An Analysis of the CARIN Model of Conceptual Combination

Barry Devereux (Barry.Devereux@ucd.ie)[1]
Phil Maguire (Phil.Maguire@ucd.ie)
Fintan Costello (Fintan.Costello@ucd.ie)
Arthur Cater (Arthur.Cater@ucd.ie)
School of Computer Science and Informatics, University College Dublin
Dublin, Ireland

Abstract

The CARIN model of conceptual combination (Gagné & Shoben, 1997) assumes that people interpret noun-noun compounds by selecting a relation to link the two constituent nouns from a fixed list of possible relations. The model uses statistical information about the frequency with which the modifiers of compounds have been associated with different possible thematic relations in past experience. The CARIN model describes how relation selection could involve competition between relations; Gagné and Shoben have shown that response strengths computed by the model correlate with people's reaction times when judging the sensibility of compounds. We present data in support of the CARIN model's assumption that people store statistical information about relation frequency: an analysis of representative compounds selected from a corpus produced frequencies that agreed with those derived by Gagné and Shoben. We also present an analysis of the equation proposed in the CARIN model, showing that it does not provide for competition among relations in the manner asserted by the theory. We propose a simple alternative mechanism for relation selection in compounds whereby response times for a given compound are proportional to the number of frequent relations that must be considered before reaching the correct relation.

Keywords: Conceptual combination; noun-noun compounds; CARIN; mathematical modelling.

Introduction

For speakers of English and many other languages, noun-noun compounds such as *volcano science* or *gas crisis* are a productive and efficient strategy for referring to novel concepts and ideas. In English, compounds consist of a modifier noun followed by a head noun: the head noun typically denotes the main category and the modifier indicates a contrast or specialization of this category (e.g. a *kitchen chair* is a type of chair typically found in kitchens). Both lexicalized and spontaneous compounds are ubiquitous phenomena in everyday language, providing a concise way to reference concepts. Such concision can be exploited when the addresser is confident that the addressee can interpret the phrase in its reduced state using their knowledge of the situation. This interpretation process can be dependent on an accompanying context or alternatively the meaning of the constituent nouns alone may be sufficient for comprehension.

There has been much interest in this phenomenon within cognitive psychology due to the fact that the study of concept combination presents a well-defined domain which retains the potential to reveal much about conceptual representation and language comprehension in general. Consequently, many different models of concept combination have been proposed (e.g. Costello & Keane, 2000; Wisniewski, 1997). Most of these models have tended to converge on the view that during the interpretation process, the basic head noun category is somehow refined or modified by the modifier concept.

One theory which has adopted a different view is the Competition Among Relations In Nominals (CARIN) model (Gagné & Shoben, 1997). According to the CARIN model, the interpretation of a compound occurs when a person identifies a relation that exists between the modifier and the head noun. The number of potential relation types (called *thematic relations*) is restricted to a set of about 16, including basic relation types such as *located* and *for*. According to the theory, the accessibility of thematic relations determines the ease with which the compound can be understood. Accessibility of thematic relations was estimated by determining the frequencies with which different thematic relations co-occurred with different modifiers and heads in a large corpus of compounds. This allowed them to identify which thematic relations were most common for the different modifiers and heads. For example, the modifier *mountain* was found to combine with the *location* relation far more than with any other relation type (in compounds such as *mountain cloud*, *mountain stream*, and *mountain goat*). After analyzing response times for a sample of compounds incorporated into a sensibility judgment task, it was found that the distribution of thematic relation preferences of the modifier noun influenced response time but that this was not the case for the head noun. In other words, Gagné and Shoben's findings suggest that the fact that *mountain* was frequently associated with the *location* relation appears to influence response times for compounds involving *mountain* as a modifier, whereas the fact that *magazine* was frequently associated with the *about* relation does not appear to have any effect for compounds involving *magazine* as a head. At first blush, this modifier primacy effect seems surprising given that previous theories have emphasized the importance of both constituents, and have usually viewed conceptual combination as a modification of the head noun concept.

