

# Modelling semantic transparency

Melanie J. Bell<sup>1</sup> · Martin Schäfer<sup>2</sup>

Received: 28 April 2015 / Accepted: 9 March 2016 / Published online: 13 April 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** We present models of semantic transparency in which the perceived transparency of English noun–noun compounds, and of their constituent words, is predicted on the basis of the expectedness of their semantic structure. We show that such compounds are perceived as more transparent when the first noun is more frequent, hence more expected, in the language generally; when the compound semantic relation is more frequent, hence more expected, in association with the first noun; and when the second noun is more productive, hence more expected, as the second element of a noun–noun compound. Taken together, our models of compound and constituent transparency lead us to two conclusions. Firstly, although compound transparency is a function of the transparencies of the constituents, the two constituents differ in the nature of their contribution. Secondly, since all the significant predictors in our models of compound transparency are also known predictors of processing speed, perceived transparency may itself be a reflex of ease of processing.

**Keywords** Semantic transparency · Compound noun · Compound processing · Semantic relation · Semantic shift · English

## 1 Introduction

The term ‘semantic transparency’ is often used in the linguistic literature, and variables intended to measure semantic transparency are regularly included in psycholinguistic experiments and theories, yet the nature of that transparency itself is not well

---

✉ M.J. Bell  
[melanie.bell@anglia.ac.uk](mailto:melanie.bell@anglia.ac.uk)  
M. Schäfer  
[post@martinschaefer.info](mailto:post@martinschaefer.info)

<sup>1</sup> Anglia Ruskin University, Cambridge, UK

<sup>2</sup> Friedrich-Schiller-Universität, Jena, Germany

understood. The aim of the present study is to address this gap by exploring the factors that correlate with more or less perceived transparency in the semantics of English compound nouns.

In linguistic morphology, the term ‘semantic transparency’ is used to describe how transparent the end product of a morphological process is with regard to its meaning. A textbook case of semantic transparency is exemplified by words like *readable* or *manageable* because, so writes Plag (2003: 46), ‘their meaning is predictable on the basis of the word-formation rule according to which they have been formed’. For compounds, which will be the focus of our study, we find that the criterion of meaning predictability is sometimes replaced by one of meaning relatedness. For example, Zwitserlood (1994: 344) deems a compound to be semantically transparent if its meaning is merely ‘synchronically related to the meaning of its composite words’. Note that *milkman*, the compound that Zwitserlood (1994: 344) uses to illustrate semantic transparency, is only transparent on the meaning relatedness view, not on the meaning predictability view. In this paper, we view semantic transparency as a scalar notion, which makes it possible to combine the two types of definition given above. We view semantic transparency as falling on a continuum, with meaning predictability constituting one end of the scale and total semantic opacity, i.e. no discernible synchronic relation between the meaning of a complex word and the meaning of any of its constituents, constituting the other end of the scale. Our central hypothesis is that the perceived semantic transparency of a compound can be understood as the degree of expectedness in its internal semantic structure: specifically, we show that human ratings of semantic transparency as a continuous variable can be partially predicted on the basis of quantitative measures of such expectedness.

Semantic transparency plays an important role in psycholinguistic models of the representation and processing of complex words, e.g. Marslen-Wilson et al. (1994) for derived forms and Libben (1998) for compounds. It also plays an important role in psycholinguistic experimentation. In particular, by manipulating semantic transparency, researchers can investigate the relationship between morphological and semantic levels of representation and the extent to which accessing the meaning of a morphologically complex word also involves the meanings of its constituents. Compounds are particularly well-suited for such investigations because there is considerable naturally-occurring variation in the transparency of compound meanings relative to the meanings of their constituents in isolation; compounds therefore allow researchers to avoid a confound between form and meaning more easily than with inflected or derived complex words (Zwitserlood 1994). Yet, despite the fact that semantic transparency plays such an important role in current psycholinguistic methodology and theorising, surprisingly little is known about this property from a linguistic perspective.

Most studies involving semantic transparency use it to investigate some other phenomenon, and only a very few works aim at explaining or modelling semantic transparency itself. However, so far as the semantic transparency of compounds is concerned, the psycholinguistic literature does point to the importance of the semantic transparency of individual compound constituents in producing the effects observed (e.g. Libben et al. 2003; El-Bialy et al. 2013). This is corroborated in the computational linguistic literature by Reddy et al. (2011), who show that human judgements

of the overall semantic transparency of compounds are highly correlated with judgments about the semantic transparency of their constituents. Arguably then, the first step in understanding what makes a compound more or less semantically transparent is to understand what factors make a compound's constituents more or less semantically transparent. In testing our hypothesis that semantic transparency can be understood as degree of expectedness, we will therefore focus initially on compound constituents. In particular, we investigate how far the perceived semantic transparency of compound constituents can be modelled on the basis of expectations about the specific senses of the constituents and the semantic relation between them.

Some points are in order here about terminology. Firstly, although compounds and their constituents can be more or less transparent in a number of ways, at very least semantically and phonologically, the focus of this study is on semantic transparency. For ease of exposition, in the remainder of the paper we will therefore use the terms 'transparent' and 'transparency' as synonymous with 'semantically transparent' and 'semantic transparency' respectively. Secondly, we should consider the distinction, if any, between 'transparency' and 'compositionality'. For Zwitserlood (1994: 366), even transparent compounds are usually not compositional since 'the meaning of the compound as a whole is often more than the meaning of its component words'. On this view, compositionality seems to be equivalent to the extreme meaning predictability end of our transparency continuum described above. For others, e.g. Reddy et al. (2011), compositionality is a gradient phenomenon, equivalent to our overall notion of transparency. Rather than committing ourselves to either of these views, we will simply try to point out the different usages where relevant in our discussion.

The rest of this paper is organised as follows: Sect. 2 gives an overview of the current state of research on the effects and nature of semantic transparency, culminating in the detailed predictions to be tested in the present study; Sect. 3 discusses the methodology of the study and Sect. 4 presents the resulting statistical models of transparency; Sect. 5 concludes.

## 2 The effects and nature of compound semantic transparency

### 2.1 Compound transparency and the mental lexicon

There is significant evidence that when someone reads a known compound word, representations of both the whole compound and its constituents are activated in the mental lexicon of the reader, irrespective of whether the compound is semantically transparent or opaque. This evidence comes from a variety of languages and experimental paradigms, usually involving some form of priming or eye-tracking. For example, Zwitserlood (1994: experiment 1) showed that reaction times in a lexical decision task for Dutch simplex words were significantly faster when subjects had previously seen a compound in which the word occurred as a constituent. The priming occurred for both transparent and opaque constituents in any position in the compound. Monsell (1985) and Libben et al. (2003), using constituents to prime for compounds, found similar results for English: reaction times to both transparent and opaque compounds were faster following exposure to either of the constituents. Jarema et al.

(1999) used the same method and found comparable results for French and Bulgarian, with the exception of fully opaque compounds in Bulgarian. Moving to the study of eye movements, Pollatsek and Hyönä (2005) showed that the time taken for initial processing during the reading of long compound words in Finnish was significantly inversely correlated with the frequency of the first constituent, i.e. compounds with more frequent first constituents were read faster. This effect was constant across transparent and opaque compounds, and independent of whole-compound frequency. The priming effects of constituents on compounds and vice versa can be explained by assuming that there is some form of link between them in the mind; similarly, the effects of constituent frequency on reading times can be explained if compounds are decomposed during recognition.<sup>1</sup> All these results can therefore be taken to support the view that, when a compound is read, mental representations of the constituents are activated regardless of how transparent the compound is. However, the results presented so far do not distinguish between representation of morphological form and representation of meaning; to tease apart these putatively different levels of representation it is necessary to use slightly different experimental techniques that more fully exploit variation in semantic transparency.

Several studies have attempted to separate out semantic or conceptual effects from purely morphological effects in the processing of compound words. One way of doing this is to use semantic priming, in which the prime is semantically associated with the target word, but is not identical: for example, reaction time to *ice* will be reduced following presentation of *snow*. This technique was used, for example, by Sandra (1990), who combined semantic priming of Dutch compound constituents with a lexical decision task. Distinguishing between transparent, opaque, and pseudocompounds, he found priming effects only for transparent compounds. Zwitserlood (1994: experiment 2) also used a combination of lexical decision and semantic priming to study the representation of Dutch compounds, this time with compounds as primes and monomorphemic words as targets. She distinguished between transparent, partially opaque and fully opaque compounds and found facilitatory priming effects only for the transparent and partially opaque ones. Zwitserlood (1994) interprets these results as showing that, although all compounds are mentally represented as morphologically complex, only transparent and partially transparent compounds are linked to the semantic representation of their constituents.

The view that transparent and opaque compounds are differently represented and processed at a semantic level is supported by evidence from eye-tracking. For example, Juhasz (2007) used eye-tracking to investigate the effects of semantic transparency on the reading of English compounds. Although her results largely replicated those of Pollatsek and Hyönä (2005) in finding no effect of transparency on measures of early processing, she did find a main facilitatory effect of transparency on go-past duration. Since go-past duration reflects the later stages of word reading, Juhasz (2007) argues that this result is analogous to those for semantic priming discussed above. She interprets the results as evidence that, although both transparent and opaque compounds are morphologically decomposed in the early stages of

<sup>1</sup> For a different explanation, see the results presented in Bowers et al. (2005) and the non-decompositional models in Baayen et al. (2016).

recognition, only transparent compounds are linked to the representations of their constituents at a semantic level; such connections at the conceptual level between a transparent compound and its constituents are assumed to lead to the observed processing advantage. Further support for this view comes from studies that attempt to inhibit direct access to the stored meanings of whole compounds. For example, Frisson et al. (2008), also using eye-tracking, found virtually no effect of semantic transparency on the processing of English unspaced compounds during normal reading. However, the same compounds written with a space did show an effect of transparency, such that opaque compounds took longer to read. The longer reading time in the spaced condition was taken as evidence that the meaning of opaque compounds is only associated with the whole-word forms and has no semantic link to the constituents. Overall, these studies suggest that the constituent meanings of opaque compounds are either not activated during reading or are very quickly suppressed by the whole-word meaning.

A problem for any theory that postulates that the semantic representations of constituents are not activated during the processing of opaque compounds is to account for how the system could know in advance that a constituent is opaque, i.e. in time to prevent its semantic activation (cf. Ji et al. 2011). This problem is avoided in the theoretical framework of conceptual combination (Gagné and Spalding 2004), in which semantic activation of constituents is assumed to occur during the processing of all compounds, opaque as well as transparent. Working in this framework, Ji et al. (2011) compared lexical decision times for transparent and opaque two-constituent English compounds with those for monomorphemic words matched for whole-word frequency and length. When the compounds were presented as single orthographic words, reaction times to both transparent and opaque types were faster than to the monomorphemic words. This is taken to reflect purely lexical activation of the compound constituents, which are likely to have a higher frequency than the compounds and frequency-matched monomorphemes. However, the processing advantage of the opaque compounds disappeared when the compounds were presented spaced or with the two constituents written in different colours. Under these conditions, which were intended to promote morphological decomposition, the opaque compounds patterned with monomorphemic words. The authors argue that the results can be explained if reading any compound leads to activation not only of lexical but also of semantic representations for both the whole compound and its constituents, with the system automatically trying to compute a compound meaning from the latter. Under normal conditions, the whole-word meaning of known opaque compounds will be quickly available, suppressing any computed meaning. However, when the experimental conditions more strongly favour the compositional route, two possible meanings may become available simultaneously. According to this theory, the time taken for the system to evaluate and decide between these possible meanings is what slows processing.

Another recent study that argues for a central role of semantic transparency in compound processing is Marelli and Luzzatti (2012). These authors report two experiments, one using lexical decision and the other using eye-tracking, whose results they analyse using mixed effects regression. Their target language is Italian, which allows them to manipulate the headedness of the compounds as well as semantic

transparency and measures of frequency and length.<sup>2</sup> In both experiments they found that significant predictors of processing time included interactions between semantic transparency, constituent frequency and head position. Furthermore, contrary to the previous eye-tracking studies reviewed above, effects of semantic transparency were found in very early stages of processing. Marelli and Luzzatti (2012) interpret their results as supporting a multi-route model of processing, in which conceptual combination as described by Gagné and Spalding (2004) is one possible route. In this multi-route model, lexical and semantic factors interact at all stages, with semantic transparency playing a key role in modulating the relative weights of different routes. Marelli and Luzzatti (2012) suggest that previous studies may have missed some effects of semantic transparency either because they used low-power factorial designs or failed to include the relevant interactions.

In summary, there is ample psycholinguistic evidence that, as asserted by Libben (1998: 35), ‘semantic transparency should play a prominent role in any model of compound representation and processing’. In the following sections we will therefore consider in more detail what exactly is meant by ‘semantic transparency’ in the context of compounds, and look at previous attempts to measure or model it.

## 2.2 What is semantic transparency?

### 2.2.1 Psycholinguistic perspectives

Compound semantic transparency is usually established through human judgements, most often using rating scales. In some cases, participants are explicitly asked to rate transparency, either the transparency of whole compounds (e.g. Pollatsek and Hyönä 2005; Juhasz 2007; Ji et al. 2011) or of compound constituents within specific compounds (e.g. Frisson et al. 2008). In other cases, participants are asked to rate aspects of meaning hypothesised to reflect transparency. For example, Zwitserlood (1994) asked her subjects to rate the extent to which the meaning of each compound was related to the meaning of its second constituent. Others have used a combination of two different questions, one targeting the compound and one targeting the individual constituents. For example, Libben et al. (2003) asked their participants to what extent the meanings of compounds were predictable from the meanings of their constituents and to what extent the constituents retained their original meanings within the compounds. Similarly, Marelli and Luzzatti (2012) asked how far the meaning of each compound could be predicted from the meaning of its constituents and how far the meaning of each constituent contributed to the meaning of the compound. Occasionally, even less direct tasks are used. For example, Sandra (1990) had subjects write out definitions for each compound and counted the number of times the individual constituents occurred. If a compound’s constituents never occurred in the definitions given for it, or occurred only very rarely, that compound was classed as opaque.

