Running Head: A Corpus Study of Semantic Patterns in Compounding

A Corpus Study of Semantic Patterns in Compounding

Phil Maguire

Department of Computer Science,

NUI Maynooth, Co. Kildare, Ireland

Edward J. Wisniewski

Department of Psychology,

University of North Carolina at Greensboro, USA

Gert Storms

Department of Psychology

University of Leuven, Belgium

Abstract

Studies of noun compounds have indicated that they tend to follow regular semantic patterns (e.g. Downing, 1977; Warren, 1978). The results of several psycholinguistic studies have supported the hypothesis that people rely on statistical knowledge about how nouns tend to be used in combination in order to facilitate the interpretation of novel compounds (e.g. Gagné & Shoben 1997, Storms & Wisniewski, 2005, Maguire, Maguire & Cater, 2010). The authors conducted a series of corpus analyses in order to establish the salience and reliability of semantic patterns in English compounds. These analyses demonstrated that similar concepts tend to appear in combination with similar sets of nouns. In addition, categorizing combinations according to the semantic category of the modifier and head revealed salient regularities in productivity reflecting the likelihood of plausible relationships. These findings support the idea that statistical knowledge about semantic patterns in compounding can be used to facilitate the interpretation of novel compounds. The implications for existing theories and models of conceptual combination are discussed.

1. Introduction

The combination of two nouns is a technique commonly adopted by speakers in order to communicate novel concepts and ideas. This strategy allows people to succinctly refer to concepts for which no suitable one word expressions exists (e.g. *jug accident, cinema food*). People display a natural propensity for generating and interpreting combinations. For example, children as young as two are able to understand novel compounds in isolation and by the age of six are able to produce them without grammatical errors (Clark, Gelman & Lane, 1985). However, accessing the meaning of a combination is not a trivial process, requiring an in-depth understanding of the constituent concepts, the context, and the addresser's communicative goals. Frequently, a modifier-noun compound reflects knowledge that is not typically referenced by the constituent concepts in isolation (e.g. *pet bird*; Hampton, 1987). Studying how people interpret combinations efficiently can yield valuable insights into how concepts are represented and how the meaning of words is affected by context.

In English, a language in which compounding is particularly productive, the simplest combinations consist of a modifier followed by a head noun. Typically, the head denotes the main category of the combined concept while the modifier is used to indicate a contrast or specialization of that category (e.g., a *plum sauce* is a type of sauce, but more specifically it is a type of sauce *made with* plums). Levi (1978) suggested that phrases of this type can be viewed in terms of a deletion, whereby a compound represents the short form of a more complex phrase. Here, the enabling condition for the deletion is the assumption on the part of the addresser that the addressee holds the prerequisite knowledge for the identification of the referent. Given this assumption, the relationship between the two constituents need not be mentioned explicitly, since it can be inferred. Regarding *plum sauce*, people know that sauces contain ingredients and also that plums are a fruit that can plausibly function as an ingredient for a sauce. They may even have experience of having cooked or tasted a plum sauce. As a result of this knowledge, the use of the two nouns *plum* and *sauce* in combination is sufficient for constraining the interpretation and thus conveying the intended referent.

1.1 Theories of combination interpretation

A variety of cognitive models of conceptual combination have been proposed (e.g. Costello & Keane, 2000; Estes & Glucksberg, 2000; Gagné & Shoben, 1997; Murphy, 1988; Wisniewski, 1997). These models have tended to converge on the view that, during the interpretation process, the basic head noun category is somehow refined or specialized by the modifier concept. The concept specialization model (Murphy, 1988) and dual-process theory (Wisniewski, 1997) are centered on a two-stage interpretation process. The first stage involves a slot-filling mechanism where the modifier is inserted into a slot in the head noun schema to form an interpretation (e.g. in *plastic chair*, the concept *plastic* is inserted into the <made of> slot of the concept *chair*). The second stage constitutes an elaborative mechanism whereby world knowledge is used to expand these interpretations (e.g. plastic chairs can be used as garden furniture). Wisniewski's dual-process theory suggests a further alignment and comparison mechanism which can account for property-based and hybrid interpretations (e.g. a *robin snake* as a snake with a red breast).

Although these schema-based theories make accurate predictions about the type of interpretations that are produced for combinations, it is not clear how people initially identify the correct slot to be modified. According to Murphy (2002), “people use their general background knowledge to choose the slot that seems best” (p. 453). However, the amount of background knowledge associated with any concept is considerable and most of it will be irrelevant to interpreting a particular combination. For example, in order to understand *plastic chair*, one does not need to know anything about the shape of chairs or how plastic is manufactured. This raises the question of whether people can selectively activate conceptual knowledge so that only the most relevant information is brought to mind.

Studies of noun compounding have indicated they tend to be regular, leading linguists to propose that most combinations can be satisfactorily ascribed to a limited set of forms (e.g. Downing, 1977; Levi, 1978). In light of these regularities, it has been suggested that people might exploit their knowledge of typical combination use to streamline the interpretation process and activate conceptual knowledge selectively. For example, Warren (1978) posited that people are able to restrict the range of interpretation of a novel combination by applying their knowledge of how concepts tend to be related, thus facilitating the process of identifying a semantic relationship between a modifier and a head. For example, in the case of *plastic chair*, simply knowing that *plastic* is a substance and that *chair* is an object is good grounds for assuming that a <made of> relationship holds between the two concepts.

The results of several psycholinguistic studies have supported the idea that people exploit statistical regularities when interpreting novel compounds. Gagné and Shoben (1997) identified a set of 16 possible relations that can be used to connect a modifier and a head (e.g. <made of>, <during>, <for>, <about>). In a series of experiments, they found that combinations involving modifiers which were more frequently associated with the appropriate relation were easier to interpret then those involving modifiers not typically associated with the appropriate relation. For example, a combination like *plastic equipment* was easier to interpret than *plastic crisis*, because the modifier *plastic* is more frequently associated with the <made of> relation than it is with the <about> relation. This effect was also replicated by Storms and Wisniewski (2005) using combinations in Indonesian, a language in which the order of the modifier and head is reversed relative to English. A study by Maguire et al. (2010) provided additional evidence in support of a statistical effect, elaborating on Gagné and Shoben's original theory. They found that the influence that a given modifier has on ease of interpretation depends on the semantic category of the head with which it is paired, suggesting an interactional statistical effect in which both constituents play a role.

A central feature of these statistic-based theories is the assumption that the relationship between a pair of concepts can be predicted without having to activate full representations of individual constituents. However, although previous studies have hinted at the presence of regular patterns in noun compounding, the evidence presented thus far has been based on subjective surveys of relatively small samples of text (e.g. Downing, 1977; Levi, 1978; Warren, 1978). While these studies have provided qualitative descriptions of regular forms of combination, they have not provided detailed statistics regarding the scope or consistency of such patterns. This kind of information is critical for substantiating the central assumption of statistic-based theories, namely that regularities in compounding can be exploited for facilitating the interpretation of novel compounds.

Psycholinguistic studies examining the validity of statistic-based models of interpretation have derived statistics from small, contrived samples of combinations which are unrepresentative of combination use in general. For example, Gagné and Shoben (1997) cross-paired two random sets of 91 nouns and based their statistics on the 3,239 plausible combinations that emerged from this process. Maguire, Devereux, Costello & Cater (2007) demonstrated that combinations generated in this way are extremely atypical: they found that 93.7% of Gagné and Shoben's set failed to appear in a sample of compounds taken from the British National Corpus (BNC). Storms and Wisniewski (2005) used an alternative technique, deriving statistics based on participant-generated combinations. However, this strategy is also unlikely to provide a representative set, as participants tend to repeatedly identify the most accessible compounds, while failing to reproduce the natural variety encountered in everyday communication. In light of this, Maguire et al. (2007) argued that the only way in which to obtain reliable statistics regarding combination use is to extract a large representative set from a corpus.

