PSIHOLOGIJA, 2013, Vol. 46 (4), 455–478

UDC 159.955.5/.6.07 ; 81'37 ; 81'23
DOI: 10.2298/PSI1304455P

Semantic relations and compound transparency: A regression study in CARIN theory

Hien Pham[1] and R. Harald Baayen[1,2]

[1] *University of Alberta, Canada*
[2] *Eberhard Karls University, Germany*

According to the CARIN theory of Gagné and Shoben (1997), conceptual relations play an important role in compound interpretation. This study develops three measures gauging the role of conceptual relations, and pits these measures against measures based on latent semantic analysis (Landauer & Dumais, 1997). The CARIN measures successfully predict response latencies in a familiarity categorization task, in a semantic transparency task, and in visual lexical decision. Of the measures based on latent semantic analysis, only a measure orthogonal to the conceptual relations, which instead gauges the extent to which the concepts for the compound's head and the compound itself are discriminated, also reached significance. Results further indicate that in tasks requiring careful assesment of the meaning of the compound, general knowledge of conceptual relations plays a central role, whereas in the lexical decision task, attention shifts to co-activated meanings and the specifics of the conceptual relations realized in the compound's modifier family.[1]

Keywords: *conceptual relations, semantic transparency, relative entropy, morphological processing, compounds, CARIN theory.*

humbug

1. something designed to deceive and mislead
2. a willfully false, deceptive, or insincere person
3. an attitude or spirit of pretense and deception
4. nonsense, drivel
5. British : a hard usually mint-flavored candy

http://www.merriam-webster.com

Most derived words and compounds have one or more meanings that are not fully predictable, or even totally unpredictable from the meanings of their

---

Corresponding author: harald.baayen@uni-tuebingen.de

1 This research was supported in part by an Alexander von Humboldt research award to the second author. The authors are indebted to Petar Milin, Victor Kuperman and Christina Gagné for valuable discussion.

constituents. The entries of the complex words *worker* and *handbag* as given by the Merriam Webster's at http://www.merriam-webster.com in Table 1 illustrate this phenomenon. How does a reader arrive at the meaning 'a bag hung from a shoulder strap for carrying small personal articles' when reading *handbag*?

According to subsymbolic connectionist models (Seidenberg & Gonnerman, 2000; Gonnerman & Anderson, 2001; Harm & Seidenberg, 2004), word meanings are represented by patterns of activation over banks of units. The more two meanings are similar, the more similar their patterns of activations over the semantic units should resemble each other. In this approach, semantic similarity can be quantified by means of the cosine similarity between activation vectors (see, e.g., Moscoso del Prado Martín, 2003). Thus, the meanings of *bag* and *handbag* are supposed to have smaller cosine distances than the meanings of *bag* and *worker*. Connectionist networks typically learn the mapping of orthographic form to meaning better if words that share form also share meaning. This allows connectionist models to account for graded effects of semantic similarity in, for instance, studies using the primed visual lexical decision task (Seidenberg & Gonnerman, 2000; Gonnerman & Anderson, 2001).

Table 1. Dictionary entries for *worker* and *handbag* in the online Merriam Websters.

worker	
1.	one that works especially at manual or industrial labor or with a particular material
2.	a member of the working class
3.	any of the sexually underdeveloped and usually sterile members of a colony of social ants, bees, wasps, or termites that perform most of the labor and protective duties of the colony
handbag	
1.	suitcase
2.	a bag held in the hand or hung from a shoulder strap and used for carrying small personal articles and money

The idea that word meanings can be quantified by means of vector spaces is the conceptual cornerstone of latent semantic analysis (Landauer & Dumais, 1997). Instead of a word's semantics being defined as a pattern of activation over a conceptual feature space, a word's semantics is derived from its co-occurrence statistics across different documents. This operationalization of semantic similarity has been found to be predictive across a wide range of different tasks. Several studies have made use of latent semantic vector spaces to characterize degrees of semantic transparency (Moscoso del Prado Martín & Sahlgren, 2002; Rastle et al., 2004; Moscoso del Prado Martín et al., 2005; Jones et al., 2006; Gagné & Spalding, 2009) in studies addressing morphological processing.

Table 2. Semantic relation coding schema, adapted from Shoben(1991). The last five conceptual relations marked with an asterisk are extensions to the original set of relations considered in the CARIN theory. H: head; M: modifier.

Relation	Example	Relation	Example
H CAUSES M	*flu virus*	H DERIVED FROM M	*peanut butter*
M CAUSES H	*job tension*	H ABOUT M	*budget speech*
H HAS M	*colledge town*	H DURING M	*summer clouds*
M HAS H	*lemon peel*	H USED BY M	*servant language*
H MADE OF M	*chocolate bar*	M LOCATION IS H	*murder town*
H MAKES M	*honey bee*	H BY M	*student vote*
H LOCATION IS M	*office friendships*	M LIKES H*	*age-long*
H FOR M	*plant food*	H OF M*	*bombshell*
H IS M	*canine companion*	H MADE BY M*	*anthill*
H USES M	*machine translation*	H RESEMBLE M*	*arrow-root*

Latent semantic analysis captures semantic similarity with a high degree of success. This is possible thanks to the fact that objects and actions that co-occur in the world and in our experience tend to co-occur in written documents. However, what is not explicitly written (or talked) about, escapes detection. As a consequence, the method can be quite insensitive to obvious differences between word meanings. For instance, the LSA score for *mailbag* and *punching bag* is 0.17, which is also the score for *mailbag* and *tea bag*. Yet a *tea bag* is a small bag holding tea, whereas a *punching bag* is ‘a stuffed or inflated bag usually suspended for free movement and punched for exercise or for training in boxing’ (http://www.merriam-webster.com). LSA scores inform us that meanings are similar, but not in what way they are similar.

The goal of the present study is to clarify to what extent semantic transparency and its processing consequences can be approximated and operationalized more precisely than is possible with latent semantic analysis for the case of noun-noun compounds in English. More specifically, we will pit LSA scores against quantitative measures building on the CARIN theory of conceptual interpretation.

Building on earlier work by Levi (1978), Warren (1978), and Shoben (1991), Gagné and Shoben (1997); Gagné (2001); Gagné et al. (2005) build on the assumption that conceptual relations such as listed in Table 2 capture important aspects of compound meanings. Central to the theory are the additional assumptions that there is a relatively small number of generic conceptual relations, and that the same general conceptual relation can be actualized across a great many different compounds. For example, *mountain cabin*, *milk virus*, and *water bird* are, according to CARIN theory, all interpreted with the same general H LOCATION IS M relation: A *mountain cabin* is a cabin located in the mountains.

