Chapter 2

Do semantic features capture a syntactic classification of compounds? Insights from compositional distributional semantics

Sandro Pezzelle
Institute for Logic, Language and Computation, University of Amsterdam

Marco Marelli
Department of Psychology, University of Milano-Bicocca

Classifying compound words has been the ultimate goal of much research in formal linguistics. A popular, cross-linguistically applicable classification (Bisetto & Scalise 2005) distinguishes three main types of compounds, namely Subordinate, Attributive, and Coordinate on the basis of the underlying *syntactic* relation between the compound elements. Similar tripartitions have also been proposed in cognitive psychology by works exploring conceptual combination. Focusing on the type of *semantic* interpretation assigned to novel combinations, three main classes have been traditionally described, namely Relation-linking, Property-mapping, and Hybrid or Conjunctive (see Wisniewski 1996). Based on these commonalities, we conjecture that syntax-based compound types might also be explained by means of the semantic properties of the compound and its constituents. Using a compositional model of distributional semantics (cDSM), we show that (a) the contribution of each constituent in determining the meaning of the compound and (b) the semantic similarity between the two constituent words are significant predictors of these classes. These findings suggest that the various compound types identified by syntactic criteria can also be predicted by means of semantic features. On the one hand, this confirms the validity of the proposed linguistic categorization. On the other hand, we bring further evidence proving the effectiveness of cDSMs in describing linguistic phenomena.

Sandro Pezzelle & Marco Marelli. 2020. Do semantic features capture a syntactic classification of compounds? Insights from compositional distributional semantics. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 33–60. Berlin: Language Science Press. DOI:10.5281/zenodo.3598556

1 Introduction

1.1 Classifying compounds

Compounding, namely the mechanism by which two independent words (e.g. *pet*, *food*) combine together to form a novel morphologically-complex word (e.g. *petfood*), is one of the most extensively covered topics in the literature of word formation.[1] On the theoretical level, many linguists have been particularly interested in classifying *compounds* according to various criteria, such as "headedness" (roughly speaking, the position and the characteristics of the compound *head*, the dominant word in the compound, e.g. *food* in *petfood*) (Bloomfield 1933; Fabb 1998); the presence of a verb or a deverbal noun (Marchand 1969); the kind of underlying relation between the *constituent* words, either at a syntactic level (Bloomfield 1933; Bally 1950; Lees 1960; Bisetto & Scalise 2005; Baroni et al. 2009; Dressler 2006; Scalise & Bisetto 2009) or at a semantic level (Levi 1978; Warren 1978; Fanselow 1981). Though different and pertaining to somehow diverse levels of analysis, these criteria have been traditionally explored and mixed together within the same classification framework (see among others Bauer 2001; Haspelmath 2002; Booij 2005). As a consequence, many influential proposals distinguish various classes of compounds on the basis of several overlapping properties that often generate an inconvenient number of subclasses and special cases.

To overcome this issue, Bisetto & Scalise (2005) proposed a cross-linguistic (and nowadays widely accepted) classification framework based on a single, homogeneous criterion, that is, the underlying syntactic relation between the compound constituents. Three main classes of compounds are isolated, namely Subordinate, Attributive, and Coordinate. To illustrate, the compound *doghouse* belongs to the Subordinate class, since the syntactic relation subtending *dog* and *house* is that of subordination. Indeed, the compound can be paraphrased as 'the house of the dog'. In contrast, *swordfish* is labeled as Attributive, given that the first constituent, *sword*, acts as an attribute of *fish* (a *swordfish* is 'a fish whose nose is shaped like a sword'). Finally, Coordinate compounds are formations like *comedy-drama*, where the first and the second constituent are linked by the underlying conjunction 'and'.

1.2 From *word* combination to *conceptual* combination

Interestingly, a similar tripartition has been proposed in the cognitive psychology literature by works on *conceptual combination* (Wisniewski 1996; Costello &

[1]For a complete and exhaustive overview of compounding, see Lieber & Štekauer (2009).

Keane 2000), where the focus is on the type of interpretations provided by people to novel combinations. By analyzing the circumlocutions produced by speakers to interpret novel compounds like *zebra-horse*, in fact, three main classes have been traditionally isolated, namely Relation-linking, Property-mapping, and Hybrid or Conjunctive. The first class includes interpretations involving a relation between the two concepts, i.e. a *zebra-horse* is 'a horse that preys zebras'. In the second, a property of one concept is mapped to the other, i.e. a *zebra-horse* is 'a striped horse'. In the third, the novel concept is interpreted as a hybrid or conjunction of the constituent concepts, i.e. a *zebra-horse* is 'a creature having many properties of both horses and zebras'. Though the aim of these works is to study the various interpretations to novel conceptual combinations, without any interest in recognizing classes of *lexicalized* compound words, the types they identify are reasonably comparable to the linguistic ones proposed by Bisetto & Scalise (2005). In particular, Relation-linking interpretations correspond to compounds included in the Subordinate class, Property-mapping to Attribute, and Hybrid/Conjunctive to Coordinate.

A notable difference is that the *linguistic* classification accounts for lexicalized (or familiar) compounds, whereas the *cognitive* one describes novel combinations which still lack a single, well-defined interpretation. However, we can easily assume that lexicalized compounds are the linguistic realization of a conceptual combination process, in a way that all compounds start out as novel formations and become lexicalized with usage in time (Gagné & Spalding 2006). Consistent with this claim is recent evidence showing that, in the processing of both novel and familiar compounds, an active combination of constituent meanings is routinely in place (Gagné & Spalding 2009; Ji et al. 2011; Marelli & Luzzatti 2012; Marelli et al. 2014). This would suggest that the difference between novel and familiar compounds is merely in their degree of lexicalization. While the former can still be interpreted by speakers in various ways, the latter have only one possible interpretation, that the classification by Bisetto & Scalise (2005) describes in terms of a fixed syntactic relation between the compound's constituents.

The second important difference is that interpretations of novel combinations pertain to the conceptual level, that is, they describe relations between the concepts being combined together. As such, the tripartition described above is essentially *semantic*. In contrast, the linguistic classification considered here is based on a purely *syntactic* criterion. Based on the commonalities highlighted above, however, it might be that the two levels of analysis are not mutually exclusive, but possibly related and somehow overlapping. Lexical semantic approaches corroborate this conjecture. Lieber (2009), for example, proposed that the different compound types identified by Bisetto & Scalise (2005) depend, at least in part, on

the intrinsic semantic features of the compound constituents. Moreover, classifications of compounds based on taxonomies of semantic relations reveal a certain degree of overlap between the syntactic and the semantic analysis (Levi 1978). For example, the semantic relation `AND` seems hardly distinguishable from the purely syntactic relation of coordination, which is subtended by the underlying conjunction 'and'.

