Processing English compounds: Investigating semantic transparency

Christina L. Gagné, & Thomas L. Spalding
Department of Psychology, University of Alberta, Canada
Kelly A. Nisbet
Western University, Canada

Semantic transparency is widely believed to affect the processing of compound words. It has been described as the degree to which the meaning of the constituent is retained in the meaning of the whole compound, but also as the degree to which the meaning of the compound is predictable from the meaning of the constituents. Furthermore, semantic transparency has been operationalized in various ways (e.g., Libben 2010; Libben et al. 2003; Sandra 1990). We describe a study in which transparency was measured based on: 1) linguistic criteria used by informed judges, 2) participant ratings of a) how predictable a compound's meaning was from its parts, and b) the extent that each constituent retains its meaning in the compound, 3) Latent Semantic Analysis (LSA; Landauer & Dumais 1997) scores for the compound and each constituent. We used these measures to test the claim that meaning retention ratings reflect the semantic similarity between a compound's meaning and the constituent meaning, whereas the predictability ratings indicate the degree of semantic compositionality of the compound's concept (see Marelli & Luzzatti 2012). We did not find support for these specific conceptualizations of semantic transparency. We then investigated the relationships among these different measures of semantic transparency to determine whether they reflect the same underlying construct, and in particular, the extent to which the LSA scores and participant ratings can predict the classification by informed judges using linguistic criteria. Finally, we used the various measures to predict typing times (a measure of processing) of compound words. The results from these various analyses indicate that the various methods of measuring semantic transparency do not reflect the same underlying aspects of semantic transparency.

Keywords: *semantic transparency, compound words, latent semantic analysis, psycholinguistics*

1. Introduction

There is ongoing debate in psycholinguistic research about whether morphologically complex words are represented and processed in terms of their morphemes (for an overview of the key issues in linguistic research on compounding, see Lieber & Štekauer 2009 and Scalise & Vogel 2005). Theories have spanned the entire range of possibilities. Some theories propose that each word has an independent representation and that words are accessed via these whole-word representations (e.g., Bradley 1980, Butterworth 1983, Manelis & Tharp 1977) and others have taken the opposite stance and argue that words are accessed via their morphological constituents (e.g., Taft 1985). Still others (e.g., Chialant & Caramazza 1995, Frauenfelder & Schreuder 1991) take a position midway between these two endpoints. In terms of the debate concerning compound words, the issue of semantic transparency is particularly relevant because some theories have proposed that whether words are represented and processed in terms of their morphemes depends on whether the constituents are semantically transparent. Semantic transparency refers to the extent to which the meaning of the compound can be derived from the constituents, as well as to the extent to which the meanings of the constituents are related to the meaning of the compound (see Libben 1998).

In this paper, we begin by providing an overview of findings demonstrating that semantic transparency influences ease of processing and discussing ways in which theories have accounted for the role of semantic transparency. Then, we describe three ways in which this construct has been measured. Finally, we present a series of analyses in which we examine how well the various methods of measuring semantic transparency predict each other and how well the measures of semantic transparency predict ease of processing in a typing task.

1.1. Empirical evidence for semantic transparency

Previous research has indicated that semantic transparency influences the ease with which compound words can be processed. For example, lexical decision latencies are longer for compounds with opaque heads than for transparent compounds (e.g., Ji, Gagné, & Spalding 2011, Libben, Gibson, Yoon, & Sandra 2003). In addition, manipulations that aid morphological decomposition (such as having the two constituents in different colours or separated by a space) slowed the processing of opaque compounds but did not affect the processing of transparent compounds (Ji et al. 2011).

Other evidence for the influence of semantic transparency comes from priming experiments. In a priming experiment, the experimenter manipulates which of several primes is presented prior to a target word and examines whether the processing of the target is differentially affected by the primes. Past research has found that whether the processing of a compound is influenced by a prior exposure to a word that is semantically related to one of the constituents depends on semantic transparency. Sandra (1990) found that transparent compounds benefited from prior exposure to a semantically related prime (e.g, *milkman* was aided by exposure to *workman*), but that opaque compounds did not (e.g, *butterfly* was not aided by exposure to the word *bread*). Similarly, Zwitserlood (1994) did not find evidence of semantic priming for fully opaque compounds. These results suggest that transparency of the individual constituents influences ease of processing. More recent evidence has suggested that the ease of processing a compound is influenced not only by the opacity of the constituents but also by whether the constituents have similar semantic transparency; El-Bialy, Gagné and Spalding (2013) found that a semantically related prime (e.g., *ear* as a prime for *eyetooth* and *eyesight*) facilitated processing when the transparency of the first and second constituents matched (i.e., when both were transparent or when both were opaque) but not when the transparency of the constituents differed (i.e., for partially opaque compounds).

Research examining written production has also found influences of semantic transparency. One way of measuring written production is by recording typing latencies for each letter of a compound as participants type the word using a computer keyboard. Libben and

Weber (2014, see also Sahel, Nottbusch, Grimm, & Weingarten 2008) found that the typing latencies are longer for the letter after the morpheme boundary than for the letter prior to the morpheme boundary. Importantly, this delay was smaller for opaque-opaque (OO) compounds than for transparent-transparent (TT) and opaque-transparent (OT) compounds. That is, fully opaque compounds showed a smaller boundary effect than did compounds with transparent head nouns. Gagné & Spalding (2014a) found that production was more difficult for compounds with opaque first constituents; it took longer to type the first letter of the word for compounds with an opaque first constituent than for compounds with a transparent first constituent. In addition, a prime word that was semantically related to the first constituent of the target compound aided the production of the compound when the head of the compound was transparent but not when the head was opaque (Gagné & Spalding 2014d).

In sum, evidence for the role of semantic transparency during the processing of compound words has come from both comprehension and production tasks.

1.2. Semantic transparency in theoretical accounts of compound word processing

Theories have taken two primary approaches in terms of incorporating semantic transparency (for an overview, see Gagné & Spalding 2009). Some have built the distinction into the architecture of the system whereas others account for semantic transparency in terms of processing.