A key insight of CARIN theory is that the distribution of conceptual relations in the set of compounds sharing a given target compound’s modifier, henceforth the target’s modifier family, is an important determinant of compound interpretation and compound processing speed. If the target’s modifier family is characterized by a highly heterogeneous set of conceptual relations, there is greater uncertainty about the conceptual relation appropriate for that target. By contrast, if the set of conceptual relations instantiated in the target's modifier family is dominated by a particular conceptual relation, then there is less uncertainty: the distributionally most strongly supported conceptual relation becomes the maximum likelihood choice for interpretation. CARIN theory therefore predicts that the compound *mountain stream* is faster to interpret than *mountain magazine* because the LOCATION relation used to interpret *mountain stream* is also used in a great many other compounds which have *mountain* as their modifier (e.g., *mountain cabin*, *mountain goat*, and *mountain resort*. By contrast, the ABOUT relation used in *mountain magazine* is not commonly used with other compounds which have *mountain* as their modifier, and hence *mountain magazine* is predicted to require more processing time. Interestingly, Gagné and colleagues observed that the relative frequency of that relation within the set of relations attested for the compound's modifier reaches significance across a variety of psychometric tasks. The question of interest to us is whether quantitative measures derived from CARIN theory allow us to predict the costs of semantic processing with greater precision than measures based on LSA vector semantics.

In what follows, we consider three measures based on CARIN theory. The first measure, C, evaluates the relative frequency of the compound's conceptual relation in the compound's modifier family $\mathcal{M}$. Let s_i denote the conceptual relation instantiated in the i-th compound, and let $n(s_i)$ denote the type count of compounds with the same conceptual relation in the set of compounds $\mathcal{M}$. Furthermore, let j range over the set $r(\mathcal{M})$ of different conceptual relations that are realized within $\mathcal{M}$. Then the CARIN relative frequency C_i for compound i is defined as

$$C_i = \frac{n(s_i)}{\sum_{j \in r(\mathcal{M})} n(j)}, \tag{1}$$

where $n(j)$ denotes the number of compounds in $\mathcal{M}$ with semantic relation j. This measure is closely related to the strength metric proposed by Gagné and Shoben (1997). There are two differences in our operationalization, however. Whereas Gagné and Shoben (1997) take into account (in addition to the compound's own conceptual relation), only the three highest ranked competing relations for the denominator, we take all conceptual relations into account. Furthermore, Gagné and Shoben (1997) make use of exponential decay functions, we use straightforward probabilities.

The second measure gC does not restrict itself to the set of compounds realized within the compound's modifier family, but generalizes the measure by evaluating the compound's conceptual relation against the full set of conceptual relations in the lexicon $r(\mathcal{L})$. Let s_i denote, as before, the conceptual relation instantiated for compound i, let $m(s_i)$ denote the number of compounds in the language which share this conceptual relation, and let $m(j)$ denote the number of compounds in the language with conceptual relation j. We now define gC as

$$gC_i = \frac{m(s_i)}{\sum_{j \in r(\mathcal{L})} m(j)}. \tag{2}$$

The third measure evaluates the difference between the probablity distribution p of the conceptual relations within the modifier family $\mathcal{M}$ and the probability distribution q for the lexicon $\mathcal{L}$ by means of the Kullback-Leibler divergence or relative entropy measure

$$reC = D(p||q) = \sum_i p_i \log_2(p_i/q_i), \tag{3}$$

a measure which Milin et al. (2009a) have shown to be predictive for inflectional paradigms in Serbian, Baayen et al. (2008b) for inflectional paradigms in Dutch, and Milin et al. (2009b) and Kuperman et al. (2010) for derivation in English. For the calculation of *reC*, it is necessary to back off from zero, which we did by adding 0.5 to the counts underlying the probabilities p_i and q_i.

These three CARIN measures are pitted against three measures obtained with latent semantic analysis (Landauer & Dumais, 1997), using the default settings for pairwise comparison at http://lsa.colorado.edu/. The first measure, Modifier-Compound LSA Similarity, is the cosine similarity of the modifier and the compound. The second measure, Head-Compound LSA Similarity, is the cosine similarity of the head and the compound, and the third measure, Modifier-Head LSA Similarity, is the cosine similarity of the modifier and the head. Of these measures, those involving the modifier are the best candidates for capturing aspects of the conceptual relations in which the modifier is entangled. The Head-Compound LSA Similarity may capture exemplar-category relations, such as *top soil* being a kind of *soil*, and a *honey bee* being a kind of *bee*. This semantic dimension is orthogonal to that of the conceptual relation between head and modifier in the compound. Given the positive results obtained in other studies with vector semantics, it is conceivable that the CARIN measures turn out to be predictive and capture effects of processing conceptual relations, and that of the LSA measures, only the Head-Compound LSA Similarity measure turns out to be predictive.

Before we present the details on the database used to calculate the CARIN measures, and on how these measures perform as predictors for experimental data, we first clarify the general framework within which we conceptualize the core ideas of CARIN theory.

We begin with noting that the meaning of a compound is not very well predictable from the meaning of its constituents, even if the conceptual relation between modifier and head is given. For instance, how exactly the meaning of *hand* functions in the meaning 'a bag hung from a shoulder strap and used for carrying small personal articles and money' is unclear. A *handbag* can be described as a 'a bag held in the hand' (see Table 1, using the H FOR M or H LOCATION is M conceptual relations), but a shopping bag is usually also held in the hands, and yet is not referred to as a handbag. As a consequence, in order to understand what a handbag is from speech or writing, the co-occurrence of *hand* and *bag* in this order is crucial. How exactly the orthographic form of *handbag* mediates understanding the meaning of *handbag* is hotly debated.

Early models posited that a complex word is accessed through its constituents (see, e.g., Taft and Forster, 1976; Taft, 1985, 1994). More recently, reading has been argued to involve obligatory blind early morphographic decomposition Rastle et al. (2004); Rastle and Davis (2008), but see Feldman et al. (2009) for counterevidence. As there is no rule that can compositionally derive the meaning 'a bag hung from a shoulder strap and used for carrying small personal articles and money' from the meanings of *hand* and *bag*, the role of constituents in this class of models reduces to that of hash codes. The compound *handbag* is broken up into the orthographic hash codes **`hand`** and **`bag`**, which jointly provide a pointer to the rich meaning of *handbag*.

This class of models is challenged, however, by the presence of whole-word frequency effects in the first fixation durations on compounds in Dutch, Finnish, and Japanese (Kuperman et al., 2008, 2009, 2010; Miwa, 2013). For instance, Kuperman et al. (2009) reported an eye-tracking experiment with 2,500 polymorphemic Dutch compounds presented in isolation for visual lexical decision while readers' eye movements were registered. They observed simultaneous effects of compound frequency, left constituent frequency, and left family size for the first fixation duration (i.e., before the whole compound has been scanned). Effects of right constituent frequency and right family size emerged only at later fixations. The time course of compound reading that emerges from the eye-tracking record is incompatible with models positing an obligatory first stage at which all the hash codes would be parsed out, followed by a subsequent stage at which the hash codes would provide access to the compound's meaning.

Our working hypothesis is that the (early) compound frequency effect indexes the activation of the compound's specific meaning. The greater the frequency of a compound, the faster this meaning (and the specific conceptual relation instantiated with this meaning) becomes available. The constituents of the compound likewise activate their meanings, and do so faster when they have a higher frequency of occurrence.