1.3 Aim of the work

Based on this concurring evidence, we conjecture that various classes of compounds defined at the syntactic level may be also explained in terms of the semantic properties of the compounds and their constituents. In particular, our hypothesis is that measures quantifying the semantic role played by each constituent in contributing to the overall compound meaning, as well as the degree of semantic similarity between the constituents, should be effective in predicting different classes. Moreover, we expect these semantic measures to be able to capture different, syntax-based classes without relying on other non-semantic properties of compounds. Crucially, we do not claim that the distinction is thus purely semantic, making superfluous any categorization focusing on the syntactic relation between the compound constituents. Rather, we believe that the theoretically motivated and widely accepted *discrete* classifications proposed by linguists can be also described in terms of the *continuous*, quantitative aspects of the meaning of compounds and their constituents. In other words, we expect the *quantitative* semantic properties to parallel the *qualitative* grammatical distinctions, thus demonstrating, at the same time, the effectiveness of our proposal and the validity of the linguistic theory.

We experiment with a dataset of English compounds for which annotation based on the classification by Bisetto & Scalise (2005) (Subordinate, Attributive, Coordinate) is available. To predict each class, we use several semantic variables such as the degree of similarity between the constituents and the individual contribution of each constituent word in determining the meaning of the whole compound. We quantify these measures by using a compositional model of distributional semantics (Baroni & Zamparelli 2010; Guevara 2010; Mitchell & Lapata 2010; Zanzotto et al. 2010), following recent evidence proving the effectiveness of this approach in modeling morphological processes such as composition and derivation (Marelli & Baroni 2015; Günther & Marelli 2016; Marelli et al. 2017).

1.4 Computational models of meaning

Based on the core notion that similar words occur in similar contexts (Harris 1954; Firth 1957), distributional semantic models (henceforth, DSMs) represent lexical meanings by means of vectors encoding the contexts in which words appear in a large corpus. The intuition is that words that occur in similar linguistic contexts (e.g., *cat* and *dog*) should be semantically more similar than words that do not. Typically, this geometric representation is used to quantify the degree of distributional similarity between two words. Given the corresponding vectors, the similarity is computed in terms of their geometric distance, typically the cosine of the angle (Turney & Pantel 2010). In particular, the closer two vectors in the semantic space (i.e., the space populated by all the linguistic vectors), the higher their similarity. Traditional DSMs, such as the pioneering Latent Semantic Analysis (LSA; Landauer & Dumais 1997), have been largely used to obtain quantitative estimates of important semantic variables such as the degree of conceptual or topical similarity between two words (Padó & Lapata 2007; Gagné & Spalding 2009; Kuperman 2009; Wang et al. 2014).

1.5 Distributional semantics and compounds

In the domain of compounds, distributional semantic approaches have been extensively applied to two main tasks: noun-noun compound interpretation (Van de Cruys et al. 2013; Dima & Hinrichs 2015; Dima 2016; Shwartz & Dagan 2018; Fares et al. 2018) and compositionality prediction (Reddy et al. 2011; Schulte im Walde et al. 2013; Salehi et al. 2014; 2015; Cordeiro et al. 2016). The former task, usually tackled as a classification problem, aims at automatically predicting the *semantic* interpretation of the compound (i.e., the semantic relation between the constituents). Given the compound *street protest*, for example, a system is trained to predict that the relation holding between the nouns is 'locative'. Several datasets of compounds annotated with different numbers of semantic relations have been released for the tasks (Ó Séaghdha 2007; Tratz & Hovy 2010), and various systems capitalizing on distributional representations (usually obtained with neural network architectures; see Section 2.1) have been recently proposed. Overall, this approach has been proved to be successful in the task, though the performance is shown to be dependent on the number and granularity of semantic relations. As for the latter task, it is focused on predicting the degree of compositionality of a noun-noun compound, namely the extent to which the meaning of the whole depends on the meaning of the constituent words. Various datasets annotated with human judgments have been proposed through time (Reddy et

al. 2011; Roller et al. 2013; Farahmand et al. 2015), and extensive explorations of DSMs in the task have been carried out. Crucially for the purpose of this study, distributional measures of similarity obtained with compositional approaches were found to be highly predictive of human judgments in this task (Reddy et al. 2011; Schulte im Walde et al. 2013; Salehi et al. 2015; Cordeiro et al. 2016).

1.6 A compositional approach to compounds

Of great interest for the present work, Lynott & Ramscar (2001) were the first to employ distributional semantic models to study novel compounds (e.g. *zebra-horse*). In particular, the aim of that work was to test whether a measure of semantic similarity between compound constituents (quantified with LSA) was predictive of both (a) the ease of novel compound comprehension and (b) the distinction between Relation-linking and Property-mapping combinations. To do so, they experimented with novel compounds and their corresponding interpretations as provided by previous works on conceptual combination (Wisniewski & Love 1998; Gagné 2000). Overall, the model was shown to perform remarkably well in all the tasks. Lynott & Ramscar (2001), however, claimed that current distributional models like LSA were not capable of modeling the whole process of conceptual combination. Since they can only quantify the similarity between independent, free-standing words (e.g. *zebra* and *horse*), they are not informative at all about the relation between these words and the resulting compound. As such, they represent static, word-based models of lexical semantics which do not account for the potentially infinite linguistic productivity.

Compositional DSMs (hence, cDSMs) tackle precisely these issues. Aimed at accounting for the compositional nature of language (Baroni, Bernardi, et al. 2014), these models capitalize on DSM vectors and perform either simple (Mitchell & Lapata 2010) or more complex, theoretically inspired operations (Baroni & Zamparelli 2010; Guevara 2010; Zanzotto et al. 2010) to *compose* existing lexical entries. By exploiting simple operations (sum, multiplication) or being trained with distributional information about combinations that are already observed in the source corpus, these models can indeed be used to generate meaning representations for both novel and lexicalized formations. Recently, this approach was shown to be effective in modeling morphological processes such as derivation and compounding (Marelli & Baroni 2015; Günther & Marelli 2016; Marelli et al. 2017). Closely related to the present study, recent work (Günther & Marelli 2016; Marelli et al. 2017) exploited cDSMs to generate compositional representations of compounds. Marelli et al. (2017), in particular, explored whether a simple but effective regression-based compositional method (Guevara 2010) can capture the

variability in semantic relations between the constituents of novel compounds. This system was shown to be remarkably effective and flexible in capturing relational information. Based on this evidence, in the present work we employ the same model and test it in the task of predicting theoretically motivated, syntax-based classes of compounds.

2 Experiment

The present experiment investigates whether different, syntax-based classes of compound words (Subordinate, Attributive, and Coordinate) can be captured by means of semantic properties of the compound and its constituents. To quantify these properties, (a) we generate compositional representations of compounds and obtain similarity scores assessing the role of each constituent in contributing to the overall meaning; (b) we measure the degree of similarity between the first and second constituent.

A note on the terminology used in the paper. Until this point, we used the neutral terms "first constituent" and "second constituent" to refer to, respectively, *dog* and *house* in *doghouse*. As briefly mentioned in Section 1.1, one constituent usually plays a dominant role compared to the other since it acts as the "head" of underlying phrase. In this example, the head is clearly *house* (indeed, *doghouse* is 'the house of the dog'). Consistently, this element determines the syntactic category of the phrase and, semantically, it represents a hyperonym of the compound. By default, in English compounds the second constituent acts as the compound "head", whereas the first acts as the compound "modifier" (Bauer 2009). We stick with this arguably simplified terminology[2] and, from now on, we interchangeably use the terms "first constituent" or "modifier" to refer to the leftmost element, "second constituent" or "head" to refer to the rightmost one.

