

# Semantic transparency: challenges for distributional semantics

Melanie J. Bell  
Anglia Ruskin University  
melanie.bell@anglia.ac.uk

Martin Schäfer  
Friedrich-Schiller-Universität Jena  
post@martinschaefer.info

## Abstract

Using data from Reddy et al. (2011), we present a series of regression models of semantic transparency in compound nouns. The results indicate that the frequencies of the compound constituents, the semantic relation between the constituents, and metaphorical shift of a constituent or of the compound as a whole, all contribute to the overall perceived level of transparency. While not proposing an actual distributional model of transparency, we hypothesise that incorporating this information into such a model would improve its success and we suggest some ways this might be possible.

## 1 Introduction

Recently, a number of studies in distributional semantics have addressed the semantics of NN and AN compounds and phrases. Under the heading of compositionality, they often discuss phenomena that in the psycholinguistic and morphological literature are cast as issues of semantic transparency. In this paper, we are not proposing an actual distributional model of semantic transparency, but rather making some linguistic observations which have consequences for distributional models that attempt to capture the phenomenon. In Section 2, we introduce the notion of semantic transparency and its relation to compositionality in distributional semantics, and in Section 3 we present our descriptive framework for the semantics of complex nominals. Sections 4 and 5 describe the method and results, respectively, of our empirical study based on the data from Reddy et al. (2011), which we recode and use to build four regression models with semantic transparency as the dependent variable. Finally, Section 6 discusses the implications of our results for distributional models.

## 2 Semantic transparency and compositionality

The term ‘semantic transparency’ aims to capture the intuitive difference felt between compounds like *hogwash*, meaning ‘nonsense’, and a compound like *milkman*. In the literature, semantic transparency is defined in two main ways. One is the idea that it can be linked to meaning predictability. Plag (2003, 46) states that words are semantically transparent if “[...] their meaning is predictable on the basis of the word-formation rule according to which they have been formed.” According to this definition, *hogwash* is clearly not transparent. But the meaning of *milkman* also does not seem predictable. Assuming the standard dictionary definition *man who delivers milk to people’s houses*, are we to assume that there is a word formation rule of the kind *x who delivers y to people’s houses*? This kind of definition seems excessively restrictive. The second kind of definition uses analysability rather than predictability. A classic example is Zwitserlood (1994, 344), who writes that “[t]he meaning of a fully transparent compound is synchronically related to the meaning of its composite words [...]”. In this sense, *milkman* clearly is transparent because any possible usage will allow linking the interpretation in some way to the meanings of the constituent parts. But here the problem seems to be that the definition is too wide. Even in cases like *buttercup*, the name for the small flower with the yellow head, the meaning is related to the meanings of its composite words, because *butter* stands for the colour and *cup* for the shape of the actual flower.

While we know of no work that gives empirical correlates for establishing semantic transparency in terms of the meaning predictability approach, many psycholinguistic studies develop classification

schemes that correspond to the second approach to transparency, e.g. Zwitserlood (1994), Libben et al. (2003). Besides psycholinguistics, the term semantic transparency also occurs in standard linguistic works on compounds, e.g. the discussion of anaphoric islands in Ward et al. (1991).

In lieu of the term semantic transparency, some psycholinguistic and linguistic studies use the term ‘semantic compositionality’ to refer to similar phenomena, and this tradition of using semantic compositionality also occurs in some studies within distributional semantics. In formal semantics, however, semantic compositionality is typically used to describe sentence-level semantic processes, namely, that the meaning of a complex expression is composed of the meanings of its constituent expressions and the rules used to combine them. However, if we accept underspecified semantic representations, then almost all meanings are compositional. For example, taking *milkman* again, if its meaning is composed by combining the two predicates MILK(x) and MAN(x) with the help of the underspecified template in (1), where R represents an underspecified relation, this is technically semantically fully compositional.

$$(1) \quad \lambda B \lambda A \lambda y \lambda x [A(x) \ \& \ R(x,y) \ \& \ B(y)]$$

On this view, semantic transparency can still be seen as a compositionality issue in so far as it correlates with the amount of additional input that is involved in arriving at the meaning of a complex expression.