There is variation not only in the nature of the tasks used to establish transparency, but also in the number of transparency levels established. In some studies, semantic

<sup>2</sup>In Italian, noun–noun and adjective–noun compounds can occur with the head in either the left-hand or right-hand position. Marelli and Luzzatti’s (2012) noun–noun examples include e.g. left-headed *pesce.spada* ‘fish.sword’ ‘swordfish’ and right-headed *astro.nave* ‘starship’.

transparency is treated as a binary property of whole compounds, which are simply classed as either transparent or opaque (e.g. Sandra 1990; Juhasz 2007). Others make a ternary distinction between fully transparent, partially opaque and fully opaque compounds (e.g. Zwitserlood 1994), where partially opaque types have one constituent that is more transparent than the other. Perhaps most influentially, Libben et al. (2003) asked whether transparency is best viewed as a property of whole compounds or as a property of constituent morphemes. They used the four-fold distinction shown in (1):

- (1) TT (transparent–transparent) (e.g. *car-wash*)  
OT (opaque–transparent) (e.g. *strawberry*)  
TO (transparent–opaque) (e.g. *jailbird*)  
OO (opaque–opaque) (e.g. *hogwash*)

Libben et al. (2003: 53)

Using a variety of priming and lexical decision tasks, Libben et al. (2003: 53) found differences in processing between these four classes of compounds. They concluded that the semantic transparency of a compound as a whole is related to the transparency of its individual constituents, their position in the string, and their morphological and semantic roles in the compound, including whether or not they function as the head. The same four-way categorisation has successfully been used in several other studies, including e.g. Jarema et al. (1999), Frisson et al. (2008) and El-Bialy et al. (2013). Nevertheless, there is some recognition in the literature that even these four levels of transparency may not fully capture the phenomenon. For example, Libben (1998) argues that eight levels of transparency are required to explain the key experimental findings: for each of the four categories mentioned above he includes endocentric and exocentric variants. Furthermore, human ratings of transparency actually produce continuous variation; the need to reduce these ratings to discrete levels comes from factorial experimental designs, which, as Marelli and Luzzatti (2012) point out, may obscure some effects. In this study, we will therefore treat the semantic transparency both of compound constituents and of whole compounds as a continuous variable rated on a numerical scale.

The aim of this paper is to model the perceived transparency of compounds and their constituents on the basis of the expectedness of their semantic properties. In this respect, a further consideration comes from the work on conceptual combination carried out by Gagné and her colleagues, e.g. Gagné and Spalding (2004). Using priming and lexical decision, these authors showed that the semantic relation between constituents is used routinely in the access and use of both novel and familiar compounds. It is therefore possible that the expectedness of this semantic relation, either in absolute terms or in combination with particular constituents, is one of the factors that contribute to perceived transparency. Some initial evidence that the semantic transparency of a compound is a function not only of the transparency of its constituents but also of the semantic relation between them is presented in Bell and Schäfer (2013). In the present study, we therefore test the hypothesis that perceived transparency is correlated both with the expectedness of the constituents themselves and with the expectedness of this relation.



### 2.2.2 *Distributional semantics*

In the field of computational linguistics, where the term ‘compositionality’ is sometimes used for what we have been calling ‘transparency’, attempts have been made to model the property using distributional semantics. This approach rests on the assumption that, since the meaning of a word can be learnt by encountering it in a series of contexts, word meaning itself can be represented in terms of the contexts in which a word occurs in a large corpus of language. In this kind of model, a word’s meaning is represented in terms of its frequency of co-occurrence with a set of reference words. This is operationalised as a multi-dimensional vector, in which each dimension represents the co-occurrence of the word in question with one of the reference words. The distance between the overall vectors of two words is taken to reflect their degree of semantic similarity: the closer the vectors, the closer the meanings of the words. Lenci (2008), Sahlgren (2008) and Erk (2012) provide introductory overviews of this approach.

A notable distributional model of compound transparency is that of Reddy et al. (2011), who also created a database of human transparency ratings. Because we use these ratings as the dependent variable in the statistical models presented in this paper, we will describe the database and Reddy et al.’s (2011) study in some detail. The database is publically available<sup>3</sup> and consists of 30 independent transparency ratings for each of 90 two-part English compound nouns and their constituents: a total of 8100 ratings. The sample of compounds was selected semi-randomly in such a way as to maximise the probability that it included different degrees of semantic transparency. However, compounds were only included if they occurred at least 50 times in the ukWaC corpus (Ferraresi et al. 2008), a corpus of 2 billion word tokens. As a result, all the compounds in the dataset show some level of semantic lexicalisation, irrespective of their transparency rating. The annotators were recruited using the Amazon Mechanical Turk crowd-sourcing service<sup>4</sup> and tasks were assigned to annotators randomly. Each task consisted of two parts. Firstly, the annotator was presented with possible definitions of a given compound (AB) and asked to choose the definition that applied most frequently in five example sentences containing it. The definitions were based on WordNet (Fellbaum 1998) and the example sentences were selected at random from ukWaC. Secondly, the annotator was asked to use the example sentences and chosen definition to answer one of the following three questions, selected at random: either ‘how literal is the phrase AB?’ or ‘how literal is the use of A in the phrase AB?’ or ‘how literal is the use of B in the phrase AB?’ Thus, literality ratings were obtained for each compound as a whole and for each of the constituents in the context of that compound. Because of the random distribution of tasks, the same annotator did not necessarily answer all three questions (i.e. give all three ratings) for any given compound. In all cases, literality was rated on a scale of 0–5, with 0 meaning ‘not to be understood literally at all’, and 5 meaning ‘to be understood very literally’. The assumption on which we base the present study is that these ratings of literality are

<sup>3</sup>The Reddy et al. (2011) database is available at [http://sivareddy.in/papers/files/ijcnlp\\_compositionality\\_data.tgz](http://sivareddy.in/papers/files/ijcnlp_compositionality_data.tgz) [last accessed 22 February 2016].

<sup>4</sup>Amazon Mechanical Turk can be found at <https://www.mturk.com> [last accessed 22 February 2016].



equivalent to ratings of semantic transparency. The basis for this assumption is that the tasks used by Reddy et al. (2011) resemble those used in many psycholinguistic studies where, as discussed in Sect. 2.2.1 above, transparency is operationalised by asking human subjects to make some form of judgement about the relationship between the semantics of a compound and the semantics of its constituents. Since we assume that these judgements of literality reflect semantic transparency, we will henceforth refer to them as transparency ratings.

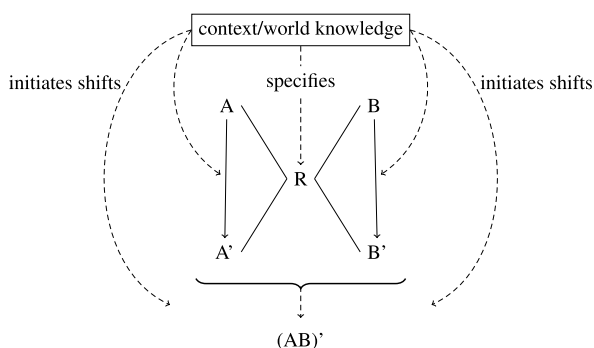
Before developing their distributional semantic models, Reddy et al. (2011) investigated the extent to which the mean transparency scores of the compounds in their database could be predicted from the mean scores of their constituents, either singly or in combination. They found that the combined transparency scores of the two constituents predict compound ratings much more successfully than the scores of either constituent alone, irrespective of whether the two scores are added or multiplied. The best results of all are obtained by combining the sum and the product of the constituent scores. Using this function, Reddy et al. (2011) report an extremely high level of correlation ( $R^2 = 0.955$ )<sup>5</sup> between human transparency ratings of compounds and their constituents. This suggests that if one could successfully model the transparency of a compound's constituents, one would also be able to successfully model the transparency of the compound as a whole.

To test the hypothesis that compound transparency is best modelled as a function of constituent transparencies, Reddy et al. (2011) modelled the human transparency ratings in their database using two different distributional semantic approaches: constituent-based and composition-function-based. In the constituent-based models, separate co-occurrence vectors are calculated for each compound and its constituents. The transparency of a constituent is then taken to be the semantic distance between its co-occurrence vector and that of the compound. Finally, compound transparency is calculated as a function of the two constituent values. In the composition-function-based models, on the other hand, the co-occurrence vectors of the constituents are first combined to give a predicted vector for the compound. This is then compared with the actual co-occurrence vector of the compound, and the distance between them is taken as a measure of compound transparency. Reddy et al. (2011) evaluated how successfully the two types of models predicted the human judgements of compound transparency. Among the constituent-based models, the best results were obtained by using the sum of the constituent vectors, either alone or in combination with their product ( $R^2 = 0.613$  and  $0.615$ , respectively). Even better results were obtained by adding the constituent vectors in the composition-function-based approach ( $R^2 = 0.620$ ). Reddy et al. (2011: 217) hypothesise that the slightly better performance of the composition-function-based model is due to the fact that 'while constituent based models use contextual information of each constituent *independently*, composition function models make use of collective evidence from the contexts of both the constituents *simultaneously*' (italics in the original). However, the performance of the two models is actually very comparable, and overall this work offers

---

<sup>5</sup> $R^2$  is a measure of the goodness of fit of a statistical model, i.e. how much of the variation in the dependent variable is explained by the predictors in the model.  $R^2$  normally takes values between 0 and 1. An  $R^2$  value of 0 would indicate that the model predicts none of the variation in the dependent variable, whereas an  $R^2$  value of 1 would mean that the model predicts the correct value of the dependent variable in all cases.

**Fig. 1** Scheme for AB combinatorics (Bell and Schäfer 2013: 2)



support for the idea that compound semantic transparency can be modelled on the basis of constituent transparency.

### 2.2.3 A linguistic account

In Bell and Schäfer (2013), we modelled the human transparency ratings in the Reddy et al. (2011) dataset from a theoretical-semantic rather than distributional-semantic perspective. The analysis is based on the scheme in Fig. 1, where A and B are the constituents of any complex nominal AB. The scheme incorporates two important aspects of compound meaning. The first of these is the idea that there is an underspecified semantic relation between the constituents, represented in the diagram as R. In the semantic and pragmatic literature, it is a commonplace that the two constituents of a noun–noun compound are connected by some semantic relation that needs to be inferred. For example, according to Levinson (2000: 147), the semantic relation between nominal compound constituents ‘is no more than an existentially quantified variable over relations’. As shown in Fig. 1, the value of this variable is determined by context and world knowledge: this is how the compound *drum sticks*, for example, can represent sticks for hitting drums, while *bread sticks* represents sticks made out of bread. The second important aspect of compound meaning shown in Fig. 1 is the fact that compounds and their constituents are prone to exhibit meaning shifts, the results of such shifts being represented in the diagram as (AB)', A' and B' respectively. These shifts can be either metaphoric or metonymic, and can take place either before the constituents are combined, in which case they are shifted independently of one another, or after they are combined, in which case the compound is shifted as a whole. Consider, for example, the compound *buttercup*, which denotes a plant whose flower has a colour like butter and a shape that, at least to some extent, resembles a cup. In this case, *butter* is metaphorically shifted to mean ‘having the colour of butter’ and *cup* is metaphorically shifted to mean ‘having the shape of a cup’. The whole combination is then metonymically shifted so that the colour and shape of the flower come to stand for the whole plant and indeed for the species. According to the Bell and Schäfer (2013) scheme, *buttercup* therefore has metaphoric shifts of constituents A and B to A' and B', as well as a metonymic shift of A'B' to (A'B')'. As with the semantic relation R, the necessary shifts in constituent and compound meaning are inferred from context and world knowledge.

Despite general recognition that the semantic relation in compounds is underspecified, attempts have been made to list and classify possible values of this variable, for example by Lees (1970), Levi (1978), Warren (1978), Fanselow (1981) and Ó Séaghdha (2008). Amongst such taxonomies, Levi's (1978) proposal has been perhaps the most influential. Although her proposed relations do not exhaust all possible interpretations nor lead to descriptions of compound meanings that are as specific as actual meanings (see e.g. Downing 1977 and Fanselow 1981 for such criticisms), they do 'intuitively seem to comprise the right kind of relations for capturing compound semantics' (Ó Séaghdha 2008: 30). Levi's (1978) classification is based on the idea that compound semantic relations reflect underlying predicates that are not overtly expressed in the compound form. One of the reasons for the success of her analysis may be that Levi (1978) identifies only nine such predicates. Although some of these predicates are associated with more than one relation, depending on which constituent functions as the subject, the resulting set of relations is manageable in size and therefore relatively easy to use for coding data. The full set of these relations, with examples from Levi (1978: 76–77) and paraphrases illustrating the intended interpretation, are shown in (2):

- (2)
- |        |  |
|--------|--|
| CAUSE1 | <i>tear gas</i> (gas that causes tears)          |
| CAUSE2 | <i>drug deaths</i> (deaths caused by drugs)      |
| HAVE1  | <i>picture book</i> (book that has pictures)     |
| HAVE2  | <i>government land</i> (land the government has) |
| MAKE1  | <i>honeybee</i> (bee that makes honey)           |
| MAKE2  | <i>daisy chains</i> (chains made from daisies)   |
| USE    | <i>steam iron</i> (iron that uses steam)         |
| BE     | <i>soldier ant</i> (ant that is a soldier)       |
| IN     | <i>morning prayers</i> (prayers in the morning)  |
| FOR    | <i>horse doctor</i> (doctor for horses)          |
| FROM   | <i>olive oil</i> (oil from olives)               |
| ABOUT  | <i>tax law</i> (law about tax)                   |

For Bell and Schäfer (2013), we used this set of relations to code the compounds in the Reddy et al. (2011) dataset for semantic relation. We also coded the dataset for metaphoric and metonymic shifts of the constituents or the compounds as a whole, treating each kind of shift as a binary variable. Thus every constituent and every compound received two codings for shiftedness: 'yes' or 'no' for metaphoric shift and 'yes' or 'no' for metonymic shift. Each of us coded both semantic relation and shiftedness for all items in the dataset, then we compared our ratings and attempted to resolve any discrepancies by discussion. For two compounds, we could not reach agreement, and these items were therefore excluded from the analysis. We used the coded semantic relations and meaning shifts as predictors in three regression models, in which the dependent variables were, respectively, the transparency of the first constituent, the transparency of the second constituent and the transparency of the whole compound, as rated by the human participants in Reddy et al. (2011).