2.1 Semantic space

Following Baroni, Dinu, et al. (2014), who demonstrated that DSMs generated using feedforward neural network models largely outperform traditional count-based architectures in many tasks, we built a state-of-the-art `CBOW` semantic space using the `word2vec` toolkit by Mikolov et al. (2013), with all the parameters that turned out to be best-predictive in Baroni, Dinu, et al. (2014). In particular, the

[2] Without going into much detail, it should be mentioned that this picture is indeed less straightforward than it may appear. For instance, in the English compound *singer-songwriter* the two constituents play a similar role, in a way that they could be both considered as the compound "head" (and the compound as "double-headed") (Bauer 2009).

vectors have 400 dimensions and were built using (a) a context window of 5 words to either side of the target word, (b) a subsampling procedure which penalizes high-frequency words in the training phase ($t = 1 \times 10^{-5}$), (c) 10 negative samples. The vectors were trained using a corpus of written English containing around 2.8-billion tokens (a concatenation of BNC, ukWaC, and a 2009-dump of Wikipedia), the same used in Baroni, Dinu, et al. (2014). To avoid sparsity effects, we experimented with the vectors corresponding to the 300k most frequent words in the corpus.

2.2 Materials

We experimented with a sample of the MorBoComp database including 163 English compounds. MorboComp is a large, multilingual database of compounds that has been developed to study compounding from a typological perspective.[3] Each compound in the database is richly annotated (i.e., it is provided with information about headedness, compound and constituents' grammatical category, compound structure, etc.) and, crucially for our purposes, it is classified as Subordinate (hence, SUB), Attributive (hence, ATT) or Coordinate (hence, CRD) on the basis of the classification and terminology proposed by Bisetto & Scalise (2005). To illustrate, *schoolteacher* is tagged as SUB, *keyword* as ATT, and *king-emperor* as CRD.

Consistent with the criteria outlined in Bisetto & Scalise (2005), the 163-item sample contained cases of both "phrasal" compounds (*do-it-yourself illustration*, *around-the-world flight*) and "neoclassical" formations (*bibliography*, *theology*). In addition, a handful of items labeled with OTH (i.e., Other) were found. However, since this label was used by the annotators for either unresolved or idiosyncratic cases, we decided not to consider them in our investigation. Similarly, we removed neoclassical formations since their constituents can be affixes and suffixes rather than free-standing, independent words (e.g. *biblio-*). As a consequence, in our distributional semantics approach we could not have a vector representation for these items. Finally, additional 9 compounds were discarded since one of their constituents turned out not to be included in the 300k-vector semantic space. Specifically, 8 out of 9 of the missing items were first constituents of phrasal compounds, e.g. *all-goes-well* (*in all-goes-well atmosphere*) or *floor-of-a-birdcage* (*in floor-of-a-birdcage taste*), whereas in one case (*well-deserver*) the missing items was the second constituent (*deserver*). After this filtering process, our resulting dataset included 132 compounds (67 SUB, 49 ATT, 16 CRD), that we used for our experiment.

[3]For further details, see: http://morbocomp.sslmit.unibo.it/

2.3 Generating composed representations

For each of the 132 compounds in the list, we generated a composed representation using the vectors described in Section 2.1 and the compositional model by Guevara (2010). As previously mentioned, one of the main strengths of compositional DSMs is their ability to produce meaning representations also for combinations that are not attested in the source corpus. That is, given a novel or unattested compound, we are able to represent it as an independent vector on the basis of the meanings of its constituents (*zebra* and *horse*). This aspect was of crucial importance in our experiment, where 60 out of the 132 compounds extracted from MorBoComp turned out not to be present in the source semantic space. That is, almost half of the compounds were not among the 300k most frequent words in the corpus and, consequently, did not have a distributional representation. By using a compositional model capitalizing on the representations of the two constituents, however, we were able to overcome this limitation of traditional DSMs and generate a meaning representation for all the items, regardless of whether they had a "static" semantic representation or not.

The method used in the present study, in particular, was implemented by Guevara (2010) to model compositionality as depending on the semantic relation instantiated in the syntactic structure. As such, it looks particularly suitable for the case of compounds, which embed a modifier-head structure. Indeed, previous work proved this model to be very effective in generating composed representations for compounds (Marelli et al. 2017). Technically, the composed representations are obtained with the combinatorial procedure depicted in Figure 1: given two vectors $\vec{u}$ and $\vec{v}$ each representing one of the constituent words, their composed representation can be computed as $\vec{c} = \mathbf{M}\vec{u} + \mathbf{H}\vec{v}$, where $\mathbf{M}$ and $\mathbf{H}$ are weight matrices estimated from training examples. These matrices are trained using least squares regression,[4] having the vectors of the constituents as independent words (*dog*, *house*) as inputs and the vectors of example compounds (*doghouse*) as outputs. The two matrices are thus optimized so that the similarity between the weighted sum of the two constituent vectors (the composed vector) and the compound vector extracted from the semantic space (the observed vector) is maximized. Or, in other words, the composed vector obtained by means of the compositional model is built in a way that closely approximates the original one.

[4] As reported by Guevara (2010), this method is commonly employed to approximate functions in problems of multivariate multiple regression with a small number of observations and a greater number of variables, that is a similar condition to the one involving high-dimensionality vectors representing word meanings and (relatively) limited data.

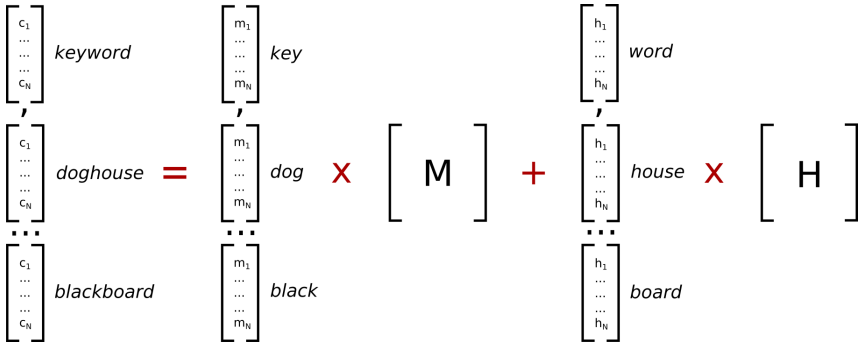

Figure 1: Representation of the training phase of the compositional method used in the study (adapted from Marelli et al. 2017).

In the present study, we trained the compositional model with a list of English noun-noun compounds extracted from the CELEX English Lexical Database (Baayen et al. 1995). By default, we treated all compounds as written in *solid* form, that is, without whitespaces or hyphens between the two constituents. When the solid compound was not found in our semantic space, we looked for it in its hyphenated form. The training set included 2174 triplets ⟨modifier, head, compound⟩, none of which was also present in the dataset we obtained from MorBoComp. We then used the estimated weight matrices for generating composed representations for each of the 132 compounds in our sample.