## 2.1 Compositionality in distributional approaches to XN semantics

Distributional studies of compositionality differ in what they actually try to model. Of most relevance here are composition models that try to model human judgements about XNs with the help of the vectors of their constituents and some compositionality function. Mitchell and Lapata (2010), for example, try to model human responses to a compound noun similarity task. Marelli et al. (2012) investigate the relation between distribution-based semantic transparency measures of compounds and constituent frequency effect in lexical decision latencies. Reddy et al. (2011) is a very good example where compositionality clearly corresponds to semantic transparency. While the term ‘semantic transparency’ does not occur in the paper, Reddy et al. (2011, 211) adapt the following definition of compound compositionality proposed in Bannard et al. (2003, 66): “[...] the overall semantics of the MWE [multi word expression] can be composed from the simplex semantics of its parts, as described (explicitly or implicitly) in a finite lexicon.” This is reminiscent of Plag’s definition of semantic transparency, and the link to semantic transparency becomes even clearer when looking at their operationalisation of the term. For the purposes of their paper, compositionality is equated with literality, and the aim of their models is to predict human ratings of compound literality. The compound literality ratings were elicited by asking the subjects to give a score ranging from 0 to 5 for how literal the phrase XY is, with a score of 5 indicating ‘to be understood very literally’, and a score of 0 indicating ‘not to be understood literally at all’. Since we use their data for the models presented here, we will simply adopt their view and treat their literality ratings as compositionality or, in our terms, semantic transparency measures.

To model the literality ratings of their subjects, Reddy et al. (2011) used a vector space model of meaning and compared the performance of constituent based models and composition function based models. For the constituent based models, literality scores for the constituents were computed, where literality was defined as similarity between compound and constituent co-occurrence vectors. The compound literality was then calculated by using 5 different functions, including additive and multiplicative functions. In contrast, for the composition function based models, a vector for the compound was composed from the vectors of its constituents. The compositionality score was then measured by comparing the resulting compound score with the vector of the compound calculated from the corpus. All the models were then evaluated against the human judgments on the compound literality (the constituent based models also against the constituent literality judgements). Among the constituent based models, those that used an additive or a combinatorial function performed best, but not as good as the composition function based models. Reddy et al. (2011, 217) hypothesize that “[t]he reason could be because 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*.”

### 3 A descriptive framework for semantic transparency

In order to capture and classify the internal semantic relations involved in semantic transparency, we start from the underspecified predicate logic notation in (2), which repeats (1), where A stands for the first part of a complex nominal, and B for the second part.

$$(2) \quad \lambda B \lambda A \lambda y \lambda x [A(x) \ \& \ R(x,y) \ \& \ B(y)]$$

We assume that an underspecified relation R links the denotations of A and B in a given construction. Based on this, we developed the scheme given in figure 1, where, for reasons of perspicuity, we omitted the arguments of the predicates (note that in a full model, they are needed, because they can be shifted independently from the predicates).

