered in Step 1; word length, frequency, and morphological family size were entered in Step 2; and semantic transparency scores were entered in Step 3.

Table 2 Correlations between the full set of predictors and dependent variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1. LDT-ZRT	—	-.706 ^{***}	.769 ^{***}	-.489 ^{***}	-.138 ^{***}	-.132 ^{***}	-.020	.117 ^{***}	.066 ^{***}	.179 ^{***}	.265 ^{***}	.183 ^{***}	.155 ^{***}	-.541 ^{***}	-.271 ^{***}	-.228 ^{***}	-.167 ^{***}	-.180 ^{***}	
2. LDT-Acc	—		-.635 ^{***}	.601 ^{***}	.190 ^{***}	.177 ^{***}	.151 ^{***}	-.123 ^{***}	-.080 ^{***}	.021	.074 ^{***}	.058 ^{***}	.034	.377 ^{***}	.272 ^{***}	.195 ^{***}	.145 ^{***}	.142 ^{***}	
3. Pronunciation-ZRT	—			-.495 ^{***}	-.049 ^{***}	-.087 ^{***}	.041	.025	.050 ^{***}	.181 ^{***}	.351 ^{***}	.294 ^{***}	.157 ^{***}	-.432 ^{***}	-.381 ^{***}	-.228 ^{***}	-.304 ^{***}	-.133 ^{***}	
4. Pronunciation-Acc	—				.116 ^{***}	.104 ^{***}	.063	-.084 ^{***}	-.050 ^{***}	-.031	-.019	-.051 ^{***}	.016	.234 ^{***}	.185 ^{***}	.147 ^{***}	.107 ^{***}	.090 ^{***}	
5. ST-c1 (present study)	—					.242 ^{***}	.626 ^{***}	-.494 ^{***}	-.049 ^{***}	-.013	.075 ^{***}	.140 ^{***}	-.016	.073 ^{***}	-.071 ^{***}	.073 ^{***}	-.162 ^{***}	.111 ^{***}	
6. ST-c2 (present study)	—						.557 ^{***}	-.161 ^{***}	-.483 ^{***}	.014	.070 ^{***}	.001	.122 ^{***}	.029	.000	-.120 ^{***}	.039 ^{***}	-.105 ^{***}	
7. ST-Juhasz	—							-.411 ^{***}	-.362 ^{***}	.024	.084 ^{***}	.087 ^{***}	.086 ^{***}	-.063	.028	.007	-.081 ^{***}	.039 ^{***}	
8. Dist-c1 (Mandera)	—								.291 ^{***}	-.008	-.036	-.092 ^{***}	.026	-.218 ^{***}	.157 ^{***}	-.031	.179 ^{***}	-.075 ^{***}	
9. Dist-c2 (Mandera)	—									-.020	-.029	.019	-.087 ^{***}	-.195 ^{***}	-.021	.131 ^{***}	-.012	.133 ^{***}	
10. Number of syllables	—										.505 ^{***}	.395 ^{***}	.246 ^{***}	-.019	-.047 ^{***}	-.038 ^{***}	-.005	-.041 ^{***}	
11. Length (compound)	—											.715 ^{***}	.588 ^{***}	-.134 ^{***}	-.195 ^{***}	-.154 ^{***}	-.164 ^{***}	-.079 ^{***}	
12. Length (c1)	—												-.027	-.063 ^{***}	-.288 ^{***}	.022	-.316 ^{***}	.145 ^{***}	
13. Length (c2)	—													-.103 ^{***}	.022	-.296 ^{***}	.099 ^{***}	-.308 ^{***}	
14. Freq-LgSUBTLCD	—														.187 ^{***}	.176 ^{***}	.028	.077 ^{***}	
15. Freq-LgSUBTLCD1	—															.124 ^{***}	.486 ^{***}	-.070 ^{***}	
16. Freq-LgSUBTLCD2	—																.007	.513 ^{***}	
17. Morph size (c1)	—																	-.143 ^{***}	
18. Morph size (c2)	—																		

Table 3 Standardized RT and accuracy regression coefficients from Steps 1 to 3 of the item-level regression analyses for speeded pronunciation and lexical decision performance for compound words (based on the present study's ratings)