2.4 Semantic variables

For each vector obtained compositionally, we computed four composition-based semantic measures, namely (1) similarity between the composed representation of the compound and its modifier (e.g. between *keyword* and *key*), (2) similarity between the composed representation of the compound and its head (e.g. between *keyword* and *word*), (3) neighborhood density, that is, the average cosine similarity between the composed vector and its top-10 nearest neighbor vectors in the semantic space (all these 3 measures have been introduced by Vecchi et al. 2011), and (4) entropy, that is a measure of vector quality firstly introduced by Lazaridou et al. (2013).

By operationalizing the similarity between the composed compound vector and either constituents, in particular, we aimed at quantifying the extent to which each single word contributes to the overall meaning obtained compositionally. Although operationalized in terms of the cosine of the angle between the compound vector and either constituents (in the same way as standard DSMs

do), indeed, these measures genuinely describe the morphological process itself rather than merely taking into account its start and end points. Based on these properties, such measures have been recently used in studies with compound words. For example, they have been shown to be effective in predicting meaningfulness ratings on novel combinations (Günther & Marelli 2016) and in capturing relational information in compounds (Marelli et al. 2017).

As far as neighborhood density and entropy are concerned, both of them have been proposed to provide information about the meaningfulness of vectors encoding new concepts. The rationale of the former is that meaningful vectors should live in a region of the semantic space that is densely populated by vectors representing many related concepts, while meaningless vectors should be way more isolated. For the latter, the intuition is that meaningful vectors should have a skewed distribution, with few dimensions (corresponding to the salient semantic features of the word) being highly activated, i.e. having large values. In contrast, meaningless vectors should have a more uniform distribution, which would be a proxy for a less defined, fuzzier meaning. As a consequence, entropy would be inversely correlated with meaningfulness.

A (5) fifth semantic but non-compositional measure was introduced following Lynott & Ramscar (2001), who employed Latent Semantic Analysis (LSA) to quantify the degree of similarity between the first and the second constituent of a compound. Here, we took the compound constituent vectors (e.g. the vectors of *key* and *word*) from the source semantic space (see Section 2.1) and simply computed their cosine similarity. This measure might be helpful in distinguishing between different compound classes, based on the evidence that in both theoretical linguistics (see Lieber 2009) and conceptual combination literature (see Wisniewski 1996) this factor has been considered as explanatory of different classes/interpretations.

2.5 Non-semantic variables

In addition to the 5 semantic variables described above, we also included in our experiment a number of non-semantic control variables. For each compound and its constituent words we extracted word-form frequency from the source corpus (i.e., the number of times a word is encountered in the corpus in that exact form, regardless of its grammatical category). Compound frequency was calculated by summing the occurrences of the given compound in both solid and hyphenated orthographic form (*blackboard* and *black-board*, respectively). All frequency values, namely (6) compound frequency, (7) modifier frequency and (8) head frequency were subsequently log-transformed following standard practice in psycholinguistics (Brysbaert et al. 2018).

Table 1: Mean and standard deviation of all the predictors included in the experiment. MCsim: modifier-compound similarity. HCsim: head-compound similarity. MHsim: modifier-head similarity. Comp length: compound length. Comp freq: compound frequency. Mod freq: modifier frequency. Head freq: head frequency.

Predictor	SUB		ATT		CRD		Total	
	mean	sd	mean	sd	mean	sd	mean	sd
MCsim	0.21	0.10	0.18	0.09	0.24	0.13	0.20	0.10
HCsim	0.25	0.11	0.21	0.10	0.24	0.10	0.23	0.11
MHsim	0.13	0.10	0.14	0.09	0.34	0.15	0.16	0.12
Density	0.41	0.07	0.40	0.09	0.39	0.08	0.40	0.08
Entropy	4.50	0.05	4.99	0.06	5.01	0.05	4.50	0.06
Comp length	10.40	2.54	10.90	3.37	10.90	2.94	10.70	2.90
Comp freq	1.87	1.17	2.25	1.38	2.31	0.72	2.06	1.22
Mod freq	5.26	0.86	5.31	1.38	5.32	0.48	5.29	1.05
Head freq	4.97	0.85	4.92	0.89	5.12	0.65	4.97	0.84
PMI	3.53	4.00	4.83	4.80	4.34	3.47	4.11	4.27
Num items	67		49		16		132	

In addition, we computed (9) Pointwise Mutual Information (PMI) between the constituents as a measure of compound lexicalization. This widely-used association measure (Church & Hanks 1990) compares the probability of co-occurrence of two words in the source corpus with the probability of the two words co-occurring by chance. To illustrate, although the word pair ⟨the apple⟩ is likely much more frequent than ⟨apple juice⟩, the PMI of the latter will be higher, since the determiner *the* is likely to co-occur very frequently with any noun in the corpus, thus being less informative compared to the pair ⟨apple juice⟩, whose mutual association is intuitively strong. In particular, the higher the degree of lexical association between two words, the higher the PMI value.

Finally, we included (10) compound length measured as the number of characters making up the string (e.g., *blackboard* has length 10). When present, hyphens were not counted. Descriptive statistics including mean values and standard deviations for all the predictors used in the present experiment are reported in Table 1.

2.6 Data analysis

Our hypothesis is that various, syntax-based classes of compounds might be predicted on the basis of semantic features. If this is correct, our semantic variables will turn out to be reliable predictors of one class over the others. In order to test our hypothesis, we included all the predictors reported in Table 1 in a series of logit regression models that individually estimated the probability of one class over the other. That is, we tested three separate models in the task of predicting one compound type against each of the others: (1) ATT vs SUB, (2) ATT vs CRD, (3) CRD vs SUB.

All analyses were carried out within the R statistical computing environment. We adopted a backward procedure to progressively simplify each statistical model. Starting from a full-factorial model including all the independent variables, predictors were removed one by one when their absence did not significantly lower the overall model fit. At each step, the removal procedure was attempted for the predictor with the largest p-value. The contribution of each parameter to be removed was checked with a goodness-of-fit chi-square test. Finally, atypical outliers were identified and removed using as a criterion 2.5 standard deviation of the residual errors.

3 Results

For the better presentation of results, we summarize them in tables and discuss each model in a separate section. In the leftmost part, each table reports the list of variables included in the full-factorial version of each model. In the central part, model-simplification procedure (*Removal order*), chi-square goodness-of-fit test (*Chi-square*) and its results (p) are reported. The rightmost part shows the effects of the variables included in the final model.

3.1 ATT vs SUB

The first model, testing ATT (*halfprice*) against SUB (*bus-stop*) compounds, reliably distinguishes the two classes on semantic bases. As shown in Table 2, SUB is predicted against ATT by the higher semantic similarity between the compound and either the modifier ($p = 0.0182$) or the head ($p = 0.0355$). That is, the meaning of SUB compounds such as *bus-stop* is found to be more strongly determined by the individual meanings of its constituents compared to ATT compounds like *halfprice*, since both the modifier and the head contribute to the overall meaning to a greater extent than either constituents of ATT compounds do. Therefore, the

Table 2: Results of the logit model opposing ATT (1) to SUB (0).