For the first approach, the influence of semantic transparency is described in terms of whether the meaning of the constituent is able to influence the activation of the compound's representation. For example, in the theoretical approaches proposed by Libben (1998), Sandra (1990), and Zwitserlood (1994) compounds are represented as morphologically complex at the lexical level and, consequently, during processing can be decomposed into the constituent morphemes. Thus, *blue* and *berry* are available during the processing of the transparent compound *blueberry*, and, likewise, *straw* and *berry* are available during the processing of the partially opaque compound *strawberry*. The difference between transparent and opaque compounds is explained in terms of the links between the constituent and compound at the semantic levels. Zwitserlood (1994) proposed that at the semantic level opaque compounds behave as monomorphemic words. That is, at the semantic level, an opaque compound has a single semantic representation that is not linked with the semantic representations of its constituents, whereas fully- and partially-transparent compounds have their own semantic representations that are linked to the semantic representations of their components (see Schriefers, Zwitserlood, & Roelofs 1991, for a similar approach). Thus, by this view, *blue* is linked to *blueberry* at the conceptual level, but *straw* is not linked to *strawberry* at the conceptual level. Libben (1998; see also Libben et al. 2003) also explains transparency in terms of links among semantic representations. However, in his theory, opaque constituents are connected via inhibitory links to the compound rather than, as in the case of Zwitserlood's approach, being unconnected.

In each of these approaches, priming with words that are semantically related to a constituent should result in faster processing for the compound when the constituent is semantically transparent but not when the constituent is opaque. That is, the presence or absence of links between the constituent and the compound's semantic representations accounts for the finding that transparent compounds but not opaque compounds benefited from exposure to a prime that was semantically related to one of the constituents. According to these theoretical approaches, exposure to a semantically related prime also activates the constituent due to spreading activation to semantically related words, which then boosts the activation of the compound due to the link between the semantic representations of the constituent and the compound for semantically transparent constituents. To illustrate, accessing the word *woman* also activates *man* due to a facilitatory connection between the semantic representations for these two words. Because *man* is linked to *milkman*, the compound's semantic and lexical representations also become more activated. In contrast, the lack of a facilitatory link between an opaque constituent and the compound does not allow the activation of the constituent's semantic representation to increase the activation of the compound's semantic representation and, thus, no benefit from the semantically related prime is observed for opaque compounds.

For the second theoretical approach, semantic transparency is not represented in the architecture but rather arises out of the processing. Theoretical perspectives such as the Competition-Among-Nominals (CARIN) theory (Gagné & Shoben 1997) and the Relational Interpretation Competitive Evaluation (RICE) theory (Spalding, Gagné, Mullaly & Ji, 2010; see also Gagné & Spalding 2014b) propose that constituents become available during processing regardless of the degree of semantic transparency. This assumption is consistent with Libben's (2010) maximization of opportunity view. The system then attempts to integrate the constituents (e.g, Inhoff, Radach & Heller 2000) using semantic integration (Fiorentino & Poeppel 2007, Gagné & Spalding 2004, 2009, Spalding et al. 2010). During the integration process, the system attempts to construct a meaning. Previously (e.g., Gagné & Spalding 2004, 2009), we have argued that this meaning construction process relies heavily on relation-based conceptual structures. Thus, one way in which the influence of semantic transparency emerges is during meaning construction processing. The constructed meaning for compounds with opaque constituents will be inconsistent with the conventional meaning. For example, a literal interpretation of *hogwash* is "a wash for pigs" which is incompatible with the conventional meaning "rubbish or nonsense". This inconsistency between the meanings needs to be resolved and, thus, slows the processing of opaque compounds. However, in the case of compounds with transparent constituents, the constructed meaning will typically be similar to or the same as the conventional meaning and, thus, should not slow processing.

Several findings are compatible with these predictions. For example, opaque compounds (e.g., *humbug*) were responded to more slowly in a lexical decision task than were frequency-matched transparent compounds (e.g., *snowball*) and manipulations that aided decomposition, such as presenting the constituents in different colours or separated by a space, slowed processing of opaque compounds but not of transparent compounds (Ji et al. 2011). Jarema, Busson, Nikolova, Tsapkini, and Libben (1999) have also found that responses in a lexical decision task are faster for semantically transparent compounds than for semantically opaque compounds.

This construction-based approach also accounts for the finding that transparent but not opaque compounds benefited from exposure to a prime that was semantically related to one of the constituents by assuming that both facilitatory and inhibitory processes are occurring during the task. In some cases, such as in the case of opaque compounds, these processes might offset each other such that the net effect is that no difference in processing time is observed. For example, a semantically related prime might facilitate the activation of a constituent which would generally aid the activation of a compound. However, in the case of an opaque compound, the increased ease of accessing the opaque constituent might speed the meaning construction process which would increase the difficulty of processing the compound due to the increased conflict between the constructed and conventional meanings. Accounting for priming in this way, rather than via properties of the architecture of the mental lexicon, is more flexible in that it can more readily account for adaptation effects that are observed in priming experiments, as demonstrated in Gagné & Spalding (2014a).

Another finding that supports the claim that morphological decomposition and meaning construction occurs during the processing of compound words regardless of semantic transparency is that increased relational diversity slowed processing of both opaque and transparent compounds (Gagné & Spalding 2014c, 2014d, Schmidtke, Kuperman, Gagné, & Spalding 2015). Relational diversity reflects the extent to which the possible interpretations concentrate on a small set of relations. High relational diversity reflects a higher degree of relation competition than does low relational diversity.

1.3. Operationalizing semantic transparency

As shown in the preceding sections, semantic transparency plays a central role in theories concerning compound words and has been shown to influence processing in a variety of experimental tasks. However, semantic transparency is a theoretical construct: What is the best way of operationalizing this construct? This question is especially relevant given that some researchers, such as Libben, Curtiss, and Weber 2014 (see also Libben 2010; Libben & Weber 2014) have argued that psychocentricity should play a larger role in determining which linguistic constructs are valid. By this view, some aspects of a word are properties of the human language user, rather than properties of the word itself; according to Libben et al. (2014, p. 1) "Language resides in the minds of the individuals" and therefore specific characteristics of words (including structural and semantic characteristics) are influenced by individual experience. Therefore, it is useful to consider how semantic transparency has been used as a variable in empirical research.