However, one problem with this evidence is that it is

---

[1]The first two listed authors should equally be regarded as first author

based on relation frequencies which are not derived from a representative sample of compounds. The CARIN theory offers no reasons why one would expect frequencies derived in the manner of Gagné and Shoben's study to approximate true statistical frequencies. Storms and Wisniewski (in press) have pointed out that the compounds on which Gagné and Shoben based their frequencies are not compounds which would be expected to occur naturally (e.g. *flu candy*, *rain lake*, *headache season*). Storms and Wisniewski attempted to rectify this by using compounds generated by experimental participants. However, participants tend to generate mostly obvious compounds (with more common relations) than less obvious ones (with rarer relations), meaning that these too do not reflect a naturally occurring sample of compounds. To our knowledge, no study so far has collected representative frequency data from a sample of compounds occurring in everyday language. In this article we address this oversight by means of a corpus study and evaluate the implications of our findings for the CARIN theory.

Estimating Relation Frequency

In Gagné and Shoben's (1997) original study relation frequency data were gathered as follows. First, 91 heads and 91 modifiers taken from Levi (1978) were paired in every possible combination, producing 8,281 compounds, of which 3,239 were judged to be sensible. These were classified into the 16 relation categories used with the CARIN model. The frequency with which a relation was associated with a given noun was determined by calculating the percentage of sensible compounds having that noun and relation. For example, for the head noun *cloud*, the *located* relation had a frequency of 13%: 13% of the sensible compounds with the head noun *cloud* used the *located* relation.

However, it is doubtful whether such a sample is representative enough to accurately reflect the relation type distribution of a given noun. It would seem intuitive that different modifiers often combine with quite different heads (and vice versa). (For example, the head noun *debate* may occur with modifiers which are a type of social concept (e.g. *war*, *government*, or *globalization*) far more often than heads in general do. Such nuances are not taken into account in Gagné and Shoben's pairing methodology. Furthermore, such arbitrary pairings are likely to include a large proportion of unfamiliar compounds.

Unfortunately, no study investigating CARIN thus far has examined the true incidence of combinational variance in everyday life. We therefore derived a fully representative, corpus-based set of relation frequencies with which to test the robustness of the CARIN theory. We believed that deriving more accurate frequencies would allow us to assess the accuracy of previous methods in estimating the actual frequencies of relations.

Corpus Study

In order to carry out this analysis we needed to consult a large corpus that was representative of naturally occurring language and which could be expected to yield a sufficient number of compounds for given heads and modifiers. We used the British National Corpus (BNC) World Edition, a tagged corpus containing over 100 million words. Noun-noun sequences were identified using GSearch (Corley et al., 2001), a chart parser that detects syntactic patterns in a tagged corpus via a user-specified context-free grammar and a syntactic query.

Method

In our study, we aimed to generate new frequency data for a subset of the 91 heads and 91 modifiers that Gagné and Shoben used in their corpus study, namely the 19 heads and 19 modifiers that are used in Gagné and Shoben's (1997) Experiment 1. To this end, we extracted from the BNC a random sample of 100 modifier-noun phrases for each of the 19 modifiers and each of the 19 heads. In selecting a sample of 100 valid compounds for each noun, we were forced to resolve the issue of noun ambiguity. For example, terms such as *plant* and *album* are ambiguous and were found to combine in different senses (e.g. *nuclear plant* vs. *flower plant*, *photo album* vs. *rock album*). However, as the materials in Gagné and Shoben's experiments suggest that noun ambiguity was ignored when materials were being created for their experiments, we also ignored this issue. To resolve potential ambiguities in compound meaning we presented each compound within its BNC sentential context. A random sample of 100 valid compounds was extracted from each sample for the purpose of deriving relation type frequencies and in the seven cases where there were less than 100 valid compounds, all were used. In total, 1,832 modifier and 1,669 head compounds were used to derive the frequencies which compares favorably with the 1,037 modifier and 878 head compounds used by Gagné and Shoben for the same materials.

The first two authors classified each of these 3,501 compounds into the most appropriate thematic relation categories. We did not find the task of classifying the compounds into relation categories to be straightforward, however. We found that it was rarely the case that there was a single clearly appropriate thematic relation for a given compound. For example, it is difficult to judge whether family activities is best classified as belonging to the *has* relation (i.e. "family has activities"), the *located* relation (i.e. "activities in a family"), or indeed the *for* ("activities for families"), *causes* ("families cause activities"), or *by* ("activities by families") relations. This suggests a significant difficulty with the thematic relation approach; thematic relations are too vague and ill-defined to serve as adequate representations of the specific relationships that are instantiated during conceptual combination. However, since our rating task required us to follow Gagné and Shoben's methodology of classifying compounds into one and only one thematic relation category, this necessitated selecting a single relation in a way that was often quite uncertain. We computed inter-rater reliability using a sample of 10 combinations for each of the 38 nouns. Given that multiple relations were often equally appropriate,

the level of agreement was reasonably high at 68%, with Cohen's (1960) $\kappa = 0.65$ ($p < .001$).

