A Corpus Analysis of Conceptual Combination

Phil Maguire (pmaguire@cs.nuim.ie)
Department of Computer Science, NUI Maynooth
Co.Kildare, Ireland

Edward J. Wisniewski (edw@uncg.edu)
Department of Psychology,
University of North Carolina at Greensboro, USA

Gert Storms (gert.storms@psy.kuleuven.ac.be)
Department of Psychology,
University of Leuven, Belgium

Abstract

Although various theories of conceptual combination have been proposed in the past, these have addressed interpretation issues rather than the circumstances in which combinations are used. As a result, existing theories make no explicit predictions about the types of nouns that will combine most frequently. We address this issue by conducting two separate studies designed to reveal patterns in compounding. In the first, we categorize combinations in the BNC according to 25 different noun types. In the second, we investigate whether similar concepts tend to combine with similar constituents. The results of both studies reveal that conceptual content has a large influence in determining how a noun is used in combination. We discuss the significance of our findings for theories of conceptual combination.

Keywords: Conceptual combination; noun-noun compounds; BNC; WordNet similarity; web as a corpus.

Introduction

The combination of two words is a technique commonly adopted by speakers in order to refer to novel concepts and ideas (e.g. *holiday tension*, *picnic bee*). Although people have a well developed means of understanding these novel compounds, the associated comprehension process is not trivial, requiring many levels of understanding. Accordingly, the study of conceptual combination is important, both because it is intimately associated with the generativity and comprehension of natural language and because it is important for understanding how people represent concepts. In English, a language in which compounding is particularly productive, combinations consist of a modifier followed by a head noun. Usually, the head noun denotes the main category while the modifier implies a relevant subcategory or a modification of that set's typical members. In this way, a *kitchen chair* is interpreted as a particular type of chair, and more precisely as the type that is located in kitchens.

Traditionally, theories of conceptual combination have centered on explaining how combinations are interpreted. In contrast, little focus has been directed towards understanding the circumstances in which combinations are used. Slot-filling theories (e.g. Wisniewski, 1997) propose that a combination is interpreted by filling the modifier concept into an appropriate slot of the head schema (e.g. the combination *kitchen chair* is interpreted by filling the concept *kitchen* into the <located> slot of *chair*). According to this view, combination use for a head should therefore be influenced by the set of available slots associated with that concept. Similarly, combination use for a modifier should be influenced by its capacity to act as a filler. The slot-filling view implies that similar nouns will combine in similar ways since they will have many slots in common. For example, *stew* will combine in a similar way to *soup* since both are dishes and are hence associated with a slot relating to ingredients. Also, *plastic* will combine in a similar way to *metal* since both are substances and thus have the ability to fill the <made of> slot for a wide variety of object concepts.

Gagné and Shoben's (1997) Competition Among Relations Among Nominals (CARIN) theory suggests the opposite. This theory proposes that there is a fixed, relatively small taxonomy of standard relations that can be used to link the concepts in a combination. According to Gagné and Shoben, the most available standard relation is the one most frequently used to interpret other compounds containing that same modifier. For instance, the modifier *mountain* is most often associated with the <located> relation, thus making the combination *mountain stream* easier to interpret than *mountain magazine*, which uses the <about> relation. Gagné and Shoben's theory proposes that people store statistical knowledge about now often each relation has been used with a modifier concept in the past. The assumption that separate distributions must be stored with every noun implies that trends in combination use cannot be inferred from a noun's conceptual content. If similar nouns combined in similar ways (e.g. *stew* and *soup*) then storing individual relation frequency distributions for them would be unnecessary.

In the following sections we present a series of corpus-based studies of the English language which examine if combination use is affected by conceptual content. The first classifies the combinations in the British National Corpus

(BNC) into different categories while the second considers a sample of 50 common concepts and examines whether their similarity is correlated with that of their most frequent combining nouns. The findings of these experiments are used to differentiate between the slot-filling and relation-based views.

Study 1

The idea of this study was to extract a large sample of combinations and to separate these into different categories based on the conceptual content of the constituent nouns. Previous efforts have attempted to label combinations by hand but the scalability of this approach is limited by the sheer effort involved. In the following section we describe a novel automated approach for categorizing a large number of combinations.