In sum, although anecdotal evidence has been provided in support of semantic patterns in compounding, it has yet to be established whether these patterns are consistent and reliable enough to facilitate interpretation. Addressing this question, we examined combinations appearing in the BNC and on the web in order to ascertain whether the assumptions of statistic-based theories of combination-interpretation are sound. The BNC is a tagged, annotated corpus containing over 100 million words. It is designed to represent a wide cross-section of modern English and therefore includes a comprehensive sample of both written and spoken language (Burnard, 1995). The web contains billions of text documents, making it another valuable resource for identifying patterns in language use (see Lapata & Keller, 2005). Two individual studies were carried out. In the first study we sought to answer the preliminary question: is there any association between conceptual content and combination use? In the second study we examined this association in greater detail, investigating the extent to which the pairings of semantic categories used in combinations are predictably distributed.

2. Study 1: Conceptual content and combination use

Although schema-based theories (e.g. Murphy, 1988, Wisniewki, 1997) do not provide a role for statistical knowledge in the interpretation process, they do suggest that a concept's features will influence how it tends to be used in combination. Specifically, the modifiers used to specialize a head noun will be those which can plausibly fill slots in its schema. Given that similar heads tend to have similar slots, then one would expect them to be specialized by the same kind of modifiers. For example, modifiers like *cheese*, *tomato* or *ham* can plausibly act as modifiers for both *sandwich* and *pizza*, given that both have an <ingredients> slot. In the same way, similar modifiers should tend to specialize similar head nouns. For example, the fact that *gold* and *silver* have similar properties means that they can fill the <made of> slot for similar sets of head nouns (e.g. *gold ring*, *silver ring*). If nouns from the same semantic category do indeed combine in similar ways, then this might result in the kind of regularities in combination use that are assumed by statistic-based theories. The following study evaluated this possibility by examining the extent to which concept similarity and similarity of combination use are associated. If every concept exhibits a unique pattern of combination use, then no useful statistical information could be extracted for the purpose of interpreting novel compounds. On the other hand, the observation that similar concepts are used in similar ways would support the fundamental premise of statistic-based theories.

2.1 Procedure

We examined how a sample set of common concepts are used in combination. Fifty nouns that occurred at least fifty times as both a modifier and a head within the BNC were chosen from Battig and Montague's (1969) database of category norms. These nouns were taken from the following artifact, natural kind and activity categories: *body part*, *dwelling*, *food*, *furniture*, *insect*, *kitchen utensil*, *mammal*, *natural earth formation*, *plant*, *profession*, *tool*, *vegetable*, *vehicle*, *weapon* and *weather* (see Appendix 1 for the full list).

A two-dimensional similarity matrix for the 50 nouns was derived using Seco, Veale and Hayes's (2004) WordNet similarity metric. WordNet is a semantic lexicon for the English language and has been used extensively to support automatic text analysis and artificial intelligence applications (see Miller, 1995). In this lexicon, English words are grouped into sets of synonyms called synsets, for which short, general definitions are provided. The semantic relations between these synonym sets is also recorded (e.g. hypernymy, hyponymy, meronymy etc.). Several measures have been proposed which rely on the structure of the WordNet hierarchy to provide ratings of semantic distance for pairs of nouns (e.g. Resnik, 1999; Jiang & Conrath, 1997; see Budanitsky & Hirst, 2006, for a review). While these measures require additional corpus frequency data to quantify the probability of occurrence of a given concept, Seco et al.'s (2004) metric has the added advantage of deriving all necessary information from the WordNet hierarchy. Seco et al. report a correlation value of .84 between human and machine similarity judgments, which is close to the theoretical upper bound of .88 proposed by Resnik (1999).

The central premise of Seco et al.'s metric is that similarity can be estimated by the amount of information two concepts have in common. This overlap can be determined by the most specific common generalisation that subsumes both concepts in the WordNet hierarchy. If one does not exist, then the two concepts are maximally dissimilar. For example *dog* is similar to *cat*, because both are animals and only a small proportion of nouns contained in the WordNet lexicon are animals. On the other hand, *dog* is very dissimilar to *ladder* since the most specific common abstraction for these nouns is [object], of which there are very many examples in WordNet. Accordingly, the similarity ratings derived for *dog* and *cat* (0.48) and *dog* and *rain* (0.08) reflect the negative log of the proportion of WordNet synsets subsumed by their most specific common generalisation (see Seco et al., 2004). The similarity values were used to populate a two-dimensional similarity matrix, denoting the pairwise similarity for each permutation of the 50 concepts under investigation.

In order to verify the accuracy of these automated similarity ratings, four human participants made the same judgments. Each participant rated the similarity of the 2,500 concept pairings and the four ratings were averaged. The correlation between Seco's

WordNet similarity metric and the participant generated ratings was .78. This correlation rose to .82 when correcting for the unreliability of the participant-generated ratings using correction for attenuation (see Lord & Novick, 1968). These results support the idea that WordNet-based similarities can be relied on to closely approximate human judgments.

Subsequently, a ‘combination profile’ was generated for each of the 50 concepts under investigation by identifying the 10 most frequent combination types involving that noun as a modifier and as a head in the BNC (e.g. *train journey*, *train service*, *train station* for *train* as a modifier). In cases of a tie in frequency, the remaining types were selected randomly. The profiles for *cat*, *dog* and *ladder* as modifiers are provided in Appendix 2.

Initially, a direct comparison was performed between the similarity of the 50 concepts and the similarity of the concepts in their combination profiles. As a measure of profile similarity, the average maximum similarity between the nouns in each profile was computed, again using Seco et al.’s (2004) WordNet similarity metric. For example, in comparing the profiles for *dog* and *cat*, we considered each of the nouns in the profile for *dog* and computed its maximum similarity with any of the nouns in the profile for *cat* (e.g. *faeces* and *dirt* obtained a similarity of .36). These ten values were then averaged to obtain an overall measure for profile similarity and the resulting values were used to populate a two-dimensional ‘profile similarity’ matrix.

The concept similarity and profile similarity matrices were then compared. For concept use as a modifier, the correlation between concept similarity and profile similarity was .32. For concept use as a head, the correlation between concept similarity and profile similarity was .20. These correlations were .33 and .19 using the participant-generated similarity ratings (all *p*s < .001).

A significant limitation of this initial analysis was that it was based on only the top ten most frequent combining nouns for a given concept. Because a concept can be plausibly combined with thousands of other nouns, there is no guarantee that the ten most frequent of these will provide a representative sample. Often, the most common combining types for a given concept are idiosyncratic and thus unsuitable for comparison with those of other concepts. For example, *tabby cat*, *pussy cat* and *tom cat* were among the most common modifiers for *cat* as a head. Because these combination types are lexicalised and hence specific to *cat*, they are unlikely to be used with any other head concepts (e.g. *tabby dog*). Another problem associated with using a limited sample of combination types is that a certain type may not feature, even though it is highly plausible. For example, although *dog basket* is not among the most frequent combination types for *dog*, it is far more acceptable than say, *ladder basket*. This fact is not reflected by a limited sample of combination types.