Considering first the meaning shifts: metaphoric shifts of the compound and both constituents survive as significant predictors in all three models reported in Bell and Schäfer (2013). In the model for whole compound transparency ( $R^2 = 0.459$ ), all

shifts are correlated with a decrease in perceived transparency. However, in the models of constituent transparency ( $R^2 = 0.499$ ;  $R^2 = 0.498$ ), the pattern is more complicated. A metaphoric shift in the meaning of the compound as a whole is always correlated with opacity, but whereas a shift in the meaning of either constituent decreases the perceived transparency of that constituent itself, in both cases the perceived transparency of the other constituent is increased. We interpret this as evidence that the perceived transparency of a constituent is relative: if the meaning of either constituent is metaphorically shifted, the other constituent seems more transparent by comparison. Concerning the effects of semantic relation on perceived transparency, the relations IN and FOR are both positively correlated with an increase in transparency of the compound as a whole. However, in the model for transparency of the first constituent (N1), the only relation found to be significant is IN, while in the model for transparency of the second constituent (N2), the only significant relation is FOR. We connect these results to work on prosodic prominence in English noun–noun combinations. For example, Plag et al. (2007) report that the locative relation (IN) is correlated with stress on N2 while the purpose relation (FOR) is correlated with stress on N1. Furthermore, Bell and Plag (2012) show that stress tends to fall on the most informative constituent. Thus, we argue in Bell and Schäfer (2013) that, if the relation IN correlates with greater transparency of N1, and we assume that the more transparent a constituent is, the less informative it is, this might explain why in such compounds stress tends to fall on N2. A similar reasoning applies to FOR. Since informativity can be conceptualised as the inverse of expectedness, this supports our present hypothesis that perceived transparency is a reflex of expectedness.

## 2.3 Expectedness in the semantic structure of compounds

### 2.3.1 *Expectedness and constituent frequency*

In the models presented in Bell and Schäfer (2013), the frequencies with which the compound constituents occur in the language generally are strong predictors both of compound and constituent transparency. This finding partly underlies our hypothesis that perceived transparency is related to expectedness, since constituent frequency is a measure of the expectedness of encountering a constituent word in the language as a whole.

In the present study, we want to explore whether transparency is also related to the expectedness of a constituent occurring in the relevant position, i.e. as either the modifier or head of a compound. This expectedness can be assessed using the notion of constituent family sizes. Every noun–noun (NN) compound has two positional constituent families. The modifier positional family consists of all other NN compounds that share the same modifier, and the head positional family consists of all other NN compounds that share the same head. These two constituent families are the positional families since the nouns used to create them occur in the same position in the target compound as they do in all other members of the family. It is also possible to create reverse families, in which the constituent in question occurs in the alternative position compared with the target compound. For example, the N1 positional family

for *bank account* includes e.g. *bank emergency*, *bank fraud* and *bank index*, while the N2 positional family includes e.g. *summary account*, *insider account* and *police account*. The N1 reverse family for *bank account* includes e.g. *asset bank*, *Beirut bank* and *blood bank*, while the N2 reverse family includes e.g. *account balance*, *account handler* and *account number*. The respective family sizes are the numbers of different compound types in each of these families. To measure the tendency of constituents to occur as either compound heads or modifiers, it is possible to use family size ratios (Bell and Plag 2013). Family size ratio is the log of the positional family size divided by reverse family size; it therefore indicates the preference of a constituent to appear either in its current or in the alternative position, and hence its expectedness in either position.

### 2.3.2 Expectedness and compound semantic relations

Evidence about the psychological nature of compound semantic relations has come largely from work within the framework of conceptual combination (discussed above in Sect. 1 and Sect. 2.2.1) and especially from studies using the paradigm of relational priming. Robust relational priming effects have been found for compounds that share the same modifier constituent and the same semantic relation, for example *teacup* and *teapot*. Reaction times to a particular compound are faster when a subject has previously seen a compound with the same modifier and semantic relation. This is true both for novel compounds (Gagné 2001; Gagné and Shoben 2002) and for established and lexicalised compounds (Gagné and Spalding 2004; Gagné et al. 2009; Spalding and Gagné 2011). Relational priming has also been found when the modifier of the prime is semantically similar but not lexically identical to that of the target (Gagné 2002), for example *scholar accusation* might prime for *student vote* (Gagné and Spalding 2014: 107). This suggests that relational interpretation takes place at the conceptual rather than lexical level. These studies did not find relational priming in cases where the prime and target shared the same head, except for ambiguous compounds with two equally plausible interpretations (Gagné and Shoben 2002). Gagné and Shoben (2002) interpret this finding as evidence that the head is primarily involved in evaluating relations suggested by the modifier.

The studies discussed in the previous paragraph provide strong evidence that compound semantic relations are somehow associated with particular compound constituents. Although some authors have suggested that these relations might have a psychologically real existence independent of particular lexical items, and have reported relational priming across compounds with unrelated constituents (Estes 2003; Estes and Jones 2006), this finding has not been consistently replicated. Gagné et al. (2005) suggest that the results reported in Estes (2003) might have been due to semantic similarity between primes and targets, and that those reported in Estes and Jones (2006) might have been due to a bias towards a particular relation in the dataset. Overall, the evidence that particular semantic relations are accessed through particular compound constituents is very strong, while the evidence that they have independent mental representations is much weaker.

There is evidence from a variety of experimental studies not only that particular semantic relations are accessed through particular compound constituents, but also that

people are sensitive to the frequency with which a particular relation occurs with a particular constituent. Gagné and Shoben (1997) constructed a corpus of compounds and coded them for semantic relation using the Levi (1978) classification. They then calculated the proportion of compounds with each head or modifier that had each relation. If a particular relation occurred in 60% or more of the compounds with a particular head or modifier, it was classed as a strong competitor for that constituent. Using a sense-nonsense decision task, Gagné and Shoben (1997) showed that the time taken to decide whether a compound made sense was determined by the strength of relation for the modifier but not the head. However, although strength of relation for the head showed no effect in the sense-nonsense task, it did have an effect when subjects were asked simply to verify whether a particular interpretation made sense for a given compound (Spalding et al. 2010). Spalding et al. (2010) suggest that this difference in results arises because subjects in sense-nonsense tasks are required to come up with their own interpretation of the compounds presented and are therefore strongly influenced by the relations available to the modifier. The interpretation verification task, on the other hand, taps into the process of evaluating a given relation. The fact that performance on this task is sensitive to the strength of relation for the head therefore supports the hypothesis that the role of the head in compound interpretation is to evaluate relations suggested by the modifier.

The results described in this section underpin the Relational Interpretation Competitive Evaluation (RICE) theory of conceptual combination (Spalding et al. 2010), according to which, compounds are interpreted in a three-step process. Firstly, the concept encoded by the modifier word (N1 in our data) suggests possible semantic relations, each of which has a certain strength based on its availability for that modifier concept. This strength is determined both by the frequency with which the relation occurs in combination with the modifier in the language generally, and by how recently they have occurred together in the experience of the person interpreting the compound. The second stage of interpretation involves evaluation of the possible relations on the basis of their strength of association with both the modifier and head concepts as well as their plausibility in the given context. At this stage, the system arrives at a gist interpretation of the compound and its semantic relation, which is further elaborated with the help of pragmatics and world knowledge during the third and final stage of interpretation. If this theory is correct, it has two main implications for our study. Firstly, semantic relations are associated with particular modifiers and heads and may not have any existence independently of the constituent concepts. Secondly, people are sensitive to the frequency with which particular relations are associated with particular constituents. The expectedness of a relation for a constituent might therefore contribute to the general expectedness of a compound's semantic structure and hence, according to our hypothesis, to the level of perceived transparency. In the present study, we attempt to code the compounds in our data for the expectedness of the relevant relation for each constituent, operationalised as the proportion of times the relation occurs in compounds with the given constituent in the given position.

### 2.3.3 Expectedness and meaning shifts

In the analysis presented in Bell and Schäfer (2013), we found that metaphoric shifts of the compound or either constituent were significant predictors in all models of compound or constituent transparency. This is perhaps not surprising, as encoding whether a constituent or a compound is shifted from its original meaning might involve the very same task that the Reddy et al. (2011) annotators had to perform, i.e. to judge a constituent or a compound for its degree of literality, which we take to reflect perceived transparency. It could therefore be argued that all we did was to use our own transparency judgements to model the transparency judgements of the subjects. The question then seems to be not so much why the shifts were significant predictors, but rather why including them did not lead to much better models. Looking at the semantic coding used in Bell and Schäfer (2013),<sup>6</sup> several possible reasons can be found for the relatively poor performance of the coded semantic shifts in predicting transparency. Firstly, even if it is correct that our coding for shifts mirrors the annotators' coding for literality, the binary choice between shift and no shift is less fine-grained than the six-point scale used by the annotators. Secondly, in the models reported in Bell and Schäfer (2013), metonymic shifts were not included as predictors because they occurred too rarely in the data to achieve significance. Finally, a closer look at the relationship between the items coded as having been metaphorically shifted and the transparency ratings given by the annotators points to a discrepancy between etymologically-informed judgements of shiftedness and naïve perceptions of literality.

To explore the relationship between semantic shifts and perceived transparency, let us consider the constituent transparency ratings for two first constituents categorised in Bell and Schäfer (2013) as shifted, *gravy* in *gravy train* and *web* in *website*, and two second constituents categorised as shifted, *card* in *credit card* and *candy* in *eye candy*. These items are instructive because, despite the relevant constituents all being coded as metaphorically shifted, their average transparency ratings are quite different: 0.15 for *gravy* vs. 2.7 for *web* and 4.9 for *card* vs. 0.7 for *candy*. Looking first at the two constituents with low ratings, we see that the annotators gave an even lower rating to *gravy* in *gravy train* than to *candy* in *eye candy*. We speculate that this difference is due to the nature of the shifts involved. The shift for *candy* seems rather straightforward: candy is sweet, and sweet things are pleasing. The shift for *gravy*, in contrast, seems less obvious and, as far as we know, the exact etymology of *gravy train* is unclear. Turning now to the two constituents with relatively high transparency ratings, an average rating of 4.9 out of 5 indicates that the majority of annotators felt that *card* in *credit card* was used literally and thus transparent. That is, while we judged *card* to be metaphorically shifted, due to the fact that credit cards are made out of plastic and not out of pasteboard, for the annotators, a small piece of stiff plastic could be literally a card. Although not perceived to be quite so transparent as *card*, *web* in *web site* sits in the middle of the rating scale. According to the OED, the first sense of *web* was that of a woven fabric, with the cobweb meaning also being

<sup>6</sup>Bell and Schäfer's (2013) semantic codings are available at [http://www.martinschaefer.info/publications/TFDS-2013/TFDS-2013\\_Bell\\_Schaefer.zip](http://www.martinschaefer.info/publications/TFDS-2013/TFDS-2013_Bell_Schaefer.zip) [last accessed 22 February 2016].



attested very early. Its usage in *web site* is etymologically clearly metaphorical. Why then the rating of 2.7? We speculate that the relatively high ratings for items such as *card* and *web* result from current usage, where the most frequent and therefore most expected sense of a word might well be one that, from an etymological point of view, would be viewed as shifted. If this reasoning is correct, that shifted senses can become standard senses as a result of their frequencies in actual language use, then there is no clear binary criterion for classifying constituents or whole compounds as synchronically shifted or not.

If the coding of shiftedness in Bell and Schäfer (2013) is problematic, what alternative method can be used to assess the contribution of meaning shifts to compound semantics? Given the preceding discussion of *web site* and *credit card*, an obvious possibility, at least for the constituents, is to consider not whether the sense is shifted, but rather the extent to which it is expected. In the present study, we therefore attempt to code our data for the expectedness of the relevant sense of each constituent. In line with the method for relational coding, discussed above, we operationalise this expectedness as the proportion of times the sense occurs in compounds with the given constituent in the given position. We assume that these proportions can be used as alternative measures for shiftedness of the individual constituents. It is less immediately obvious what measure can be used for the whole compound shifts. However, on the assumption that a shift in whole compound meaning is associated with strong semantic lexicalisation, a variable intended to measure degree of compound lexicalisation could also be used as an indication of shiftedness. For compounds in English, one such variable is ‘spelling ratio’: the number of times a compound is written unspaced or hyphenated in a large corpus divided by the number of times it is written spaced in the same corpus (cf. Bell and Plag 2012).

## 2.4 Predictions

We are now in a position to spell out in detail the predictions made by our hypothesis that perceived transparency is a reflex of expectedness. In general, we predict that the more expected a compound’s semantic structure, the greater will be the perceived transparency both of the whole compound and of its constituents. We take semantic structure to include the concepts represented by the constituents, both in general and with the specific sense of the compound, the semantic roles of the constituents (head or modifier), and the semantic relation between them. As described above, these various aspects will be operationalised in our models respectively as constituent frequency, the proportion of the positional constituent family with the same constituent sense, family size ratio, and the proportion of the positional constituent family with the same semantic relation. We also include spelling ratio as a measure of whole compound shift.

### 2.4.1 *How will a constituent’s properties affect the perceived transparency of that constituent?*

We predict that:

1. A constituent will be perceived as more transparent the more frequent it is, i.e. the more expected it is in the language in general.

2. A constituent will be perceived as more transparent, the more expected its particular sense within the positional family, i.e. the more likely it is to occur with that sense as the head (for N2) or modifier (for N1) of a compound.

3. A constituent will be perceived as more transparent, the more expected it is as the head (for N2) or modifier (for N1) of compounds in general, i.e. the more characteristic the relevant role for the constituent in question.

4. A constituent will be perceived as more transparent, the greater the proportion of compounds in its positional family that share the same semantic relation as the compound in question, i.e. the more expected the relevant semantic relation with that constituent.

#### *2.4.2 How will a constituent's properties affect the perceived transparency of the other constituent?*

We predict that:

5. A constituent will be perceived as more transparent the more frequent the other constituent, i.e. the more expected it is in the language in general. This is what we found in the study reported in Bell and Schäfer (2013); we expect to replicate this result.

6. A constituent will be perceived as more transparent, the less expected the relevant sense of the other constituent within its positional constituent family. In Bell and Schäfer (2013) we reported that a semantic shift in either constituent was associated with greater perceived transparency of the other constituent. If, as we hypothesise, sense frequencies can be used to estimate semantic shiftedness, then we would expect to find the same effect.