Method

In order to obtain a sufficient number of combinations, we availed of the British National Corpus (BNC), a tagged, annotated corpus containing over 100 million words. The BNC is designed to represent a wide cross-section of modern English and therefore includes a comprehensive sample of both written and spoken language. In order to increase the accuracy of compound noun identification, we obtained a version that had been parsed using the Charniak parser and extracted all compound noun phrases consisting of two nouns. Additional filtering was used to remove acronyms, misspellings, common nouns and errors (see Maguire, Wisniewski & Storms, 2006 for a more detailed description). This process yielded a total of 252,127 different combination types, involving 16,878 (57%) of the 29,617 nouns appearing in the BNC.

In order to analyse the data, we required some way of classifying nouns into a limited number of broad types. The WordNet online lexicon contains definitions for all common nouns in the English language. These nouns happen to be arranged into 25 separate "lexicographer files" corresponding to such general categories as animal, plant, time period etc. The main obstacle to using this classification was that many nouns have multiple senses, with 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, we cannot assume that the noun *dog* will always refer to the animal sense when used in combination.

In order to circumvent this issue, 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 (i.e. "a sleeveless garment worn underneath a coat" or "a collarless undergarment"). However, discarding all multi-sense nouns would have potentially biased our sample. In light of this, we selected an arbitrary threshold for diagnosticity which allowed a reasonable balance between accuracy and ecological validity. Any noun was included where the dominant sense accounted for at least 90% of the occurrences of that noun, as determined by the Senseval sense frequencies provided in WordNet. For example, we were able to include the canine sense of *dog* 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.

Some errors in classification arose due to WordNet anomalies while others were caused by unrepresentative Senseval frequencies. For example, *builder* has three senses, two of which are "a person who creates a business" and "someone who supervises construction". However, both of these have a Senseval frequency of 0 while the sense of "a substance added to soaps or detergents to increase their cleansing action" has a frequency of 11. The combination *yacht builder* was therefore classified as artifact-substance. In contrast, *notepad* has a single sense in WordNet, but is classified as a paper 'substance' rather than as an artefact. As a result, the combination *desk notepad* was classified as artifact- substance.

We analysed a random sample of 100 diagnostic combinations in order to ascertain the reliability of the classifications. Of these, 3 were not genuine combinations, namely *mother rose*, *word hippodrome* and *suspicion falls*. The inclusion of these phrases was a result of parsing errors by the Charniak parser and the fact that the verbs *falls* and *rose* are also included as nouns in WordNet. Of the remaining 97 combinations, only 4 included nouns that were incorrectly classified. The head *colours* in *rose colours* was categorised as an artefact because this plural form has a WordNet entry relating to "an emblem". Similarly, the plural head *hearts* in *leftist hearts* has a WordNet entry relating to "a form of whist". The head *acoustic* in *stereo acoustic* was categorised as artefact because the only sense in WordNet is that of "a remedy for hearing loss". Finally, the head *court* in *practice court* was incorrectly classified using the justice sense, since this has a Senseval frequency of 831, compared to 36 for the appropriate sports sense. Based on this analysis, we concluded that the level of accuracy was sufficient for identifying patterns of combination between the various noun types contained in the sample.

Results

We report our results in terms of combination types and tokens. A combination type is one particular pairing of a modifier and head (e.g. *kitchen chair* constitutes a single combinational type) while the token count reflects the total number of occurrences. In our first analysis, we considered the use of our diagnostic nouns in combination throughout the BNC. One marked trend was that these nouns appeared more frequently as modifiers than as heads. While there are obviously the same number of modifiers and heads in the BNC, our nouns appeared as modifiers 479,440 times but only 383,210 times as heads (123,221 versus 95,970 types respectively). Maguire et al. (2006) found that specific concepts are more likely to appear in the modifier role. Due to our selection process, whereby only the least ambiguous nouns were included, our diagnostic sample may have contained more concepts of this type. In general, abstract concepts are more likely to have multiple senses and are therefore less likely to be diagnostic of one lexicographer file (e.g. *security*). On the other hand, specific concepts are more likely to have a single sense (e.g. *turtle*), therefore explaining why they might have appeared more frequently in our sample.

We obtained the ratio between modifier type frequency and head type frequency for each of the 25 noun categories. The types of noun which appeared most frequently as a modifier were time periods and substances (82% and 79% respectively) while the types of noun appearing most frequently as a head were cognition and shape (e.g. *classroom strategy*, *cathedral square*, 66% and 61% respectively). These results support Maguire et al.'s (2006) finding that less specific concepts prefer the head role. Nouns from the cognition and shape categories are relatively abstract while time periods and substances are of a more precise nature. The productivity of time periods and substances as modifiers corresponds with their potential to participate in the <during> and <made of> roles.