Depending on the task, these meanings may engage with the conceptual relations available in the language, to build a potential interpretation for the compound — the system's best guess at what the compound might mean if the compound were novel. The idiosyncratic aspects of a compound's meaning, however, are not predictable from the constituents' meanings and a conceptual relation. As a consequence, there will always be some friction between a compound's actual meaning and the meaning that can be projected from the constituents with the help of a conceptual relation.

Two opposing outcomes are logically possible for the *gC* and *C* CARIN measures. First, greater values could result in shorter processing latencies under the assumption that the relations make it easier to evaluate the compound's meaning. The more likely a conceptual relation is, the faster it would become available and the more quickly a compound's meaning would be computed. Second, it is also conceivable that larger values of *gC* and/or *C* might lead to elongated processing latencies if the availability of a relation would lead to deeper semantic evaluation, resulting in a compositional reading for the compound that would compete with the compound's idiosyncratic meaning.

When the probability distribution of the conceptual relations in the compound's modifier family differs markedly from the general distribution in the language, a conflict is predicted, the resolution of which should require more processing time. This potential conflict is gauged with *reC*, and we therefore expect *reC* to be positively correlated with processing time, and to lead to reduced ratings.

In what follows, we first describe the database with conceptual relation statistics that we compiled. We then report two experimental studies, followed by one study analyzing response latencies in the English Lexicon Project. We conclude with a discussion of the implications of our findings.

STUDY 1: A DATABASE OF CONCEPTUAL RELATIONS

Most of the studies on conceptual relations have focussed on a small set of compounds for which conceptual relations have been carefully assessed and validated. As a first step, we constructed a database of conceptual relations that is much larger and enables regression studies with the *C*, *gC* and *reC* as predictors.

For a random sample of 783 English compounds, all compounds sharing their modifier constituent with any of these 783 compounds were selected from the CELEX lexical database, resulting in 3455 compounds. For each compound, its meaning was extracted from WordNet (Fellbaum, 1998). The first author then coded all compounds with four types of information: Semantic Type, Semantic Relation, Semantic Modifier, and Semantic Head. Semantic Type specified whether the compound is Transparent, Partially Opaque, or Fully Opaque. These individual judgements of transparency were further validated with a rating experiment (Study 3, reported below).

Semantic Relations were coded following Gagné and Shoben (1997), extended with the four relations in the lower half of Table 2. For a proper evaluation of the conceptual relations, it was necessary to distinguish between the dominant meanings of the constituents in isolation (e.g., *air* in *airstrip* denoting the gas we breathe) and the meaning of the constituent within the compound. For instance, *air* in compounds such as *airstrip* and *airport* denotes, either by shortening or by metonomy, the vehicle for traveling through air, i.e., its semantics within a series of compounds is *airplane* or *aircraft*. We therefore distinguish between the semantic modifier (*airplane*) and the modifier itself (*air*). The semantic relation for *airport* is therefore defined as a port for airplanes (H FOR M). For a compound like *backlash*, we distinguish between the semantic modifier *adverse* and the semantic head *(violent) reaction*. Even though the meaning of *backlash* is not straightforwardly derivable from the meanings of *back* and *lash*, the latter words do contribute to the meaning of the compound through structural metaphors (forwards is good, backwards is bad) and partial semantic inheritence (Lakoff & Johnson, 1980). Hence, the semantic relation for *backlash* is defined as H IS M, i.e., a violent reaction that is adverse. For exocentric compounds such as *camel-hair*, the semantic modifier is the material *camel-hair* and the notional head is *cloth*. Here, the semantic relation is cloth made of camel hair (H MADE OF M).

The information about the conceptual relations for the 3455 compounds sharing a modifier with any of the 783 target compounds allowed us to calculate the three measures introduced above for gauging the role of conceptual relations in lexical processing: the CARIN strength measure *C*, the generalized strength measure *gC*, and the relative entropy measure *reC*. We then built a database for the 783 target compounds and their values for the three CARIN measures. Figure 1 summarizes the distributions of these predictors by means of histograms. The distribution of *C* is marked by a large proportion of compounds (41%) with a modifier for which only one conceptual relation is attested in the modifier family.

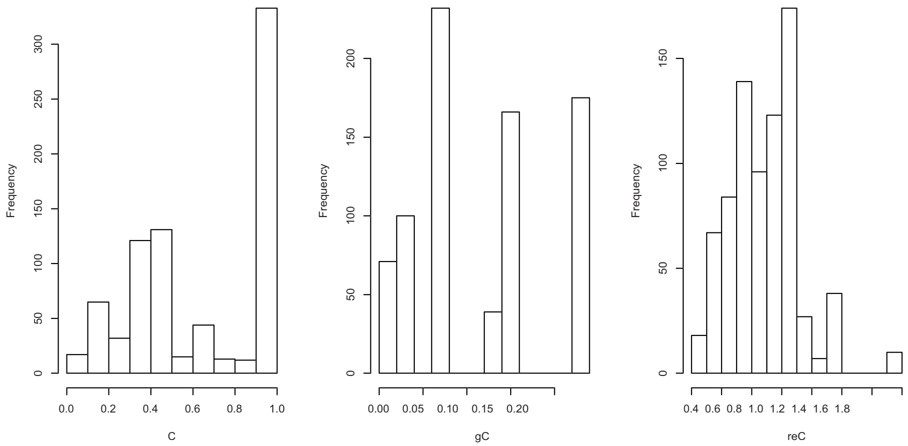

Figure 1. The distributions of the CARIN measures *C*, *gC*, and *reC*.

The data set was further enriched with the frequencies of the modifiers, the heads, and the whole compounds, which were collected from the COCA corpus Davies (2010). The degree of orthographical overlap of the modifier and head constituents was estimated with the Levenshtein distance measure (Levenshtein, 1966) (the minimal number of deletions, insertions, or substitutions that is required to transform the source string into the target string), following Baayen et al. (2011). We further calculated Compound Entropy (defined over the probability distribution of modifier and head, cf. Baayen et al. (2008b) and Baayen (2010)), and Modifier and Head Family Size (De Jong et al., 2002; Baayen et al., 2010. Three measures evaluating semantic similarity using latent semantic analysis were also added: Modifier-Head LSA Similarity, Modifier-Compound LSA Similarity, and Head-Compound LSA Similarity. Finally, each compound was assigned the value for its semantic transparency (transparent, partially opaque, or fully opaque) in our dataset of semantically analyzed compounds.

The experiment to which we now turn was designed to assess the predictivity of the CARIN measures, as well as the semantic transparency classes that we established. The experiment requested subjects to specify whether they knew the meaning of the compound. We registered both the answers and the response latencies.

STUDY 2: FAMILIARITY CLASSIFICATION EXPERIMENT

Method

Materials. A total of 1313 compounds was randomly selected from the words for which we collected the CARIN statistics described above.

Subjects. Thirty-three native English speaking undergraduate students from the University of Alberta were participated for partial course credit. All had normal or corrected-to-normal vision.