Having classified each compound into one thematic relation category, we derived relation frequencies by counting the percentage of cases falling into each category (the same manner in which relation frequencies were derived in Gagné & Shoben, 1997). These were then compared to the frequencies derived in Gagné and Shoben's study.

Results

We compared the frequencies for the 19 heads and 19 modifiers derived from the BNC sample with the frequencies obtained for those items in Gagné and Shoben's dataset. In comparing these frequencies with Gagné and Shoben's frequencies we did not include all 16 relation categories for every noun. This would have produced an artificially high correlation as many thematic relations will always receive frequencies of 0% given that some thematic relations are simply invalid for some heads and modifiers (for example, the statistic that *mountain* as a modifier does not combine using the *during* relation is uninformative). On average, the nouns occurred with compounds for only a handful of the thematic relations and so the most informative method of comparison would involve using only non-zero values. As a result frequencies of 0% in both our and Gagné and Shoben's corpus study were discarded. Given that frequencies tended not to be normally distributed we always used Spearman's ρ.

In spite of the restrictions on the statistics described above, there was a significant degree of agreement between the frequency scores derived by Gagné and Shoben and those derived in our study (for the heads, $\rho = 0.337$, $p < 0.01$, $N = 166$; for the modifiers $\rho = 0.392$, $p < 0.01$, $N = 188$). These results suggest a close relationship between the relational distribution in Gagné and Shoben's original dataset and in the BNC sample, and demonstrate that relation frequencies derived using Gagné and Shoben's arbitrary modifier and head pairing methodology are a good approximation to the relation frequencies that exist in natural language.

While the agreement between relation frequencies in our BNC corpus and in the Gagné and Shoben dataset was high, there were some interesting cases where the two datasets deviated significantly. These are worth examining in detail. One reason for deviations in frequency was due to the fact that some relation frequencies were more domain specific than others. Thus, for example, the fact that *plastic* has an affinity for the *made of* relation could be picked up using the arbitrary pairing methodology since many different things can be made of plastic. However, the modifier which deviated greatest was *servant*, which we found occurred with the *is* relation 76% of the time (typically in compounds such as *servant boy* and *servant girl*), compared with only 5% for Gagné and Shoben's corpus (where typical compounds would be *servant book* and *servant scandal*). The reason for this discrepancy is that the use of this thematic relation for *servant* is quite domain specific. Gagné and Shoben's approach to estimating frequency fails to take

account of this; the modifier *servant* is combined with exactly the same set of heads that the modifier *plastic* is, failing to account for usage patterns of real compounds.

The limitations of arbitrary pairings were also revealed in the opposite sense in that universal acceptors (i.e., heads that can combine sensibly with modifiers of any type) like *book* were forcibly paired using the universal slot more often that was representative. Although the X in the compound X *book* can be filled by nearly any subject matter, in reality terms like *accounts book* and *address book* are more common than indicated by Gagné and Shoben's frequencies.

Analysis of the CARIN Strength Equation

In the previous section we provided evidence supporting the relation frequency distributions used in Gagné and Shoben's studies of conceptual combination. In this section we turn to the equations that they put forward to represent the Competition Among Relations In Nominals (CARIN) model of compounding. The degree of competition among relations in the CARIN equation is a function of the frequency of those relations for the modifier of the compound being interpreted. We intend to show here that this equation does not actually provide for competition among relations.

The CARIN model proposes two factors influencing the ease of interpretation, and hence reaction time, for a given compound. First, 'interpretations are easier if the required relation is of high strength [high frequency]'. Second, the availability of other high-strength [relations] should slow the interpretation of conceptual combinations' (both quotes from Gagné & Shoben 1997, pp 81). In other words, given that the correct relation for some compound has a frequency f, if there is some other incorrect relation with a frequency close to f, then reaction time for the compound should be slow (because of competition between the two similarly frequent relations). However, if there is no other relation with a frequency close to f, then the reaction time for the compound should be faster (because the correct relation can be selected without competition).

As an explanatory step in presenting the CARIN model, Gagné and Shoben (1997) first present the following version of their strength equation:

$$strength = \frac{P_{selected}}{P_{selected} + P_1 + P_2 + P_3} \qquad (1)$$

where $P_{selected}$ is the frequency, or proportion of times, that the selected (i.e. most appropriate) relation is used with compounds that have the same modifier, and P_1, P_2, P_3 are the corresponding proportions for the three highest-frequency alternative relations for that modifier.