We also expressed the frequency of each diagnostic noun appearing 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). In contrast, the type of noun least frequently used in combination were attributes, shapes and feelings (10%, 9% and 5% respectively). Again, this is consistent with Maguire et al.'s (2006) finding that abstract concepts appear less frequently in combination. These are less likely to interact with other concepts (modifier productivity) and are less likely to have identifiable subclasses (head productivity).

In our second analysis we filtered our sample of combinations down to those consisting of two diagnostic nouns, yielding a total of 72,510 (28.8% of the total). We then separated these into 625 different categories based on the 25 possible permutations of modifier and head type. In order to verify that the reduced sample was still representative of overall combination patterns in the BNC, we compared the modifier and head productivity for the different noun categories. The correlations between number of modifier and head types of the reduced sample and the complete BNC sample were both .986. The correlation between the modifier/head frequency ratio of both samples was .794 (all *p*s $< .01$). These statistics indicate that combination patterns for the various noun types was not biased by our restriction of the sample to combinations consisting of two diagnostic nouns.

Of the 625 possible combination categories, the most productive by types were artefact-artefact (7.0%, *bicycle shed*), person-person (3.4%, *peasant soldier*), artefact-act (2.8%, *guitar tuning*), artefact-person (2.8%, *clarinet teacher*), and substance-artefact (2.3%, *steel pipe*). Subsequently, we examined patterns of combination use for each type of noun as a modifier and as a head. In both cases, distinctive patterns were evident for different types of noun. Regarding modifier token use, the most peaked patterns were observed for substance, food, and plant nouns. For substance modifiers, the only two types of head with an incidence greater than 10% were artefacts (34%, *plastic robot*) and other substances (27%, *wax paste*). These types of head are likely to be associated with some consistency, which substance modifiers can indicate. In contrast, heads not typically associated with a consistency obtained far lower proportions (e.g. plant 1%, animal 1%, location 1%, event, feeling and time 0%). Food modifiers combined primarily with two categories of head noun, namely artefacts (40%, *egg spoon*) and other food heads (14%, *custard pie*). These two categories correspond with the <for> and <has as ingredients> relations. Plant modifiers combined primarily with two categories of head, namely artefact (29%, *cedar staircase*) and plant (20%, *flower seed*), reflecting the <made of>, <is> and <has> relations.

Regarding head noun use by tokens, the most distinctive patterns were observed for time, substance, phenomenon and artefact nouns. Time periods as head nouns combined with three types of modifier with an incidence greater than 5%, namely act (22%, *camping holiday*), artefact (20%, *firework night*), and time (23%, *autumn afterno*on). All three of these modifier types can elaborate on a time period, describing what occurs or when it occurs. In contrast, modifier types which were not easily associated with a time period were far less frequent (e.g. substance, 2%). Substance heads combined most frequently with other substances, (51%, *wax paste*), and body parts (13%, *blood protein*). Again, modifier types with no obvious relationship to substances scored far lower (e.g. feeling 0%). Phenomenon heads combined with time modifiers, (31%, *autumn sunlight*) and artefact modifiers (21%, *pipe smoke*), reflecting the <during> and <caused by> relations. Finally, artefact heads combined with artefact modifiers (35%, *gun ammunition*), act modifiers (12%, *gardening hat*) and substance modifiers (10%, *plastic robot*). In summary, these results provide converging evidence that the way in which nouns tend to be used in combination reflects the interaction of their conceptual content.

For the above analyses, the incidence of combinations involving the different noun types was strongly influenced by the frequency of that noun type in the corpus. For example, artefacts and acts constituted 20% and 13% of the 12,960 diagnostic nouns in our sample while animals and plants made up only 1% each. Consequently, combinations involving artefacts and acts were by far the most common. In order to control for this variance in noun type frequency, we conducted a further analysis. For each of the 625 different categories of combination, we computed the ratio of the number of tokens observed versus the number that would have been expected given the frequency of the modifier and head noun types. We then extracted the top ten of these ratios in cases where that category included at least 20 different types.