Parameter	Chi-square	p	Removal order	Estimate	z-value	p
(Intercept)	–	–	Not removed	−0.8362	−0.741	0.4587
MCsim	–	–	Not removed	−6.4337	−2.361	0.0182
HCsim	–	–	Not removed	−5.5904	−2.102	0.0355
Density	–	–	Not removed	7.6332	1.916	0.0553
PMI	–	–	Not removed	0.0878	1.885	0.0594
MHsim	2.5275	0.1119	6			
Head freq	0.7637	0.3822	5			
Mod freq	0.4882	0.4847	4			
Comp length	0.8206	0.365	3			
Comp freq	0.0745	0.7848	2			
Entropy	0.005	0.9433	1			

higher the similarity between the compound and either constituent, the higher the probability to have a SUB rather than an ATT compound.

It should be noted that frequency measures, entropy, compound length and the similarity between the two constituents were progressively removed from the model. That is, their effects do not contribute to predict one class over the other. The remaining variables, namely PMI and neighborhood density, are instead included in the final model, even though their effect is only partially reliable ($p > 0.05$). Both these measures, anyway, indicate that higher values of both PMI and density are more likely to predict ATT rather than SUB compounds.

3.2 ATT vs CRD

The second model tests ATT (*halfprice*) against CRD (*comedy-drama*) compounds. As reported in Table 3, our model reliably distinguishes between the two classes on the basis of a single, highly significant semantic predictor, namely the semantic similarity between the compound constituents ($p = 0.0002$). In particular, the higher the similarity between the modifier and the head of a compound, the higher the probability of having a CRD, rather than an ATT compound. All other variables have been progressively removed from the final model since none of them significantly contribute to the overall goodness of fit.

Table 3: Results of the logit model opposing ATT (1) to CRD (0).

Parameter	Chi-square	*p*	Removal order	Estimate	z-value	*p*
(Intercept)	–	–	Not removed	5.667	4.094	0.0001
MHsim	–	–	Not removed	−18.182	−3.667	0.0002
Comp length	2.0539	0.1518	9			
Comp freq	1.2866	0.2567	8			
Head freq	0.77	0.3802	7			
HCsim	0.8033	0.3701	6			
MCsim	0.5967	0.4398	5			
Mod freq	0.7654	0.3816	4			
PMI	0.5182	0.4716	3			
Entropy	0.1344	0.7139	2			
Density	0.0346	0.8524	1			

Table 4: Results of the logit model opposing CRD (1) to SUB (0).

Parameter	Chi-square	*p*	Removal order	Estimate	z-value	*p*
(Intercept)	–	–	Not removed	−0.0843	−0.037	0.9707
MHsim	–	–	Not removed	36.3465	3.03	0.0024
HCsim	–	–	Not removed	−25.2847	−2.44	0.0146
Comp length	–	–	Not removed	−0.5323	−1.996	0.0459
PMI	0.3016	0.5828	7			
Comp freq	0.3286	0.5665	6			
MCsim	0.2825	0.595	5			
Density	0.1055	0.7453	4			
Entropy	0.2244	0.6357	3			
Mod freq	0.1236	0.7251	2			
Head freq	0.0084	0.9267	1			

3.3 CRD vs SUB

The third model opposes CRD (*comedy-drama*) to SUB (*bus-stop*) compounds. As in the other cases, the model reliably distinguishes between one class and the other on semantic bases. In particular, CRD is predicted over SUB by the degree of semantic similarity between the two constituents ($p = 0.0024$). The greater the similarity between the modifier and the head of a compound, the higher the probability of having a CRD rather than a SUB compound. Also, SUB is predicted over CRD by the degree of similarity between the compound and its head ($p = 0.0146$). That is, the head constituent contributes more to the overall meaning of SUB compounds (e.g., *stop* in *bus-stop*) than CRD compounds (e.g., *drama* in *comedy-drama*).

In addition to these semantic variables, compound length turns out to be also predictive of one class over the other. As reported in Table 4, in fact, its effect is reliable ($p = 0.0459$) and it indicates that longer compounds are more likely to be SUB than CRD. All other parameters were instead progressively removed.

3.4 Overall results

Taken together, these results indicate that the degree of semantic similarity between the compound's constituents (i.e. the modifier and the head) is a highly reliable predictor of CRD against both other classes. As shown in the barplot in Figure 2, the higher the similarity between the constituents, the more a compound is likely to be CRD rather than either ATT ($p = 0.0002$) or SUB ($p = 0.0024$). Moreover, the semantic similarity between the compound and its head is a predictive measure of SUB over both other types, as shown in Figure 3. That is, the more the head contributes to the meaning of the overall compound, the more the compound is likely to be SUB rather than either ATT ($p = 0.0355$) or CRD ($p = 0.0146$).

In order to evaluate the predictive power of each model, we further computed the accuracy with which the items under investigation were correctly assigned to the correct classes. First, we obtained the classes predicted by each logit model. Second, we computed the accuracy of each model by dividing the number of correctly predicted items by the total number of items included in the final model. As a comparison, for each model we also computed the accuracy of a majority baseline obtained by simply dividing the number of cases of the majority class by the total number of cases involved. As reported in Table 5, the best predictive model turned out to be the one opposing CRD vs SUB (0.90 accuracy) followed by ATT vs CRD (0.85) and ATT vs SUB (0.61). These numbers were in line with the

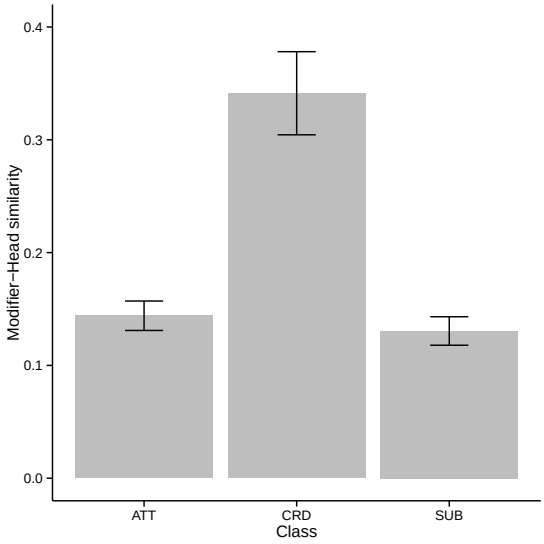

Figure 2: Similarity between modifier and head is predictive of CRD over both ATT and SUB.