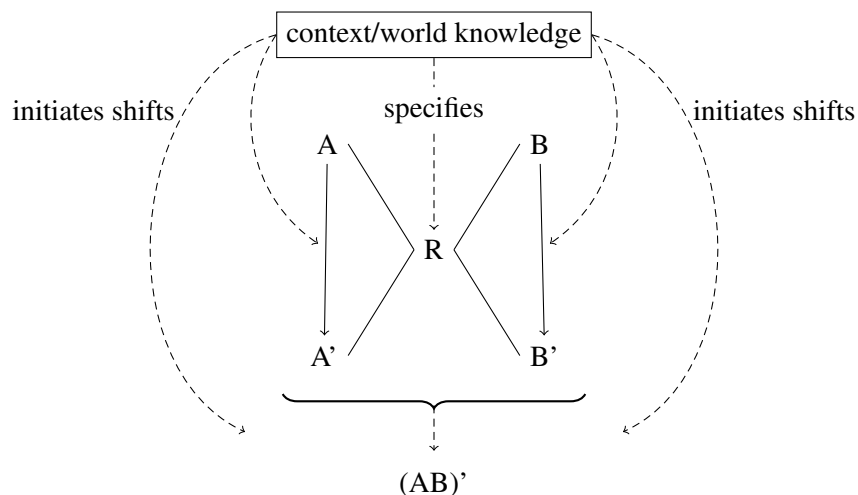


Figure 1: Scheme for A B combinatorics

As the scheme indicates, we assume that context and world knowledge are responsible for any further specification of the meaning of an AB combination. Specifically, we assume that A as well as B can be shifted from their literal meaning to a secondary meaning, labeled A' and B'. Metaphors and metonyms presents types of well-known shifts, other candidates would be e.g. the process of meaning differentiation, cf. Bierwisch (1982). However, even after a shift, they are still linked to the other part of the construction via the R relation. This kind of semantics for A B combinations therefore clearly falls into the category of radically underspecified approaches (cf. the characterization in Blutner (1998, 128)), and is much in the spirit of the ideas in Fanselow (1981) about the analysis of determinative compounds, and with him we assume that the specification of the exact relationship between the denotations as well as the shifts of the A and B parts fall into the domain of pragmatics.

The most basic configuration possible would be one where A and B retain their original meaning, and the relationship is set to identity. That is, the property expressed by A and by B hold of the very same entity, and the semantics is thus intersective. These combinations might be regarded as the most transparent AB combinations. Classic examples result from the combination of Kamp's (1975) predicative adjectives with a nominal head, e.g. *fourlegged animal*. However, even for standard examples of intersective modification further differentiation is needed, cf. the overview in Blutner (1998), and Kennedy (2007) specifically for gradable adjectives. We will give examples for shifted As and Bs in section 4.2.

**The relation R** As mentioned above, the underlying semantic format we assume is radically underspecified, and it is pragmatics and world knowledge that determine how the parameter R is specified. Since we hypothesize that the exact specification of R will have an influence on the semantic transparency of the AB combination, we need a way to distinguish between different possibilities of fixing R. Proposals for generalizations over this R relation can be taken from the large literature essentially

concerned with developing generalizations over possible relations, for English most famously in Lees (1970), Warren (1978) and Levi (1978). We chose the classification scheme from Levi (1978) (cf. also the discussion in Ó Séaghdha (2008)), fully aware that her scheme, or in fact any generalized scheme, will not allow one to reproduce the exact meaning nor all the possible meanings of AB combinations (for comprehensive criticism to this end, cf. Downing (1977); Fanselow (1981)). On the other hand, note that Gagné and Spalding (2009), in a series of priming experiments, find that the ease of deriving the meaning of a compound word ‘is mutually determined by the ease with which the constituents can be assigned to a particular role within a relational structure and by the availability of the appropriate relational structure.’ Since there is evidence that these relational structures have psychological reality, it seems likely that not only the semantics of the individual constituents, but also the relation between them, contributes to overall level of transparency.

In our scheme, we also allow for whole compound shifts. At this point, we just indicate this possibility by the (AB)’ in the scheme, without distinguishing in detail between the further internal possibilities. A very clear example of a whole compound shift is the derogative *asshole*, examples from the dataset used later include *ivory tower* and *cloud nine*. The complex possibilities can be illustrated by a combination like *buttercup*, discussed in the introduction.

## 4 Method

To test our hypothesis that the degree of semantic transparency of a complex nominal will be affected by the semantic relation between its constituents as well as shifts in meaning of the constituents or of the construction as a whole, we devised a series of regression models.

### 4.1 Dataset

We used the publicly available data set collected for and described in Reddy et al. (2011) (see the references for the download site). These authors selected a set of 90 compound nouns from the ukWaC corpus, a large web-derived corpus of English (Ferraresi et al. 2008). The sample was selected semi-randomly in such a way as to maximise the probability that it included different degrees of semantic transparency. Furthermore, all the selected compounds occurred at least 50 times in the corpus. For each of the 90 compounds, Reddy et al. obtained literality ratings from 30 raters. Importantly, the individual raters went through two distinct steps in rating each of the items. First, the rater was presented with ‘all possible definitions’ (Reddy et al. 2011) of the compound under investigation: in practice 1 or 2 definitions. The rater was asked to read through 5 randomly selected example sentences containing the compound and, for compounds with alternative definitions, decide which definition applied most frequently. In the second step, they were asked to rate either (a) how literal they perceived the compound to be, or (b) how literally the first constituent was used in the compound or (c) how literally the second constituent was used in the compound. This procedure has two important advantages: firstly, compounds are always presented in context, avoiding the artificiality associated with presenting words in isolation, and secondly, forcing the raters to settle on the most frequent definition ensures that the subsequent ratings are made for the compound with this particular reading. This elegantly avoids the usual problems that arise from the ubiquitous vagueness and ambiguity of compounds. For the purposes of this study we assume that the perceived literality of a compound or compound constituent is a measure of its semantic transparency.