In light of this, we used web data to avoid over-generalising based on the limited sample of combinations available in the BNC. The web is being increasingly used as a data source for a wide range of natural language processing tasks (Lapata & Keller, 2005). Given that search engines index several billion pages of text, we were able to obtain frequencies for novel combinations not attested in the BNC. Novel combinations were created by taking the top ten combining types in a concept's profile and substituting the 49 other concepts in its place. For example, performing this substitution for *dog breeder* yielded combinations such as *cat breeder*, *ladder breeder*, *wind breeder* etc.

Subsequently, the Google search engine was used in order to obtain frequency counts for the 490 ‘synthetic’ compounds generated in this manner. A list of frequencies for the synthetic combinations produced using the combination profiles for *cat*, *dog* and *ladder* is given in Appendix 3.

We computed the log of the number of hits for each and normalised this value according to the following formula

$$norm(c_1, c_2) = \frac{logfreq(c_1, c_2) \times |C| \times |P|}{\sum_{x \in C} logfreq(x, c_2) \times \sum_{x \in P} logfreq(c_1, x)}$$

where C is the set of 50 concepts being compared, P is the set of nouns in those concepts’ profiles and c_1 and c_2 are the modifier and head of a synthetic compound. The purpose of the normalisation process was to control for the fact that some words are more common than others and therefore more likely to take part in a greater number of combinations (as well as producing a greater number of false positives). The normalized values for each of the ten synthetic combinations produced were then averaged, and the resulting values were used to populate a two-dimensional ‘substitutability’ matrix. For example, the relatively high value of 0.52 between *dog* and *cat* reflects the fact that substituting combinations involving the modifier *dog* with the modifier *cat* yields combinations with relatively high Google hit counts (e.g. *cat owner*, *cat food*, *cat breeder*). A sample of these substitutability values is given in Appendix 4.

The concept similarity matrices and substitutability matrices were then compared. Using the WordNet-derived similarities, the correlation between the two matrices was .49 for concept use as a modifier, and .40 for concept use as a head. Using the participant-generated similarities, these correlation coefficients rose to .58 and .49 respectively, or .61 and .51 when controlling for unreliability (all *p*s < .001). The full set of intercorrelations is given in Table 1. All correlations were significant at the .001 level.

Table 1. Intercorrelations of similarity and corpus measures

	WordNet	Human	Profile_{mod}	Profile_{head}	Sub_{mod}	Sub_{head}
WordNet	-	.78	.32	.20	.49	.40
Human		.88	.33	.19	.58	.49
Profile_{mod}			-	.12	.38	.27
Profile_{head}				-	.17	.27
Sub_{mod}					-	.53
Sub_{head}						-

The strength of these correlations is notable considering the small size of the profiles used and the noisiness of the web as a corpus. Using frequency data from a search engine like Google can be problematic (see Hundt, Nesselhauf & Biewer, 2007). For example, the Google search engine is insensitive to punctuation and capitalisation, leading to false positives whenever the paraphrase match crosses a sentence boundary. Matches are also likely to include links, web addresses, names and other non-textual data. Even when two nouns do co-occur, it cannot be assumed that they form a genuine combination. False positives can result from truncated multiple-noun compounds (e.g. "…a <u>dog rain</u> jacket will keep your pet warm and dry") and other non-combinational noun collocations (e.g. "…I will walk your dog rain or shine"). In addition, duplication of documents on the web can inflate frequency counts dramatically. Kilgariff (2007) notes that the frequencies themselves are unreliable: search engines can give substantially difference counts even for repeats of the same query. The reason for this is because queries are sent to different computers, at different points in the update cycle and with different data in their caches. In light of these limitations, the correlations between concept similarity and constituent substitutability provide clear evidence that semantic content strongly influences how concepts are used in combination.

2.2 Discussion

These results reveal a significant association between semantic content and combination use, thus validating the fundamental assumption of statistic-based theories. Specifically, the various correlations provide converging evidence that similar concepts combine in similar ways and that the more similar they are, the more likely they are to combine with the same nouns.

The observed regularities in combination use reflect constraints on the manner in which concepts tend to be plausibly related. As noted by schema-based theories of conceptual combination (e.g. Murphy, 1988, Wisniewski, 1997), the modifier must plausibly fill some slot in the head noun's schema. Thus, similar concepts, which are more likely to share the same features by virtue of their similarity, will tend to be combined with similar groups of nouns via the same set of plausible relationships. For instance, the reason that satisfactory combinations emerge when substituting *dog* with *cat* is because *dog* and *cat* have many features in common. The resulting combinations are therefore more likely to involve the same plausible relationships between modifier and head (e.g. *cat basket*, *dog basket*). In contrast, when *cat* is substituted with the concept *ladder*, the same relationships do not hold and the resulting combinations are less satisfactory (e.g. *ladder basket*).

If similar modifiers and heads tend to combine in similar ways, then this implies that pairings of modifiers and heads are unlikely to be randomly distributed. Instead, such pairings should tend to fall into a number of regular semantic patterns reflecting productive relationships, as originally suggested by Warren (1978). This possibility would offer strong support for statistic-based theories of combination interpretation, as it would provide a means for inferring relations based on statistical knowledge, without needing to activate detailed representations of the individual concepts. In the following study we examined the scope and consistency of these hypothesised regularities.

3. Study 2: Distribution of semantic pairings

In this study a low granularity semantic classification was imposed on combinations appearing in the BNC, with modifiers and heads separated into 25 different semantic categories. Based on the results of the previous study, we predicted that combinations would group together in clusters reflecting productive relationships, as opposed to occurring randomly across the different modifier-head pairing categories. For example, one would expect combinations of the form [substance – artifact] to be predominantly associated with the <made of> relation, since artifacts typically have a constitution that can be denoted by a substance concept. In this case, a relatively large proportion of combinations should fall into the [substance – artifact] category. On the other hand, one would expect combinations of the form [substance – emotion] to be relatively rarer, since this category does not reflect a productive relationship: emotions are intangible and are therefore unlikely to be associated with a constitution. If a basic classification scheme can reveal predictable patterns in modifier-head pairings, then this would provide strong support for the utility of statistical knowledge.

3.1 Procedure

The BNC was again used to obtain a representative sample of combinations. Although the corpus contains part-of-speech tagging, this information alone is not adequate for separating genuine combinational phrases from other noun collocations. Lapata and Lascarides (2003) estimated that up to 30% of all noun-noun co-occurrences extracted based on the BNC's part-of-speech tagging are not genuine combinations. These inaccuracies are mostly due to errors in assigning parts of speech (e.g. "the mountain rose up before them..."), and non-combinational noun collocations (e.g. "last year houses were snapped up..."). In order to better identify genuine compounds we used a version of the BNC parsed using Charniak's (2000) parser.

All compound noun phrases consisting of two nouns were extracted from the parsed output. Some acronyms, misspellings, common nouns and errors remained and additional filtering was required to eliminate these. We discarded all combinations containing proper nouns, plural modifiers and nouns made up of fewer than three letters. We also removed any nouns containing hyphens, numerical digits or any form of punctuation. Plural heads were converted to the singular form. In order to guarantee that all remaining combinations consisted of valid nouns, we compared our set with the lexicon of nouns included in WordNet and removed any combinations consisting of unidentified words. Following this procedure, much of the remaining error could be attributed to words that, because of their nature, triggered a disproportionate number of false positives. Many of these were nouns that could double as adjectives, verbs or adverbs (e.g. “it was a light snack”, “the children dread school”, “give me my umbrella back”). Accordingly, we discarded any combinations involving nouns with part-of-speech ambiguity, as well as a further 31 nouns attracting high levels of noise (e.g. “good value meal”, “fifteenth century houses”, “second hand car”, “low risk venture”). The entire filtering process reduced the total number of combination types from 320,430 to 252,127, a reduction of 21%. Although some legitimate combinations are likely to have been removed by applying these filtering measures, we had no reason to believe that their elimination was non-random relative to the hypothesis under investigation.