Qualitatively, this equation predicts that the ease or rapidity for interpreting a compound c with a relation r is an increasing function of the frequency with which compounds which have the same modifier as c are interpreted with relation r, and a decreasing function of the corresponding frequencies for the three most common

alternative relations for those compounds. For Experiment 1 from Gagné and Shoben (1997), the correlation between the strength ratios derived from this equation and the response time data is -0.35 for the 57 items.

Though only used as an explanatory step in presenting the CARIN model, we were interested in whether this equation actually reflects competition in relation selection. Is the response strength for the correct relation affected if there is another relation with frequency close to that of the selected relation? In fact, no. Whether one of these alternative relations has a frequency close to that of the selected relation or not does not have any influence on the equation's output. For example, suppose the proportional frequency of the selected relation is 0.50. If one of the alternative relations also has a frequency of 0.50 and the other two have frequencies of 0.0 this equation will give a strength response of (0.50 / (0.50 + 0.50 + 0.0 + 0.0)) = 0.5. However, if the three most frequent relations have frequencies of 0.17, 0.17 and 0.16 (which together sum to 0.50), this equation will also give a strength response of (0.50 / (0.50 + 0.17 + 0.17 + 0.16)) = 0.5. According to the equation, these two situations are indistinguishable. However, the first case clearly should involve competition between two relations both with a frequency of 0.50 while the second case does not involve competition (the correct relation having a frequency 0.50 and the next most frequent relation having a frequency of only 0.17).

If this equation does not involve competition, how does it give a significant correlation with the response times obtained in Gagné & Shoben's experiment? If we consider a generalized form of this equation, we can see what is actually happening:

$$strength = \frac{P_{selected}}{P_{selected} + P_1 + P_2 + ... + P_N} \qquad (2)$$

where N is the total number of thematic relations being used. In this equation, the numerator is simply the frequency for the correct relation for the compound in question, and the denominator is the sum of frequencies for all relations (not just the three most frequent alternatives). Since, the sum of all frequencies totals 1, this equation simply reduces to the frequency of the correct relation: in this equation, response strength is simply equal to $P_{selected}$. The reason Equation 1 gives a significant correlation with response time is not because it involves competition among relations, but simply because it reduces to the frequency of the selected relation. An analysis of the data confirmed this relationship, giving a correlation between $P_{selected}$ and response strength as computed in Equation 1 of $r > 0.99$.

In the second, more important, version of the CARIN model presented in Gagné and Shoben (1997), each of the variables are transformed using the negative exponential function and a free parameter α:

$$strength = \frac{e^{-\alpha P_{selected}}}{e^{-\alpha P_{selected}} + e^{-\alpha P_1} + e^{-\alpha P_2} + e^{-\alpha P_3}} \qquad (3)$$

Qualitatively, this version of the model predicts that the difficulty or length of time (*not* ease and rapidity) of in-terpreting a compound c with relation r is an increasing function of the frequency with which compounds which have the same modifier as c are interpreted with relation r, and a decreasing function of the corresponding frequencies for the three most common alternative relations for those compounds. The correlation of this negative-exponential-term version of the CARIN model and the response time data is 0.45, using a value of 36 for the parameter α to optimally fit the data.

Though both of these versions of the CARIN model take the same general form, they differ in how they model response time: the linear-term version of the model measures the ease or rapidity of interpreting a compound (as indicated by the fact that that model gives a negative correlation with RT), whilst the exponential-term version measures the difficulty or slowness of interpreting a compound (as indicated by the fact that that model gives a positive correlation with RT). To further investigate how the two models differed, we calculated correlations between the variables and the strength values that they contributed to. For the linear-term model, we examined the correlation between the strength values for the 57 items in Experiment 1 and the untransformed variables (which are the terms used in that model). This revealed the variable with the strongest linear relationship to strength to be $P_{selected}$, the second strongest to be P_1, the third strongest to be P_2, and the fourth strongest to be P_3 ($r = 0.99$, $r = -0.71$, $r = -0.66$, and $r = -0.57$ respectively). This is sensible: the selected relation and its closest competitor should have the greatest influence on the model's predictions. For the exponential-term model, the corresponding correlations are between the strength values and the exponentially transformed variables (which are the terms in that model). The pattern of correlations was very different for this model: the strongest linear relationship with strength values was again found to be for the selected relation ($r = 0.59$); however, P_3 was found to be the second strongest ($r = -0.41$), with P_2 the third strongest ($r = -0.36$) and finally P_1 the fourth strongest $r = -0.18$. The order of importance of the frequencies for the alternative relations was $P1$, then $P2$, then $P3$ for the linear-term model but was $P3$, then $P2$, then $P1$ for the exponential term model.