In psycholinguistic studies, semantic transparency has been defined and operationalized in various ways. Indeed, it has been measured in three primary ways. First, items have been classified as transparent or opaque based on informed judgments by the researchers (i.e., judgements made by individuals with an expert knowledge of linguistic/psycholinguistic theories) using linguistic criteria such as whether the second constituent is the semantic head (e.g., Ji et al. 2011; Libben 2010; Libben & Weber 2014; Sandra 1990). Second, transparency has been measured by participant ratings (e.g, Fiorentino & Fund-Reznicek 2009; Juhasz, Lai, & Woodcock, 2015; Libben et al. 2003; Marelli & Luzzatti 2012). The specific details of the ratings have varied in terms of the range of the scale (e.g., a four point scale, Libben et al. 2003, or a five point scale, Zwitserlood 1994) and, more importantly, in terms of the way in which transparency was defined. The two dominant ways of conceptualizing semantic transparency have been to rate the degree to which the meaning of a compound is predictable from the constituents and the degree to which each of the constituents (rated individually) retain their meaning in the compound.

A third way of measuring transparency is by using latent semantic analysis (LSA), which is the degree of association and semantic similarity between the meanings of words based on patterns of co-occurrence in similar contexts (e.g., Kuperman 2013; Wang, Hsu, Tien, & Pompiu 2014). LSA is a technique that estimates the semantic distance between two words based on the contexts in which the two words occur in a large corpus (Landauer & Dumais, 1997). The scores are cosine values that range from -1 to 1 and larger absolute values indicate greater association between the two words. For example, the LSA score for *honey* and *honeycomb* is .61 and indicates that these two items are more closely associated than are *dumb* and *dumbbell* which have an LSA score of .05. The LSA scores were obtained from term-to-term LSA scores at http://lsa.colorado.edu.

2. Current investigation

Although various methods for operationalizing semantic transparency have been used in the literature, there has not yet been a systematic exploration of how these methods relate to each other. Therefore, the aim of our investigation was to better understand how the three ways of measuring semantic transparency relate to each other. In particular, we were interested in determining whether they measure the same aspect of semantic transparency. Our second aim was to focus on LSA values and human ratings to determine whether they accounted for behavioural data in the same way. For this aspect of our investigation, we focused on typing latencies from a written production task.

2.1. Dataset

In the current paper, we use the dataset collected by Nisbet, Gagné, and Spalding (2015) which used three measures of semantic transparency (linguistic classification, ratings, and LSA scores) for 200 compound words and 50 pseudo-compound words. The items were selected such that each item had a unique first and second constituent. The compounds were classified into four categories based on the semantic transparency of the first and second constituent: transparent-transparent (e.g., *blueberry*, N=59), transparent-opaque (e.g., *honeycomb*, N=53), opaque-transparent (e.g., *chopstick*, N=46) and opaque-opaque (e.g,. *hogwash*, N=42). Pseudo-compound words were those that orthographically consist of two English morphemes, but in which the morphemes do not play any morphemic role (e.g., *carpet* is not composed of *car* + *pet* in a morphological sense even though it does contain these two orthographic units).

The participant ratings of semantic transparency were collected from 90 participants in a two-part study. The items were displayed one at a time on a computer screen in a randomized order. In Part One, participants rated “the extent to which the word parts contribute to the overall meaning”. Participants indicated their response by manipulating a slide bar that ranged from 0% (very unpredictable) to 100% (very predictable) and pressing enter once they placed the marker at the desired percentage. In Part Two, participants rated each constituent individually on “the extent to which the word parts retain their meaning in relation to the whole word". Again, participants used a slide bar from 0% to 100%. The descriptive statistics for these ratings are provided in Table 1.

Table 1: Mean and range (i.e., minimum to maximum values) in percentages for meaning predictability and meaning retention ratings

Compound Type	Predictability Rating	Retention Rating for C1	Retention Rating for C2
TT	79 (51 to 93)	78 (34 to 93)	79 (41 to 93)
OT	60 (36 to 81)	38 (12 to 76)	80 (53 to 94)
TO	65 (37 to 86)	78 (51 to 91)	49 (17 to 85)
OO	52 (28 to 78)	49 (10 to 86)	49 (8 to 86)
Pseudo	24 (13 to 48)	17 (6 to 63)	15 (6 to 66)

In addition to these ratings, we obtained three Latent Semantic Analysis (LSA) scores (Landauer & Dumais, 1997) for each item using the database located at http://lsa.colorado.edu using the term-to-term scores for the General Reading up to first Year College topic space. One score represented the semantic relationship between the first constituent and the compound (e.g., *honey* and *honeycomb*), the second represented the relationship between the second constituent and the compound (e.g., *comb* and *honeycomb*) and the third score represented the relationship between the two constituents (e.g., *honey* and *comb*). LSA scores were available for all of the constituents, and for 156 of the compounds. The descriptive statistics for the three LSA scores are provided in Table 2.

The Pearson correlations among the ratings and LSA measures are provided in Table 3.

Table 2: Mean and range (i.e., minimum to maximum values) for LSA scores

Compound Type	C1 and Compound	C2 and Compound	C1 and C2
TT	.21 (-.06 to .95)	.18 (-.05 to .92)	.24 (.01 to .74)
OT	.15 (-.05 to .82)	.28 (.01 to .74)	.17 (0 to .57)
TO	.26 (.01 to .74)	.11 (-.07 to .40)	.14 (-.02 to .35)
OO	.09 (-.03 to .49)	.07 (-.05 to .31)	.34 (-.01 to .44)
Pseudo	.06 (-.06 to .32)	.07 (-.06 to .24)	.10 (-.02 to .52)

Table 3: Pearson correlations between ratings and LSA scores

	Predictability Rating	Retention Rating C1	Retention Rating C2	LSA: C1 and	LSA: C2 and
Retention Rating C1	0.85				
Retention Rating C2	0.83	0.50			
LSA: C1 and	0.35	0.27	0.28		
LSA: C2 and	0.39	0.41	0.27	0.31	
LSA: C1 and C2	0.24	0.05	0.39	0.31	0.23

2.2. Do LSA scores predict semantic transparency ratings?

If meaning retention ratings directly reflect strength of association between the constituent and the compound then the ratings for the first constituent should be affected by the LSA values for the compound and the first constituent (e.g., *house* - *houseboat*) but not by the other LSA measures. Likewise, ratings for the second constituent should be affected only by the LSA for the compound and the second constituent (e.g., *boat* - *houseboat*).