Of all the categories, plant-plant combinations had the highest ratio, being 27.5 times more common than the frequency of plant nouns would imply (*elm tree*, *flower bud*, *bramble leaf*). Food-food combinations had the next highest ratio, being 11.6 times more common than expected (*hamburger bun*, *kebab sauce*, *dessert beer*). These categories were followed by substance-substance combinations with 8.8 (*lithium metal*, *powder ice*, *wax paste*), animal-body with t 8.6 (*giraffe neck*, *lamb leg*, *rabbit teeth*), animal-animal with 8.1 (*terrier dog*, *rat flea*, *hen bird*), body-state with 6.9 (*eye trauma*, *kidney disease*, *muscle tension*), finance-finance with 6.7 (*pension money*, *tax profits*, *cash wage*), time-phenomenon with 6.6 (*autumn sunlight*, *dawn wind*, *winter mist*), body-body with 6.5 (*chest muscle*, *ankle tendon*, *jaw tooth*) and finally area-animal with 6.1 (*mountain cattle*, *ocean plankton*, *river bird*). In general, combinations with the same type of modifier and head were far more common than expected, demonstrating that nouns tend to combine with others from within the same domain. Many of the most productive categories were those associated with some specific relation. For example, area-animal combinations were more common than expected because of the fact that geographic areas are often used to reference the habitat of animal species.

In order to reinforce the intuitive observation that different noun types did not combine randomly, we performed a series of analyses. First, taking the log of each of the ratios, we tested the data for normality using a Wilks-Shapiro analysis and found that it was not normally distributed (p >.05). Instead, there were several very high ratios for the more productive patterns but far more lower ratios, indicating that noun types have strong preferences for combining with a small selection of other noun types. We also correlated the modifier and head ratios with each other in order to demonstrate the intuitive interaction between conceptual content and role. This correlation was not significant, $r = .04$, $p > .05$, showing that nouns combine differently depending on whether they are used as a modifier or a head. This result reveals a clear distinction between the two roles and demonstrates that the influence of conceptual content is manifested differently in each case. Finally, we physically divided our corpus of diagnostic combinations into two equal random halves. We recomputed separate ratios based on each of these sets and then compared them, obtaining a significant correlation of $r = .75$, $p < .01$. Given that both halves of the sample exhibited similar tendencies, this rules out the possibility that such trends arose randomly. Instead, our results indicate that nouns of the same type will tend to combine in predictable patterns.

Discussion

The principal finding of this experiment is that separating nouns into a small number of broad categories yields distinctive patterns in noun use as both modifiers and heads. Not only does this undermine Gagné and Shoben's (1997) proposal that the modifier is the only reliable indicator of combination use, it also undermines their assumption that every modifier will combine in a unique fashion, independent of its conceptual content. Using a large sample of combinations, we have observed that nouns of the same general type tend to combine in similar ways. In addition, the more productive pairings of noun types were those where the conceptual content of the constituents supported some specific relation. Given the association between semantics and combination use, there appears to be little motivation for storing a separate set of statistics with every concept. In the following study we explored the same issue using specific nouns as opposed to generalized noun types.

Study 2

In this study we analyzed the use of a sample set of 50 common concepts in combination. The objective was to ascertain whether similar concepts combine more frequently with the same constituents.

Method

In order to guarantee a broad sample of concepts, we consulted Battig and Montague's (1969) category norms. From this database, we selected 50 nouns that occurred at least 50 times as both a modifier and a head within the BNC. In order to ensure a broad selection, these nouns were taken from the following categories: body part, dwelling, furniture, insect, kitchen utensil, mammal, natural earth formation, profession, tool, vegetable, weapon, food, vehicle, weather and plants.

We employed two different methods for computing a 2-dimensional similarity matrix between the 50 nouns, one automatic and one participant-generated. The computational method was based on Resnik's Information Content metric as applied to the WordNet hierarchy. WordNet groups English words into sets of synonyms, as organized in a hierarchy. For example, *cat* and *dog* are hyponyms of the synset *mammal*, which itself is a hyponym of the synset *animal*. According to Resnik (1995), similarity can be determined by the amount of information two concepts have in common, as given by the most specific common abstraction that subsumes both concepts. 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 WordNet are animals. On the other hand, *dog* is very dissimilar to *rain* since the most specific common abstraction for these nouns is the *entity* node in WordNet, of which every entry is an example. Accordingly, the similarity ratings we derived for *dog* and *cat* (0.6) and *dog* and *rain* (0.0) reflect the inverse log of the proportion of WordNet synsets subsumed by their more specific common abstraction.