Procedure. Stimuli were presented on a 19-inch Lenovo CRT monitor with a refresh rate of 85 Hz and a resolution of 1,024 × 768 pixels. The monitor was controlled by a Pentium 4 3-GHz PC. Stimuli were presented in lowercase in 20-point Arial font, and they appeared as black characters on a grey background. Stimuli were presented one at a time, following a fixation point, and centered around this fixation point, using DMDX version 4.0.4.4 (Forster & Forster, 2003). Five practice items preceded the experimental items. Trial presentation was self-paced. Participants initiated a trial by pressing the space bar. Paricipants were given a short break after the practice trials and were encouraged to ask the experimenter to explain parts of the procedure that they did not understand.

Subsequently, participants were presented with 125–147 of the experimental compounds, presented in random order.

In this experiment, which took about 25 minutes to complete, participants were asked to report whether they knew the meaning of the compound, with as possible responses “Yes”, “I can guess what this means”, and “I have no idea what this means”, using the number keys 1, 2, and 3 on the number keypad. After their answer, they pressed the space bar to move to the next trial.

Participants were asked to place the fingers of their left hand on the number keys 1–3 on the top left of the computer keyboard, and to place the index finger of their right hand on the space bar key, in preparation for responding. Subjects were assigned randomly to one of the sub-experiments. Both the button presses and the response latencies were recorded.

Results and Discussion

Response Latencies. Response latencies were analysed with a generalized additive mixed model (GAMM) with subject and item as crossed random effect factors (see, e.g., Baayen et al., 2008a; Wood, 2006, 2011).

We performed a stepwise variable selection procedure in which non-significant predictors were removed to obtain a parsimonious yet adequate model. Subsequently, potentially harmful outliers, defined as data points with standardized residuals exceeding 2.5 standard deviation units, were removed from the data set. We then fitted a new model with the same predictors to this trimmed data set. This procedure was followed also for the analyses reported below.

The optimal model incorporated by-subject and by-item random intercepts, as well as by-subject random slopes for compound frequency. Other random-effect parameters were considered, but were found not to improve the model fit, and were therefore removed from the model specification. Table 3 summarizes the resulting GAMM. The upper part of the table presents the parametric coefficients, the lower part the spline (a mathematical technique for modeling wiggly lines) and tensor smooths (a mathematical technique for modeling wiggly surfaces) and the random effect components. Here, edf denotes the estimated degrees of freedom and Ref.df. the reference degrees of freedom for the F-test. For examples of generalized additive models in (psycho)linguistics, see, e.g., Baayen et al. (2010); Tremblay and Baayen (2010); Kryuchkova et al. (2012); Balling and Baayen (2012); Wieling et al. (2011) and Koesling et al. (2012).

Table 3. Generalized additive model for the familiarity categorization latencies of Study 2.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	6.5510	0.0783	83.6587	< 0.0001
Compound Frequency	-0.0679	0.0098	-6.9202	< 0.0001
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
tensor for Compound Entropy and gC: C < 1	4.7182	5.1093	4.3826	0.0005
tensor for Compound Entropy and gC: C = 1	4.9448	5.2763	2.6929	0.0178
spline for Head-Compound LSA Similarity	3.0290	3.4962	12.4040	< 0.0001
random intercepts for Item	479.7905	862.0000	1.3759	< 0.0001
random intercepts for Subject	30.3427	32.0000	28.1783	< 0.0001
by-subject random slopes for Compound Frequency	13.1893	32.0000	0.8708	0.0081

Compound frequency had the expected facilitatory effect: The more often a compound is encountered, the better it is known, and the faster its meaning can be retrieved.

A three-way, nonlinear interaction emerged involving Compound Entropy by the CARIN relative entropy measure *reC* and a factor specifying whether *C* was equal to one, as shown in the upper panels of Figure 2. When $C = 1$, only a single conceptual relation is attested in the compound's modifier family. In this case, a nearly uniform distribution for the compound is compared with the general distribution in the language. Therefore, for $C = 1$, the relative entropy measure is stretched to its limits, and the statistical support for its predictivity is restricted. For less trivial distributions of conceptual relations in the compound's modifier family, the interaction is more robust. Here we see that latencies decrease for smaller values of Compound Entropy, and that especially for these smaller values of Compound Entropy, a greater CARIN relative entropy predicts longer latencies. This processing cost associated with atypicality is consistent with costs detected with the relative entropy measure in other paradigmatic domains (see, e.g., Milin et al., 2009a,b; Baayen et al., 2011).

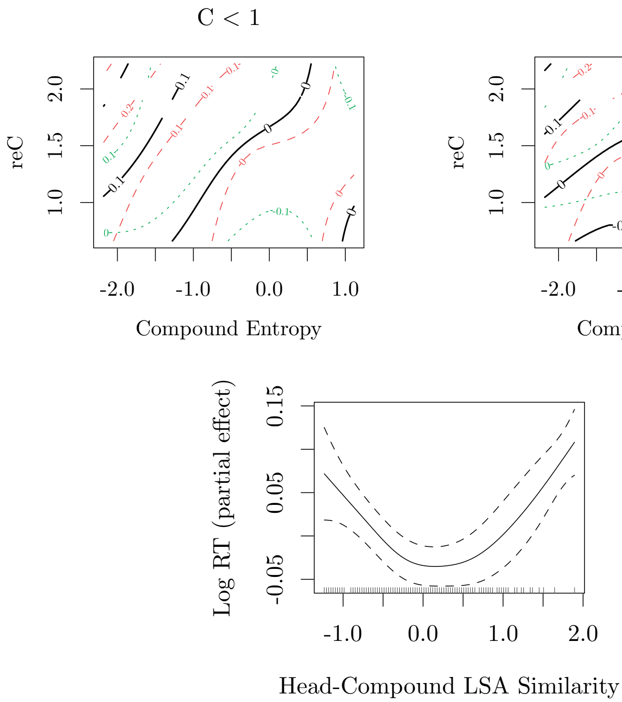

Figure 2. The tensor smooths for Compound Entropy by *gC* for compounds with multiple conceptual relations (top left) and a single conceptual relation (top right) for their modifier family, and the spline smooth of Head-Compound LSA Similarity (bottom), for the response latencies in the familiarity classification task. In the upper panels, contour lines connect points with the same log RT. Dotted lines denote 1 standard error regions up from their contour line, dashed lines denote the 1 standard error lower border.

As expected, the Modifier-Head and Modifier-Compound LSA Similarity measures were not predictive at all. However, a strong, non-linear effect of Head-Compound LSA Similarity was present, as shown in the bottom panel of Figure 2. Responses are fastest for words with intermediate values. Since the head of an (endocentric) compound typically denotes a general category of which the compound denotes a subcategory (e.g., *tree* versus *olive tree*), for highly similar head and compound pairs, the difference between the two may be difficult to discern. For instance, the compounds with the highest Head-Compound LSA Similarity in our data set are *honeybee* and *topsoil*, which are near synonyms of *bee* and *soil*. For such pairs, discrimination between the two appears to come with a cost, possibly because the modification is more difficult in case concepts that are very similary have to be distinguished. Conversely, for very low Head-Compound LSA Similarity, there is no sensible exemplar/category relation between head and compound (compare, e.g., *bug* and *humbug*). This appears to also give rise to a processing cost, see also Ji et al. (2011) for independent evidence suggesting that opaque compounds give rise to greater processing costs when the task emphasizes the importance of semantic interpretation.