The Reddy et al. dataset contains 30 ratings for each of the three tasks (a-c above) for all 90 compounds: in other words, a total of 8100 ratings. However, because tasks were assigned randomly to raters, the same rater did not necessarily perform all three tasks for any given compound. Since we wanted to use the perceived literality of the constituents to predict the perceived literality of the compound, we chose to use within-subject comparisons: this would allow us to model how well an individual’s perception of constituent literality predicts their compound literality rating. From the total dataset, we therefore extracted only those items for which the same rater had performed all three tasks. This produced a set of 1337 tokens for which literality judgements for each constituent as well as the compound as a whole

had been given by a single person. Within this set, 12 of the 90 compound types showed variation in the definition assigned, i.e. each of the possible definitions had been chosen by at least one rater. Because we were interested in the relationship between semantic structure and literality ratings, we wanted to code and analyse these different readings separately from one another. A token-based analysis allowed us to do this since, for each token, the dataset indicates the definition assigned by the rater in question.

## 4.2 Categories coded

We coded this set of compounds for a variety of semantic and frequency-based variables.<sup>1</sup> The semantic coding was definition-specific: each token was coded according to the definition chosen by the particular literality rater, so different tokens of the same compound did not necessarily receive identical coding. To encode the relations that can be used to specify the R-parameter, we used the classification system of Levi (1978) which has proven itself to be useful in computational linguistics (cf. Ó Séaghdha 2008). As far as shifts of the A and B constituents were concerned, we only distinguished between metaphorical and metonymic shifts. This coding was done by two linguists (the authors), one a trained semanticist and the other a native speaker of English: we first coded independently, and then discussed the results to reach a consensus about those items where we initially disagreed. For two compounds, *kangaroo court* and *flea market*, we were unable to reach consensus and these were therefore subsequently excluded.

The following examples from the dataset illustrate our coding scheme: *application form*, defined as *a form to use when making an application*, was classified as having unshifted first and second constituents, and the parameter R was set to FOR ('a form for an application'). In contrast, *crash course*, defined as *a rapid and intense course of training or research*, contains a metaphorical shift of the first element ('sth. fast and intense'), and R is set to BE. A metaphorical shift of the second element is exemplified by *eye candy*, where *candy* is shifted to mean *something pleasing but intellectually undemanding*. Again, the relationship is FOR. *Ground floor* exemplifies the IN-relation, which includes temporal and spatial location, and *brick wall* exemplifies the MAKE (TYPE 2) relation.

We also coded whether the compound as a whole had been shifted, as in *ivory tower* for example. *Ivory tower* as a whole stands for 'A condition of seclusion or separation from the world' (OED online), and it is not possible to synchronically decompose it further in any sensible way. However, it is clear to the native speaker that there has been a shift; otherwise it is unexplainable why, although neither *ivory* nor *tower* have anything to do with its current meaning, the concept of *tower* still shines through in expressions like *live in ivory towers/assault their ivory towers/geek atop an ivory tower*.

In addition, we extracted various frequency measures from the British National Corpus, namely the lemmatised frequencies of the individual constituents and of the whole compound written spaced and unspaced (either hyphenated or as a single word). On the basis of the last two of these measures, we calculated the 'spelling ratio' for each compound: this is the proportion of tokens that are written unspaced, which is taken to be a measure of the degree of lexicalization (Bell and Plag 2012).

## 4.3 Statistical analysis

The frequency and semantic variables were used as predictors in ordinary least squares regression analyses with literality of the compound or its constituents as the dependent variables. To alleviate the potentially harmful effects of extreme values on our statistical models, all quantitative predictors were first logarithmatised. Some of the semantic categories, including all metonymical shifts and several values of R, applied to very few compounds in the dataset. This would greatly reduce the power of any statistical analysis involving these variables: failure to reach significance could be simply the result of low frequency in this particular set of compounds or significant effects could be due to other features of those particular types. We therefore included in the analyses only metaphorical shifts and the three most frequent values of R, namely FOR, IN and BE. Each of the classes coded was represented by at least 9 types (i.e. compound senses) and 140 tokens in our data.