In order to verify the accuracy of these automated similarity ratings, four participants were asked to make the same judgments. Each participant rated the similarity of all 2,450 pairs of concepts and these values were averaged for each comparison. As participants were native Dutch speakers, concepts were therefore presented in Dutch. The estimated reliability of the resulting similarity values was .73. The correlation between the WordNet similarity metric and the participant ratings was .63, rising to .74 following correction for the unreliability of the latter ($p < .01$). These statistics show that WordNet similarities can be relied on to approximate human judgments. For each of 50 concepts, we selected 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. In the following section we describe a series of analyses which investigated whether the similarity of the concepts was related to the similarity of the nouns they combined with.

Similarity of Combining Nouns

In this analysis we computed a two-dimensional matrix reflecting the degree of overlap between the top combining nouns for our 50 concepts. Given that the analysis was based on the 10 most frequent combination types for each concept, the maximum overlap was 10. For example, the modifiers *cat* and *dog* had an overlap of three, because they had *owner*, *food* and *show* in common. However, their seven other most frequent combining nouns were different. In fact, no two concepts had an overlap of greater than three and over 90% of entries in both the modifier and head overlap matrices were zero. We computed the correlation between the overlap and similarity matrices. The correlation involving the 50 nouns as modifiers was 0.31 while that involving the 50 nouns as heads was 0.19. Using the participant ratings in place of the WordNet similarities, these correlations rose to .35 and .20 respectively (all *p*s < .01). These results demonstrate that similar concepts are more likely to combine with the same noun, although the strengths of our correlations were relatively weak. The main problem with the above analysis was the low number of shared nouns, meaning that the overlap measure was not very informative. Furthermore, in cases where two words were very similar though not identical, no overlap was registered. For example, *dog dirt* and *cat faeces* were among the most common combination types for *dog* and *cat* respectively, yet the overlap for these was still zero.

Taking into account these issues, we ran a further analysis based on the level of similarity between the sets of combining nouns for a pair of concepts. In order to obtain similarity ratings we again made use of Resnik's metric. This time, in comparing *dog* and *cat*, we considered each of *dog*'s combining nouns and computed its maximum similarity with any of *cat*'s combining nouns (e.g. *faeces* and *dirt* had a similarity of 0.36). These ten values were then averaged to obtain the overall similarity between both sets. Using this technique, we obtained pairwise similarity values for all 500 of the modifier and head combining nouns, involving a total of half a million comparisons. In order to compensate for contamination resulting from false positives, we squared all the values so that those that were relatively high (indicating very similar concepts) would have a greater influence than the lower ones. The correlation between the similarity of the 50 concepts and their set of associated combining nouns as modifiers was 0.29. The same correlation involving the 50 concepts as head nouns was 0.28. These correlations rose to .30 and .31 using the participant-generated frequencies (all *p*s < .01). Again, these correlations demonstrate that similar concepts tend to combine with similar constituents.

Although the correlations were significant, they were not as strong as we expected. One reason for this may have been the unreliability of the similarity ratings produced. Some of these were unrealistic due to peculiarities in the arrangement of the WordNet hierarchy. For example, *shop* is classified as a hyponym of the *structure* synset, yielding unrealistically high similarities with nouns such as *bridge*, *wall* and *door* (all .36). Nevertheless, we suspected that even had perfect ratings been available, the correlations would not have been much higher. The inherent problem with the above analyses is that they are based only on the top 10 most frequent combining nouns for a given concept. Because a noun can plausibly combine with thousands of other nouns, there is no guarantee that the 10 most frequent of these will provide a representative sample. Often, the most common combination types are idiosyncratic and unsuitable for comparison. For example, *tabby cat*, *pussy cat* and *tom cat* were among the most common types 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. Another problem associated with using a limited sample of combinations 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 a combination such as *ladder basket*, a fact which cannot be reflected by the current paradigm. In the following section we present an alternative method of comparison which avoids over-generalizing based on a limited sample of combination types.