The ratings, which were originally expressed as a percentage, were converted to a proportion by dividing by 100 for the regression analysis. The data were analyzed using fractional response regression because the predicted variable (rating) could take on values between 0 and 1. For each rating, we fit a separate model using the three LSA values as predictor variables and rating as the dependent variable (i.e., the to-be-predicted variable).

All three LSA values were predictive of the ratings for the first constituent. Ratings increased as the values of LSA for the compound and first constituent increased (b = 0.41, SE = .12, z = 3.50, p < .0001) and as the values of LSA for the first and second constituent increased (b = 0.25, SE = .13, z = 2.02, p = .04). As the values of LSA for the compound and second constituent increased, however, meaning retention ratings of the first constituent decreased (b = -0.31, SE = .12, z = -2.57, p = .01); the more associated the second constituent and compound were to each other (as indicated by the C2-compound LSA score), the less transparent the first constituent appeared to be to the participants when they made their judgments. The finding that the meaning retention ratings for the first constituent was predicted by all three LSA scores (i.e., C1-compound, C1-C2, and C2-compound), indicates that the meaning retention rating for the first constituent did not purely reflect the degree of association between the compound (e.g., *blueberry*) and the first constituent (e.g., *blue*).

In contrast, the meaning retention ratings for the second constituent were predicted only by the LSA score between the second constituent and the compound (b = .43, SE = .09, z = 4.82,

p < .0001) and not by the C1-compound scores nor by the C1-C2 LSA scores. This result indicates that the ratings for the second constituent were primarily affected by the degree of association between the compound (e.g., *blueberry*) and the head category (e.g., *berry*).

The predictability ratings from Part 2 (i.e., how much the meaning of the compound could be predicted by the constituents) were predicted by the LSA for the compound and the first constituent (b = 0.15, SE = .07, z = 2.02, p = .04) and by the LSA for the first constituent and the second constituent (b = 0.27, SE = .09, z = 2.81, p = .005). The ratings increased as the association between the compound and the first constituent increased and as the association between the two constituents increased. The association between the compound and the second constituent did not affect judgments of the compound meaning's predictability.

In sum, the LSA measures were predictive of the human judgments of meaning retention and predictability. However, with the exception of the ratings for the second constituent, the relation was not one-to-one. That is, the semantic transparency rating for the first constituent was not a direct function of the similarity between the constituent and the compound (as measured by LSA). Instead, the other LSA measures also played a role. This suggests that the meaning retention ratings for the first constituent do not uniquely reflect the association between the first constituent and the compound but, rather, are influenced by the association between the second constituent and the compound and by the association between the first constituent and the second constituent. In terms of the predictability rating, this rating is influenced by the similarity of the two constituents to each other and by the similarity of the first constituent to the compound, but not by the similarity of the second constituent and the compound. These results are especially important given that, in the literature, LSA measures have been used as a means of determining whether a constituent is semantically transparent or opaque.

2.3. How well do ratings and LSA predict linguistic classification?

In this section we examine how well the LSA scores and rating data can predict the four types of compounds (TT, TO, OT, and OO) and pseudo-compounds, as well as whether they can discriminate fully opaque compounds (e.g., *hogwash* and *buttercup*) from pseudo-compounds (e.g., *carpet* and *kitten*). That is, given the LSA scores and/or the three ratings, what percentage of items can we correctly classify as TT, TO, OT, OO, and pseudo-compound? To answer this question, we used canonical discriminant analysis, which is a statistical technique related to analysis of variance and regression. This analysis finds sets of linear functions of the predictor variables (e.g., LSA scores or ratings) that best predict the differences among the various groups (e.g., the four types of compounds). We used the functions produced by the analysis to classify the items into categories and examined how well the predicted categories matched the original categories.

In our first set of analyses we focused only on compounds and compared the success of a model based on the three ratings with a model based on the three LSA values. The percentage of TT, TO, OT, and OO compounds that were correctly classified were, in order, 84.8%, 66.0%, 69.6%, and 59.5% when the predictability ratings and the two meaning retention ratings were used. However, when the three LSA scores were used, the percentage of TT, TO, OT, and OO compounds that were correctly classified was much lower, 46.0%, 47.6%, 45.4%, and 51.6%.

The full set of classifications (i.e., the correct and incorrect classifications) are provided in Table 4 for the classifications based on the ratings and in Table 5 for the classifications based on the LSA scores.

Table 4: Classification of items into compound types based on discriminant dimensions derived from ratings (% based on row)

	Predicted Category			
Actual Category	TT	OT	TO	OO
TT	84.75	5.08	5.08	5.08
OT	19.57	69.57	0	10.87
TO	30.19	0	66.04	3.77
OO	9.52	11.90	19.05	59.52

Table 5: Classification of items into compound types based on discriminant dimensions derived from LSA (% based on row)

	Predicted Category			
Actual Category	TT	OT	TO	OO
TT	46.00	14.00	28.00	12.00
OT	27.27	45.45	9.09	18.18
TO	30.95	4.76	47.62	16.67
OO	25.81	3.23	19.35	51.61

In our second set of analyses we included pseudo-compounds along with the set of compounds to determine how well the model could predict the various item types. When ratings were used to classify the items, the percentage of pseudo-compounds, TT, TO, OT, and OO compounds that were correctly classified was 93.9%, 84.75%, 64.2%, 69.6%, and 40.5% (see Table 6 for the full set of classifications). Classification was best for pseudo-compounds and TT compounds.