A novel automated approach was used for categorising the large set of combinations retrieved from the BNC. We made use of the fact that definitions for common nouns in WordNet are arranged in 25 separate lexicographer files, which happen to correspond to such general categories as *animal*, *plant* and *time period*. The main obstacle to applying this classification directly was that many nouns have multiple senses, and thus have entries in multiple lexicographer files. For example, if we consider the noun *dog*, the most intuitive sense is that of the animal. However, in addition to this, we find alternative definitions in WordNet, inter alia “a dull, unattractive woman”, “a smooth, textured sausage”, and “a metal support for logs in a fireplace”. Consequently, it cannot be assumed that the noun *dog* will always refer to the animal sense when used in combination.

In order to mitigate this problem, we constrained our sample to combinations whose constituents were diagnostic of one particular lexicographer file. For instance, some nouns such as *aardvark* have only a single sense while others such as *vest* have multiple senses which all come from the same lexicographer file (e.g. “a sleeveless garment worn underneath a coat” or “a collarless undergarment”). We included any noun whose dominant lexicographic category subsumed at least 90% of its occurrences. Sense frequencies were based on the Senseval frequencies provided in WordNet (see Kilgarriff, 1998). For example, the canine sense of *dog* was included in the analysis since the Senseval frequency for this sense is 42 while the combined frequency of all other senses is 0. Applying this diagnosticity constraint yielded a total of 12,960 diagnostic nouns, or 76.8% of all nouns appearing in combination in the BNC.

In order to ascertain the reliability of the resulting classifications, we conducted an analysis based on a random sample of 100 compounds. Three of the phrases were not genuine combinations. Of the remaining 97, all but four were correctly classified, with the observed errors resulting from either inaccurate WordNet information or the use of a subdominant sense. These results suggest that level of accuracy achieved in noun classification was adequate for exposing regular patterns in compounding.

3.2 Results

In total, 11,765 different nouns were used as modifiers in the BNC and 13,550 nouns were used as heads. These numbers suggest that trends in overall modifier and head use are broadly similar. Figure 1 compares rank and frequency for modifiers and heads using a logarithmic scale. As can be seen, the distribution of frequencies for modifiers and heads is closely matched at all ranks, highlighting the generativity of combination use. These statistics establish that heads are not restricted to being modified by a limited range of common modifiers (e.g. *plastic*, *mountain*). Instead, nouns are used just as productively in both the modifier and head roles.

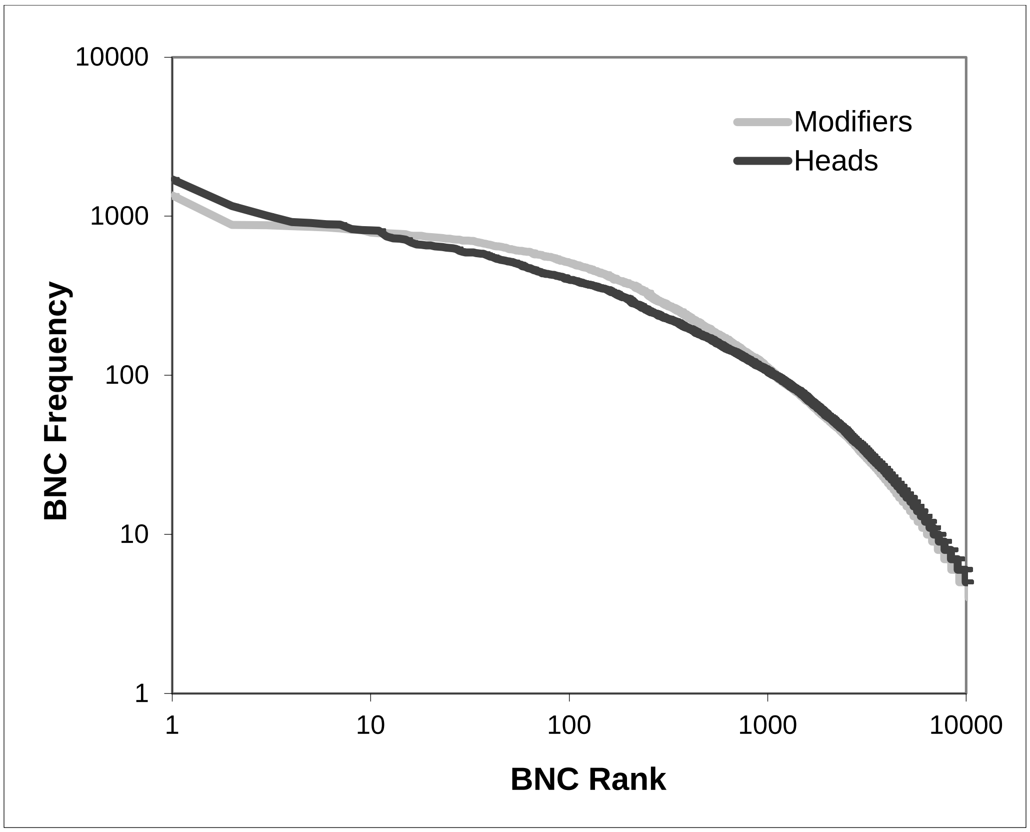

Figure 1. Rank versus frequency for modifiers and heads

We expressed the number of times each diagnostic noun appeared in combination as a percentage of that noun's occurrence in the BNC as a whole. The types of noun most likely to appear as part of a combination were substances, possessions and plants (49%, 41% and 39% of all occurrences respectively; e.g. *plastic chair*, *peasant estate*, *pine tree*). In contrast, the types of noun least frequently used in combination were attributes, shapes and feelings (10%, 9% and 5% respectively; e.g. *machine advantage*, *metal spiral*, *mob anger*). A summary of these data is presented in Figure 2.

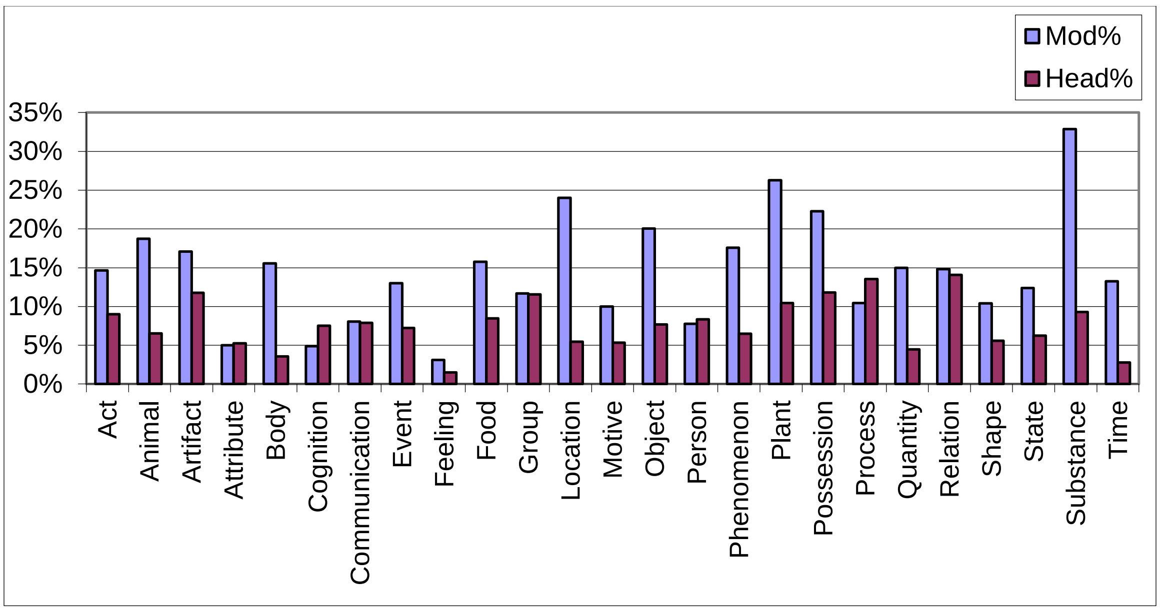

Figure 2. Total proportion of noun occurrences in combination by modifier and head percentage

We filtered the BNC combinations down to those consisting of two diagnostic nouns, yielding a total of 72,510 types (28.8% of the total). These were then separated into 625 different categories, corresponding to the different permutations of the 25 modifier and head types. A chi-square test of independence was performed to examine the relation between modifier type and head type. The relation between these variables was strongly significant, χ^2 (576, N = 258675) = 208953.5, p < .0001. Thus, pairings of modifier and head type were far from random, even using a low granularity classification of only 25 concept categories.