Familiarity Responses. An analysis of the familiarity responses, using a random forests analysis with the **party** package for **R** (Hothorn et al., 2006; Strobl et al., 2009) ranks Compound Frequency as the most important predictor, followed by Compound Entropy (see Table 4). The importance of these two predictors is also supported by logistic regression modeling with as response variable whether the response was `yes', as shown in Table 5. Other predictors did not reach signifance in this analysis.

Table 4. Variable importance of the predictors for the familiarity decision using a random forest with conditional inference trees as base learners.

Predictor	Variable Importance
C	0.0003
Transparency	0.0005
Levenshtein Distance	0.0006
Length	0.0006
reC	0.0007
LSA Head Compound	0.0020
Modifier Family Size	0.0023
gC	0.0025
Compound Entropy	0.0036
Compound Frequency	0.0040

We next consider an experiment in which subjects were asked to evaluate the transparency of compounds on a seven point Likert scale. As this task seeks explicitly to tap into knowledge of how the constituents in a compound make sense, we expected the CARIN measures to be predictive. As the head-compound exemplar/category relation is irrelevant to this task, we judge it unlikely for the Head-Compound LSA Similarity measure to be predictive.

Table 5. Coefficients of the generalized mixed-effects model fitted to the familiarity response, dichotomized to ’yes’ versus ’no’ and ’uncertain’.

	Estimate	Std. Error	z-value	p-value
Intercept	4.2119	0.1799	23.4182	<0.0001
Compound Frequency	1.2607	0.1502	8.3924	<0.0001
Compound Entropy	0.4530	0.1080	4.1952	<0.0001

STUDY 3: TRANSPARENCY RATING EXPERIMENT

Method

Materials. The materials were identical to the materials of Study 2.

Subjects. The subjects were identical to those that participated in Study 2. All subjects first completed Study 2, before taking part in Study 3.

Procedure. Subjects were taken through the list of compounds (presented in a new random order), but now a compound was presented together with a sentence describing its meaning. They were asked to rate the compound's transparency on a 7-point scale, specifically with respect to whether the constituents of the compound help to understand its meaning. Answers ranged from "Not at all" (1) to "Fully" (7). After providing an answer on the numeric keypad, they pressed the space bar to move to the next screen. Note that by presenting a sentence describing the meaning of the compound, even those subjects who were initially unfamiliar with the compound's meaning were enabled to provide a reasonably informed judgement about the transparency of the compound.

Results and Discussion

Analyses were performed on only those compounds that participants were familiar with according to their responses in Experiment 1. As a consequence, 3% of the data points were removed, resulting in a dataset with 4424 data points.

Rating latencies were long and revealed no effects of interest, and are therefore not discussed further. The rating scores were analysed with a linear mixed effects model with subject and item as crossed random effects. (Analysis with generalized additive models were carried out, but no nonlinearities were observed.) The most parsimoneous yet adequate model incorporated four parameters for the random-effects structure of the data, all of which were supported by likelihood ratio tests: standard deviations for the random intercepts for subjects and items, a standard deviation for by-subject random slopes for compound frequency, and a correlation parameter for the two by-subject random effect components.

Table 6. Coefficients of the mixed-effects model fitted to the transparency ratings. The reference level of Semantic Type is Opaque.

	Estimate	Std. Error	t value
Intercept	4.1795	0.3705	11.2804
Semantic Type: partially opaque	1.2371	0.3442	3.5939
Semantic Type: transparent	1.9426	0.3244	5.9884
gC	1.3627	0.5583	2.4409
reC	-0.3475	0.0927	-3.7467
Compound Frequency	0.1262	0.0439	2.8774
Modifier Family Size	0.0931	0.0382	2.4387
Compound Entropy	0.1075	0.0368	2.9259

The mixed-effects covariance model fitted to the transparency ratings is summarized in Table 6. Higher-frequency compounds elicited higher transparency ratings, as did compounds with larger modifier families. A greater compound entropy likewise came with increased ratings. The enhancement in the ratings is consistent with the general effect of Compound Entropy in Study 2, where a greater Compound Entropy afforded reduced response latencies. The effect of semantic type was well supported, indicating that our manual classification is well in line with the intuitions of native speakers of English.

Although including Semantic Relation (a factor with no less than 22 levels) did improve the model fit, a simpler model with a better fit was obtained by replacing Semantic Relation by two CARIN measures. The greater the probability of the compound's conceptual relation in the language, as indexed by *gC*, the higher the transparency of a compound is rated. In addition, the relative entropy measure *reC* reached significance, with a negative slope indicating that a greater difference between the compound's modifier's probability distribution of conceptual relations and the corresponding general probability distribution predicts lower (less transparent) ratings. In other words, if a modifier has an atypical distribution of conceptual relations, then it is felt to be more opaque. This dovetails well with the nding that in the familiarity categorization task, response latencies increase for increasing *reC*.

Our first experiment specifically addressed subjects' knowledge of the meaning of compounds. The rating experiment addressed the same subjects' understanding of the interpretation of the constituents in the compounds. Our analysis restricted itself to those compounds that they had reported previously as familiar, and we provided a definition of the meaning of the compound to make sure all participants are targeting the same meaning. The next and final study turns to the lexical decision task, which by its nature allows subjects to respond when there is enough evidence for a general lexicality decision, without forcing them to actually zoom in on precisely the meaning of the word in the visual input (see, e.g., Grainger & Jacobs, 1996). In the lexical decision task, we therefore expect to see more evidence for predictors tied to the constituents. For instance, lexical decisions can be informed simply by the co-activation of the meanings of the constituents. Since previous studies in the framework of CARIN theory have worked with the *C* measure, it is conceivable that in lexical decision, this measure gains importance.

STUDY 4: LEXICAL DECISION LATENCIES IN THE ENGLISH LEXICON PROJECT

Response latencies in visual lexical decision for the 783 English compounds for which we assessed the conceptual relation between modifier and head (Study 1) were extracted from the English Lexicon Project (Balota et al., 2007).

Table 7 and Figure 3 summarize the generalized additive model that we fitted to the log-transformed response latencies. As expected, longer words elicited longer latencies. Words with a greater Compound Entropy were responded to more quickly.

As expected, more constituent effects were observed than in the preceding experiments. Modifier frequency was facilitatory, and the same holds for modifier and head family size. These predictors entered into an interaction (see the upper left panel of Figure 3) suggesting that facilitation from the Modifier Family Size is present only for smaller head families. Furthermore, most of the facilitation from head family size is present for intermediate values of this measure. The effect of Levenshtein Distance (see the lower left panel of Figure 3) indicates that orthographically less confusable constituents were activated more strongly, leading to shorter response latencies.