Table 6: Classification of items based on discriminant dimensions derived from ratings (% based on row)

	Predicted Category				
Actual Category	TT	OT	TO	OO	Pseudo
TT	84.75	5.08	5.08	5.08	0
OT	19.57	69.57	0	10.87	0
TO	30.19	0	64.15	5.66	0
OO	9.52	11.90	19.05	40.48	19.05
Pseudo	0	2.04	2.04	2.04	93.88

When the three LSA measures were used, the percentage of pseudo-compounds, TT, TO, OT, and OO compounds that were correctly classified was 81.2%, 40.0%, 42.8%, 45.45%, and 0.0% (see Table 7 for the full set of classifications). When using LSA scores, classification was accurate only for the pseudo-compounds. One of the most striking findings is that when pseudo-compounds were included, no items were classified as OO compounds when the discriminant function was based on LSA scores.

Table 7: Classification of items based on discriminant dimensions derived from LSA (% based on row)

	Predicted Category				
Actual Category	TT	OT	TO	OO	Pseudo
TT	40.00	10.00	26.00	0	24.00
OT	18.18	45.45	3.03	0	33.33
TO	23.81	4.76	42.86	0	28.57
OO	19.35	3.23	16.13	0	61.29
Pseudo	12.50	0	6.25	0	81.25

Taken together, these results indicate that ratings are much better able to classify items than are the LSA values. Indeed, the success rate for functions based on the LSA measures is less than 50% for most categories types. Although classification based on the three ratings were more successful, the inclusion of pseudo-compounds reduces the ability of functions based on those ratings to distinguish the various compound types; instead, the function discriminates the most transparent compounds (the TT) from the pseudo-compounds and is only moderately successful at classifying the partially opaque and fully opaque compounds.

In the third set of analyses, we focused on the OO and pseudo-compounds to determine whether the ratings and the LSA scores could discriminate between these two types of items. Both items consist, orthographically, of two words. These words function as morphemes only for OO compounds (e.g., *buttercup* has a bi-morphemic structure) but not for the pseudo-compounds (e.g., *kitten* has a monomorphemic structure). Although a function based on the ratings was able to correctly classify the two types of words, 85.7% and 95.9% for the OO and pseudo-compounds, a function based on the LSA measures was not, 25.8% correct for the OO compounds vs. 85.4% for the pseudo-compounds. There was a large bias toward classifying items as pseudo-compounds when using the LSA measures. In sum, the rating data was able to discriminate the two item types whereas the LSA was not. This suggests that the meaning retention ratings were sensitive to the presence/absence of morphological structure whereas the LSA scores were not.

2.4. Using ratings and LSA to predict typing latencies before and after the morpheme boundary

The analyses reported in the previous sections indicate that ratings of semantic transparency and LSA scores are related to each other, but are not synonymous. In this section, we examine how well these two measures predict behavioural data. In particular, we examine whether they predict typing times for compound words. During a typing task, participants type a word on a computer keyboard as the computer records the time required to type each letter. The data from this task can then be analyzed to see whether typing time differs at key points within a word. We will focus on the time required to type the letters before and after the morpheme boundary (e.g., the letters *p* and *b* in *cupboard*).

Previous research (Gagné & Spalding 2014d; Libben, Weber & Miwa 2012; Libben & Weber 2014) found that the time to type the letter after the morpheme boundary was longer than the time to type the letter before the morpheme boundary. In addition, semantic transparency affected the size of the elevation at the boundary. Libben and Weber (2014) found that the increase at the boundary was smaller for opaque-opaque (OO) compounds than for transparent-transparent (TT) and opaque-transparent (OT) compounds. The increase at the boundary for OO and transparent-opaque (TO) compounds was equivalent. Gagné & Spalding (2014a) also found that the increase at the morpheme boundary was affected by semantic transparency and, in addition, that production of compounds with transparent second constituents (i.e., for TT and OT compounds) was affected by whether the preceding prime was semantically related to the first constituent of the compound.

For the current project, we use typing data collected by Gagné and Spalding (in press). The dataset consisted of the typing times for 200 compounds and 50 pseudo-compounds from 140 native speakers of English, which was a total of 11462 trials for which the word had been correctly typed without revisions (e.g., without deletions).

In Gagné and Spalding (in press), semantic transparency was measured using a dichotomous classification (transparent vs. opaque) of semantic transparency and the participant ratings of semantic transparency. One set of analyses focused on the morpheme boundary and used linear mixed effects regressions (Rabe-Hesketh & Skrondal, 2012) using semantic transparency of each constituent and letter position (end of C1 vs. start of C2) as predictor variables for latencies of the letter before and at the morpheme boundary and items and participants as random factors. One model was fit using the dichotomous classification and a second model was fit using the participant ratings. Both models indicated that the size of the boundary effect was influenced by the semantic transparency of the first constituent but was not strongly affected by the second constituent. The increase in typing at the boundary was larger when the first constituent was transparent (or had high transparency ratings) than when it was opaque (or had lower transparency ratings). The analyses with the ratings indicated that the influence of the first constituent's transparency was only seen during the production of the first constituent (i.e., at the last letter of the first constituent, such as the *w* in *snowball*); typing latencies were faster at this position when semantic transparency ratings for the first constituent were higher than when they were lower. Transparency did not influence the time required to type the first letter of the second constituent (e.g., the *b* in *snowball*).

For the current project, we examined whether semantic transparency as measured by LSA scores produced similar effects to the measures reported in Gagné and Spalding (in press). In our analysis, position (end of C1 vs. start of C2), the LSA score for the first constituent and the compound, and the LSA for the second constituent and the compound were entered as predictor variables for typing latency. All models included items and participants as random factors. These three variables interacted ($x^2 = 5.68$, $p = .02$). The nature of this interaction is shown in Figure 1.