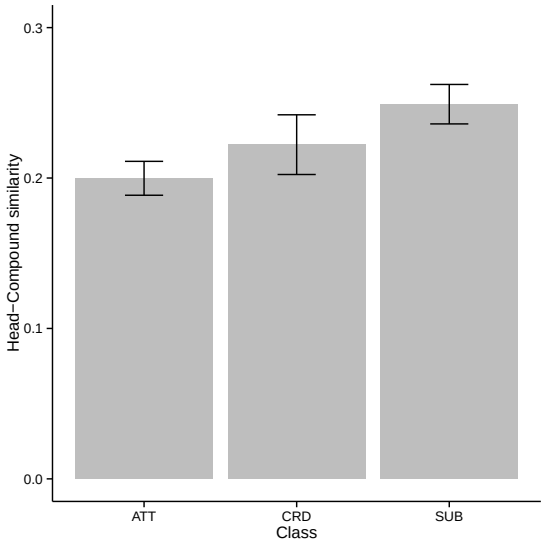

Figure 3: Similarity between the compound and its head is predictive of SUB over both ATT and CRD.

Table 5: From left to right: overall accuracy of each model as compared to the accuracy of the majority baseline, correctly predicted cases, missed cases. In brackets we report the correct class.

Model	Accuracy (baseline)	Predicted cases	Missed cases
ATT-SUB	0.61 (0.58)	*able-bodied (ATT)* *long-awaited (ATT)* *schoolteacher (SUB)* *racingclub (SUB)*	*commonroom (ATT)* *ironcurtain (ATT)* *underbody (SUB)* *apronstring (SUB)*
ATT-CRD	0.85 (0.75)	*best-equipped (ATT)* *keyword (ATT)* *father-daughter (CRD)* *comedy-drama (CRD)*	*bodypolitic (ATT)* *highschool (ATT)* *typewrite (CRD)* *subject-verb (CRD)*
CRD-SUB	0.90 (0.81)	*comedy-drama (CRD)* *blue-black (CRD)* *bus-stop (SUB)* *cutthroat (SUB)*	*schoolteacher (SUB)* *sleepwalk (SUB)* *typewrite (CRD)* *subject-verb (CRD)*

pattern of accuracy obtained by the majority baseline, which is sensible to the low number of CRD cases and therefore outputs higher scores for comparisons involving this class. Though our models always outperformed the baselines, the increase was noticeably lower in ATT vs SUB (+3%) compared to both CRD vs SUB (+9%) and ATT vs CRD (+10%). The limited number of items do not allow us to make any statistically reliable claim on the performance of the classifier. However, our focus is on testing whether the membership in a compound class is affected by a set of theoretically-relevant variables rather than proposing an effective classification algorithm. In this light, our results provided evidence for the effectiveness of these models. At the same time, they suggested that experimenting with more data would be desirable to further validate their power.

Besides accuracy, Table 5 reports some cases of correctly predicted and missed compounds for each of the models.

4 Discussion

The present study investigated whether various, syntax-based classes of compounds (Subordinate, Attributive, Coordinate) can be described in terms of the quantitative, continuous properties of the meaning of the compounds and their

constituents. To obtain these semantic measures, we generated cDSM representations for a list of compounds for which such classification was available. By running a series of logit models including both semantic and non-semantic factors as independent variables, we showed that our models are able to reliably capture different classes by means of semantic features.

4.1 On the modifier-head similarity

In particular, we showed that Coordinate compounds like *comedy-drama* are predicted over either Subordinate (*busstop*) or Attributive (*halfprice*) by the higher semantic similarity between the head and the modifier. This finding is consistent with previous evidence from both theoretical linguistics and psychology. Within the lexical semantics approach, Lieber (2009) indeed proposed that Coordinate compounds are generated when the two constituents share almost identical "bodies" and "skeletons", that is, when the words to be combined have highly similar meanings.

Also, our finding is in line with several theories of conceptual combination, according to which Hybrid or Conjunctive interpretations would be produced by people for novel combinations which involve highly similar concepts, e.g. *moose-elephant* (see among others Wisniewski 1996). Accordingly, and consistent with our results, Relation-linking interpretations (roughly equivalent to Subordinate compounds) would be instead produced for semantically highly dissimilar pairs, e.g. *apartment-dog*. Since in our model the similarity between the constituents also distinguishes between Coordinate (Hybrid/Conjunctive) and Attributive compounds (Property-mapping), we argue that this result is consistent with the graded description proposed in many conceptual combination theories, where the difference between Property-mapping and Hybrid combinations would be due to an increasing number of both "commonalities" and "alignable differences" between the concepts to be combined (Wisniewski 1996).

4.2 On the semantic role of compound constituents

Second, we showed that Subordinate compounds are predicted against Attributive on the basis of the higher similarity between the compound and either constituent. That is, in compositionally obtained Subordinate compounds both the modifier and the head contribute to a greater extent to the overall meaning than in Attributive ones. Moreover, the similarity between the compound and its head is a reliable predictor of Subordinate over both other classes.

First of all, these findings are again consistent with the lexical semantics literature (Scalise et al. 2005; Lieber 2009). In it, Subordinate compounds are typically

characterized by a structure in which the head *selects* its argument. Therefore, the head contributes more to the overall meaning in this kind of compounds compared to the other classes, where a formal relation between the elements is absent. Also, these results are consistent with the different mechanisms proposed in the conceptual combination literature for Relation-based interpretations (capitalizing on a "slot-filling" procedure) and Property-based ones (where an "alignment" process is routinely carried out) (Wisniewski & Gentner 1991; Wisniewski 1996). In a nutshell, the slot-filling procedure would imply a bigger role of the compound *head* compared to the other competing mechanism since, during combination, the head would be just *filled* in one of its "slots" by the modifier concept.

Interestingly, these findings are also consistent with evidence from embodied cognition (Louwerse 2008). In particular, the embodied conceptual combination theory (ECCo) by Lynott & Connell (2010) proposes that the great majority of relational interpretations (corresponding to Subordinate compounds) are "non-destructive", namely, they result from the combination of constituent concepts that are left intact during the meshing of their "affordances". To illustrate, in this approach the compound *picture book* (i.e. 'a book that has pictures') is non-destructive, since the *pictures* in question are still intact entities in the pages of the *book*. Simplifying somewhat, the combinatorial procedure that leads to Relation-based interpretations (Subordinate) does not modify heavily the meaning of the original constituents, whereas Property-mapping ones (Attributive) are almost always destructive, that is, they involve the "destruction" of (part of) the constituent concepts. Using an example from Lynott & Connell (2010), the compound *icicle fingers* would reduce *icicle* to a representation of 'coldness' and 'stiffness'. At the same time, the representation of the head (*fingers*) would be switched toward a more figurative, metaphorical meaning, less similar to its prototypical representation (see also, e.g., *iron curtain*). In this light, the similarity between either constituent as an independent word and the compound will be generally higher in Relation-based (Subordinate) compared to Property-based interpretations (Attributive), given that the combinatorial procedure of the former type does not heavily modify the meaning of the original constituents. Moreover, this observation provides indirect evidence that meaning representations extracted from texts via distributional semantics models can encode grounded information, at least to some extent (Louwerse 2011).