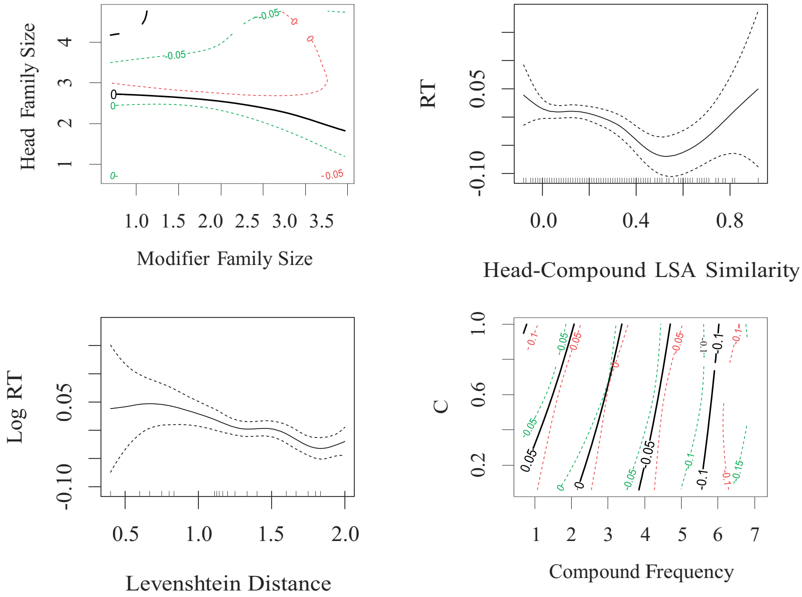

Figure 3. Partial effects of the tensor and spline smooths in the generalized additive model fitted to the lexical decision latencies to noun compounds in the English Lexicon Project (Study 4). Dotted lines denote 1 standard error regions up from their contour line, dashed lines denote the 1 standard error lower border.

Table 7. Generalized additive model for the visual lexical decision latencies to 783 compounds from the English Lexicon Project

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	6.62	0.04	164.11	< 0.0001
Compound Entropy	-0.05	0.02	-3.24	0.00
Word Length	0.01	0.00	3.41	0.00
Modifier Frequency	-0.01	0.00	-2.41	0.02
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
tensor smooth for modifier and head Family Size	6.25	7.76	3.44	0.00
tensor smooth for Compound Frequency and C	3.05	3.10	32.38	< 0.0001
spline smooth for Head-Compound LSA Similarity	4.80	5.87	3.75	0.00
spline smooth for Levenshtein Distance	4.88	5.86	3.31	0.00

An effect of Head-Compound LSA Similarity was present, and roughly U-shaped (see the upper right panel of Figure 3), replicating the U-shaped pattern for the familiarity classification response latencies. However, whereas the effect in familiarity classification is quite symmetric, the effect in lexical decision is mostly facilitatory, with broad confidence intervals in the right tail of the distribution. Discriminating between *honey bee* and *bee*, and between *top soil* and *soil*, appears to be less of a problem, consistent with the lexical decision task not requiring subjects to identify the precise meaning of the compound.

Of the CARIN predictors, the measure estimating the probability of the compound's conceptual relation in its modifier family reached significance in interaction with Compound Frequency, as shown in the lower right panel of Figure 3. A strong effect of Compound Frequency is modulated by *C* such that for higher values of *C*, more facilitation occurs (the contour lines are closer together). At the same time, an inhibitory effect of *C* is present that decreases with increasing compound frequency. The inhibitory effect of *C* may indicate that a lexical decision is more dificult to make when two readings of the compound become available at the same time: the compound's idiosyncratic meaning, and the meaning predicted from its constituents and the semantic relation. The lower the frequency of the compound, the more the compositional meaning appears to intrude, leading to elongated lexical decisions..

The modifier-specific effect of the CARIN strength *C* emerges together with strong evidence for the activation of the meaning of the modifier, provided by the modifier frequency and family size effects — effects that are absent in the familiarity categorization latencies. The full pattern of results suggest that in the lexical decision task, but not in the familiarity categorization task, the meaning of the modifier is considered on its own, as evidence for lexicality, in conjunction with its (modifier) family. The probability of the compound's conceptual relation given its modifier family, as estimated by *C*, comes into play, and more so when compound frequency is low. For high-frequency compounds, the compound's meaning, including how modifier and head relate to each other (if they do so), becomes available quickly, obviating interpretational processes based on the modifier and its properties. As compound frequency decreases, the role of the modifier and its conceptual relation becomes more important.

GENERAL DISCUSSION

This study addressed the role of semantic transparency and the role of conceptual relations in compound processing. Whereas many previous studies have sought to assess semantic transparency through human ratings or similarity scores based on latent semantic analysis (e.g., Rastle et al., 2004; Moscoso del Prado Mart´ın et al., 2005; Jones et al., 2006; Gagné & Spalding, 2009; Libben et al., 2003), we have approached this question by means of key insights from CARIN theory (Gagné & Shoben, 1997). According to CARIN theory, the conceptual relation between modifier and head is an important part of a compound's meaning. We hypothesized that knowledge of how the modifier and the head relate to each other may be more important than the bare knowledge that modifier and the head, or the modifier and the compound, have a high probability of co-occurring across the same documents.

Three experimental studies, familiarity classification, transparency rating, and visual lexical decision, provide strong support for the importance of conceptual relations. Of the three LSA measures that we considered, only Head-Compound LSA Similarity reached significance, and this in only two out of three experiments. (For a summary overview of the effects in our experiments, see Table 8.) The effect of Head-Compound LSA Similarity was U-shaped in both familiarity classification and in lexical decision. We interpreted this U-shaped effect as a trade-off between exemplar/category discrimination (when head and compound are too similar in meaning, discrimination becomes more difficult) on the one hand, and a problematic exemplar/category relation (as in the case of opaque compounds such as *humbug*) on the other hand. Processing is optimal when the meanings of head and compound are not too similar and not too dissimilar. We have argued that the Head-Compound LSA Similarity measure is orthogonal to the conceptual relation between modifier and head. We included Modifier-Compound LSA Similarity and Modifier-Head LSA Similarity as measures that might help predict the semantic interpretation of compounds, but these measures turned out not to be helpful. As explained in the introduction, latent semantic analysis tends to capture how similar word meanings are, but it may fail to capture the specific way in which modifier and head are conceptually related.

We developed three CARIN measures that did turn out to be predictive: the probability of the compound's conceptual relation *C* in the modifier family, the generalized probability *gC* of the compounds' conceptual relation in the language, and the relative entropy *reC* gauging the extent to which the modifier-conditional and the unconditional probability distributions are dissimilar (see also Milin et al., 2009a). The generalized measure *gC* was predictive in the semantic transparency rating task. In the lexical decision task, the modifier-specific measure (*C*) reached significance. The relative entropy measure *reC* was predictive for both the transparency ratings and the similarity decision latencies.

Higher values of the relative entropy measure *reC* predicted lower transparency ratings. The more atypical a compound's modifier makes use of the conceptual relations in the language, the less transparent a compound is judged to be. This results fits well with other typicality effects traced with relative entropy measures (see, e.g., Milin et al., 2009a,b; Baayen et al., 2011).