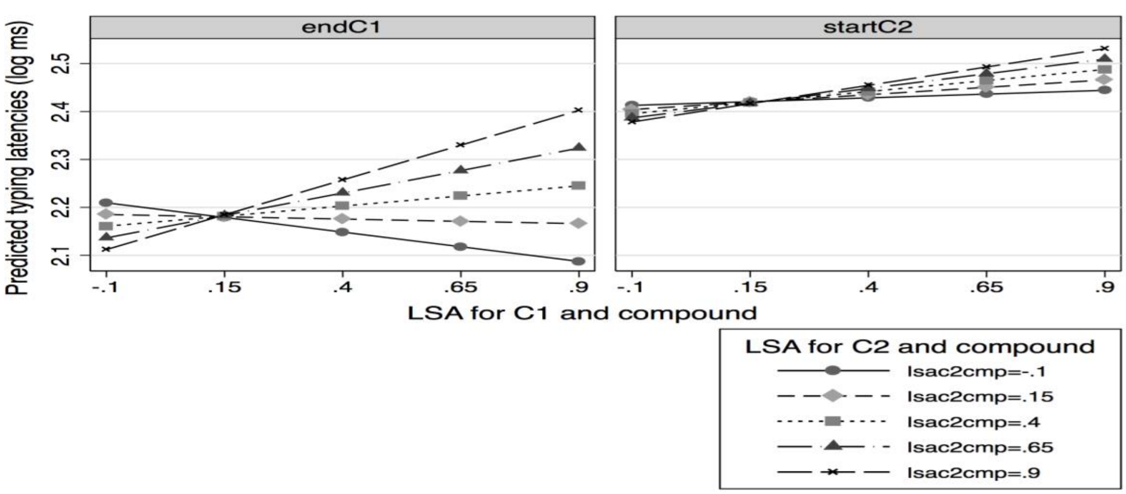

Figure 1: Interaction between C1-compound LSA and C2-compound LSA by position

Due to the interaction, we subsequently conducted separate analyses for the two positions to examine the influence of the two LSA variables. For the end of C1 position, the two LSA scores interacted ($z = 2.18$, $p = .03$). As can be seen in Figure 1, the influence of the LSA score for the second constituent and the compound was larger when LSA for the first constituent and the compound (LSA C1-compound) was higher than when LSA C1-compound was lower. Typing latencies at the end of the first constituent were fastest when the LSA C1-compound was higher and LSA C2-compound was lower. An alternate way to describe the interaction is that when the second constituent and the compound were highly associated with each other and the first constituent and the compound were highly associated with each other, then typing latencies at the end of the first constituent were slower than when the second constituent and compound were not strongly associated.

However, at the start of the second constituent, neither variable influenced typing times. Thus, as was the case for the dichotomous classification and participant ratings of semantic transparency, the LSA scores only affected the processing of the final letter of the first constituent but had no influence on the first letter of the second constituent. However, unlike the other two measures of semantic transparency, the LSA score involving the second constituent influenced ease of processing.

2.5. Using ratings and LSA to predict typing latencies at all letter positions

Gagné and Spalding (in press) found that typing time became faster across the word. That is, the typing of each letter became faster towards the end of the word. However, this speedup was slower for compound words than for monomorphemic words. In addition, analyses with the dichotomous measure of semantic transparency (e.g., transparent vs. opaque) indicated that the speedup across letter position was slower when the first constituent was transparent than when it was opaque. In other words, transparent compounds seemed to impair the normal speedup across the word. Transparency of the second constituent did not affect the speedup in typing latencies across letter positions.

In terms of the ratings, there was a two-way interaction between letter position (e.g., first letter, second letter, … , last letter) and the rating for the first constituent, $z = 4.79$, $p < 0001$; consistent with what was found for the dichotomous measure of semantic transparency, increased semantic transparency between the first constituent and the compound slowed down the speedup in the typing latencies across the word. The speedup in typing latencies across letter position was unaffected by transparency ratings for the second constituent, $z = 1.10$, $p = .27$.

For the current investigation, we examined whether the LSA scores show the same pattern as the dichotomous measure of semantic transparency. As was the case for the analysis with the ratings, we included an interaction term for letter position and each of the two measures of semantic transparency (i.e., one for the first constituent and one for the second constituent). The analyses revealed an interaction between letter position and the LSA scores for the first constituent and the compound, $z = -3.40$, $p = .001$). However, the nature of the interaction was the opposite to what had been found for the dichotomous measure and the ratings. The slope across letter position became steeper as the value of LSA increased; that is, increased LSA scores (i.e., increased similarity between the first constituent and the compound) were associated with an increased rate of speedup across the word. The speedup in typing latencies across letter position was unaffected by LSA scores for the second constituent and the compound, $z = 1.05$, $p = .29$.

In sum, the LSA scores, although predictive of the speedup in typing latencies across the word, did not behave as the linguistic classification and rating measures of semantic transparency. Higher LSA scores aided the output of the letters across the word, whereas higher values for the other two measures of semantic transparency hindered the output across the word. This difference in terms of how the measures impact typing latencies suggests that LSA scores are not measuring the same theoretical construct as are the linguistic classification and the human ratings.

3. Conclusions

Semantic transparency has played a central role in theories of complex word processing and there have been several ways of measuring semantic transparency. Our analyses indicate that linguistic classification, human ratings, and LSA scores, although related to each other, do not reflect the same underlying constructs. Interestingly, the relation among the measures depended on morphological structure — that is, on whether the constituent is in the first or second (i.e., the head) position. Only the C2-compound LSA scores predicted the semantic transparency ratings for the second constituent, whereas C1-compound, C2-compound, and C1-C2 LSA scores predicted semantic transparency ratings for the first constituent.

In terms of the question of whether ratings or LSA scores can be used to re-capture the linguistic classifications (i.e., the dichotomous classification of opaque vs. transparent), neither measure appears to be well-suited for this task. The LSA scores in particular were unable to successfully classify the items. The meaning retention ratings were more successful, especially for classifying the fully transparent compounds, but still performed poorly for items that had opaque constituents.

Furthermore, only the human ratings were able to distinguish the OO compounds from the pseudo-compounds (i.e., items that have two embedded morphemes but actually lack a compound structure). The LSA measure could not accurately distinguish these two compound types. This aspect of our results indicates that people are sensitive to morphological structure, in a way that the LSA measures do not capture.