7. A constituent will be perceived as more transparent, the more expected is the other constituent as the head (for N2) or modifier (for N1) of compounds in general, i.e. the more characteristic is the relevant role for the other constituent. This is because we hypothesise that a more readily available semantic structure will lead to an increased perception of transparency all round.

8. A constituent will be perceived as more transparent, the greater the proportion of the other constituent's positional family that shares the same semantic relation as the compound in question, i.e. the more expected is the other constituent with the relevant semantic relation. This is because we hypothesise that the more easily accessible the semantic relation, the greater the perceived transparency all round.

#### *2.4.3 How will constituent properties affect the perceived transparency of the whole compound?*

In general, we expect that factors that increase the perceived transparency of either constituent will increase the perceived transparency of a compound as a whole, since we know from Reddy et al. (2011) that compound transparency can be modelled as a function of constituent transparencies. Specifically, we predict that:

9. A compound will be perceived as more transparent the more frequent is either constituent i.e. the more expected it is in the language in general. Greatest perceived transparency will occur when both constituents are frequent.

10. A compound will be perceived as more transparent, the more expected is either constituent in the relevant role, i.e. as the head (for N2) or modifier (for N1) of compounds in general. Greatest perceived transparency will occur when both constituents occur in their characteristic roles.

11. A compound will be perceived as more transparent, the greater the proportion of either constituent's positional family that shares the same semantic relation as the compound in question, i.e. the more expected is either constituent with the relevant semantic relation. Greatest perceived transparency will occur when the relation occurs in a high proportion of both families.

12. The effect of constituent senses on whole compound transparency will be less pronounced than the effects on individual constituents, and less pronounced than the effects of the other predictors on whole compound transparency. This follows from our hypotheses that compound transparency is a function of constituent transparencies, and that a high sense proportion of a given constituent increases the perceived transparency of that constituent while decreasing the perceived transparency of the other.

#### 2.4.4 *Semantic shifts of the whole compound*

We predict that:

13. Both constituents will be perceived as less transparent, the greater the spelling ratio of the compound, i.e. the more frequently it occurs with non-spaced orthography relative to its frequency with spaced orthography. Bell and Schäfer (2013) found that semantic shifts of the compound as a whole were associated with lower perceived transparency of both constituents. We take spelling ratio to be a measure of the degree of semantic lexicalisation of a compound (after Bell and Plag 2012, 2013) and hypothesise that it can therefore be used to replicate the effect of whole-compound semantic shift.

14. The compound will be perceived as less transparent, the greater its spelling ratio, i.e. the more frequently it occurs with non-spaced orthography relative to its frequency with spaced orthography. As described above, we hypothesise that high spelling ratio is a correlate of whole-compound semantic shift.

### 3 Methodology

#### 3.1 The dataset

We used the publically available dataset created by Reddy et al. (2011), which consists of 30 human transparency ratings for each of 90 two-part English compounds and their constituents, making a total of 8100 ratings (cf. Sect. 2.2.2). From this total, Reddy et al. (2011) excluded the ratings of any annotator whose ratings were negatively correlated with the others' and accepted the work of annotators with an overall positive correlation coefficient greater than 0.6. For the remaining annotators, individual annotations were accepted provided they were within  $\pm 1.5$  standard deviations of the mean rating for the task in question. We used only the ratings accepted by these criteria, leaving a total of 7717 ratings. From this set, we

removed five compound types in which the modifier had an exclusively adjectival reading, and one in which the head occurred only as a verb in the WordNet lexical database (Fellbaum 1998). This was to facilitate the later stages of processing the data, in which we would search for the constituent families of each compound: restricting ourselves to compounds in which both constituents are normally tagged as nouns meant we could search for constituent family members by searching for strings of two nouns, and this kept the data within manageable proportions. Since we wanted to restrict the study to two-part compounds, we also removed three types in which one of the constituents was itself a compound. This left us with a total of 6952 ratings, constituting 2307 ratings of whole compound transparency, 2317 ratings of N1 transparency and 2328 ratings of N2 transparency for 81 compounds.

### 3.2 Obtaining constituent families

In order to calculate the expectedness of particular word senses and semantic relations for our compounds, we first needed to access their constituent families. For this, we used the British National Corpus (BNC XML Edition 2007). The BNC is a 100 million-word corpus containing 90% written and 10% spoken data, representing a cross-section of British English from the late 20th century. We accessed the BNC via the web-interface provided by Lancaster University: the CQP-edition (Version 4.3) of BNCweb<sup>7</sup> developed by Sebastian Hoffmann and Stefan Evert (cf. Hoffmann et al. 2008). We used the CQP query syntax to extract all strings of two nouns that followed a definite article and were not themselves followed by another noun, an adjective, or a possessive marker. In this way, we excluded noun–noun strings that were part of larger complex nominal constructions. We used both the spoken and written parts of the BNC, but excluded any strings from the spoken part that contained pauses, unclear portions or other paralinguistic events. From the resulting set of NN strings, we selected those that shared a constituent lemma with a compound in our dataset. In doing this, we took into account both the position of the constituent in the original item, and whether it occurred in the same position in the family member. In this way, we produced a positional constituent family and a reverse constituent family for both constituents of every item in the dataset. Note that we did not make any attempt to exclude proper names. First of all, we do not assume that proper names are processed by humans in a principally different way from common nouns. Secondly, there is no clear formal distinction between proper names and their complement in English, and in fact many doublets exist due to conversion in either direction (Huddleston et al. 2002: 516). Thirdly, and most importantly, it is well known that proper names interact in notable compound patterns, e.g. the classic stress contrast observed for NN combinations headed by either *avenue* or *street*, where the relevant constituent families include a considerable number of proper names (Plag et al. 2008).

---

<sup>7</sup>BNCweb is available at <http://bncweb.lancs.ac.uk/> [last accessed 22 February 2016].

It is not possible to search the BNC for NN combinations written as one word, since these items are tagged as single nouns. This had two significant implications for the creation of our dataset. Firstly, NN strings extracted as described in the preceding paragraph might actually turn out to have more than two nominal constituents, if at least one of the nouns in the string was itself an unspaced compound. Secondly, our search algorithm would not detect compounds that occur only unspaced in the BNC. To address the first of these issues, we filtered our search results against all English compounds and simplex words in the English part of the CELEX lexical database (Baayen et al. 1995). We left in our dataset all compounds that consisted only of simplex words, but excluded those with one or more compound constituents. Compounds with constituents not occurring in CELEX were checked manually, at which point we also excluded compounds in which either constituent consisted of an abbreviation. To address the second issue, we added to our constituent families all the unspaced compounds in CELEX that shared a constituent with any of the 81 compounds in our core data.

Despite the care taken with the corpus search, the resulting set of noun–noun strings included many that were not in fact compounds. In an attempt to reduce this noise in our data, as well as to keep the dataset to a manageable size, we decided to exclude types with very low frequencies. However, because of the low textual frequencies of compounds in general (Plag et al. 2008: 776), a comparatively small corpus like the BNC contains a very large number of compounds that occur only once, and it is difficult to distinguish between those with relatively high and low frequencies. In order to address these points, we decided to get frequencies for the compounds in our constituent families from a much larger corpus, and for this we used the reduced redundancy USENET corpus (Shaoul and Westbury 2013), containing over 7 billion tokens. Because this corpus is not lemmatised, we searched for all inflectional variants of the compounds in question, as well as all spelling variants (spaced, hyphenated and unspaced in British and American English), and summed these frequencies to get the lemma frequencies. We then restricted our constituent families to only those items that occurred with a lemma frequency of at least 5. This left a total of 2893 compound lemmas in the N1 positional constituent families and 6425 compound lemmas in the N2 positional constituent families.

### 3.3 Coding semantic structure

#### 3.3.1 Constituent frequency and family size ratio

Lemmatised constituent frequencies were calculated by summing the frequencies of all inflectional variants of the constituents in the USENET corpus.

Family size ratios were calculated as shown in (3):

$$(3) \quad \text{family size ratio} = \log(\text{positional family size}/\text{reverse family size})$$

Family size ratio is the log of the positional family size divided by reverse family size and therefore indicates the preference of a constituent to appear either in its current or in the alternative position. We calculated family size ratios both on the basis of the

entire corpus, and using a lemmatised frequency baseline of at least 5 occurrences, the same frequency baseline we used for relation and sense proportions. The latter values were found to be better predictors of perceived transparency and are therefore used in the models reported here.

### 3.3.2 Relation proportion

In order to calculate the proportion of compounds in each constituent family with each semantic relation (henceforth ‘relation proportion’), every compound in the positional constituent families had to be coded for semantic relation. We used the relations in the Levi (1978) taxonomy, discussed in Sect. 2.2.3, for their ease of application and to ensure a similar granularity of classification in each of the modifier or head families. However, within this scheme or any other, we do not believe that there is necessarily a single correct classification for any given compound; in the literature on compound semantics it is well established that alternative coding possibilities can exist for a compound, even given the very same denotation. For example, Levi (1978: 90) discusses *chocolate bar*, which might be analysed alternatively as either MAKE (bar which is made of chocolate) or BE (bar which is chocolate). In the present context, the main aim was to establish reliable relation proportions for each constituent. To give an example, in the case of *knife handle* it would be more important to recognise that this relation for *handle* is comparable to the relation in e.g. *fork handle* and *spoon handle* than to decide whether the label should be FOR as in ‘handle for a knife’ or HAVE as in ‘handle that a knife has’ or the general locative IN as in ‘handle on a knife’. Similarly, because our primary aim was to achieve consistency of classification within constituent families, we took a principled decision that the coding should all be done by a single person. There is evidence in the literature that inter-annotator agreement for compound relations is usually quite poor. One of the most successful schemes is that of Ó Séaghdha (2008), but even this achieves only 66.2% agreement between two trained annotators on a test set of 500 items (Ó Séaghdha 2008: 45). The options in cases of disagreement are either to discard cases where annotators do not agree or to attempt to resolve the disagreements through discussion. Neither of these options was desirable for the present study. We did not want to discard cases where we might disagree, both because it could have resulted in losing a large proportion of our data and hence reducing the statistical power of our analysis, and because it would have skewed the data in favour of the more obvious examples. However, the risk of using a system that aims to resolve disagreements through discussion is that it could lead to an inconsistency in the classification overall, depending on which annotator’s view prevails in each case. We therefore took the decision that the rating should be done by just one of us, with the overriding concern that the classification should be internally consistent.

The coding was undertaken by the second author, who has a PhD in theoretical semantics and near-native fluency in English. In addition to Levi’s (1978) categories, the relation VERB was used for deverbal heads with an argument in N1 position, and the category OTHER was used for cases that did not fit any of the other relations. Coding examples from our positional families are shown in (4):

(4)	CAUSE1	<i>cost centre, collision course</i>
	CAUSE2	<i>night blindness, rocket trail</i>
	HAVE1	<i>video arcade, diamond watch</i>
	HAVE2	<i>video title, search complexity</i>
	MAKE1	<i>diamond factory, engine plant</i>
	MAKE2	<i>silver replica, number sequence</i>
	USE	<i>rocket plane, video conference</i>
	BE	<i>video clip, zebra crossing</i>
	IN	<i>night life, radio discussion</i>
	FOR	<i>video gear, think tank</i>
	FROM	<i>research data, bank pamphlet</i>
	ABOUT	<i>spelling game, speed freak</i>
	VERB	<i>climate change, credit check, engine driver</i>
	OTHER	<i>bank holiday, lotus hotel, eye tooth</i>

In view of the psycholinguistic evidence that relations are associated with particular constituents (Sect. 2.3.2), and because the aim was to have codings that formed plausible subclasses within individual constituent families, the compounds were coded in comparison with the other compounds in each positional constituent family. This meant that each compound was classified twice, once in the context of its modifier family and once in the context of its head family. In some cases, this resulted in different relational labels being assigned to the same compound in the N1 and N2 families. An example is *face value*, which in the N1 family was coded with the relation HAVE2 (value that the face has), along with e.g. *face price*, and in contrast to IN, which was used for e.g. *face ache* and *face wound*. On the other hand, in the N2 family *face value* was coded as IN (value on the face), as was e.g. *market value*, while examples for HAVE2 are e.g. *pixel value* and *probability value*. Note that we do not think this means that such compounds involve more than one relation, nor that the coding is wrong, simply that a given relation can sometimes have more than one acceptable label, as described above for *chocolate bar* and *knife handle*. Since the aim of the coding was to establish relation proportions, the important considerations were that coding was consistent within families and that the level of taxonomic granularity, rather than the specific label used, was comparable between families.

When the coder could not determine a plausible classification based on the compound alone, the combination was checked in its sentential context in the BNC. In these cases, the combination often turned out not to be a compound, either because the two nouns actually belonged to different syntactic constituents or because of mistagging in the corpus. Such cases were excluded from further analysis.

The coding of the positional constituent families for semantic relations was used to calculate two variables for each constituent of the 81 compounds in our core dataset. These are shown in (5):

- |     |                      |   |
|-----|----------------------|---|
| (5) | relation proportion: | the proportion of positional family members that share a constituent's semantic relation  |
|     | relation rank:       | the frequency-based ranking of a constituent's semantic relation in its positional family |



An example will help to make this clear. *Application* in *application form* is coded with the relation FOR, and there are 27 further types in the *application* N1 family with this relation. Since the family itself has 42 members, FOR occupies the first rank, occurring in  $28/42 = 67\%$  of family members. Hence, *application form* is coded with an N1 relation proportion of 0.67 and an N1 relation rank of 1. All our semantic coding is publically available online.<sup>8</sup>

### 3.3.3 Synset proportions

In order to capture different usages of the constituent nouns in the positional families, the individual constituents where coded using the different word senses (synsets) available in WordNet (Fellbaum 1998). For the reasons described in Sect. 3.3.2, this coding was undertaken by a single expert judge, again the second author. In the default case, the coding was done on the basis of the compound in isolation, but where necessary for clarification, reference was made to the sentential context or contexts in the BNC. Some example codings from the constituent family of *line* in N2 position are shown in (6).