---

<sup>1</sup>Our semantic codings are available at [www.martinschaefer.info/publications/TFDS-2013/TFDS-2013\\_Bell\\_Schaefer.zip](http://www.martinschaefer.info/publications/TFDS-2013/TFDS-2013_Bell_Schaefer.zip)

	Coef	S.E.	t	Pr(>  z )
Intercept	-0.5861	0.3207	-1.83	0.0678
logFreqN1	0.2830	0.0243	11.63	<0.0001
logFreqN2	0.1535	0.0283	5.42	<0.0001
spellingRatio	-0.1240	0.0249	-4.98	<0.0001
Ametaphor=Yes	-0.6397	0.0939	-6.82	<0.0001
Bmetaphor=Yes	-0.4841	0.0920	-5.26	<0.0001
ABmetaphor=Yes	-1.8411	0.0910	-20.23	<0.0001
In=Yes	0.6041	0.1273	4.75	<0.0001
For=Yes	0.2363	0.0882	2.68	0.0074

Table 1: Final model for compound literality using semantic and frequency-based predictors,  $R^2 = 0.459$

	Coef	S.E.	t	Pr(>  z )
Intercept	-0.8117	0.2211	-3.67	0.0003
literality of A	0.4558	0.0179	25.43	<0.0001
literality of B	0.4147	0.0180	23.03	<0.0001
logFreqN1	0.0804	0.0179	4.50	<0.0001
logFreqN2	0.0506	0.0196	2.58	0.0100
Ametaphor=Yes	-0.2361	0.0720	-3.28	0.0011
Bmetaphor=Yes	-0.2059	0.0726	-2.84	0.0046
ABmetaphor=Yes	-0.1849	0.0752	-2.46	0.0141

Table 2: Final model for compound literality including constituent literality ratings,  $R^2 = 0.739$

## 5 Results

**Model 1** We first modelled the overall literality of the compound, as given by the human raters, using our semantic and frequency-based variables as predictors. Table 1 shows the final model, from which all non-significant predictors have been removed step-wise, following standard procedures of model simplification. In all tables in this paper, positive coefficients indicate a tendency towards higher literality, i.e. transparency, while negative coefficients indicate a tendency towards lower literality, i.e. opacity.

It can be seen that both types of predictor, semantic and frequency-based, were found to be statistically significant. Literality increases with increasing frequency of either constituent and, as might be expected, falls as the proportion of unspaced tokens increases (i.e. as lexicalization increases). Literality rating is lower when either constituent, or the whole compound, is metaphorical. Most significantly, however, certain semantic relations (FOR and IN) are associated with greater literality. On the assumption that literality is a measure of semantic transparency, this is the first evidence that the relation between constituents, as well as the semantics of the constituents themselves, contributes to transparency.

**Model 2** We next included the human ratings for constituent literality as predictors, alongside those used in the previous model. Reddy et al. (ibid.) show that there is a strong correlation between the average literality scores for the compounds and those for their constituents, so we expected that the constituent literality scores would be highly significant predictors in our model. Furthermore, on the assumption that the properties of a constituent contribute to its degree of transparency, we hypothesised that the constituent literality ratings would subsume our other constituent-based variables, namely frequency and semantic shift. We therefore expected that these variables would become less significant or even insignificant in the presence of constituent literality. On the other hand, we expected that the effects of semantic relations and whole-compound metaphorical shifts would remain significant, since they are properties of the whole compound, rather than either constituent.