The various measures also differ in terms of how they account for behavioural measures. We demonstrate that, at least in the case of a typing task, ratings and LSA scores provide different perspectives into the question of how semantic transparency affects ease of processing. For example, only LSA scores, and not the ratings or linguistic classification, influenced typing times for the second constituent. As another example, even though LSA scores and the ratings affected the rate of speed-up across the typing of a word, their effects were in the opposite direction. Increased semantic transparency as measured by LSA scores between the first constituent and compound was associated with a faster speed-up, whereas increased semantic transparency as measured by either the linguistic classification or by the meaning retention ratings were associated with a slower speed-up. Clearly, the two sets of measures (LSA and human judgments) are not measuring the same aspect of semantic transparency. The implication of this is that researchers must use caution when comparing research results from studies that use different measures of semantic transparency. Measures of semantic transparency are simply not interchangeable, and general conclusions about semantic transparency are not possible without proper caveats with respect to the measures of semantic transparency used in the research. It remains to be seen whether other methods of distributional semantics which take into consideration morphological structure (such as Marelli & Baroni, 2015) will be more successful than LSA.

Taken together, the analyses indicate that the various measures of semantic transparency are not directly interchangeable. That is, there is not a strict one-to-one mapping among the linguistic classification, the ratings, and the LSA scores. More importantly, semantic transparency of a constituent does not solely reflect similarity between the compound and the constituent, as is typically assumed. Instead, semantic transparency of a constituent, especially the first constituent, is affected by information about the other constituent. Our analyses indicate that semantic transparency is multi-faceted; it appears that no single measure (e.g., LSA or Ratings) fully captures this theoretical construct.

These results have direct theoretical implications in terms of evaluating previous psycholinguistic work and in terms of future explorations of semantic transparency. In particular, the results show that studies that use different measures of semantic transparency are not directly comparable. Furthermore, not all measures are equally predictive of the typing data. This is particularly important to keep in mind before concluding that semantic transparency does not exert an influence. Another implication is that a compound's semantic transparency is not a simple function of the similarity between the constituents and the compound, which suggests that transparency cannot be easily represented in terms of the presence or absence of links between a constituent representation and a compound representation in the Mental Lexicon.

Furthermore, although this research focused on psycholinguistic processing, our findings have direct implications for linguistic structure because they provide insight into the role of the constituents. For example, much linguistic work has been done concerning the structure of multi-morphemic words (e.g., see Lieber & Štekauer 2009; Scalise & Vogel 2005). One key issue in this work has been the role of morphology and part of this question centers around the role of semantic transparency (see Section 1.2). Indeed, semantic transparency is central to the issue of compositionality (Dressler 2005). The concept of semantic transparency and morphemic structure is also directly tied to theoretical issues concerning lexical semantics and, in particular, whether (and if so, how) semantic properties of the constituents are related to the whole compound (e.g., Lieber 2004, 2009; for a discussion of this issue see Gagné & Spalding 2015). Our data suggest that the two constituents (of a bi-morphemic compound) do not equally contribute to determining the transparency of the entire compound. Furthermore, our data suggest that semantic transparency is not a direct reflection of the semantic overlap between a constituent and the whole word, and that the semantic transparencies of the constituents are not independent of each other. In a broader sense, our data are consistent with Libben's (2005) suggestion that semantic transparency is not a property of the words and constituents, but rather can only be understood in terms of how people process multi-morphemic words; instead, it is a psycholinguistic effect. If so, then this would suggest the linguistic treatments of semantic transparency also need to consider the psycholinguistic aspects of this issue.

Acknowledgements

This research was supported by NSERC Discovery Grant (05100) to the first author and by NSERC Discovery Grant (250028) to the second author.

References

Bradley, Dianne. 1980. Lexical representation of derivational relation. In Aronoff, M. & Kean, M.-L. (eds.), *Juncture*, 37–55. Saratoga: Anma Libri.

Butterworth, Brian. 1983. Lexical representation. In Butterworth, B. (ed.), *Language Production*, 257–294. San Diego: Academic Press.

Chialant, Doriana & Caramazza, Alfonso. 1995. Where is morphology and how is it processed? The case of written word recognition. In Feldman, L. (ed.), *Morphological aspects of language processing*, 57–76. Hillsdale: Erlbaum.

Dressler, Wolfgang U. 2005. Word-formation in natural morphology. In Štekauer, P. & Lieber, R. (eds.), *Handbook of word-formation*, 267-284. Dordrecht: Kluwer.

El-Bialy, Rowan, Gagné, Christina L. & Spalding, Thomas L. 2013. Processing of English compounds is sensitive to the constituents' semantic transparency. *The Mental Lexicon* 8. 75–95.

Fiorentino, Robert & Fund-Reznicek, Ella. 2009. Masked morphological priming of compound constituents. *The Mental Lexicon* 4. 159–193.

Fiorentino, Robert & Poeppel, David. 2007. Compound words and structure in the lexicon. *Language and Cognitive Processes* 22. 953–1000.

Frauenfelder, Ulli H. & Schreuder, Robert. 1991. Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In Booji, G. E. & van Marle, J. (eds.), *Yearbook of morphology*, 165–183. Kluwer: Dordrecht.

Gagné, Christina L. & Edward J. Shoben. 1997. Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory and Cognition* 23. 71–87.

Gagné, Christina L. & Spalding, Thomas L. 2004. Effect of relation availability on the interpretation and access of familiar noun-noun compounds. *Brain & Language* 90. 478–486.

Gagné, Christina L. & Spalding, Thomas L. 2009. Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory & Language* 60. 20–35.

Gagné, Christina L. & Spalding, Thomas L. 2014a. Adaptation effects in lexical processing. *Suvremena Lingvistika (Contemporary Linguistics)* 40. 127–149.

Gagné, Christina L. & Spalding, Thomas L. 2014b. Conceptual composition: The role of relational competition in the comprehension of modifier-noun phrases and noun-noun compounds. In Ross, B. (ed.), *The psychology of learning and motivation*, 97–130. New York: Elsevier.

Gagné, Christina L. & Spalding, Thomas L. 2014c. Relational diversity and ease of processing for opaque and transparent English compounds. In Rainer, F., Gardani, F., Luschutzky, H. C. & Dressler, W. U. (eds.), *Morphology and meaning*, 153–162. Amsterdam & Philadelphia, PI: John Benjamins.