Of the 625 possible modifier-head pairings, the most productive (by number of combination types) were [artifact – artifact] (7.0%, *bicycle shed*), [person – person] (3.4%, *peasant soldier*), [artifact – act] (2.8%, *guitar tuning*), [artifact – person] (2.8%, *clarinet teacher*), and [substance – artifact] (2.3%, *steel pipe*). Substance modifiers exhibited the most skewed distribution, combining predominantly with artifacts (34%, *plastic robot*) and other substances (27%, *wax paste*). These head types are likely to be associated with a constitution, which substance modifiers can indicate. In contrast, heads not typically associated with a constitution had much lower proportions for substance modifiers (e.g. plant 1%, animal 1%, location 1%, event, feeling and time 0%).

The number of combinations in each modifier-head pairing category was strongly affected by the number of nouns subsumed by the constituent semantic categories. For example, artifacts and acts constituted 20% and 13% of the 12,960 diagnostic nouns in our sample, while animals and plants made up only 1% each. Controlling for this factor, we computed the ratio of the number of tokens observed versus the number that would have been expected taking into account the frequency of the constituent types. The following formula was used

$$ratio = \frac{N \times N_{mod-head}}{N_{mod} \times N_{head}}$$

where N is the total number of combination tokens included in the study, $N_{mod\text{-}head}$ is the number of tokens in the given modifier-head category, and N_{mod} and N_{head} are the total number of tokens with the same modifier and head categories respectively.

Of the 137 modifier-head pairing categories involving at least 300 tokens, the [plant – plant] category had the highest ratio, with these combinations occurring 54.5 times more often than expected (*elm tree*, *flower bud*, *bramble leaf*). The [substance – substance] category had the next highest ratio, with these combinations appearing 24.4 times more often than expected (*lithium metal*, *powder ice*, *wax paste*). The ten highest ratios are detailed in Table 2. Overall, combinations with the same category of modifier and head were 2.03 times more common than expected, highlighting that concepts are more likely to interact with others from within the same domain than would be expected by chance.

Table 2. Ten most productive combination type patterns in BNC

Type	**Ratio**	**Examples**
plant – plant	54.5	*elm tree*, *flower bud*, *bramble leaf*
substance – substance	24.4	*lithium metal*, *powder ice*, *wax paste*
location - group	13.2	*city police*, *dockyard authorities*, *site personnel*
food – food	10.8	*hamburger bun*, *kebab sauce*, *dessert beer*
animal – animal	7.9	*terrier dog*, *rat flea*, *hen bird*
time – phenomenon	7.3	*autumn sunlight*, *dawn wind*, *winter mist*
body – state	7.1	*eye trauma*, *kidney disease*, *muscle tension*
body – substance	6.5	*blood glucose*, *hair dye*, *liver protein*
time – time	5.7	*autumn afternoon*, *midnight hour*, *winter day*
time – food	5.4	*evening meal*, *morning coffee*, *winter feed*

In order to ascertain whether semantic content affects modifier use and head use differently, we correlated the modifier and head ratio statistics for the 25 semantic categories involving at least 300 tokens (e.g. comparing the ratio for [substance – artifact] with that for [artifact – substance]). The correlation was not significant, $\rho(80) = .128$, $p > .05$, indicating that the relationship between semantic content and combination use differs according to role. In other words, the probability that a noun from a given semantic category will be combined in a particular way differs according to whether it is being used as a modifier or a head.

3.3 Discussion

The principal finding of this study is that separating nouns into a small number of broad semantic categories is sufficient for revealing consistent patterns in modifier and head use. The results show that the spread of combinations does not reflect a random pairing of semantic categories: some modifier-head pairings occur a lot more frequently than expected while most permutations appear less frequently than would be expected based on a random distribution. This variation reflects differences in the potential of modifier-head pairings to capture productive relationships. For example, the [substance – artifact] category is more common than expected (by a factor of 5.2) because substances can fill the <made of> slot for a wide range of artifacts. On the other hand, the [substance – feeling] category is less common than expected (by a factor of 3.3) because feelings do not tend to have a dimension which can be filled by substance modifiers.

These results offer support for statistic-based theories of combination interpretation because they indicate that the semantic category of a combination's constituent concepts can be used to narrow down the range of possible relations. For instance, [substance – artifact] combinations are predominantly associated with the <made of> relation (68% from a random sample of 100 combinations), [time period – event] combinations with the <during> relation (89%) and [area – animal] with the <located> relation (91%). These associations support the idea that patterns in compounding can be exploited for the purpose of facilitating combination interpretation.

3.4 Implications for statistic-based theories

Gagné and Shoben's (1997) original statistic-based model of combination interpretation focuses on the word level, in that it proposes that people maintain statistical knowledge for individual modifier words. However, the patterns revealed in Study 2 reflect predictable pairings of semantic categories rather than words. This observation has two important implications for statistic-based theories. First, it indicates that patterns in combination use can be generalized to the level of semantic categories. Second, it indicates that modifier type and head type are strongly dependent, suggesting that they should not be modeled separately.

In order to investigate the potential differences in accuracy which emerge from using word-level versus semantic category based statistics, we considered the combination *student doctor*, one of the materials used in Gagné and Shoben's (1997) study. *Student doctor* is of type [person – person], insofar as both *student* and *doctor* are contained in WordNet's *person* lexicographer file. Relation frequency distributions were computed for combinations of type [person – person], [person – *] (any combination with a modifier of type person) and [* – person] (any combination with a head of type person).

In each case, 100 combinations were randomly sampled from the BNC. These combinations were then ascribed to one of the 16 relations identified by Gagné and Shoben (1997) and the overall proportion of combinations using each relation type was calculated. Figure 3 illustrates the relation frequency distributions for the three generalised modifier-head type pairings. Figure 4 illustrates the relation frequency distributions originally calculated by Gagné and Shoben (1997) based on the assumption that statistics are stored independently for individual modifier and head words.