Predictors	Lexical Decision		Speeded Pronunciation	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets				
R-square	.031***	.021***	.146***	.025***
Step 2: Lexical variables				
R-square	.395***	.243***	.436***	.104***
	$\Delta R^2 = .364$	$\Delta R^2 = .222$	$\Delta R^2 = .291$	$\Delta R^2 = .079$
Step 3: Semantic transparency				
c1	-.091***	.123***	-.043**	.082***
c2	-.098***	.118***	-.085***	.064**
R-square	.416***	.277***	.447***	.117***
	$\Delta R^2 = .021$	$\Delta R^2 = .034$	$\Delta R^2 = .010$	$\Delta R^2 = .013$

c1: first constituent; c2: second constituent.
 $\dagger p < .1$, * $p < .05$; ** $p < .01$; *** $p < .001$

The present scores Table 3 presents the results of the regression analyses for the effect of our collected semantic transparency scores on the four dependent measures after surface and lexical predictors were controlled. The control variables accounted for significant variance across the dependent measures—response latencies and accuracies—in both speeded pronunciation and lexical decision performance. Semantic transparency scores at the constituent level accounted for significant additional unique variance in the dependent measures, over and above the variance already accounted for by the surface and lexical variables (2.1% and 3.4%, respectively, for RTs and accuracy in lexical decision; 1% and 1.3%, respectively, for RTs and accuracy in speeded pronunciation). This provides clear-cut evidence that constituent based semantic transparency plays a role in compound processing.

Whole-word-based semantic transparency (using Juhasz et al.'s, 2015, scores) Table 4 presents the results of the same regression analyses using Juhasz et al.'s (2015) whole-word-based semantic transparency scores. After controlling both the surface and lexical variables, Juhasz et al.'s (2015) semantic transparency accounted for relatively less unique variance than did our ratings across the dependent measures (0.2% and 2.1%, respectively, for RTs and accuracy in lexical decision; 0.1% and 0.4%, respectively, for RTs and accuracy in speeded pronunciation), and only significantly accounted for accuracy in lexical decision.

Table 4 Standardized RT and accuracy regression coefficients from Steps 1 to 3 of the item-level regression analyses for speeded pronunciation and lexical decision performance for compound words (based on ratings from Juhasz et al., 2015)

Predictors	Lexical Decision		Speeded Pronunciation	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets				
R-square	.065***	.032*	.173***	.027
Step 2: Lexical variables				
R-square	.405***	.246***	.462***	.099***
	$\Delta R^2 = .339$	$\Delta R^2 = .214$	$\Delta R^2 = .214$	$\Delta R^2 = .072$
Step 3: Semantic distance				
Overall rating	-.049	.150***	-.026	.066
R-square	.408	.267***	.462	.103
	$\Delta R^2 = .002$	$\Delta R^2 = .021$	$\Delta R^2 = .001$	$\Delta R^2 = .004$

c1: first constituent; c2: second constituent.
 $\dagger p < .1$, * $p < .05$; ** $p < .01$; *** $p < .001$

Corpus-based calculation data (using Mandera et al.'s, 2017, cosine distances) Table 5 presents the results of the same regression analyses using Mandera et al.'s (2017) corpus-based computed pairwise cosine distances for semantic transparency scores, based on the CBOW model trained on 300 dimensions, six windows, UKWAC, and the subtitle corpus. After controlling the surface and lexical variables, Mandera et al.'s semantic distance

Table 5 Standardized RT and accuracy regression coefficients from Steps 1 to 3 of the item-level regression analyses for speeded pronunciation and lexical decision performance for compound words (based on semantic distance from Mandera et al., 2017)