Gagné, Christina L. & Spalding, Thomas L. 2014d. Typing time as an index of morphological and semantic effects during English compound processing. *Lingue e Linguaggio* XIII. 241–262.

Gagné, Christina L. & Spalding, Thomas L. (in press). Effects of morphology and semantic transparency on typing latencies in English compound and pseudo-compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Gagné, Christina L., & Spalding, Thomas. L. 2015. Semantics, concepts and meta-cognition: Attributing properties and meanings to complex concepts. In Bauer, L., Körtvélyessy, L. & Štekauer, P. (eds.), *Semantics of complex words*, 9–25. New York: Springer.

Inhoff, Albrecht W. & Radach, Ralph & Heller, Dieter. 2000. Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory & Language* 42. 23–50.

Jarema, Gonia, Busson, Céline, Nikolova, Rossitza, Tsapkini, Kyrana & Libben, Gary. 1999. Processing compounds: A cross-linguistic study. *Brain & Language* 68. 362–369.

Ji, Hongbo, Gagné, Christina L. & Spalding, Thomas L. 2011. Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory & Language* 65. 406–430.

Juhasz, Barbara J., Lai, Yun-Hsuan, & Woodcock, Michelle 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.

Kuperman, Victor. 2013. Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology* 4 (203). 1–10.

Landauer, Thomas K. & Dumais. Susan 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.

Libben, Gary. 1998. Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain & Language* 61. 30–44.

Libben, Gary 2010. Compound words, semantic transparency, and morphological transcendence. In Olson, S. (ed.), *New impulses in word-formation* (Linguistische Berichte Sonderheft 17), 212–232. Hamburg: Buske.

Libben, Gary, Curtiss, Kaitlin & Weber, Silke. 2014. Psychocentricity and participant profiles: Implications for lexical processing among multilinguals. *Frontiers in Psychology* 5 (557). 1–13.

Libben, Gary, Gibson, Martha, Yoon, Yeo & Sandra, Dominiek. 2003. Compound fracture: The role of semantic transparency and morphological headedness. *Brain & Language* 84. 50–64.

Libben, Gary & Weber, Silke & Miwa, Koji. 2012. P3: A technique for the study of perception, production, and participant properties. *The Mental Lexicon* 7. 237–248.

Libben, Gary & Weber, Silke. 2014. Semantic transparency, compounding, and the nature of independent variables. In Rainer, F., Gardani, F., Luschutzky, H. C. & Dressler, W. U. (eds.), *Morphology and meaning*, 205–222. Amsterdam & Philadelphia, PI: John Benjamins.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge, MA: Cambridge University Press.

Lieber, Rochelle. 2009. A lexical semantic approach to compounding. In Lieber, R. & Štekauer, P. (eds.), *The Oxford handbook of compounding*, 78–104. Oxford: Oxford University Press.

Manelis, Leon & Tharp, David A. 1977. The processing of affixed words. *Memory & Cognition* 5. 690–695.

Marelli, Marco & Baroni, Marco. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122. 485–515.

Marelli, Marco & Luzzatti, Claudio. 2012. Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory & Language* 66. 644–664.

Nisbet, Kelly A., Gagné, Christina L. & Spalding, Thomas L. 2015. A comparison of three measures of semantic transparency in English compounds. Manuscript in preparation.

Rabe-Hesketh, Sophia & Anders Skrondal. 2012. *Multilevel and longitudinal modeling using stata.* College Station: Stata Press Publication.

Sahel, Said, Nottbusch, Guido, Grimm, Angela & Weingarten, Rüdiger. 2008. Written production of German compounds: Effects of lexical frequency and semantic transparency. *Written Language & Literacy* 11. 211–227.

Sandra, Dominiek. 1990. On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology* 42A. 529–567.

Scalise, Sergio & Vogel, Irene. 2005. *Cross-disciplinary issues in compounding*. Amsterdam & Philadelphia, PI: John Benjamins.

Schmidtke, Daniel, Kuperman, Victor, Gagné, Christina L. & Spalding, Thomas L. 2015. Competition between conceptual relations affects compound recognition: The role of entropy. *Psychonomic Bulletin & Review.* Advance online publication.

Schriefers, Herbert & Zwitserlood, Pienie. 1991. The identification of morphologically complex spoken words: Continuous processing or decomposition? *Journal of Memory & Language* 30. 26–47.

Spalding, Thomas L., Gagné, Christina L., Mullaly, Allison & Ji, Hongbo. 2010. Relation-based interpretation of noun-noun phrases: A new theoretical approach. In Olson, S. (ed.), *New impulses in word-formation* (Linguistische Berichte Sonderheft 17), 283–315. Hamburg: Buske.

Spalding, Thomas L & Gagné, Christina L. 2014. Relational diversity affects ease of processing even for opaque English compounds. *The Mental Lexicon* 9. 48–66.

Taft, Marcus. 1985. Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes* 1. 297–308.

Wang, Hsueh-Cheng, Hsu, Li-Chuan, Tien, Yi-Min& Pomplun, Marc. 2014. Predicting raters' transparency judgments of English and Chinese morphological constituents using latent semantic analysis. *Behavioral and Research Methods* 46. 284–306.

Zwitserlood, Pienie. 1994. The role of semantic transparency in the processing and representation of Dutch compounds. *Language & Cognitive Processes* 9. 341–368.

Christina Gagné
Department of Psychology
University of Alberta
P217 Biological Sciences Building
Edmonton Alberta Canada
T6G 2E9
e-mail@ualberta.ca

Thomas Spalding
Department of Psychology
University of Alberta
P217 Biological Sciences Building
Edmonton Alberta Canada

T6G 2E9
e-mail: spalding@ualberta.ca

Kelly A. Nisbet
University of Western Ontario
Department of Psychology and Brain & Mind Institute
London, Ontario, Canada
N6A 5C2
knisbet2@uwo.ca

In SKASE Journal of Theoretical Linguistics [online]. 2016, vol. 13, no.2 [cit. 2016-09-11]. Available on web page http://www.skase.sk/Volumes/JTL32/pdf_doc/01.pdf. ISSN 1336- 782X.