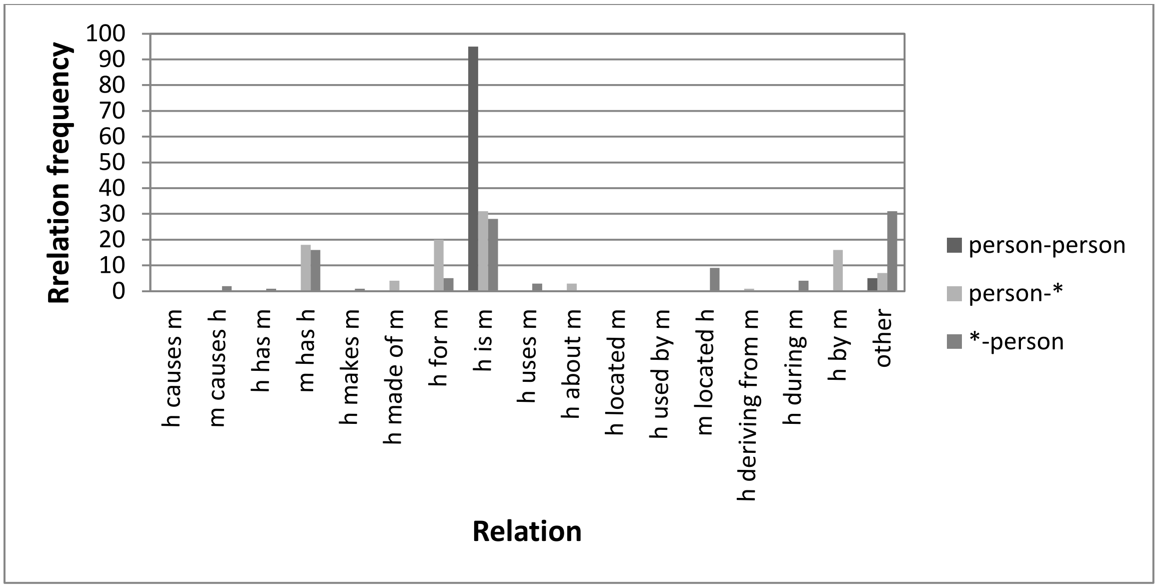

Figure 3. Relation frequencies for [person - person], [person - *] and [* - person]

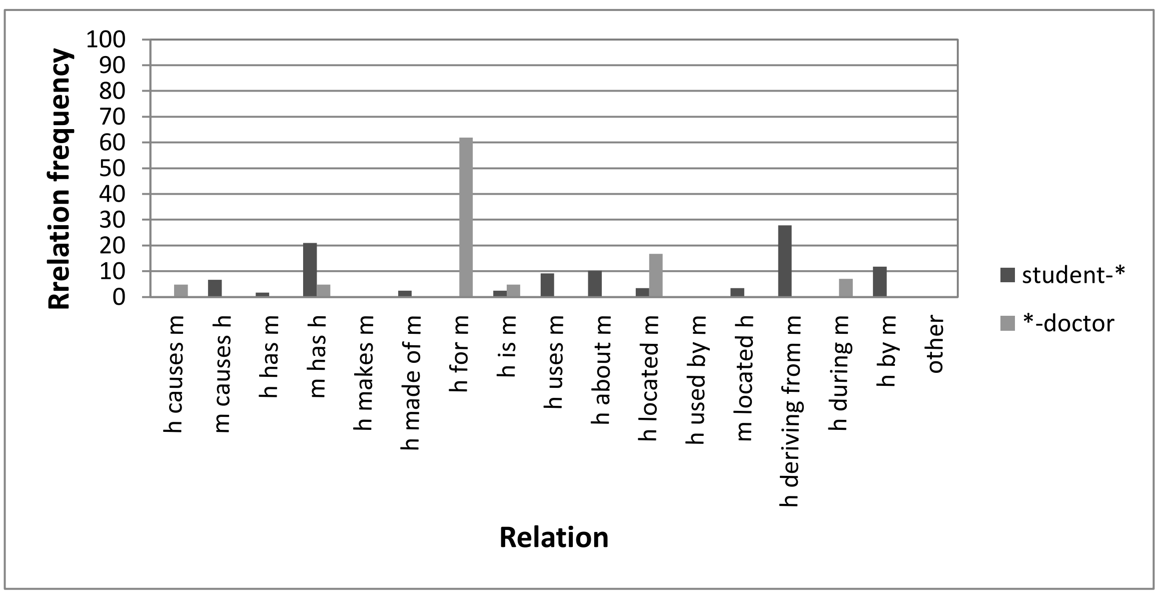

Figure 4. Relation frequencies for [*student* - *] and [* - *doctor*] from Gagné and Shoben's (1997) study

As can be seen from these diagrams, the most accurate prediction of the relation is obtained when the semantic categories of both constituents are taken into account. Specifically, 95% of combinations involving the modifier-head pairing of [person – person] involve the <is a> relation (the exceptions in this case being the nominalisations *slave mimic, consultant assessor, messenger director, pilot interviewer* and *defendant employer*). When the semantic category of only one constituent is taken into account (i.e. [person - *] and [* - person]), the relevance of the resulting relation frequency distribution is diminished. Gagné and Shoben's relation frequencies, which are based on a single noun as opposed to a generalized category, are even poorer predictors. Neither the statistics for [*student* - *] nor those for [* - *doctor*] provide support for the appropriate <is a> relation. Indeed, if one were to apply only the relation frequency of the individual modifier word, as Gagné and Shoben's theory proposes, then no relation could be assumed with greater than 28% confidence. These observations imply that statistics based on the interaction of these categories are more informative than statistics based on the relation preference of individual words.

4. General discussion

While schema-based theories (e.g. Murphy, 1988; Wisniewski, 1997) make accurate predictions about the types of interpretations that are produced for combinations, they assume that a full conceptual schema is activated each time a noun is encountered in the head role. However, much of this information is irrelevant for the purpose of interpreting a combination and its activation would therefore impair rather than aid comprehension (McElree, Murphy & Ochoa, 2006). Statistic-based models (e.g. Gagné & Shoben, 1997;

Maguire et al., 2009) propose that people can activate conceptual knowledge selectively by exploiting regular patterns that exist in compounding, thus avoiding the consideration of irrelevant information. For example, Gagné and Shoben (1997) suggested that "using the modifier's relational distribution to determine a suitable relation may be a means of constraining the amount of elaboration that is needed to obtain a more detailed interpretation of a phrase… people can identify that a *mountain bird* is a sensible phrase that uses the relation "*noun located modifier*" (a bird located in the mountains) before knowing in detail what a mountain bird is like" (p. 83-84).

Although qualitative descriptions have been provided of patterns in compounding, no rigorous large-scale analysis of compounding had previously been carried out. We have addressed this lacuna by examining a representative sample of combinations in the English language. Our corpus analyses have confirmed the fundamental assumption made by statistic-based theories, namely that the manner in which concepts are paired in combination is strongly constrained by the relationship that can plausibly link them, leading to productive patterns which are strongly associated with particular forms of interpretation. The corpus data indicate that knowing the basic semantic categories of the modifier and head can often be sufficient for identifying an appropriate relation, obviating the need to activate fine-grained features which may not be appropriate to the combination. For example, people will realize that *stone squirrel* matches the pattern [substance – object] and will be guided towards the <made of> relation. With this in mind, they can tailor their representations for *stone* and *squirrel* accordingly and avoid activating inappropriate features such as 'runs' or 'is brown' (though McElree et al., 2006, provide evidence that some inappropriate features are still activated during interpretation; see also Swinney et al., 2007).