- |     |  |   |
|-----|--|---|
| (6) | <i>soldier line</i><br><i>chalk line</i><br><i>payee line</i><br><i>intersection line</i><br><i>propaganda line</i><br><br><i>electricity line</i> | a formation of people or things one beside another<br>a mark that is long relative to its width<br>a formation of people or things one behind another<br>a length, straight or curved, without breadth or thickness<br>a course of reasoning aimed at demonstrating a truth or falsehood<br><br>a conductor for transmitting electrical or optical signals or power |
|-----|--|---|

The coding of the WordNet senses involved a number of non-trivial coding decisions. One issue was how to deal with ambiguity. Our motivation for using WordNet senses was to have an objective way of classifying constituent meanings that were most likely related to each other via meaning shifts, as in the examples for *line* above. However, our constituent families also contained word meanings that were not related via meaning shifts. In fact, they might not be etymologically related at all, for example *rock* in its stone sense (derived from a Romance noun) vs. *rock* as a genre of music (derived from a Germanic verb). Because we have no principled basis for distinguishing between etymologically related and unrelated word senses, all members of a constituent family were retained regardless of their sense (see Maguire et al. 2007 for a similar procedure). While typically the specific word sense applicable to a given compound is clear, e.g. *rock music* vs. *rock arch*, in some cases the ambiguity might persist, e.g. *rock mix*, which could in principle mean a mix of different kinds of stones or a musical mix in the rock style. In these cases, which are rare, the meaning occurring in the BNC was used or, if both meanings occurred, the one that occurred most frequently. A second difficulty the coder encountered were WordNet senses he deemed impossible to distinguish, either conceptually or in application to the data.

<sup>8</sup>The full dataset is available at [http://martinschaefer.info/publications/download/Bell\\_and\\_Schaefer\\_2016\\_semantic-coding.zip](http://martinschaefer.info/publications/download/Bell_and_Schaefer_2016_semantic-coding.zip).

In these cases, the senses were collapsed into one sense. Finally, some compounds involved constituent senses that did not occur in WordNet, in which case the missing senses were provided; often these missing senses involved proper names, e.g. *Ring* referring to the set of four operas by Richard Wagner.

The coding of the positional constituent families for WordNet senses was used to calculate two variables for each constituent of the 81 compounds in our core dataset. These are shown in (7):

- (7)      synset proportion:    the proportion of positional family members that share a constituent's WordNet sense  
              synset rank:        the frequency-based ranking of a constituent's WordNet sense in its positional family

For example, *application* in *application form* is coded with WordNet sense 2, and there are 19 other types with this WordNet sense in a family of 42 members. Hence the N1 synset proportion for *application form* is  $20/42 = 0.48$ . There are in total 4 WordNet senses of *application* in this constituent family, of which sense 2 is the most frequent, hence *application form* gets an N1 synset rank of 1.

### 3.3.4 Spelling ratio

The spelling ratio of a compound was calculated as shown in (8):

- (8)      spelling ratio =  $\log((\text{unspaced frequency} + \text{hyphenated frequency}) / \text{spaced frequency})$

This variable has been used as a measure of lexicalisation (e.g. Bell and Plag 2012). We assume that the degree of lexicalisation is positively correlated with whole compound meaning shifts.

## 3.4 Statistical modelling

The variables described above were used as predictors in regression models of the human transparency ratings of the compounds in our dataset and their constituents. Because the transparency ratings include multiple contributions by each annotator, as well as multiple ratings on each compound type, the individual data points are not statistically independent of one another. In order to nevertheless be able to make full use of the data, we used linear mixed-effects regression models including crossed random effects for annotators and items. This type of model allows us to treat effects due to specific items and specific annotators as random effects, thus ensuring that the remaining effects in the resulting models, the fixed effects, are truly due to the semantic and distributional predictors of interest and not due to any peculiarity of particular items or annotators (for an introduction to these types of linear mixed-effects models, see Baayen et al. 2008). For example, there may be idiosyncratic reasons, unrelated to our predictors, why a particular compound is perceived as being more or less transparent than others. To allow for this possibility, we included a by-item random intercept in our models; in other words, we allowed baseline transparency to vary according to the compound being rated. Similarly, individual annotators may tend to

give higher or lower ratings overall, or may differ in their sensitivity to particular predictors. To accommodate these possibilities, respectively, we included by-annotator random intercepts and slopes for the various predictors.

All statistical analysis was done with R (R Core Team 2015), using the lme4 package (Bates et al. 2015) and the lmerTest package (Kuznetsova et al. 2015) for the regression modelling. In order to reduce the risk of extreme values adversely affecting our models, all frequencies were logarithmised and the resulting variables were centred on their means. Including synset and relational ranks as well as the corresponding proportions led to unacceptable levels of collinearity; since the proportions produce more fine-grained distinctions than the ranks, the latter were dropped from the analysis. As well as the predictors individually, we also included an interaction between the relation proportions of N1 and N2: if the RICE theory of conceptual combination (Spalding et al. 2010) is correct, then we might expect some interaction between the strength of association of the relation with N1 and the strength of its association with N2. The appropriate complexity for the random effects was selected by carrying out a series of likelihood ratio tests on sequences of models with increasingly complex random effects structures (cf. Baayen et al. 2008) and the non-significant fixed effects were progressively removed from the models by stepwise elimination. For confirmation of the final models, this manual analysis was checked against the results of the step function in lmerTest, which performs automatic backward elimination on random and fixed effects in a linear mixed effects model. Marginal and conditional  $R^2$  values were calculated using the r.squaredGLMM function in the MuMIn package (Bartón 2016). For mixed effects models, marginal  $R^2$  values are those associated with the fixed effects, and conditional  $R^2$  values are those associated with the fixed effects plus the random effects (Nakagawa and Schielzeth 2013).

## 4 Results

### 4.1 Transparency of first constituents

The final model for transparency of the first constituent is shown in Table 1 and represented graphically in Fig. 2. The top section of Table 1 shows the random effects: the model includes random intercepts for items, as well as by-annotator random intercepts and slopes for the effects of spelling ratio and N1 frequency. The lower half of the table shows the fixed effects. A positive coefficient in the first column of numbers indicates that the relevant predictor variable is positively correlated with N1 transparency, whereas a negative coefficient indicates negative correlation. Thus N1 transparency is positively correlated with both N1 frequency and N1 relation proportion, and negatively correlated with synset proportion of N2, spelling ratio and N2 frequency. The final column in the table indicates the level of statistical significance of each effect.

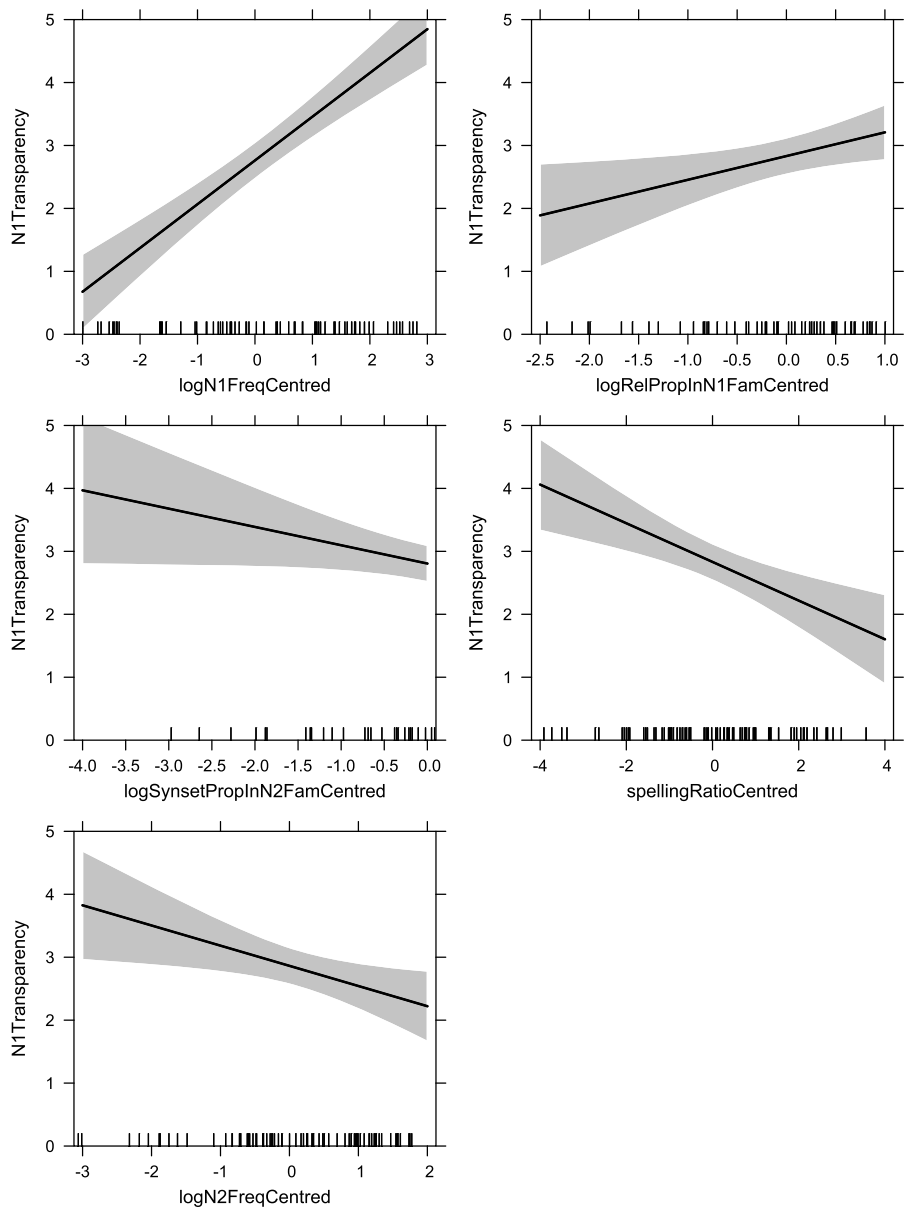
In Fig. 2, the vertical axis of each plot represents N1 transparency and the horizontal axes of the five plots represent the five significant predictors. The grey shaded areas around the regression lines show the 95% confidence intervals and the short vertical black lines directly above the horizontal axes represent the distribution of the

**Table 1** Final mixed effects model for N1 transparency, marginal  $R^2 = 0.37$ , conditional  $R^2 = 0.77$ 

Random effects:						
Groups	Name	Variance			Std. Dev.	
annotator	(intercept)	0.046443			0.21551	
	spellingRatioCentred	0.003116			0.05582	
annotator.1	(intercept)	0.018355			0.13548	
	logN1FreqCentred	0.003735			0.06111	
item	(intercept)	1.627959			1.27592	
residual		0.978290			0.98909	
Number of obs: 2317, groups: annotator, 114; item, 81						
Fixed effects:						
	Estimate	Std. Error	df	t value	Pr(> t )	Signif.
(intercept)	2.84159	0.14830	81.75	19.161	<2e-16	***
logN1FreqCentred	0.69523	0.08641	76.32	8.046	8.78e-12	***
logRelPropInN1FamCentred	0.37687	0.15973	75.02	2.359	0.020904	*
logSynsetPropInN2FamCentred	-0.29082	0.14541	75.00	-2.000	0.049126	*
spellingRatioCentred	-0.30717	0.08391	76.04	-3.661	0.000462	***
logN2FreqCentred	-0.32067	0.13186	75.00	-2.432	0.017407	*
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

data within the range of each variable: where the lines merge together to create black blocks, it indicates that the data is clustered around those values. The top left-hand plot shows that the more frequent N1, the more transparent it is perceived to be. This is in accordance with our Prediction 1: more frequent items are more expected in the language generally, and so a positive correlation between frequency and transparency also represents a correlation between expectedness and transparency. The top right-hand plot shows that the greater the relation proportion in the N1 family, the more transparent N1 is perceived to be. This is in keeping with Prediction 4: the greater the N1 relation proportion, the more expected is that relation in compounds with that particular modifier, so this effect represents a positive correlation between the expectedness of a compound's internal semantic structure and the perceived transparency of the modifier; it shows that the compound context itself influences transparency judgements on the constituents.

The plots on the second row of Fig. 2 represent relationships between N1 transparency and predictors we take to be representative of semantic shiftedness. The left-hand plot shows that the lower the synset proportion of N2, i.e. the less expected the relevant sense of N2 within its positional constituent family, the more transparent N1 is perceived to be. Recall that we are using synset proportions to operationalise semantic shift of individual constituents; we assume that frequent senses of a constituent are perceived as central, and less frequent senses as shifted. In the analysis reported in Bell and Schäfer (2013), we found that a semantic shift in either con-



**Fig. 2** Partial effects in final model for N1 transparency

stituent was associated with greater perceived transparency of the other constituent. The negative correlation between transparency of N1 and synset proportion of N2 mirrors this earlier finding, and is in keeping with our Prediction 6. There seem to be two possible explanations for this effect. Firstly, it could be that perceived transparency is relative, so that when one constituent has a very unexpected and opaque

**Table 2** Linear regression of N1 transparency by N2 frequency

	Estimate	Std. Error	t value	Pr(> t )	Signif.
(intercept)	2.796101	0.042914	65.155	<2e-16	***
logN2FreqCentred	0.009546	0.036318	0.263	0.793	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

sense, the other seems more transparent by comparison. Alternatively, it could be that very unexpected and opaque constituents can only form interpretable compounds with partners that are relatively transparent in more absolute terms. The right-hand plot shows that spelling ratio is also negatively correlated with the perceived transparency of N1. This is in accordance with our Prediction 13. We take spelling ratio to be a measure of semantic lexicalisation, equivalent to whole compound shift in the scheme in Fig. 1. We further assume that lexicalisation of the compound as a whole is associated with meaning specialisation or shift in the constituents, and that shifted meanings are less expected than non-shifted meanings. Hence, our hypothesis predicts that compounds with higher spelling ratios have less transparent constituents.