The final model, from which all non-significant predictors have been eliminated, is shown in Table 2. As expected, the literality ratings of the constituents are highly significant predictors of overall literal-

	Coef	S.E.	t	Pr(>  z )
Intercept	-0.3791	0.3418	-1.11	0.2676
logFreqN1	0.3406	0.0262	12.99	<0.0001
logFreqN2	0.0953	0.0305	3.13	0.0018
spellingRatio	-0.0674	0.0268	-2.51	0.0122
Ametaphor=Yes	-1.7234	0.1003	-17.19	<0.0001
Bmetaphor=Yes	0.8728	0.0987	8.85	<0.0001
ABmetaphor=Yes	-1.8728	0.0939	-19.95	<0.0001
In=Yes	0.9275	0.1344	6.90	<0.0001

Table 3: Final model for literality of constituent A,  $R^2 = 0.499$

	Coef	S.E.	t	Pr(>  z )
Intercept	1.2383	0.3448	3.59	0.0003
logFreqN1	0.1224	0.0259	4.73	<0.0001
logFreqN2	0.1443	0.0304	4.75	<0.0001
spellingRatio	-0.1563	0.0264	-5.93	<0.0001
Ametaphor=Yes	0.8382	0.1009	8.31	<0.0001
Bmetaphor=Yes	-1.6511	0.0989	-16.70	<0.0001
ABmetaphor=Yes	-2.0563	0.0978	-21.02	<0.0001
For=Yes	0.2241	0.0929	2.41	0.0160

Table 4: Final model for literality of constituent B,  $R^2 = 0.498$

ity: in each case, the more literal the constituent, the more literal the compound. Surprisingly, however, the other constituent-based variables remain significant even in the presence of the constituent literality ratings: though the effects are much weakened, an increase in frequency of either A or B still leads to greater overall transparency, while metaphorical shift of either constituent leads to greater opacity. It might be argued that the strong effects in our models of metaphorical shifts are a result of the data collection method: asking subjects to rate literality may have led them actually to rate the presence or absence of metaphor. However, if this were all they rated, we would not expect the effects of metaphorical shift of A or B to survive in model 2 alongside the constituent literality ratings, since both types of predictor would be accounting for the same portion of the variance. An even more unexpected finding is that, once constituent literality ratings are included in the model, lexicalisation and semantic relations become insignificant as predictors of overall transparency. This suggests that these relations are correlated with the literality of the constituents, so that they account for the same portion of the overall variation.

**Models 3 and 4** To test the hypothesis that the semantic relation between compound constituents influences the extent to which the constituents are perceived as having literal readings, we constructed two models with the literality ratings of A and B respectively as the dependent variables, and our semantic and frequency-based variables as the predictors.

Table 3 shows the final model for literality of constituent A, with non-significant predictors removed. It can be seen that one semantic relation, IN, is indeed associated with an increase in perceived literality, and constituent A is also perceived as more literal as the frequency of *either* constituent increases. On the other hand, when the compound is more highly lexicalized (as indicated by a higher spelling ratio), or when the whole compound has undergone metaphorical shift, constituent A is perceived as less literal; similarly, when A itself has shifted metaphorically, it is perceived as less literal. However, in contrast to the frequency effects, metaphorical shift of B leads to A being perceived as more literal, presumably relative to B. Table 4 shows the final model for literality of constituent B, again with non-significant predictors removed. This is very similar to the model for A except that here the relation FOR is associated with an increase in perceived literality.

It is interesting both that the effect of semantic relation on compound transparency is mediated

through the transparency of the constituents, and that each constituent is associated with a different relation in this respect. The results tie in with recent work on prosodic prominence in the English NN. Plag et al. (2008), for example, demonstrate that the FOR relation is correlated with stress on N1, whereas IN is correlated with stress on N2. Furthermore Bell and Plag (2012) show that stress tends to fall on the most informative constituent. If FOR is associated with greater transparency of N2, that might explain why in such compounds stress tends to fall on N1, the assumption being that the less transparent constituent is also the more informative. The reverse pattern would hold in the case of compounds with R set to IN: N1 is more transparent, hence N2 is relatively more informative, hence prone to be stressed.

## 6 Consequences for distributional semantics

The findings described in this paper pose challenges for a distributional account of semantic transparency. In particular, if the aim is to use distributional semantics as a tool to understand human language processing, then those semantic factors that play a role in human processing should be reflected in distributional models. And, although the models of human literal ratings tested by Reddy et al. (2011) are not unsuccessful, there is still room for improvement. The strong effects of constituent frequency in our models, for example, suggest that it would be worth experimenting with different ways of introducing frequency-based weightings. We also hypothesise that taking into account the internal semantic structure of the data could further improve model performance. In this respect, we see two promising directions that can be explored, one concerning shifts, the other concerning semantic relations.