Determining Frequency using Web Counts

The internet is being increasingly used as a data source in a wide range of natural language processing tasks. Given that it represents a corpus of some 100 billion words, we were able to obtain frequencies for combinations which would have been too unusual to be represented in a smaller corpus such as a BNC. We created novel combinations by taking the top 10 combination types for a given concept and substituting the 49 other concepts in its place. For example, performing this substitution for *cat breeder* yielded combinations of the form *dog breeder*, *ladder breeder*, *wind breeder* etc. Subsequently, the Google search engine was used in order to obtain frequency counts for the 490 artificial compounds generated for each concept. We computed the log of the number of hits for each compound and normalized this value according to the frequency of the nouns involved. The purpose of the normalization 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 yielding more false positives. In order to obtain an overall 'interchangability' score for each entry in our two dimensional matrix, we averaged the values for each of the ten novel combinations produced. Thus, the value of 0.5 between *dog* and *cat* reflects the fact that substituting the modifier *cat* in place of *dog* yielded combinations with relatively high frequencies in Google (e.g. *cat owner*, *cat food*, *cat breeder*). The main problem with the use of the web as a corpus is the level of noise associated with the data. The Google search engine ignores punctuation and capitalization, leading to false positives whenever the paraphrase match crosses a sentence boundary (e.g "…he called his dog. Rain had started to fall…"). Matches are also likely to include links, web addresses, names and other non-textual data. Furthermore, web data is not tagged with part of speech information, meaning that a significant portion of hits for noun-noun compounds will inevitably involve the use of the constituent nouns as verbs or adjectives (e.g. "she watched the cat hunt the mouse").

The two-dimensional interchangability matrix was correlated with the original similarity matrix. Using the WordNet similarity values, the correlation between these two matrices for the 50 concepts as modifiers was .49 while that involving the 50 concepts as heads was .40. Using the participant-generated similarity ratings, these correlations rose to .53 and .47 respectively or .63 and .55 assuming perfectly reliable ratings (all *p*s < .01). Given the unreliability of Google frequencies, these results are particularly impressive and demonstrate that conceptual content has a strong bearing in determining how nouns are used in combination.

Discussion

Our results highlight a relationship between conceptual content and the way that nouns tend to combine as both modifiers and heads. This association was observed both in comparing the top ten most frequent combining types for a pair of concepts and in obtaining web counts for combinations in which one concept was substituted with another. These results offer converging evidence that similar nouns combine in similar ways and that the more similar they are, the more likely they are to combine with the same nouns.

The difference between the modifier and head correlations for the overlap and web count analyses was significant using both the participant generated and WordNet similarities (all *p*s < .01 using two-tailed *z*-tests). However, this discrepancy may simply be an artifact of our sample. The 50 nouns we selected were not particularly representative since many denote superordinate categories for which many subtypes exist (e.g. *grass*, *rice*, *cat*). As a result of this, many of the most frequent combination types for these nouns as heads were of a sub-type super-type nature (e.g. *marram grass*, *pilau rice*, *tabby cat*). Such modifiers are unlikely to overlap with any others and substitutions of the head concept are unlikely to result in sensible combinations (e.g. *marram tree*, *pilau potato*, *tabby dog*). As a result, the difference in the correlations for the modifier and head matrices cannot be interpreted as evidence that conceptual content is a better predictor of modifier use. On the contrary, our results reveal that the semantics of a noun influence its use as both a modifier and a head.

General Discussion

In recent years, many theories of conceptual combination have been proposed, yet no large-scale analysis of combination use had hitherto been conducted. We have addressed this issue, providing converging evidence that conceptual content strongly influences how a noun will be used in combination.

Our results can be interpreted as supporting the slot-filling view. On the other hand, they contradict the main principles of the CARIN theory. First, both modifiers and head nouns revealed distinctive patterns of combination, undermining the notion of modifier primacy. Second, if different nouns of the same type exhibit similar trends in combination then there should be no need to store separate statistical distributions for each. The CARIN theory fails to acknowledge the intuitive link between relation frequency and noun properties, a link which we has been consistently reinforced throughout our study. As a result, the statistics on which the theory bases itself may be measured at the wrong level of abstraction and relation frequency may actually be an epiphenomenon of conceptual content.

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.

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.

Maguire, P., Wisniewski, E. J. & Storms, G. (2006). An Investigation of the Productivity of Abstract and Concrete Nouns in Combination. In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, Vancouver, Canada.

Resnik, P. (1995). Semantic similarity in a taxonomy. An information based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, *11*, 95-130.

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