In particular, the nature of the semantic patterns revealed in Study 2 suggest that statistical knowledge about how concepts tend to be used in combination can be generalized to the level of semantic categories. The results from several psycholinguistic studies have supported the idea that people are sensitive to patterns of this type. Maguire et al. (2010) found that participants were liable to misinterpret combinations whose modifier-head type category suggested an alternative relation. For example, participants were prone to interpreting *leather needle* as *needle* <made of> *leather*, suggesting that they were responding to the semantic pattern [substance – artifact] as opposed to activating the precise knowledge that needles are sharp while leather is soft. In another study, Maguire, Maguire and Cater (2007) investigated the time taken to reject implausible combinations using a speeded sensibility task. They found that implausible combinations belonging to a productive modifier-head type category took longer to reject than those belonging to an unproductive modifier-head type category. For example, *daffodil tail* was more quickly rejected than *frog tail*. Participants were more likely to view *frog tail* as well-formed, based on the productivity of the [animal – body part] category and its strong association with the <has> relation. Fine-grained features which ruled out the plausibility of certain combinations (e.g. knowing that amphibians do not have tails) only became available at a later stage of processing. These experimental observations support the view that people represent and exploit semantic patterns when interpreting novel compounds, and are in line with the current findings.

4.1 Implications for computational models of conceptual combination

The capacity to correctly interpret noun-noun compounds is often crucial to understanding a passage of text. Although some compounds have lexicalized definitions, many are effectively unique: of the 400,000 types in the BNC, Lapata and Lascarides (2003) found that almost 70% occurred only a single time. The key to understanding these novel combinations lies in identifying a relationship between the modifier and head. Various computational methods have been applied to infer this relation automatically (e.g. Costello & Keane, 2000; Cater & McLoughlin, 2000; Kim & Baldwin, 2005; Lapata & Keller, 2005; Lauer, 1995).

One approach is to use a detailed knowledge base of conceptual features. Costello and Keane's (2000) C3 model represents concepts as a complex predicate structure which includes attributes, roles and relations. Interpretations are constructed by combining sets of diagnostic predicates from both concepts and elaborating with co-occurring predicates from other similar concepts. Although the C3 model is capable of providing detailed interpretations, it suffers from a lack of precision: for a single compound, the model produces an average of 4,000 interpretations (Costello & Keane, 2000). In addition, because of the difficulty of providing exhaustive real world knowledge, the model often fails to output the most intuitive interpretation.

Another approach to automatically identifying the modifier-head relationship is to extract information from corpora. Lauer's (1995) model is based on the idea that semantic relations between heads and modifiers can be diagnosed based on the prepositions which are used to express them in a corpus. For example, the likelihood of the <for> relation for the combination *dog food* can be estimated based on the frequency of diagnostic paraphrases such as “food for a dog” or “food for dogs”. Due to a lack of sufficient corpus data, Lauer (1995) implemented a weaker version of the model, obtaining an overall accuracy of 40% given a baseline majority-class frequency of 33%. Lapata and Keller (2005) later implemented the stronger version of Lauer’s paraphrasing paradigm using the web as a corpus and obtained an accuracy of 56%.

Although paraphrasing is a promising technique for automatically assigning relations to combinations, it suffers from several limitations. For a start, many relations cannot be expressed using a simple connective. In addition, truly novel combinations unlikely to be attested more than a few times, even in a corpus as large as the web (Lapata and Lascarides, 2003). An alternative to relying on corpus statistics is to use a semantic hierarchy to recognize similarities between a target combination and those in an annotated training set. For example, Kim and Baldwin’s (2005) model works by calculating the WordNet similarity of a test case to all of the combinations in a training set of annotated compounds. The item in the training set which is most similar to the test instance is then selected and its relation chosen as the output. For example, in order to interpret *beef stew*, Kim and Baldwin’s model consults the training set, finds that *chicken soup* is most similar, and outputs its associated <made of> relation. Kim and Baldwin (2005) reported an accuracy of 53% for their model, with the baseline majority-class, in this case the <topic> relation, receiving a total of 43%. A disadvantage of similarity-based models is that they assume that similar combinations will always be interpreted using the same relation, an assumption which is not always valid (e.g. *metal tube* and *mercury tube*). The accuracy of such models also depends on having a sufficient number of annotated examples in the training set.

Our current findings suggest that relation frequency statistics based on semantic category pairings might provide a reliable means of predicting relations. A computational model developed by Cater and McLoughlin (2000) uses this approach. The model works by partitioning combinations into different relation zones according to the location of modifier and head concepts in the WordNet hierarchy. Boundaries of relation zones are delineated in the hierarchical space using a set of 929 annotated combinations extracted from the Suzanne corpus. These boundaries are defined as the lowest, or, most specific link covering a group of combinations involving the same relation. For example, the combinations *cat tail*, *dog tongue* and *horse leg* might be grouped under the super-ordinate link [mammal – body part] and labeled with the <has> relation. If other training examples violate this rule, other more specific rules will be added which will override this generalization. The interpretative process of Cater and McLoughlin's model works by assigning the relation of the nearest super-ordinate link to a test case.

Cater and McLoughlin found that, although their model discovered many reliable interpretative links, the overall accuracy was only 55%, with a baseline majority-class frequency of 23%. One problematic issue they identified is that many intuitive forms of semantic category are not represented in the WordNet hierarchy. For example, if the head noun *bag* is modified by a concept that can be contained in a bag, then the resulting combination can be interpreted using the <for> relation (e.g. *mail bag*, *coin bag*, *sweet bag*). However, WordNet does not contain a category for 'a collection of small things that can be stored in a bag'. Similarly, although *seat*, *mirror*, *chain* and *brake* can all describe part of a motorbike, these concepts are not grouped together under a single category but are instead scattered throughout the WordNet tree. Cater and McLoughlin concluded that in many cases, the type of information needed to interpret combinations is not reflected by the arrangement of semantic categories in a lexical hierarchy. Although a concept's position in a hierarchy reveals important information, it often fails to reflect key features which can distinguish that concept from others within a larger domain (e.g. size, shape, association etc.)

The goal of the present study has been to investigate whether semantic patterns are evident in compounding and whether knowledge of such patterns could be used to facilitate the interpretation process. Although these questions have been answered in the affirmative, the analyses we have described have not exhausted the range of possible patterns that could potentially be observed, nor the precision with which they could be represented. As reinforced by the relatively disappointing performance of Cater and McLoughlin's (2000) WordNet-based model, further research is required to identify the statistical patterns which best facilitate interpretation.

5. Conclusion

The fact that people can quickly and reliably interpret novel combinations suggests that they are able to filter out inappropriate information and quickly home in on a promising interpretation. While schema-based theories of combination interpretation can make accurate predictions regarding the type of interpretations that are produced, an additional component is required to explain how this process is carried out efficiently. Although schema-based theories acknowledge constraints on how modifier and head concepts can be related, they do not acknowledge the possibility that such constraints might lead to regular semantic patterns in compounding. Statistic-based theories assume that such patterns exist, and that these regularities are exploited for the purpose of streamlining interpretation.

Through a series of corpus analyses we have provided converging evidence of broad semantic patterns in compounding, thus supporting the premise of statistic-based theories. Specifically, we have shown that the semantic content of a concept strongly influences how it is used in combination. As a result, generalized information regarding the semantic categories of a modifier and head can often be useful in diagnosing the relationship between them. However, further research is required to clarify the precise nature of the statistical knowledge that people maintain and the manner in which it is applied.