Vecchi et al. (2011) use distributional semantics to characterise semantic deviance in ANs. Unattested ANs were rated by two of the authors using a 3-point scale (deviant, intermediate or acceptable), where the two endpoints marked ‘semantically highly anomalous, regardless of effort’ vs. ‘completely acceptable’. Only those items with inter-rater agreement on ‘deviant’ or ‘acceptable’ were included in the test set. They investigated the ability of three measures to distinguish between deviant and non-deviant ANs: length of the AN vectors, cosine similarity between vectors for AN and N, and the average cosine with the top 10 nearest neighbours (density). Of these three indices, only the first and the last yield significant results for AN classification. The authors hypothesize that a wide angle between N and AN might not be a measure of deviance, but rather a common feature of a number of types of non-deviant ANs, among them metaphorical constructions. If they are correct, it should be possible to use cosine similarity on acceptable AN combinations in order to identify shifts.

However, it is not clear to what extent the distinction between unattested metaphorical and deviant types is real. The examples given by Vecchi et al. (ibid.) suggest that the difference may be a matter of degree and related to semantic transparency. Combinations were rated as deviant if the authors found them ‘semantically highly anomalous no matter how much effort one put in’, but there was very low inter-rater agreement, and all four examples of deviant types given in the paper seem to us effortlessly interpretable: e.g. you might suffer with an *academic bladder* if you need to urinate every 50 minutes, *blind pronunciation* could be the attempted pronunciation of a word in a language you don’t recognise, *sharp glue* could be glue with a pH less than 7 and, by analogy with *couch potato*, a *parliamentary potato* could be a parliamentarian who spends a lot of time on the benches but makes little contribution to the proceedings. However, all these interpretations, though readily available, involve semantic shifts and are therefore relatively opaque according to our model. Vecchi et al.’s (ibid.) acceptable examples, on the other hand, are relatively transparent. Three of these, *vulnerable gunman*, *huge joystick* and *blind cook* have obvious literal interpretations where there is no shift and R is set to identity, in addition to any possible metaphorical meanings. The fourth, *academic crusade* also has fairly obvious possible meanings involving a shift only of *crusade* (a crusade by academics, for example) and the shift of *crusade* from its original meaning is now so frequent that it is debatable whether it is a shift at all. Deciding at what point a diachronically shifted meaning becomes central was one of the difficulties we had when coding our data cf. *china* in *china clay* and *web* in *web site*. In a similar vein, a reviewer points out that one reason for the correlation between transparency and constituent frequency might be that shifted meanings are more likely to be taken as literal/lexicalised with more frequent words. Vecchi et al. (ibid.)



themselves acknowledge that ‘semantically deviant’ expressions might be interpretable metaphorically, and suggest that distributional measures might ‘naturally lead to a gradient notion of semantic anomaly’. It seems that such a notion of semantic anomaly is in fact very close to our notion of semantic opacity, and it would therefore be very interesting to devise distributional models similar to those used by Vecchi et al. (ibid.) for the data collected by Reddy et al. (2011). If we are right that the two notions are similar and that a single model of transparency can encompass all complex nominals, then we would expect to find similar results.

With regard to semantic relations, the situation is a bit more challenging. A fairly straightforward first step might be to use a compound classification algorithm on the data.<sup>2</sup> Ideally, this should be combined with automatic selection of the relevant senses of the compound constituents using the methodology described in Reddy et al. (2011), i.e. by using either static or dynamic prototypes. Another relevant study is Boleda et al. (2012) who evaluate different composition functions with respect to their ability to model the distinction between three types of adjectival modifier: intersective, subsective, and intensional. These are exemplified in the AN combinations *white towel*, *white wine* and *former bassist* respectively. In terms of our model, the three types can be seen as representing a cline in transparency, with intersective types the most transparent and intensional types the most relatively opaque. Boleda et al. (ibid.) find that the cosine between the A and AN vectors differs significantly between the three groups, being highest for intersective types, lowest for intensional types and intermediate for subsective. There was little difference between the groups when A was compared with N, or AN was compared with N. Likewise, Reddy et al. (2011) obtained much better results for compound literalness modelled on the basis of N1 alone, than on the basis of N2 alone. These results suggest that, in a distributional model of AB semantic transparency, the vector for A should be given more weighting than the vector for B. This might also partly explain why Vecchi et al. (2011) did not get significant results using the cosine between AN and N.