4.3 On attributive compounds

Third, compositionally-derived Attributive compounds are characterized by both a weaker contribution of the constituents in determining the overall meaning compared to Subordinate and a lower similarity between the constituents compared to Coordinate. This pattern of results is again consistent with Lieber (2009), who proposes that Attributive formations emerge when the semantic features of the constituents are too disparate to be interpreted in a Coordinative way and lack the argument structure that is typical of Subordinate compounds. Accordingly, Attributive compounds would represent a last-resort strategy used when the typical semantic features of the other classes are not satisfied (Lieber 2009). This description, according to which Attributive compounds would result when no discriminative features are present, is in line with evidence from conceptual combination showing that acceptability judgements for Property-based (Attributive) interpretations to novel compounds (e.g., a *whale boat* is 'a large boat') are slower compared to Relation-based (Subordinate) interpretations (e.g. a *whale boat* is 'a boat for hunting whales') (Gagné 2000). According to Gagné & Spalding (2015), indeed, this would suggest that Relation-based interpretations are the product of an initial compositional process that, in the absence of the features that lead to either a relational interpretation (Subordinate) or a coordinate interpretation (Coordinate), leads to Property-mapping interpretations.

4.4 On the methodology

On the methodological level, it should be mentioned that we used a compositional model to generate representations for a list of compounds whose constituents were nouns, verbs, adverbs, adjectives, etc. even though in the training phase only noun-noun compounds from CELEX were used. This could have represented a weakness for the system, causing the model to be biased toward noun-noun combinations. By looking at the results, however, we observed a similar, remarkably good performance of the model in all items, regardless of the grammatical category of the constituents. This is also clear by inspecting the examples in Table 5, where it can be noted that the parts-of-speech are almost uniformly distributed. However, it might be still possible that a richer training set might lead to even better results, perhaps achieving a better performance in generating meaning representations for less systematic, more opaque compounds. Indeed, we hypothesize that the lower accuracy obtained by the model opposing Attributive vs Subordinate compared to the others might be possibly due to this issue. Finally, we believe that the effectiveness of such an approach might

be further validated by testing it on a larger (and possibly balanced with respect to compound type) set of annotated compounds. This, on the one hand, would strengthen the predictive power on the prediction task. On the other hand, it would allow more extensive, fine-grained analyses on the successes and failures of the models. We plan to further investigate this issue in future work.

4.5 On the effectiveness of cDSMs in predicting compound relations

The effectiveness of our approach in the proposed task is in line with previous work showing that compositional models of distributional semantics are successful in capturing relational information between the constituents of a compound. In particular, our task is related to that of predicting compound semantic interpretation (see Section 1.5), where compositionally-obtained representations have been used to assign the correct semantic relation to noun-noun expressions. By experimenting with a number of cDSMs (including the one adopted in this study by Guevara 2010), for example, Dima (2016) obtained results comparable to state-of-the-art in 2 popular datasets (Ó Séaghdha 2007; Tratz & Hovy 2010). Compared to SoA methods, however, Dima (2016) only exploited information from word embeddings, thus proving the effectiveness of both distributed representations and compositional methods. In quantitative terms, our results are not directly comparable due to both the different experimental setting (we did not tackle the task as a classification problem) and the number of relations involved (3 vs either 6 or 43). Moreover, our results cannot be compared with previous work since, to our knowledge, we are the first in proposing this task. However, these studies jointly show that compositional representations are successful in predicting compound relations defined on either *semantic* or *syntactic* bases.

4.6 Final remarks

In conclusion, this study suggests that different compound types identified on syntactic bases can be also defined in terms of continuous, quantitative features of the meaning of the compound and its constituents. We believe that discrete and continuous approaches are two faces of the same coin, the former representing a theoretically motivated, cross-linguistically valuable framework aimed at describing complex linguistic phenomena, the latter providing an interesting way to quantitatively test them. As indicated by our results, compositional models of distributional semantics present a flexible and powerful way to capture many of these phenomena.

Acknowledgements

Sandro Pezzele did most of the work reported in the present article while employed at CIMeC, University of Trento. We are grateful to Marco Baroni and Laura Vanelli for their valuable feedback during the early stages of the project. We thank Sergio Scalise for providing the MorBoComp subset used in the experiment. We are also grateful to the participants of the First Quantitative Morphology Meeting (Belgrade, July 2015) for the helpful questions and discussion.

References

Baayen, R. Harald, Richard Piepenbrock & Léon Gulikers. 1995. *The CELEX lexical database (CD-ROM).* University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Bally, Charles. 1950. *Linguistique générale et linguistique française.* 3rd edition. Berne: A. Francke.

Baroni, Marco, Raffaela Bernardi & Roberto Zamparelli. 2014. Frege in space: A program of compositional distributional semantics. *Linguistic Issues in Language Technologies* 9(6). 5–110.

Baroni, Marco, Georgiana Dinu & Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 1, 238–247.

Baroni, Marco, Emiliano Guevara & Vito Pirrelli. 2009. Sulla tipologia dei composti N+N in italiano: Principi categoriali ed evidenza distribuzionale a confronto. In Ruben Benatti, Giacomo Ferrari & Monica Mosca (eds.), *Linguistica e modelli tecnologici di ricerca (Atti del 40esimo Congresso della Società di Linguistica Italiana)*, 21–23. Rome: Bulzoni.

Baroni, Marco & Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1183–1193. Boston.

Bauer, Laurie. 2001. Compounding. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, 695–707. Berlin & New York: De Gruyter.

Bauer, Laurie. 2009. Typology of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, chap. 17, 343–356. New York: Oxford University Press.

Bisetto, Antonietta & Sergio Scalise. 2005. The classification of compounds. *Lingue e linguaggio* 4(2). 319–332.

Bloomfield, L. 1933. *Language*. Chicago: University of Chicago Press.

Booij, Geert. 2005. *The grammar of words: An introduction to linguistic morphology* (Oxford linguistics). Oxford: Oxford University Press.

Brysbaert, Marc, Paweł Mandera & Emmanuel Keuleers. 2018. The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science* 27(1). 45–50.

Church, Kenneth W. & Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16(1). 22–29.

Cordeiro, Silvio, Carlos Ramisch, Marco Idiart & Aline Villavicencio. 2016. Predicting the compositionality of nominal compounds: Giving word embeddings a hard time. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, vol. 1, 1986–1997.

Costello, Fintan J. & Mark T. Keane. 2000. Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science* 24(2). 299–349.

Dima, Corina. 2016. On the compositionality and semantic interpretation of English noun compounds. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, 27–39.

Dima, Corina & Erhard Hinrichs. 2015. Automatic noun compound interpretation using deep neural networks and word embeddings. In *Proceedings of the 11th International Conference on Computational Semantics*, 173–183.

Dressler, Wolfgang U. 2006. Compound types. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 23–44. Oxford: Oxford University Press.

Fabb, Nigel. 1998. Compounding. In A. Spencer & A.M. Zwicky (eds.), *Handbook of morphology*, 66–83. Oxford: Blackwell.

Fanselow, Gisbert. 1981. *Zur Syntax und Semantik der Nominalkomposition*. Tübingen: Niemeyer.

Farahmand, Meghdad, Aaron Smith & Joakim Nivre. 2015. A multiword expression data set: Annotating non-compositionality and conventionalization for English noun compounds. In *Proceedings of the 11th Workshop on Multiword Expressions*, 29–33. Denver, Colorado, USA.