Appendices

Appendix 1: 50 nouns used in Study 1

Category	Nouns
Body part	*eye*
Dwelling	*apartment, house, tent*
Food	*cheese, bread, pie, sandwich*
Furniture	*bed, chair, desk, table*
Insect	*ant, bee, butterfly*
Kitchen utensil	*knife, pot*
Mammal	*cat, cow, dog, horse, lion*
Natural earth formation	*hill, mountain, river, rock, valley*
Plant	*bush, flower, grass, tree*
Profession	*doctor, lawyer, teacher*
Tool	*drill, hammer, ladder*
Weapon	*bomb, gun, rifle, sword*
Vegetable	*potato, rice, salad*
Vehicle	*bike, train, truck*
Weather	*rain, snow, wind*

Appendix 2: Combination profiles for *cat*, *dog* and *ladder* as modifiers

Mod	Head
cat	food
cat	owner
cat	family
cat	litter
cat	book
cat	basket
cat	breeding
cat	faeces
cat	show
cat	woman

Mod	Head
dog	owner
dog	food
dog	breeder
dog	warden
dog	collar
dog	dirt
dog	hotel
dog	show
dog	track
dog	walk

Mod	Head
ladder	stile
ladder	bridge
ladder	climb
ladder	climber
ladder	firm
ladder	pitch
ladder	rail
ladder	safety
ladder	stairs
ladder	work

Appendix 3: Substitution Google frequencies for *cat*, *dog* and *ladder*

Mod	Head	Log Freq
cat	food	6.4
cat	owner	5.5
cat	family	5.4
cat	litter	6.1
cat	book	5.4
cat	basket	4.7
cat	breeding	4.9
cat	faeces	4.3
cat	show	5.6
cat	woman	5.6
cat	owner	5.5
cat	food	6.4
cat	breeder	5.3
cat	warden	2.2
cat	collar	5.4
cat	dirt	4.2
cat	hotel	4.5
cat	show	5.6
cat	track	4.8
cat	walk	5.5
cat	stile	0.3
cat	bridge	2.7
cat	climb	2.8
cat	climber	2.9
cat	firm	2.1
cat	pitch	2.0
cat	rail	2.4
cat	safety	4.8
cat	stairs	4.0
cat	work	4.6

Mod	Head	Log Freq
dog	food	6.7
dog	owner	6.1
dog	family	6.2
dog	litter	5.0
dog	book	5.5
dog	basket	4.5
dog	breeding	5.8
dog	faeces	4.8
dog	show	6.3
dog	woman	4.9
dog	owner	6.1
dog	food	6.7
dog	breeder	6.0
dog	warden	5.7
dog	collar	6.3
dog	dirt	4.6
dog	hotel	4.7
dog	show	6.3
dog	track	5.6
dog	walk	5.4
dog	stile	1.7
dog	bridge	2.7
dog	climb	2.6
dog	climber	1.3
dog	firm	2.3
dog	pitch	2.1
dog	rail	2.4
dog	safety	5.1
dog	stairs	4.4
dog	work	4.9

Mod	Head	Log Freq
ladder	food	2.0
ladder	owner	1.9
ladder	family	2.5
ladder	litter	1.2
ladder	book	3.0
ladder	basket	2.1
ladder	breeding	0.3
ladder	faeces	0.0
ladder	show	4.3
ladder	woman	2.0
ladder	owner	1.9
ladder	food	2.0
ladder	breeder	0.3
ladder	warden	0.0
ladder	collar	0.6
ladder	dirt	1.4
ladder	hotel	1.4
ladder	show	4.3
ladder	track	3.0
ladder	walk	4.0
ladder	stile	2.9
ladder	bridge	2.9
ladder	climb	4.8
ladder	climber	4.3
ladder	firm	1.8
ladder	pitch	2.7
ladder	rail	4.1
ladder	safety	5.3
ladder	stairs	2.9
ladder	work	4.6

Appendix 4: Substitutability scores for *cat*, *dog* and *ladder* as modifiers

	cat	dog	ladder
cat	.68	.69	.26
dog	.52	.62	.29
ladder	.35	.37	.88

Author Note

We would like to thank Arthur Cater for his assistance with data filtering, Mirella Lapata for providing a parsed version of the BNC and Rebecca Maguire for her helpful comments and advice. Correspondence should be addressed to Phil Maguire at Department of Computer Science, NUI Maynooth, Co. Kildare, Ireland. Email: pmaguire@cs.nuim.ie.

References

Battig, W.G., & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, *80*, 1-127.

Budanitsky, A. & Hirst, G. (2007). Evaluating WordNet-based measures of semantic distance. *Computational Linguistics*, *32*(1), 13-47.

Burnard, L. (1995). *Users Reference Guide for the British National Corpus*. Oxford: Oxford University Press.

Cater, A. & McLoughlin, D. (2000). Cluster strategy for choice of compound noun interpretation rules. In Griffith, J. and O' Riordan, C. (Eds), *Proceedings of the 11th Conference on Artificial Intelligence and Cognitive Science*, 25-34.

Charniak, E. (2000). A maximum-entropy-inspired parser. *Proceedings ANLP-NAACL' 2000*, Seattle, Washington.

Clark, E.V., Gelman, S.A., & Lane, N.M. (1985). Compound nouns and category structure in young children. *Child Development*, *56*, 84–94.

Costello, F., & Keane, M.T. (2000). Efficient creativity: Constraints on conceptual combination. *Cognitive Science, 24,* 299-349.

Downing, P. (1977). On the creation and use of English. compound nouns. *Language, 53,* 810-842.

Estes, Z., & Glucksberg, S. (2000). Interactive property attribution in concept combination. *Memory & Cognition*, *28*, 28-34.

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

Hampton, J.A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory and Cognition*, *15*, 55-71.

Hundt, M., Nesselhauf, N. & Biewer, C. (2007). Corpus Linguistics and the Web. Rodopi, Amsterdam.

Kilgarriff, A. (1998). Gold standard datasets for evaluating Word Sense Disambiguation programs. *Computer Speech and Language*, *12(3)*, 453-472.

Kilgarriff, A. (2007). Googleology is bad science. *Computational Linguistics*, *33*(1), 147-151.

Kim, N.S. & Baldwin, T. (2005). Automatic interpretation of noun compounds using WordNet similarity. *Second International Joint Conference on Natural Language Processing*, 945-956.

Lapata, M. & Keller, F. (2005). Web-based models for natural language processing. *ACM Transactions on Speech and Language Processing*, *2:1*, *1-31*.

Lapata, M. & Lascarides, A. (2003). Detecting novel compounds: The role of distributional evidence. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, 235-242.

Lauer, M. (1995). Corpus statistics meet the compound noun: Some empirical results. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA.

Levi, J. (1978). *The Syntax and Semantics of Complex Nominals.* Academic Press, New York.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison Wesley, Reading, MA.

Maguire, P., Devereux, B., Cater, A.W. & Costello, F. (2007). A Re-Analysis of the CARIN model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, *4*, 811–821.

Maguire, P., Maguire, R., & Cater, A. (2007). In Search of the Frog's Tail: Investigating the Time Course of Conceptual Knowledge Activation. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Maguire, P., Maguire, R. & Cater, A. W. (2010). Interactional factors influencing the interpretation of noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. In press.

Miller, G.A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38(11),* 39- 41.

McElree, B., Murphy, G. L., & Ochoa, T. (2006). Time-course of retrieving conceptual information: A speed-accuracy tradeoff study. *Psychonomic Bulletin & Review, 13,* 848-853.

Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science*, 12, 529-562.

Murphy, G. L. (2002). *The Big Book of Concepts.* Cambridge, MA: MIT Press.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 448–453.

Seco, N., Veale, T., & Hayes, J. (2004). An intrinsic information content metric for semantic similarity in WordNet. *Proceedings of the 16th European Conference on Artificial Intelligence*.

Storms, G. and Wisniewski, E. J. (2005). Does the order of head noun and modifier explain response times in conceptual combination? *Memory and Cognition*, *33*(5), 852-861.

Swinney, D., Love, T., Walenski, M., Smith, E. E. (2007). Conceptual combination during sentence comprehension: Evidence for compositional processes. *Psychological Science*, *18*, 397-400.

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

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, 4, 167-183.