Of all the plots in Fig. 2, the only one not in keeping with our predictions is the bottom left-hand plot, which shows that the greater the frequency of N2, the lower the perceived transparency of N1, exactly the opposite of what we predicted (Prediction 5). A possible interpretation of this finding is that the effect of N2 frequency on N1 transparency is similar to the effect of N2 synset proportion, i.e. that less expected, and we assume therefore more opaque, heads are somehow associated with greater perceived transparency of the associated modifiers. Although this seems plausible, the reason we expected the opposite result is because that is what we found in our previous analysis of the data: in Bell and Schäfer (2013) we found that increased frequency of either constituent was associated with greater perceived transparency of both constituents. There are several possible explanations for this discrepancy, since the two studies used different subsets of the data, different frequency measures and different modelling techniques. However, careful comparison of the two analyses, as well as a series of further exploratory models, revealed that the difference can be attributed mainly to the inclusion of random effects in the current study. This is illustrated in Tables 2 and 3. Table 2 shows a simple linear regression model, with N1 transparency as the dependent variable and N2 frequency as the only predictor. It can be seen that the estimated coefficient of N2 frequency is positive, although in the absence of other predictors the effect is small and statistically insignificant. Table 3 shows the same model with the inclusion of the random effects from our full model. The estimated coefficient of N2 frequency is now negative, although in the absence of the other fixed effect predictors it remains small and statistically insignificant.

The random effects in our final model of N1 transparency are illustrated in Figs. 3 and 4. Figure 3 shows how the effect of spelling ratio on N1 transparency varies according to which annotator performed the task. In the left-hand plot, the dots represent the intercept for each annotator. The further to the right a dot appears, the

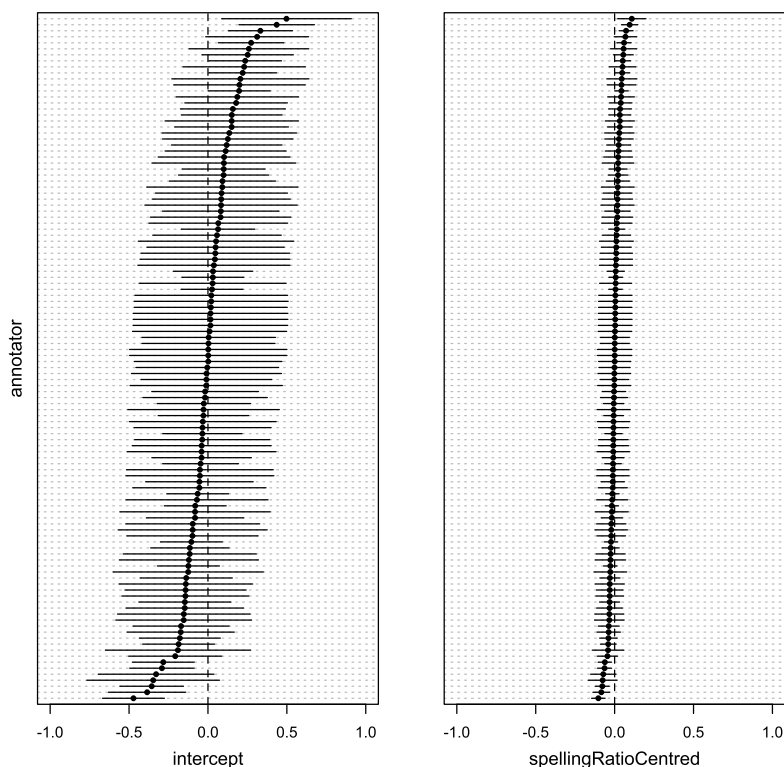
**Table 3** Mixed effects model of N1 transparency by N2 frequency

Random effects:						
Groups	Name	Variance			Std. Dev.	
annotator	(intercept)	0.046100			0.21471	
	spellingRatioCentred	0.003126			0.05591	
annotator.1	(intercept)	0.019008			0.13787	
	logN1FreqCentred	0.003855			0.06209	
item	(intercept)	3.153580			1.77583	
residual		0.978060			0.98897	
Number of obs: 2317, groups: annotator, 114; item, 81						
Fixed effects:						
	Estimate	Std. Error	df	t value	Pr(> t )	Signif.
(Intercept)	2.810467	0.202340	82.65	13.890	<2e-16	***
logN2FreqCentred	−0.003716	0.168883	78.75	−0.022	0.982	
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						

greater the tendency of that annotator to give high transparency ratings. The horizontal lines represent the 95% confidence intervals for these intercepts; if the confidence interval crosses zero, we cannot be entirely confident that the annotator in question exhibits any significant bias. Nevertheless, at least 12 annotators, more than 10%, do exhibit a clear tendency towards either higher or lower ratings. The right-hand plot shows how annotators vary in their sensitivity to spelling ratio. Although this effect is small in magnitude, it is nevertheless significant in the overall model. Plots of by-annotator random intercepts and slopes for the partial effect of N1 frequency on N1 transparency (not presented here) reveal similarly small but significant effects.

Figure 4 shows the random intercepts for different compound types in our final model for N1 transparency, i.e. the variation in baseline transparency according to the compound being rated. For example, the bottom two dots, furthest to the left, represent the intercepts for *diamond wedding* and *face value*, while the top two dots, furthest to the right, represent the intercepts for *engine room* and *climate change*. The further to the right a dot appears, the greater the baseline N1 transparency rating for that compound. The horizontal lines again represent the 95% confidence intervals for these intercepts. By comparing the scale on the horizontal axis with the corresponding scales in Fig. 3, it can be seen that this effect is much larger than the by-annotator effects; furthermore, few of the confidence intervals cross the origin, so we can be confident that these items really do vary in baseline transparency. It is easy to understand how including this random variation in our models could alter the apparent fixed effects of other predictors. In the present study, our variables of interest emerge as significant predictors of the individual transparency ratings given by different annotators, even after taking into account this random variation between compounds.



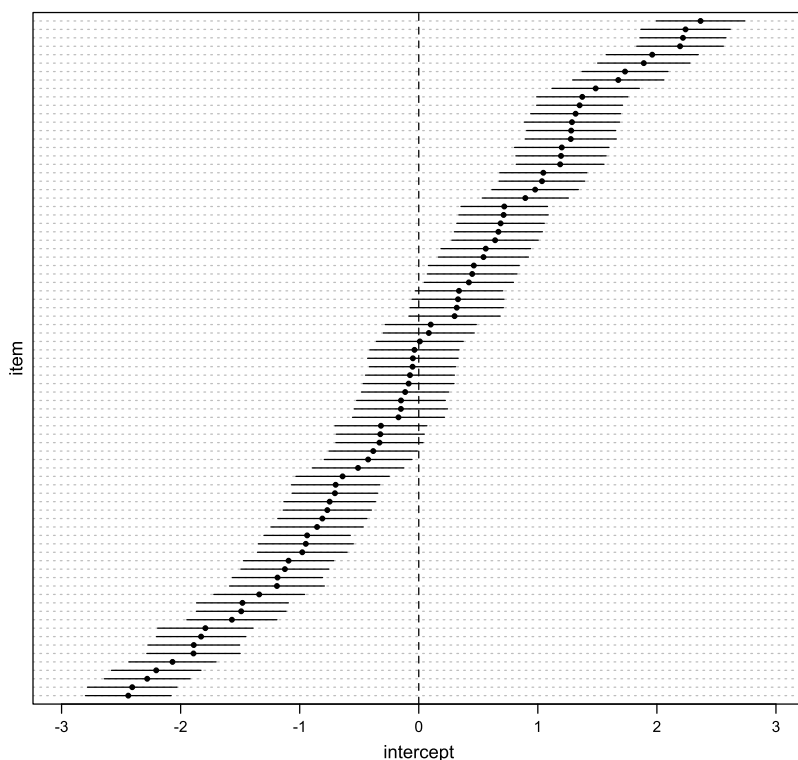


**Fig. 3** By-annotator random intercepts and slopes for the partial effect of spelling ratio in final model for N1 transparency

The fact that the fixed effect predictors emerge as significant, even in the presence of such a strong random effect, strengthens our confidence in their validity.

## 4.2 Transparency of second constituents

Table 4 shows the final model for transparency of the second constituent. It includes random intercepts for items, as well as by-annotator random intercepts and slopes for the effects of spelling ratio and N2 frequency. In the bottom half of the table, positive coefficients in the first numerical column again indicate a positive correlation between the predictor and transparency, whereas negative coefficients indicate negative correlation. There are significant effects of spelling ratio, N2 frequency and N1 relation proportion, all in accordance with our hypothesis. These effects are shown graphically in Fig. 5. Looking first at the top left-hand plot, we see that the greater the frequency of N2, the greater its perceived transparency. This accords with our Prediction 1 that the more expected a constituent in the language generally, the more transparent it is perceived to be. The top right-hand plot shows that, similarly, the greater the relation proportion in the N1 family, the more transparent N2 is perceived to be. This is in keeping with our Prediction 8 that the more easily accessible the



**Fig. 4** Random intercepts for items in final model for N1 transparency

semantic relation, the greater the perceived transparency of all elements of the compound. The greater the N1 relation proportion, the more expected is that relation in compounds with that particular modifier, so a high N1 relation proportion represents a high degree of expectedness in the internal structure of the compound. The positive correlation between N1 relation proportion and N2 transparency represents a correlation between transparency of the head and expectedness of the relation given the modifier. Finally, the bottom left-hand plot shows that the greater the spelling ratio, the lower the perceived transparency of N2. This is in keeping with our Prediction 13 that greater semantic lexicalisation of a compound is associated with lower perceived transparency of its constituents.

### 4.3 Transparency of compounds

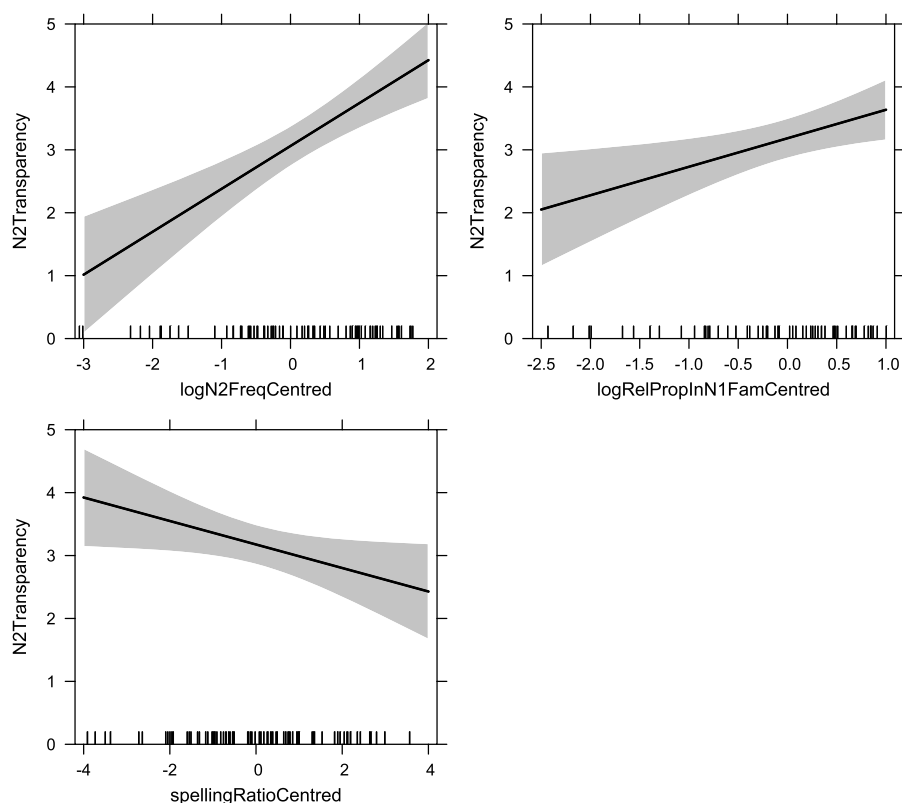
Table 5 shows the final model for transparency of the whole compound, and this is represented graphically in Fig. 6. The significant random effects include random intercepts for items, as well as by-annotator random intercepts and slopes for the effect of spelling ratio. It can be seen that there are significant fixed effects of N1 frequency, N1 relation proportion, N2 synset proportion and spelling ratio. One of the most striking things about this model is the extent to which it represents a combination of the two models for constituent transparencies. This can be

**Table 4** Final mixed effects model for N2 transparency, marginal  $R^2 = 0.26$ , conditional  $R^2 = 0.77$ 

Random effects:						
Groups	Name	Variance	Std. Dev.			
annotator	(intercept)	0.056162	0.23699			
	spellingRatioCentred	0.002847	0.05336			
annotator.1	(intercept)	0.003501	0.05917			
	logN2FreqCentred	0.012510	0.11185			
item	(intercept)	2.009099	1.41743			
residual		0.953022	0.97623			
Number of obs: 2328, groups: annotator, 108; item, 81						
Fixed effects:						
	Estimate	Std. Error	df	t value	Pr(> t )	Signif.
(intercept)	3.09783	0.16357	82.40	18.938	<2e-16	***
logN2FreqCentred	0.68191	0.14249	78.74	4.786	7.83e-06	***
logRelPropInN1FamCentred	0.45301	0.17608	77.00	2.573	0.0120	*
spellingRatioCentred	-0.18680	0.08969	77.86	-2.083	0.0406	*
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

clearly seen by comparing Figs. 2, 5 and 6. Firstly, there is a positive correlation between N1 relation proportion and the transparency of each constituent, and this carries over straightforwardly to the transparency of the compound as a whole. Similarly, there is a negative correlation between spelling ratio and the semantic transparency of each constituent, which also carries over to the model for compound transparency. In contrast, the correlation of N2 frequency with N1 transparency is the inverse of its correlation with N2 transparency, and these effects seem to cancel one another out in the compound, so that N2 frequency does not emerge as a significant predictor of compound transparency. However, the correlations of N1 frequency and N2 synset proportion with N1 transparency, positive and negative respectively, do survive in the model for compound transparency. We conclude that factors that increase the perceived transparency of either constituent will increase the perceived transparency of a compound as a whole, provided they do not simultaneously reduce the perceived transparency of the other constituent. Overall, these models support previous theories and empirical findings that compound transparency is best understood as a function of the constituent transparencies (Libben et al. 2003; Reddy et al. 2011).