## Acknowledgements

We thank the three anonymous reviewers for their advice and comments, and we especially thank Aurelie Herbelot for a most fruitful abundance of the same.

## References

- Bannard, C., T. Baldwin, and A. Lascarides (2003). A statistical approach to the semantics of verb-particles. In *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment - Volume 18*, MWE '03, Stroudsburg, PA, USA, pp. 65–72. Association for Computational Linguistics.
- Bell, M. J. and I. Plag (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics* 48(3), 485–520.
- Bierwisch, M. (1982). Formal and lexical semantics. *Linguistische Berichte* (80), 3–17.
- Blutner, R. (1998). Lexical pragmatics. *Journal of Semantics* 15(2), 115–162.
- Boleda, G., E. M. Vecchi, M. Cornudella, and L. McNally (2012). First-order vs. higher-order modification in distributional semantics. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Juhu Island, Korea, pp. 1223–1233. Association for Computational Linguistics.
- Downing, P. (1977). On the creation and use of english compound nouns. *Language* 53(4), 810–842.
- Fanselow, G. (1981). *Zur Syntax und Semantik der Nominalkomposition*, Volume 107 of *Linguistische Arbeiten*. Tübingen: Niemeyer.

---

<sup>2</sup>This was suggested by Aurelie Herbelot.

- Ferraresi, A., E. Zanchetta, M. Baroni, and S. Bernardini (2008). Introducing and evaluating ukwac, a very large web-derived corpus of english. In *Proceedings of the WAC4 Workshop at LREC 2008*, Marrakech. ELRA.
- Gagné, C. L. and T. L. Spalding (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language* 60, 20–35.
- Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal Semantics for natural languages*, pp. 123–155. Cambridge, UK: Cambridge University Press.
- Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30, 1–45.
- Lees, R. B. (1970). *The Grammar of English Nominalization*. The Hague: Mouton.
- Levi, J. N. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
- Libben, G., M. Gibson, Y. B. Yoon, and D. Sandra (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language* 84, 50–64.
- Marelli, M., G. Dinu, R. Zamparelli, and M. Baroni (2012). Semantic transparency and the distributional origin of constituent effects in compound processing. Poster presented at the conference Architectures and Mechanisms for Language Processing (AMLAP) 2012, Riva del Garda, Italy, September 6-8.
- Mitchell, J. and M. Lapata (2010). Composition in distributional models of semantics. *Cognitive Science* 34(8), 1388–1429.
- Ó Séaghdha, D. (2008). Learning compound noun semantics. Technical Report 735, Computer Laboratory, University of Cambridge.
- Plag, I. (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.
- Plag, I., G. Kunter, S. Lappe, and M. Braun (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in english. *Language* 84.4, 760–794.
- Reddy, S., I. P. Klapaftis, D. McCarthy, and S. Manandhar (2011, November). Dynamic and static prototype vectors for semantic composition. In *Proceedings of The 5th International Joint Conference on Natural Language Processing 2011 (IJCNLP 2011)*, Chiang Mai, Thailand.
- Reddy, S., D. McCarthy, and S. Manandhar (2011). An empirical study on compositionality in compound nouns. In *Proceedings of the 5th International Conference on Natural Language Processing*, Chiang Mai, Thailand, pp. 210–218. AFNLP. All data for the paper is available from the following site: [http://sivareddy.in/papers/files/ijcnlp\\_compositionality\\_data.tgz](http://sivareddy.in/papers/files/ijcnlp_compositionality_data.tgz).
- Vecchi, E. M., M. Baroni, and R. Zamparelli (2011). (linear) maps of the impossible: Capturing semantic anomalies in distributional space. In *Proceedings of the DISCO (Distributional Semantics and Compositionality) Workshop at ACL 2011*, East Stroudsburg PA, pp. 1–9. ACL.
- Ward, G., R. Sproat, and G. McKoon (1991). A pragmatic analysis of so-called anaphoric islands. *Language* 67(3), 439–474.
- Warren, B. (1978). *Semantic patterns of noun-noun compounds*. Number 41 in Gothenburg studies in English. 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.