The transparency ratings were also partially predictable from *gC,* with a greater general probability of the conceptual relation giving rise to higher transparency ratings. Since the modifier-specific CARIN measure *C* was not predictive for the ratings by itself — although it was relevant, albeit indirectly, through the relative entropy measure — we offer as a tentative hypothesis that in the rating task subjects primarily bring their higher-level knowledge of the distributional properties of conceptual relations in the language (reflecting their distribution of noun-noun affordances in the world as we experience it) into play, and do not condition specifically on the distribution of conceptual relations instantiated in the modifier family. In other words, we think that when subjects are evaluating how well a compound's constituents help understand its meaning, they evaluate its conceptual relation against the full range of possible conceptual relations in the language, without restricting themselves to the sub-world of the modifier. For instance, when evaluating *chocolate bar* (H MADE OF M), compounds such as *granola bar* (with the same conceptual relation) are co-evaluated when judging semantic transparency. What emerges from this study is that a transparency judgement does not appear to be a measure that is isolated to a particular compound, but rather appears to be based on a larger set of compounds.

For the familiarity classification task, the CARIN relative entropy measure *reC* interacted with Compound Entropy. In theory, Compound Entropy assesses the amount of information carried by the constituent meanings in the restricted semantic space spanned by the compound. In the visual lexical decision task, strongly co-activated, and potentially relationally uninterpreted and thus competing meanings of modifier and head provide evidence for lexicality, enabling shorter response latencies. Recall that in the semantic transparency rating task, having the meanings of modifier and head equally well available is, apparently, optimal for evaluating how well the (idiosyncratic) meaning of the compound makes sense. Returning to the familiarity classification task, we first note that the interaction of Compound Entropy by *reC* provides a strong indication that the effect of Compound Entropy is indeed a semantic effect, as Compound Entropy interacts with a semantic factor — *reC*. Interestingly, the tensor surface for the interaction of Compound Entropy by *reC* indicates that the effect of *reC* is more prominent for words with smaller values of Compound Entropy. For smaller values of Compound Entropy, the availability of modifier and head is unbalanced, possibly delaying access to the compound's idiosyncratic meaning, and allowing for a more prominent role of the evaluation of the conceptual relation between modifier and head. The greater the difference between the local and the global probability distributions of the conceptual relations is, the more idiosyncratic the compound is, and the longer it takes to evaluate whether it is familiar.

In the visual lexical decision task, *C* interacted with Compound Frequency, showing its strongest (inhibitory) effect for lower-frequency words. This suggests that when the mapping of form to meaning is strong enough, the compound's conceptual relation becomes available as part of its meaning, and further conceptual interpretation is obviated. For lower-frequency compounds, conceptual interpretation processes come into play that seek to predict the compound's meaning from the conceptual relations in its modifier family. These conceptual processes then lead to elongated responses. This pattern of results is reminiscent of a trade-off between compound frequency and modifier frequency in lexical decision and eye-tracking (Baayen et al., 2010), with facilitation from modifier frequency only for lower-frequency compounds. This pattern emerges already, albeit in diminutive form, in the first fixation, and becomes stronger in modifier subgaze durations, and strongest in the gaze durations (see their Figure 4 for further details). In other words, the less well the compound is known, the longer processing takes because interpretation comes to depend more on the modifier and its distribution of conceptual relations.

Inspection of Table 8 shows that the lexical decision task picks up more constituent frequency and family size effects than the familiarity classification task. Since the familiarity classification task targets precise knowledge of the word, we think that the constituent effects in lexical decision, including the inhibitory effect of C, may be the result of a multiple read-out strategy (Grainger and Jacobs, 1996) specific to the lexical decision task. Whereas the tasks that require focusing attention on the meaning of the compound itself (familiarity categorization and transparency rating) show few constituent effects but effects of gC and reC, the lexical decision task shows abundant constituent effects and a modifier-specific effect of C. This fits well with the requirements of this task: speeded decisions on whether the stimulus is likely to be a word, for which any semantic activation counts as evidence for lexicality.[2]

Table 8. Summary of predictors considered and the experiments, if any, in which they reached significance.

	Fam. Class. RT	Transparency Rating	Visual Lexical Decision
Compound Frequency	*	*	*
Compound Length	•	•	*
Modifier Frequency	•	•	*
Head Frequency	•	•	•
Modifier Family Size	•	*	*
Head Family Size	•	•	*
Head-Compound LSA Similarity	*	•	*
Head-Modifier LSA Similarity	•	•	•
Modifier-Compound LSA Similarity	•	•	•
C	•	•	*
gC	•	*	•
reC	*	*	•
Compound Entropy	*	*	*
Levenshtein Distance	•	•	*
Transparency Class	•	*	•

2 A reviewer suggested that it is conceivable that in the lexical decision task, due to the inclusion of compound nonwords combining a nonword constituent with a word constituent, it would be advantageous to strategically avoid the activation of constituents. However, the observed effects of modifier frequency as well as those of modifier and head family size indicate that participants are not able to selectively ignore constituent activation. Conversely, it might also be possible that participants are made aware of constituents precisely because existing words appear in nonword compounds. Finally, it is also conceivable that the observed pattern of results reflects a balance between becoming aware of constituents and seeking to ignore them in nonword compounds.

These considerations lead to the following perspective on compound processing in reading. We assume that the goal of reading a compound is to understand what that compound means. The more frequent the compound is, the better the meaning of the compound has been learned, allowing shorter response latencies. In tasks that direct attention to the specific meaning of the compound, the meanings of the constituents receive little or no attention. Responses and response latencies are shaped by general knowledge of conceptual relations and the relations in the world of which they are our experiental reflection. The more a compound's meaning makes sense against the backdrop of this general knowledge, the more we interpret it as transparent. At the same time, the more idiosyncratic the use of semantic relations in the compound's modifier family is, the longer it takes to decide on whether the compound's meaning is known, and the lower its transparency is judged to be.

However, in a task such as lexical decision, attention fans out to constituent meanings, the meanings of morphologically related words (Bertram et al., 2000; Moscoso del Prado Mart´ın et al., 2004; Dijkstra et al., 2005), and to the meanings of embedded words such as *hoe* in *shoelace* (Bowers et al., 2005; Baayen et al., 2007). Although the meanings of the many words that are fully or partially compatible with the orthographic input will tend to be at odds with the meanings relevant for interpretation — this holds even for the meanings of compounds' constituents, see Libben (2010) the co-activation of many meanings provides evidence for lexicality that speeds lexical decision latencies (Grainger & Jacobs, 1996). However, when attention fractionates over many meanings, the risk is that interpretation becomes chaotic. This predicts that conceptual processes mitigating the costs of chaos should come into play. This is exactly what we think is indexed by the inhibitory effect of C in the visual lexical decision task: An attempt is made at interpreting the compound given the conceptual relations in the modifier family.