Predictors	Lexical Decision		Speeded Pronunciation	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets				
R-square	.032***	.023***	.151***	.027***
Step 2: Lexical variables				
R-square	.393***	.244***	.435***	.103***
	$\Delta R^2 = .361$	$\Delta R^2 = .221$	$\Delta R^2 = .284$	$\Delta R^2 = .076$
Step 3: Semantic distance				
c1	.032 \dagger	-.071***	.004	-.067**
c2	-.022	-.007	.006	-.001
R-square	.394	.249**	.435	.107**
	$\Delta R^2 = .001$	$\Delta R^2 = .005$	$\Delta R^2 = .000$	$\Delta R^2 = .004$

c1: first constituent; c2: second constituent.
 $\dagger p < .1$, * $p < .05$; ** $p < .01$; *** $p < .001$

scores accounted for relatively less unique variance than did our data across the dependent measures (0.1% and 0.5%, respectively, for RTs and accuracy in lexical decision; 0% and 0.4%, respectively, for RTs and accuracy in speeded pronunciation), and only significantly accounted for accuracy in both lexical decision and speeded pronunciation. Note that significant regression coefficients were found only for the first constituent scores.

Present scores with the items that are common with Juhasz et al. (2015, N = 621) Because the numbers of compound words in our data and in Juhasz et al. (2015) were not equivalent, we tested the effect of our data based only on the items included in Juhasz et al.'s (2015) database. Table 6 presents the regression analysis results. The semantic scores in the present dataset were again better at accounting for unique variance across the dependent measures, except for accuracy in speeded pronunciation (1.4% and 2.7%, respectively, for RTs and accuracy in lexical decision; 0.8% for both RTs and accuracy in speeded pronunciation).

Present scores with the items that are common with Mandera et al.'s (2017, N = 2,337) As above, we also tested the effect of our data based only on the items included in Mandera et al.'s (2017) calculation. Table 7 presents the

regression analysis results. The semantic scores in the present dataset were again better at accounting for unique variance across the dependent measures (2.1% and 3.3%, respectively, for RTs and accuracy in lexical decision; 1.1% and 1.2%, respectively, for RTs and accuracy in speeded pronunciation).

Conclusion

The present study provides human-judged semantic transparency ratings for 2,861 English compounds at their constituent level. The ratings we collected accounted for performance in the ELP's lexical decision and speeded pronunciation data. Specifically, greater transparency in either constituent was associated with both faster latencies and greater accuracy in both the lexical decision and naming tasks in the ELP. This can be seen in the consistently negative regression coefficients for RTs and the positive coefficients for accuracy. This pattern of results supports earlier findings showing transparency effects for both constituents (e.g., Fiorentino & Fund-Reznicek, 2009; Shoolman & Andrews, 2003).

More critically, our constituent-based human judgment approach to measuring semantic transparency turned out to possess more predictive power than the

Table 6 Standardized RT and accuracy regression coefficients from Steps 1 to 3 of the item-level regression analyses for speeded pronunciation and lexical decision performance for compound words (current data using only items that are common with Juhasz et al., 2015)

Predictors	Lexical Decision		Speeded Pronunciation	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets				
R-square	.065***	.032*	.173***	.027
Step 2: Lexical variables				
R-square	.405***	.246***	.462***	.099***
	$\Delta R^2 = .340$	$\Delta R^2 = .214$	$\Delta R^2 = .289$	$\Delta R^2 = .072$
Step 3: Semantic transparency				
c1	-.067†	.123**	-.040	.071†
c2	-.089**	.099*	-.079*	.051
R-square	.419**	.273**	.470**	.107†
	$\Delta R^2 = .014$	$\Delta R^2 = .027$	$\Delta R^2 = .008$	$\Delta R^2 = .008$

c1: first constituent; c2: second constituent.
† $p < .1$, * $p < .05$; ** $p < .01$; *** $p < .001$

Table 7 Standardized RT and accuracy regression coefficients from Steps 1 to 3 of the item-level regression analyses for speeded pronunciation and lexical decision performance for compound words (current data using only items that are common with Mandera et al., 2017)

Predictors	Lexical Decision		Speeded Pronunciation	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets				
R-square	.032***	.023***	.151***	.027***
Step 2: Lexical variables				
R-square	.393***	.244***	.435***	.103***
	$\Delta R^2 = .361$	$\Delta R^2 = .221$	$\Delta R^2 = .284$	$\Delta R^2 = .076$
Step 3: Semantic distance				
c1	-.092***	.125***	-.047**	.079***
c2	-.099***	.115***	-.085***	.063**
R-square	.414**	.278***	.446***	.115***
	$\Delta R^2 = .021$	$\Delta R^2 = .033$	$\Delta R^2 = .011$	$\Delta R^2 = .012$

c1: first constituent; c2: second constituent.
† $p < .1$, * $p < .05$; ** $p < .01$; *** $p < .001$