Turning now to the individual effects in the model of compound transparency, the top left-hand plot in Fig. 6 shows that the greater the frequency of N1, the more transparent the compound is perceived to be. This provides partial evidence for our Prediction 9, that a compound will be perceived as more transparent the more frequent and therefore expected is either constituent. However, because N2 frequency is not significant in the model, there is no direct evidence for the second part of Prediction 9,



**Fig. 5** Partial effects in final model for N2 transparency

namely that greatest perceived transparency will occur when both constituents are frequent. Similarly, the top right-hand plot shows that the greater the relation proportion of N1, the more transparent the compound is perceived to be. This provides partial evidence for Prediction 11, that a compound will be perceived as more transparent the more expected the relevant semantic relation for either constituent. But again, there is no evidence for the second part of Prediction 11, namely that greatest perceived transparency will occur when the relation occurs in a high proportion of both modifier and head constituent families. The bottom right-hand plot shows that, in accordance with Prediction 14, the greater the spelling ratio of a compound, the lower the perceived transparency. As described above, we hypothesise that high spelling ratio is a correlate of semantic lexicalisation and especially of whole-compound semantic shift. In other words, the greater the spelling ratio, the less expected the sense of the compound relative to the dominant senses of the constituents in isolation.

Of all the plots in Fig. 6, the only one not in keeping with our predictions is the bottom left-hand plot, which shows that the greater the synset proportion in the N2 family, the lower the perceived transparency of the whole compound. This is contrary to our Prediction 12 that the effects of semantic shifts on constituent transparencies would cancel one another out in the model of compound transparency. At first sight,

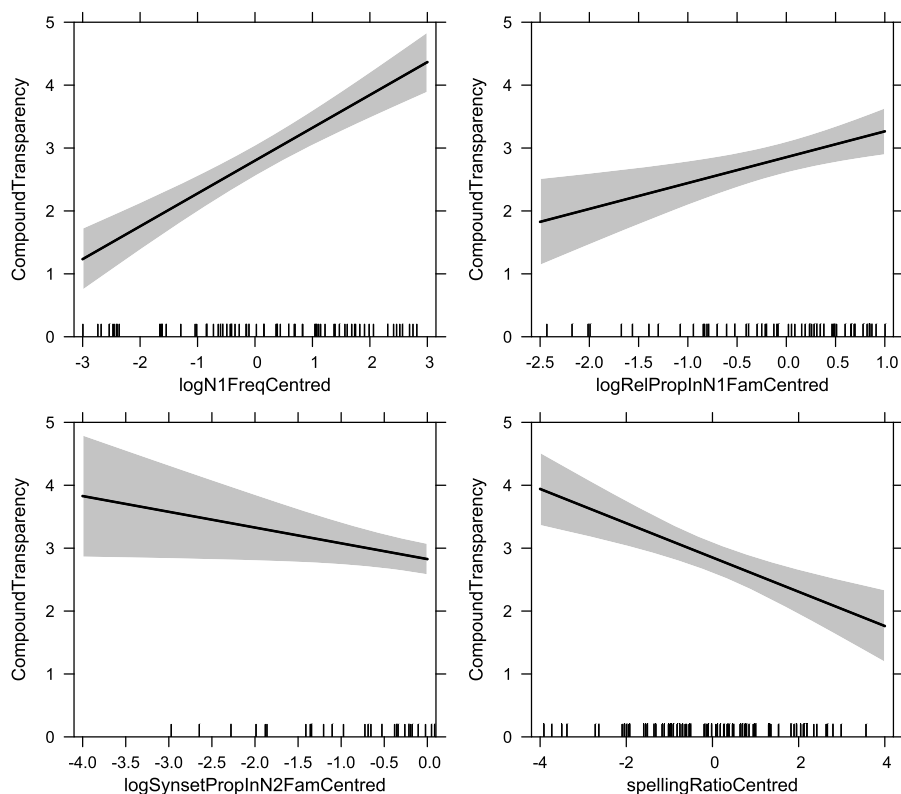
**Table 5** Final mixed effects model for compound transparency, marginal  $R^2 = 0.33$ , conditional  $R^2 = 0.72$ 

Random effects:						
Groups	Name	Variance			Std. Dev.	
annotator	(intercept)	0.150915			0.38848	
	spellingRatioCentred	0.004018			0.06339	
item	(intercept)	1.145611			1.07033	
residual		0.930349			0.96455	
Number of obs: 2307, groups: annotator, 119; item, 81						
Fixed effects:						
	Estimate	Std. Error	df	t value	Pr(> t )	Signif.
(intercept)	2.83314	0.12886	95.57	21.986	<2e-16	***
logN1FreqCentred	0.52198	0.07116	75.97	7.336	2.04e-10	***
logRelPropInN1FamCentred	0.41104	0.13442	76.04	3.058	0.003076	**
logSynsetPropInN2FamCentred	−0.25046	0.12056	76.10	−2.077	0.041134	*
spellingRatioCentred	−0.27239	0.06674	77.83	−4.081	0.000108	***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						

the negative correlation of N2 synset proportion with compound transparency seems to follow logically from its negative correlation with the transparency of N1. However, this creates a problem for our previous interpretation of the effect, whereby increased transparency of N1 was hypothesised to arise in contrast to the opacity of N2 when N2 synset proportion is low. Although this could explain the effect of N2 synset proportion on N1 transparency, it does not adequately explain the effect on compound transparency, where the opacity of N2 would be expected to neutralise the corresponding transparency of N1.

One possible explanation of the unexpected effect of N2 synset proportion is that it is actually reflecting some other underlying variable that was not included in our models. Such a variable could be the positional family size of N2. A high synset proportion can arise either with a small positional constituent family or with a large family. In the extreme case, if the family includes only one sense of the head word, then the synset proportion will be 1 irrespective of how many compounds the family contains. However, low values of synset proportion can only arise in larger families. To understand why this is, consider the situation where one compound in a family involves a unique sense of the head, and therefore has the lowest possible N2 synset proportion. If the family has two members, then each sense will have a synset proportion of 50%. If the family has five members, then the lowest possible synset proportion is 20% for a constituent with a unique sense. For a family with ten members, the lowest possible synset proportion is 10% while for a family with a hundred members, the lowest possible synset proportion is 1% and so on.

In order to investigate whether the N2 synset proportion effect was in fact reflecting an effect of positional family size, we ran a new model including positional



**Fig. 6** Partial effects in final model for compound transparency

family sizes alongside all other predictors originally included. Although this slightly increased the collinearity in the data, it was still well within acceptable limits ( $c\text{-number} = 3.777$ ).<sup>9</sup> The final model for compound transparency, including positional family sizes, is shown in Table 6 and represented graphically in Fig. 7. Again, positive coefficients in the first numerical column represent a positive correlation between the predictor and transparency while negative coefficients represent a negative correlation. It can be seen that the model is very similar to the model of compound transparency previously presented in Table 5. The random effects again include random intercepts for items, as well as by-annotator random intercepts and slopes for the effect of spelling ratio. Similarly, the significant fixed effects again include positive correlation of compound transparency with N1 frequency and N1 relation proportion, and negative correlation with spelling ratio. However, the previous effect of N2 synset proportion is now replaced with an effect of N2 positional family size. The greater the family size of N2, the more transparent the compound is perceived to be. We used the `anova` function in R to compare the fit of the two models; this revealed that the model using N2 family size rather than synset proportion is the better model of compound

<sup>9</sup>A  $c\text{-number}$  of 10 represents mild collinearity within a dataset, while  $c\text{-numbers}$  of 30 or more indicate that collinearity is so high as to be detrimental to statistical models using that data.

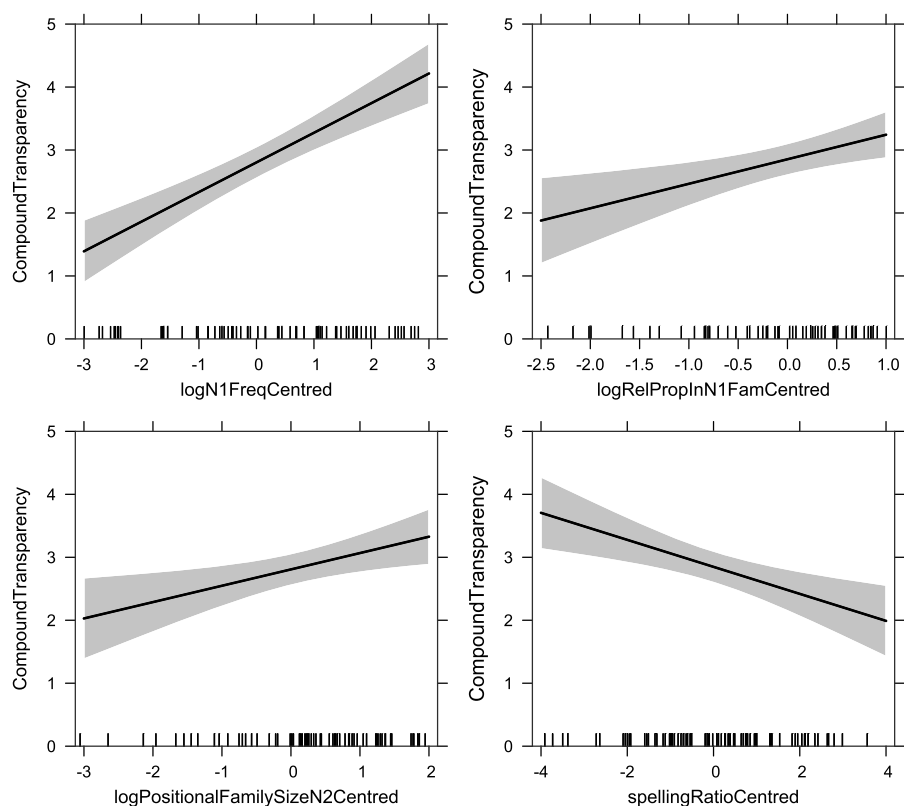
**Table 6** Final mixed effects model for compound transparency, including positional family sizes as predictors, marginal  $R^2 = 0.34$ , conditional  $R^2 = 0.72$ 

Random effects:						
Groups	Name	Variance			Std. Dev.	
annotator	(intercept)	0.150796			0.38833	
	spellingRatioCentred	0.004011			0.06333	
item	(intercept)	1.107062			1.05217	
residual		0.930388			0.96457	
Number of obs: 2307, groups: annotator, 119; item, 81						
Fixed effects:						
	Estimate	Std. Error	df	t value	Pr(> t )	Signif.
(intercept)	2.81222	0.12734	96.06	22.084	<2e-16	***
logN1FreqCentred	0.47065	0.07170	75.97	6.564	5.75e-09	***
logRelPropInN1FamCentred	0.38949	0.13259	76.04	2.938	0.00438	**
logPositFamilySizeN2Centred	0.25964	0.09793	76.01	2.651	0.00976	**
spellingRatioCentred	−0.21449	0.06521	77.87	−3.289	0.00151	**
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

transparency ( $p < 0.001$ ). This finding is compatible with our general hypothesis that transparency is a reflex of expectedness, and that factors that increase the perceived transparency of either constituent will increase the perceived transparency of a compound as a whole: the greater the positional family size of N2 the more productive it is as a compound head and therefore the more expected it is in that position, i.e. following another noun. Furthermore, the fact that N2 positional family size replaces N2 synset proportion in the final model, without affecting the significance of any other predictor, suggests that these two predictors do indeed account for the same portion of the variation in the dependent variable, compound transparency.

To investigate whether the effect of N2 synset proportion in the model for N1 transparency is also a reflex of N2 family size, we constructed a new model with N1 transparency as the dependent variable, and all predictors including the positional family sizes of N1 and N2. However, after stepwise removal of non-significant terms, the model was exactly the same as that presented in Table 1; in other words, N2 family size was not significant. There are at least two possible explanations for this. If N2 synset proportion and N2 family size do account for the same portion of the variation in perceived transparency, then the two variables may effectively compete for significance in our models, with one or the other winning out depending on the exact dataset used (recall that the three models use slightly different sets of data because of the way tasks were assigned to annotators, and how the data was subsequently cleaned). On the other hand, N2 synset proportion might actually reflect two different effects: a correlation between semantically shifted heads and transparent modifiers, most relevant for transparency of N1, and a collinear effect of N2 family size, most relevant for compound transparency.





**Fig. 7** Partial effects in final model for compound transparency, including positional family sizes as predictors

## 5 Conclusion

We set out in this paper to test the hypothesis that the perceived semantic transparency of compound constituents, and hence of compounds, is a reflex of the extent to which the internal semantic structure of the compound is expected. The predictions of this hypothesis have been at least partly borne out, although in a more complex manner than originally expected. We found some evidence for eight of our fourteen predictions. Perhaps least surprisingly, there is clear evidence that semantic lexicalisation of a compound, operationalised as spelling ratio, is associated with reduced transparency of both constituents and of the compound as a whole. We take this kind of lexicalisation to reflect a shift of the compound away from a compositional reading to one that would be less expected given the meanings of the constituents in isolation.

Another clear finding is that the perceived transparency of each constituent is strongly positively correlated with its frequency and hence with its expectedness in the language at large. However, we did not find the predicted effects of either synset proportion or of family size ratio on constituent transparency and it seems that, in terms of the relationship between the expectedness of a given constituent and its perceived transparency, the overwhelmingly important variable is how expected, i.e.

frequent, the free form of the constituent is generally. Any direct effects of the expectedness of particular senses, or of one positional compound role relative to the other, are apparently insignificant. The only significant synset proportion effect in the models of constituent transparency is an indirect effect of N2 synset proportion on transparency of N1. We had expected such indirect effects: since constituent transparency was judged in the context of particular compounds, we predicted that the perceived transparency of each constituent would be influenced by the expectedness of its partner. However, against our predictions, we did not find a positive correlation between the frequency of one constituent and transparency of the other, and nor did we find a symmetrical effect, only an effect of N2 frequency on N1 transparency. It seems that the more expected N2 is in the language generally, the lower the perceived transparency of N1; similarly, the more expected the specific sense of N2 as a compound head (as measured by synset proportion), the less transparent N1 is perceived to be.

In general, the contributions of the modifier and head to compound transparency are less symmetrical than our hypothesis predicted and, if compound transparency is a function of the transparencies of the constituents, the two constituents apparently differ in the nature of their contribution. In our model for whole compound transparency, the important predictors related to N1 are its overall frequency and its relation proportion, i.e. the expectedness of N1 in the language generally, and the expectedness of the compound semantic relation given N1. The contribution of N2 to compound transparency is best captured by its positional constituent family size, i.e. the expectedness of N2 as the head of a compound, given any N1. What seems to be important in this respect is the productivity of N2 as a head, regardless of whether or not it is also productive as a modifier (relative productivity in the two positions is captured by family size ratio, and was not found to be significant).