This perspective on compound processing is very different from theories that posit a hierarchy of form units (letters, letter n-grams, morphemes, and words) that are supposed, through an interactive activation process, to single out a single form, to the suppression of all other forms. This selected form would then provide access to semantics. Our approach, by contrast, assumes a fast mapping from orthographic input to a broad array of meanings (Baayen et al., 2011), with the hard questions about what the compound actually means being solved at the level of semantics. This is the true arena of higher-level cognition, where typically among a wealth of information, attention is often focussed on some particular detail. Our hypothesis is that in reading, attention can be directed to the meaning of the compound, to the exclusion of co-activated meanings. When reading *shoelace*, we know that a *shoelace* is not a shoe, nor lace, but a small piece of rope used to tighten shoes, and that *humbug* denotes nonsense (or, in Britain, a hard mint-flavored candy) and has nothing to do with humming or with bugs. This situation is no different from a car driver in a busy street focusing on a traffic light: only one object from a rich array of objects in the visual scene is attended.

Because the language code, forced by the linear nature of speech, overloads old strings (*hoe*, *lace*) with new significances (*shoe, shoelace*), it can leave the reader with a comet's tail of semantic debris. The danger of the lexical decision task is that the experimental evidence for the debris is easily mistaken for the comet. Complementation by tasks that target the compound's meaning, such as familiarity categorization, and by eye-tracking studies (see, e.g., Kuperman et al., 2013, for non-decompositional access in sentential reading) are crucial for bringing the comet itself back into focus.

REFERENCES

Baayen, R. H. (2010). The directed compound graph of English. an exploration of lexical connectivity and its processing consequences. In S. Olsen (Ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17)* (pp. 383–402). Hamburg: Buske.

Baayen, R. H., Davidson, D. J., & Bates, D. (2008a). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baayen, R. H., Kuperman, V., & Bertram, R. (2010). Frequency effects in compound processing. In S. Scalise & I. Vogel (Ed.), *Compounding*. Amsterdam/Philadelphia: Benjamins.

Baayen, R. H., Kuperman, V., Shaoul, C., Milin, P., and Ramscar, M. (2013). Decomposition makes things worse. a discrimination learning approach to the time course of understanding compounds in reading. *Manuscript submitted for publication*.

Baayen, R. H., Levelt, W., Schreuder, R., & Ernestus, M. (2008b). Paradigmatic structure in speech production. *Proceedings Chicago Linguistics Society, 43*(1), 1–29.

Baayen, R. H., Milin, P., Filipović Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438–481.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words. a regression study across tasks and modalities. *The Mental Lexicon*, *2*, 419–463.

Balling, L. & Baayen, R. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, *125*, 80–106.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*, 390–405.

Bowers, J., Davis, C., & Hanley, D. (2005). Automatic semantic activation of embedded words: Is there a "hat" in "that"? *Journal of Memory and Language*, *52*, 131–143.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of english. *Literary and linguistic computing*, *25*(4), 447–464.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., & Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects. *Brain and Language*, *81*, 555–567.

Dijkstra, T., Moscoso del Prado Mart´ın, F., Schulpen, B., Schreuder, R., & Baayen, R. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes*, *20*, 7–41.

Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martin, F. (2009). Early morphological processing is morpho-semantic and not simply morpho-orthographic: evidence from the masked priming paradigm. *Psychonomic Bulletin & Review*, *16*(4), 684–691.

Fellbaum, C. (1998). *WordNet: An electronic database*. Cambridge, MA: The MIT Press.

Forster, K. & Forster, J. (2003). Dmdx: A windows display program with millisecond accuracy. *Behavior Research Methods*, *35*(1), 116–124.

Gagné, C. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 236–254.

Gagné, C., & Shoben, E. J. (1997). The influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 71–87.

Gagné, C., & Spalding, T. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language*, *60*, 20–35.

Gagné, C., Spalding, T., & Ji, H. (2005). Re-examining evidence for the use of independent relational representations during conceptual combination. *Journal of Memory and Language*, *53*(3), 445–455.

Gonnerman, L., & Anderson, E. (2001). Graded semantic and phonological similarity effects in morphologically complex words. In S. Bendjaballah, W. Dressler, O. E. Pfeiffer, & M. D. Voeikova (Eds.), *Morphology 2000: Selected papers from the 9th Morphology meeting* (pp. 137–148). Amsterdam: John Benjamins.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111,* 662–720.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, *7*, 355–373.

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*(4), 406–430.

Jones, M., Kintsch, W., & Mewhort, D. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552.

Koesling, K., Kunter, G., Baayen, R. H., & Plag, I. (2012). Prominence in triconstituent compounds: Pitch contours and linguistic theory. *Language and Speech*, in press.

Kryuchkova, T., Tucker, B., Wurm, L., & Baayen, R. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*, 81–91.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*, 1089–1132.

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, *62*, 83–97.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading of multimorphemic Dutch compounds: Towards a multiple route model of lexical processing. *Journal of Experimental Psychology: HPP*, *35*, 876–895.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady*, *10*(8), 707-710.

Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.

Libben, G. (2010). Compound words, semantic transparency, and morphological transcendence. *Linguistische Berichte Sonderheft, 17*, 212–232.

Libben, G., Gibson, M., Yoon, Y., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*, 50–64.

Milin, P., Filipović Durdević, D., & Moscoso del Prado Mart´ın, F. (2009a). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language, xx*, 50–64.

Milin, P., Kuperman, V., Kostić, A., & Baayen, R. (2009b). Paradigms bit by bit: an information-theoretic approach to the processing of paradigmatic structure in inflection and derivation. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: form and acquisition* (pp. 214–252). Oxford: Oxford University Press.

Miwa, K. (2013). *Emergence of wordlikeness in the mental lexicon: Language, population, and task effects in visual word recognition*. Doctoral dissertation, University of Alberta, Edmonton.

Moscoso del Prado Martín, F. (2003). *Paradigmatic effects in morphological processing: Computational and cross-linguistic experimental studies*. MPI Series in Psycholinguistics. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 1271–1278.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Hebrew and Dutch. *Journal of Memory and Language*, *53*, 496–512.

Moscoso del Prado Martín, F., & Sahlgren, M. (2002). An integration of vector-based semantic analysis and simple recurrent networks for the automatic acquisition of lexical representations from unlabeled corpora. In A. Lenci, S. Montemagni, & V. Pirrelli (Eds.), *Proceedings of the LREC'2002 Workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data*. Paris: European Linguistic Resources Association.

Rastle, K., & Davis, M. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7–8), 942–971.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morphoorthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*, 1090–1098.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*(9), 353–361.

Shoben, E. (1991). Predicating and nonpredicating combinations. In P. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 117–135). Hillsdale: Lawrence Erlbaum Associates, Inc.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*, *14*(4), 26.

Taft, M. (1985). The decoding of words in lexical access: A review of the morphological approach. In D. Besner, T. G. Waller, & G. E. Mackinnon (Eds.), *Reading research: advances in theory and practise* (pp. 83–123). London: Academic Press.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, *9*(3), 271–294.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, *15*, 607–620.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.

Warren, B. (1978). *Semantic patterns of noun-noun compounds*. Acta Universitatis Gothoburgensis, Göteborg.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), e23613.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society(B)*, *73*(1), 3–36.

Wood, S. N. (2006). *Generalized Additive Models*. New York: Chapman & Hall/CRC.