Fares, Murhaf, Stephan Oepen & Erik Velldal. 2018. Transfer and multi-task learning for noun–noun compound interpretation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 1488–1498.

Firth, John R. 1957. *Papers in Linguistics, 1934-1951*. Oxford: Oxford University Press.

Gagné, Christina L. 2000. Relation-based combinations versus property-based combinations: A test of the CARIN theory and the dual-process theory of conceptual combination. *Journal of Memory and Language* 42(3). 365–389.

Gagné, Christina L. & Thomas L. Spalding. 2006. Conceptual combination: Implications for the mental lexicon. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, chap. 7, 145–168. New York: Oxford University Press.

Gagné, Christina L. & Thomas L. Spalding. 2015. Semantics, concepts, and metacognition: Attributing properties and meanings to complex concepts. In Laurie Bauer, Lívia Körtévlyessy & Pavol Štekauer (eds.), *Semantics of complex words*, 9–25. New York: Springer.

Gagné, Christina L. & Thomas 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(1). 20–35.

Guevara, Emiliano. 2010. A regression model of adjective-noun compositionality in distributional semantics. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics*, 33–37. Uppsala, Sweden.

Günther, Fritz & Marco Marelli. 2016. Understanding karma police: The perceived plausibility of noun compounds as predicted by distributional models of semantic representation. *PLoS ONE* 11(10). e0163200.

Harris, Zellig. 1954. Distributional structure. *Word* 10(2-3). 146–162.

Haspelmath, Martin. 2002. *Understanding morphology*. London: Edward Arnold.

Ji, Hongbo, Christina L. Gagné & Thomas L. Spalding. 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.

Kuperman, Victor. 2009. Revisiting semantic transparency in English compound words. In *6th International Morphological Processing Conference*. Turku, Finland.

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

Lazaridou, Angeliki, Eva Maria Vecchi & Marco Baroni. 2013. Fish transporters and miracle homes: How compositional distributional semantics can help NP parsing. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*.

Lees, Robert B. 1960. *The grammar of English nominalizations*. Bloomington: Research Center in Anthropology.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Lieber, Rochelle. 2009. A lexical semantic approach to compounding. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, chap. 5, 78–104. New York: Oxford University Press.

Lieber, Rochelle & Pavol Štekauer (eds.). 2009. *The Oxford handbook of compounding*. New York: Oxford University Press.

Louwerse, Max M. 2008. Embodied relations are encoded in language. *Psychonomic Bulletin & Review* 15(4). 838–844.

Louwerse, Max M. 2011. Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science* 3(2). 273–302.

Lynott, Dermot & Louise Connell. 2010. Embodied conceptual combination. *Frontiers in Psychology* 1. 212.

Lynott, Dermot & Michael Ramscar. 2001. Can we model conceptual combination using distributional information? In *Proceedings of the 12th Irish Conference on Artificial Intelligence and Cognitive Science*, 1–10.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Munich: Beck.

Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122(3). 485–515.

Marelli, Marco, Georgiana Dinu, Roberto Zamparelli & Marco Baroni. 2014. Picking buttercups and eating butter cups: Spelling alternations, semantic relatedness, and their consequences for compound processing. *Applied Psycholinguistics* 36(6). 1421–1439.

Marelli, Marco, Christina L. Gagné & Thomas L. Spalding. 2017. Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition* 166. 207–224.

Marelli, Marco & Claudio Luzzatti. 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.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. http://arxiv.org/abs/1301.3781/.

Mitchell, Jeff & Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science* 34(8). 1388–1429.

Ó Séaghdha, Diarmuid. 2007. Annotating and learning compound noun semantics. In *Proceedings of the 45th Annual Meeting of the ACL: Student Research Workshop*, 73–78.

Padó, Sebastian & Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics* 33(2). 161–199.

Reddy, Siva, Diana McCarthy & Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, 210–218. Chiang Mai, Thailand.

Roller, Stephen, Sabine Schulte im Walde & Silke Scheible. 2013. The (un) expected effects of applying standard cleansing models to human ratings on compositionality. In *Proceedings of the 9th Workshop on Multiword Expressions*, 32–41.

Salehi, Bahar, Paul Cook & Timothy Baldwin. 2014. Using distributional similarity of multi-way translations to predict multiword expression compositionality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 472–481. Gothenburg, Sweden.

Salehi, Bahar, Paul Cook & Timothy Baldwin. 2015. A word embedding approach to predicting the compositionality of multiword expressions. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 977–983. Denver, CO.

Scalise, Sergio & Antonietta Bisetto. 2009. The classification of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, chap. 3, 34–53. New York: Oxford University Press.

Scalise, Sergio, Antonietta Bisetto & Emiliano Guevara. 2005. Selection in compounding and derivation: Morphology and its demarcations. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (eds.), *Morphology and its demarcations*, chap. 14, 133–150. Amsterdam: John Benjamins Publishing Company.

Schulte im Walde, Sabine, Stefan Müller & Stefan Roller. 2013. Exploring vector space models to predict the compositionality of German noun-noun compounds. In *2nd Joint Conference on Lexical and Computational Semantics*, vol. 1, 255–265. Atlanta, GA, USA.

Shwartz, Vered & Ido Dagan. 2018. Paraphrase to explicate: Revealing implicit noun-compound relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1200–1211. Melbourne, Australia.

Tratz, Stephen & Eduard Hovy. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 678–687.

Turney, Peter D. & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37(1). 141–188.

Van de Cruys, Tim, Stergos Afantenos & Philippe Muller. 2013. MELODI: A supervised distributional approach for free paraphrasing of noun compounds. In *2nd Joint Conference on Lexical and Computational Semantics, Volume 2: Proceedings of the 7th International Workshop on Semantic Evaluation*, vol. 2, 144–147.

Vecchi, Eva Maria, Marco Baroni & Roberto Zamparelli. 2011. (Linear) maps of the impossible: Capturing semantic anomalies in distributional space. In *Proceedings of the Workshop on Distributional Semantics and Compositionality*, 1–9.

Wang, Hsueh-Cheng, Li-Chuan Hsu, Yi-Min Tien & Marc Pomplun. 2014. Predicting raters' transparency judgments of English and Chinese morphological constituents using latent semantic analysis. *Behavior Research Methods* 46(1). 284–306.

Warren, Beatrice. 1978. Semantic patterns of noun-noun compounds. *Acta Universitatis Gothoburgensis. Gothenburg Studies in English Goteborg* 41. 1–266.

Wisniewski, Edward J. 1996. Construal and similarity in conceptual combination. *Journal of Memory and Language* 35(3). 434–453.

Wisniewski, Edward J. & Dedre Gentner. 1991. On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning. *Advances in Psychology* 77. 241–284.

Wisniewski, Edward J. & Bradley C. Love. 1998. Relations versus properties in conceptual combination. *Journal of Memory and Language* 38(2). 177–202.

Zanzotto, Fabio Massimo, Ioannis Korkontzelos, Francesca Fallucchi & Suresh Manandhar. 2010. Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd International Conference on Computational Linguistics*, 1263–1271.