Why should the perceived transparency of compounds and their constituents be correlated with the expectedness of their semantic structure? We think the answer may be related to ease of processing. As described in Sect. 2.1, Pollatsek and Hyönä (2005) showed that compounds with more frequent first constituents were read faster; in other words, N1 frequency is positively correlated with ease of compound processing as well as with perceived compound transparency. Similarly, Gagné and Shoben (1997) showed that reaction time in a sense-nonsense decision task on compounds was determined by the strength of relation for the modifier; so N1 relation proportion is also positively correlated with both ease of processing and perceived transparency. In terms of N2 family size, Kuperman et al. (2009) showed an effect of right constituent family size on compound processing in a visual lexical decision task, although the impact of the right constituent was smaller than that of the left constituent and only showed up in the later stages of processing. This is in keeping with the RICE theory of conceptual combination (Spalding et al. 2010), according to which the head concept only plays a role in the later stages of compound interpretation. Finally, there is evidence that our spelling ratio effect might also be related to ease of processing. Kuperman and Bertram (2013) showed that the time taken to process concatenated compounds, both in lexical decision and naming tasks, was positively correlated with the tendency of the compound to occur with spaced orthography. In other words, the more frequently a compound occurs with one spelling format, the more difficult it is

to process when encountered in the other format. In our models, spelling ratio is the proportion of a compound's tokens that are written unspaced, whereas the annotators who rated the items for transparency were presented with the compounds written spaced (Reddy et al. 2011). Thus, assuming that the positive correlation between frequency of one spelling format and reaction time to the other format (Kuperman and Bertram 2013) applies to the reading of spaced as well as unspaced compounds, we would expect that the lower the spelling ratio for a compound in our dataset, the more easily were the annotators able to process it. Overall, there is evidence that those variables we have found to be significant predictors of perceived compound transparency are all also significant predictors of ease of processing.

If there is a three-way correlation between expectedness, semantic transparency and ease of processing, what is the likely causal relationship between them, if any? By expectedness, we essentially mean familiarity and frequency of use. It therefore seems plausible that greater expectedness will lead to greater ease of processing, since greater familiarity with an item or relational structure will mean that the language user has had more practice with it.<sup>10</sup> In the psycholinguistic studies described in Sect. 2.1, it sometimes seems to be assumed that perceived transparency is also the causal element in its relationship with ease of processing. However, the results reported in the present study suggest that the relationship might actually be the other way around, with ease of processing underlying perceived transparency. As discussed in Sects. 1 and 2, most definitions of semantic transparency make reference to the relationship between the meaning of a compound and the meanings of its constituent words. However, the meanings of constituent words are themselves a matter of interpretation. For example, the extent to which *web* in *website* is perceived as transparent will depend on the extent to which the annotator judges *web* to mean *internet*; we hypothesise that this in turn will be related not only to some folk etymology, but more significantly to the familiarity of the annotator with that sense of the word, both in and out of the compound context. The extent to which a reading is familiar to any individual will clearly depend on their own linguistic experience, both over their lifetime and perhaps especially in the recent past. In the present study we can only approximate to this familiarity, using measures that estimate the expectedness of semantic structures in the language as a whole.

The effects in our model of compound transparency are compatible with a hypothesis that the perceived transparency of a compound reflects the ease with which the appropriate sense can be retrieved or computed. When one encounters the first noun of a compound, it is not necessarily obvious that it is part of a compound, and so the most important index of expectedness is likely to be the expectedness of the word in the language overall, in our terms N1 frequency. Once the second noun occurs, however, the compound structure is much more likely, so the relevant index for N2 will be its expectedness as a compound head, i.e. the positional family size. Finally, we know from the studies described in Sect. 2.3.2 that the most important index of the expectedness of a compound's semantic relation, in terms of processing, is its association with the modifier, our relation proportion N1.

---

<sup>10</sup>There could also be a positive feedback effect, whereby users are more likely to choose to use items they themselves would find easier to process, so that ease of processing conversely leads to greater frequency of use.

It is perhaps surprising that the compound transparency ratings elicited by Reddy et al. (2011) are so clearly correlated with predictors of ease of processing, since the rating tasks were by no means performed online. Annotators could work at their own pace and could see both constituents simultaneously while making their judgement, in addition to having access to example sentences and alternative definitions. Our models for constituent transparencies may reflect more of the considered nature of the judgements, especially the model for N1 transparency, where the transparency of N1 seems to be judged partly in relation to that of N2, as indicated by the effects of N2 frequency and N2 synset proportion.

A brief note is in order about our use of synsets. Although we did not find the expected direct effects of sense frequency on constituent transparency, this may have been a result of the measure we used to operationalise it, namely synset proportion. In contrast to compound semantic relations, which only exist in the context of compounds, the different senses of compound constituents can also be found in the language more widely. Synset proportions calculated on the basis of constituent families may therefore not be representative of the expectedness of the various senses in the language generally. Furthermore, the synsets included in WordNet are by no means exhaustive and are sometimes rather idiosyncratic, reflecting the particular texts from which the WordNet vocabulary happens to have been extracted. This can lead to unevenness in the level of granularity of the classification system, so that synset proportions may not be comparable from item to item.

In summary, we have found robust effects of the expectedness of N1 generally, of the expectedness of the semantic relation given N1, and of the expectedness of N2 as a compound head, in predicting compound transparency. These findings are all the more robust because the effects survive in the presence of random effects for annotator and item. Our results support our hypothesis that perceived transparency is a reflection of expectedness in compound structure, and perhaps also of ease of processing. Studies that attempt to dissociate the effects of transparency from those of e.g. frequency may therefore be misguided unless, as suggested by Marelli and Luzzatti (2012), all critical interactions are tested.

**Acknowledgements** We are very grateful to Cyrus Shaoul and Gero Kunter, who provided the corpus frequencies for our analyses, and to Christoph Rzymiski, who provided discussion and assistance in matters related to statistics and R. We also thank Sabine Arndt-Lappe, Barbara Schlücker and three anonymous reviewers for their detailed and constructive comments on an earlier version of this paper. While the project was in progress, we received helpful feedback from Martin's colleagues in Jena as well as the audiences at several conferences where we had the chance to present our work—we thank them all. The work was partially supported by three short visit grants from the European Science Foundation through NetWordS—The European Network on Word Structure (grants 4677, 6520 and 7027), for which we are also extremely grateful.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). CELEX2. Philadelphia: Linguistic Data Consortium.

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016). Comprehension without segmentation: a proof of concept with naive discrimination learning. *Language, Cognition, and Neuroscience* 31(1), 106–128. doi:[10.1080/23273798.2015.1065336](https://doi.org/10.1080/23273798.2015.1065336).
- Bartofi, K. (2016). MuMIn: Multi-model inference. R package version 1.15.6. <https://CRAN.R-project.org>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9. <https://CRAN.R-project.org>.
- Bell, M. J., & Plag, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics*, 48(3), 485–520. doi:[10.1017/S0022226712000199](https://doi.org/10.1017/S0022226712000199).
- Bell, M. J., & Plag, I. (2013). Informativeness and analogy in English compound stress. *Word Structure*, 6(2), 129–155.
- Bell, M. J., & Schäfer, M. (2013). Semantic transparency: challenges for distributional semantics. In A. Herbelot, R. Zamparelli, & G. Boleda (Eds.), *Proceedings of the IWCS 2013 workshop: Towards a formal distributional semantics*. Potsdam: Association for Computational Linguistics (pp. 1–10). <http://www.aclweb.org/anthology/W13-0601>.
- Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: is there a “hat” in “that”? *Journal of Memory and Language*, 52(1), 131–143. doi:[10.1016/j.jml.2004.09.003](https://doi.org/10.1016/j.jml.2004.09.003).
- Downing, P. (1977). On the creation and use of English compound nouns. *Language*, 53(4), 810–842.
- El-Bialy, R., Gagné, C. L., & Spalding, T. L. (2013). Processing of English compounds is sensitive to the constituents’ semantic transparency. *The Mental Lexicon*, 8(1), 75–95. doi:[10.1075/ml.8.1.04elb](https://doi.org/10.1075/ml.8.1.04elb).
- Erk, K. (2012). Vector space models of word meaning and phrase meaning: a survey. *Language and Linguistics Compass*, 6(10), 635–653.
- Estes, Z. (2003). Attributive and relational processes in nominal combination. *Journal of Memory and Language*, 48(2), 304–319.
- Estes, Z., & Jones, L. L. (2006). Priming via relational similarity: a copper horse is faster when seen through a glass eye. *Journal of Memory and Language*, 55(1), 89–101.
- Fanselow, G. (1981). *Zur Syntax und Semantik der Nominalkomposition*. Linguistische Arbeiten (Vol. 107). Tübingen: Niemeyer.
- Fellbaum, C. (Ed.) (1998). *WordNet: an electronic lexical database*. Cambridge: MIT Press.
- Ferraresi, A., Zanchetta, E., Baroni, M., & Bernardini, S. (2008). Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the WAC4 workshop at LREC 2008*, Marrakech: ELRA.
- Frissón, S., Niswander-Klement, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology*, 99(1), 87–107. doi:[10.1348/000712607X181304](https://doi.org/10.1348/000712607X181304).
- Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 236–254.
- Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain and Language*, 81(1), 723–735.
- Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 71–87.
- Gagné, C. L., & Shoben, E. J. (2002). Priming relations in ambiguous noun–noun combinations. *Memory & Cognition*, 30(4), 637–646.
- Gagné, C. L., & Spalding, T. L. (2004). Effect of relation availability on the interpretation and access of familiar noun–noun compounds. *Brain and Language*, 90(1–3), 478–486.
- Gagné, C. L., & Spalding, T. L. (2014). Conceptual composition: the role of relational competition in the comprehension of modifier–noun phrases and noun–noun compounds. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 97–130). San Diego: Academic Press.
- Gagné, C. L., Spalding, T. L., Figueredo, L., & Mullaly, A. C. (2009). Does snow man prime plastic snow? The effect of constituent position in using relational information during the interpretation of modifier–noun phrases. *Mental Lexicon*, 4(1), 41–76.
- Gagné, C. L., Spalding, T. L., & 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.
- Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund-Prytz, Y. (2008). *Corpus linguistics with BNCweb—a practical guide*. English Corpus Linguistics (Vol. 6). Frankfurt am Main: Peter Lang.

- Huddleston, R., Pullum, G., et al. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., & Libben, G. (1999). Processing compounds: a cross-linguistic study. *Brain and Language*, 68(1–2), 362–369.
- Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, 65(4), 406–430.
- Juhász, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. P. v. Gompel (Ed.), *Eye movements: a window on mind and brain* (pp. 373–390). Amsterdam: Elsevier.
- Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: toward a multiple route model of lexical processing. *Journal of Experimental Psychology. Human Perception and Performance*, 35(3), 876–895.
- Kuperman, V., & Bertram, R. (2013). Moving spaces: spelling alternation in English noun–noun compounds. *Language and Cognitive Processes*, 28(7), 939–966.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29. <http://CRAN.R-project.org>.
- Lees, R. B. (1970). *The grammar of English nominalizations*. The Hague: Mouton.
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, 20(1), 1–31.
- Levi, J. N. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
- Levinson, S. C. (2000). *Presumptive meaning. The theory of generalized conversational implicature*. Cambridge: MIT press.
- Libben, G. (1998). Semantic transparency in the processing of compounds: consequences for representation, processing, and impairment. *Brain and Language*, 61(1), 30–44.
- Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: the role of semantic transparency and morphological headedness. *Brain and Language*, 84(1), 50–64.
- Maguire, P., Devereux, B., Costello, F., & Cater, A. (2007). A reanalysis of the CARIN theory of conceptual combination. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 33(4), 811–821.
- Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: modulation of headedness and semantic transparency. *Journal of Memory and Language*, 66(4), 644–664.
- Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101(1), 3–33.
- Monsell, S. (1985). Repetition and the lexicon. In A. Ellis (Ed.), *Progress in the psychology of language* (Vol. 2, pp. 147–195). Hove: Lawrence Erlbaum Associates Ltd.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142.
- Ó Séaghdha, D. (2008). *Learning compound noun semantics*. Cambridge: Computer Laboratory, University of Cambridge.
- Plag, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.
- Plag, I., Kunter, G., & Lappe, S. (2007). Testing hypotheses about compound stress assignment in English: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory*, 3(2), 199–233.
- Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language*, 84(4), 760–794.
- Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes* 20(1–2), 261–290. doi:10.1080/01690960444000098.
- R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.Rproject.org/>.
- Reddy, S., McCarthy, D., & Manandhar, S. (2011). An empirical study on compositionality in compound nouns. In *Proceedings of the 5th international conference on natural language processing*, AFNLP, Chiang Mai, Thailand (pp. 210–218).
- Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, 20(1), 33–54.
- Sandra, D. (1990). On the representation and processing of compound words: automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 42(3), 529–567. doi:10.1080/14640749008401236.

- Shaoul, C., & Westbury, C. (2013). A reduced redundancy USENET corpus (2005–2011), Edmonton, AB: University of Alberta, <http://www.psych.ualberta.ca/~westburylab/downloads/usenetcorpus.download.html>.
- Spalding, T. L., & Gagné, C. L. (2011). Relation priming in established compounds: facilitation? *Memory & Cognition*, 39(8), 1472–1486. doi:10.3758/s13421-011-0112-1.
- Spalding, T. L., Gagné, C. L., Mullaly, A. C., & Ji, H. (2010). Relation-based interpretations of noun–noun phrases: a new theoretical approach. In S. Olson (Ed.), *New impulses in word-formation*, *Linguistische Berichte, Sonderheft 17* (pp. 283–315). Hamburg: Buske.
- The British National Corpus (2007). BNC XML Edition. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. <http://www.natcorp.ox.ac.uk/>.
- Warren, B. (1978). *Semantic patterns of noun–noun compounds*. Göteborg: Acta Universitatis Gothoburgensis.
- Zwitserslood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9(3), 341–368.