Table 8 Comparison of the effects of the three measures of semantic transparency

Measure	Lexical Decision		Speeded Pronunciation	
	RT	Accuracy	RT	Accuracy
Constituent-based human judged measure (the present study)				
c1	-.091 ^{***}	.123 ^{***}	-.043 ^{**}	.082 ^{***}
c2	-.098 ^{***}	.118 ^{***}	-.085 ^{***}	.064 ^{**}
Whole-word-based human judged measure (Juhasz et al., 2015)				
	-.049	.150 ^{***}	-.026	.066
Constituent-based prediction model measure (Mandera et al., 2017)				
c1	.032 [†]	-.071 ^{***}	.004	-.067 ^{**}
c2	-.022	-.007	.006	-.001

c1: first constituent; c2: second constituent.

[†] $p < .1$, * $p < .05$; ** $p < .01$; *** $p < .001$

other two approaches, of Juhasz et al. (2015) and Mandera et al. (2017), as is summarized in Table 8. The differences in the proportions of variance accounted for across the dependent measures are likely due to the way transparency was scored in each study. Namely, Juhasz et al. (2015) used the whole-word approach, which seems to be limited in capturing the individual lexemes' influence on lexical processing. In that sense, our approach provides a methodological improvement that extends their study. Our results are consistent with Schmidtke et al. (2018), who also used human judgments of semantic transparency at the constituent level and showed that constituent-based transparency influenced eye movement patterns, but their effects were also moderated by individual differences in print exposure.

Additionally, our ratings outperformed Mandera et al.'s (2017) prediction-based semantic relatedness calculations. The present findings suggest that distributional semantic relatedness scores were less effective than our human-judged ratings at accounting for human performance in simple lexical decision and naming data, at least for compound words. However, caution should be taken in interpreting the regression analysis results between our human-judged data and corpus-based data. In the latter, the quality of the word embeddings for the compounds may be insufficient if there are many very-low-frequency compounds. It is possible that the resulting measures would perform better if the distributional model were trained on a larger corpus, so the low performance of the corpus-derived measures may not be due to inherent limitations of corpus-based calculations.⁴

We hope that other researchers will find this dataset useful in order to manipulate and control for semantic transparency at the constituent level. Empirical work

based on these norms can also be used to inform and constrain future computational models of compound word processing.

Author note This work was supported by a Faculty of Arts & Social Sciences Staff Research Support Scheme grant (C-581-000-222-091) to K.S.Y. and W.D.G. We thank Barbara Juhasz, Erin Buchanan, and two anonymous reviewers for helpful comments, and Ruth Poh for data collection assistance.

References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*, 814–823. doi:<https://doi.org/10.1111/j.1467-9280.2006.01787.x>
- Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects. *Journal of Memory and Language*, *25*, 726–740.
- Baayen, R. H., Lieber, R., & Schreuder, R. (1997). The morphological complexity of simplex nouns. *Linguistics*, *35*, 861–877.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316. doi:<https://doi.org/10.1037/0096-3445.133.2.283>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. doi:<https://doi.org/10.3758/BF03193014>
- Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition: Vol. 1. Models and methods, orthography and phonology* (pp. 90–115). Hove, UK: Psychology Press.
- Brybaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–900. doi:<https://doi.org/10.3758/BRM.41.4.977>

⁴ We thank an anonymous reviewer for pointing out this possibility.

- Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing, 14*, 549–568.
- DeDe, G., Ricca, M., Knilans, J., & Trubl, B. (2014). Construct validity and reliability of working memory tasks for people with aphasia. *Aphasiology, 28*, 692–712.
- Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language, 64*, 344–358. doi:<https://doi.org/10.1016/j.jml.2011.01.003>
- Dressler, W. U. (2006). Compound types. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 23–44). Oxford, UK: Oxford University Press.
- Duñabeitia, 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*, 1171–1176. doi:<https://doi.org/10.3758/BF03193108>
- El-Bialy, R., Gagné, C. L., & Spalding, T. L. (2013). Processing of English compounds is sensitive to the constituents' semantic transparency. *The Mental Lexicon, 8*, 75–95.
- Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martin, F. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review, 16*, 684–691. doi:<https://doi.org/10.3758/PBR.16.4.684>
- Fiorentino, R., & Fund-Reznicek, E. (2009). Masked morphological priming of compound constituents. *The Mental Lexicon, 4*, 159–193.
- Frisson, S., Niswander-Klement, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology, 99*, 87–107. doi:<https://doi.org/10.1348/000712607X181304>
- Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M. R., & Tan, L.-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology, 7*, 976. doi:<https://doi.org/10.3389/fpsyg.2016.00976>
- Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C.-S., . . . Buchanan, E. (2013). The Semantic Priming Project. *Behavior Research Methods, 45*, 1099–1114. doi:<https://doi.org/10.3758/s13428-012-0304-z>
- Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., & Libben, G. (1999). Processing compounds: A cross-linguistic study. *Brain and Language, 68*, 362–369.
- Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language, 65*, 406–430.
- Juhász, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. von Gompel, W. Murray, & M. Fischer (Eds.), *Eye movements: A window on mind and brain* (pp. 373–389). Boston, MA: Elsevier.
- Juhász, B. J. (2008). The processing of compound words in English: Effects of word length on eye movements during reading. *Language and Cognitive Processes, 23*, 1057–1088.
- Juhász, B. J. (2018). Experiences with compound words influences their processing: An eye movement investigation with English compound words. *Quarterly Journal of Experimental Psychology, 71*, 103–112.
- Juhász, B. J., & Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition: A multitask investigation. *Language and Cognitive Processes, 26*, 653–682.
- Juhász, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods, 47*, 1004–1019. doi:<https://doi.org/10.3758/s13428-014-0523-6>
- Juhász, 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.
- Kuperman, V., Bertram, R., & Baayen, H. R. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes, 23*, 1089–1132.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240. doi:<https://doi.org/10.1037/0033-295X.104.2.211>
- Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language, 84*, 50–64. doi:[https://doi.org/10.1016/S0093-934X\(02\)00520-5](https://doi.org/10.1016/S0093-934X(02)00520-5)
- Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language, 92*, 57–78. doi:<https://doi.org/10.1016/j.jml.2016.04.001>
- McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing, 8*, 323–339.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. ArXiv preprint. arXiv: 1301.3781.
- Monsell, S. (1985). Repetition and the lexicon. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 147–195). Hillsdale, NJ: Erlbaum.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*, 45–52. doi:<https://doi.org/10.3758/BF03193811>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes, 20*, 261–290.
- Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: an MEG study. *Cognition, 91*, B35–B45.
- Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes, 15*, 507–537. doi:<https://doi.org/10.1080/01690960050119689>
- Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology, 42A*, 529–567.
- Schmidtko, D., Van Dyke, J. A., & Kuperman, V. (2018). Individual variability in the semantic processing of English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 421–439.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Shoolman, N., & Andrews, S. (2003). Racehorses, reindeer, and sparrows: Using masked priming to investigate morphological influences on compound word identification. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: The state of the art* (pp. 241–278). New York, NY: Psychology Press.

- Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 8, 411–416. doi:<https://doi.org/10.1111/j.1467-9280.1997.tb00453.x>
- Sze, W. P., Yap, M. J., & Rickard Liow, S. J. (2015). The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis. *Quarterly Journal of Experimental Psychology*, 68, 1541–1570.
- Wang, H.-C., Hsu, L.-C., Tien, Y.-M., & Pomplun, M. (2014). Predicting raters' transparency judgments of English and Chinese morphological constituents using latent semantic analysis. *Behavior Research Methods*, 46, 284–306. doi:10.3758/s13428-013-0360-z
- Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60, 502–529.
- Zwitserslood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9, 341–368. doi:<https://doi.org/10.1080/01690969408402123>