Why is the third most frequently selected relation having such a disproportionate influence on strength as calculated using the exponential-term version of the CARIN strength equation? It seems contrary to our intuitions about competition among relations. The effect is observed because of the use of the negative exponential in this version of the model. For example, the term e^{-36*P} evaluates to less than 0.04 for all values of P between about 0.09 and 1.00 and evaluates to between 1 and 0.04 for the small range of values from 0.00 to 0.09. Since P_1 is typically a relatively large number (its average value in the data is 0.31) the corresponding exponential term tends to have no influence on the denominator of the strength equation (i.e. it is effectively zero). P_3 is typically a relatively small number (its average value in the data is 0.08) and therefore the corresponding ex-

ponential term tends to range over a much wider range of values and is much more likely to be a non-negligible term in the denominator.

In fact, the CARIN model of Equation 3 can be approximated by the formula $e^{-\alpha P_{selected}}/e^{-\alpha P_{lowest}}$ indicating that usually only the lowest frequency occurring in the denominator (i.e. P_{lowest}) is influential (for the Experiment 1 data, the correlation between this formula and the formula of Equation 3 is significant, with $r = 0.97$). In other words, the output of the CARIN model can be approximated using only two relation frequencies, namely $P_{selected}$ and P_{lowest}. Since no other frequencies tend to influence the output of the formula, it clearly cannot support competition among relations.

In Gagné and Shoben (1997), the CARIN model is presented as being an application of Luce's (1959) choice function. Luce's choice function formulates the probability of selecting a particular item from a pool of alternatives (in the context of the CARIN model, the particular item is the selected relation, and the pool of alternatives is all the thematic relations in the CARIN model's taxonomy). However, a problem with how Luce's choice function is applied in the CARIN model is how the exponential decay function is used to transform the frequency variables. Gagné and Shoben report that they are following previous applications of Luce's choice rule (e.g. Sadler & Shoben, 1993) when they utilize the exponential decay function; however, there are important differences between how the exponential decay function is used in the CARIN model and how it is used in these previous applications. In Sadler and Shoben (1993), for example, the probability that an item i is selected from a pool of n alternatives in an analogical reasoning task is given by

$$p(i) = \frac{e^{-\alpha d_i}}{\sum_{j=1}^{n} e^{-\alpha d_j}} \qquad (4)$$

where d_i is the distance between item i and some ideal solution to the analogy problem. The probability that item i is selected is therefore a decreasing function of distance: the probability that i is selected decreases as the distance between i and the ideal solution point increases. In this and other models, the negative exponential transformation of the distance measure is used because distance transformed in this way is taken to be a measure of psychological similarity (following Shepard, 1958). In terms of this similarity measure, Luce's choice function asserts that the probability that item i is selected is an increasing function of similarity.

In the CARIN model the relation frequencies correspond to the notion of similarity, not distance: if the frequency of the selected relation is high, then the probability that that relation should be selected should also be high. If the CARIN model is to be specified as an application of Luce's choice rule, therefore, the correct formulation of it is the linear-term version that we have presented above, as the probability that a given relation is selected is an increasing function of the untransformed frequency measure for that relation. The exponential-term version of the model cannot be considered a correct application of Luce's choice rule.

Leaving this issue aside, does the exponential-term model implement a competition amongst relations approach to conceptual combination? As done above for the linear-term version of the model, we investigate this issue by examining whether the response strength for the correct relation is affected if there is another relation with frequency close to that of the selected relation. We found again that the model failed to pass this test. For example, with $\alpha = 36$ as in Gagné and Shoben (1997), suppose that $P_{selected} = 0.2$. If $P_1 = 0.2$ also and P_2 and $P_3 = 0.05$, then the calculated strength response is $0.0008/(0.0008 + 0.0008 + 0.1653 + 0.1653) = 0.00225$. However, if the three most frequent relations have frequencies of 0.7, 0.05 and 0.05 the strength response is $0.0008/(0.0008 + 1.14 \times 10^{-11} + 0.1653 + 0.1653) = 0.00225$. According to the equation, these two situations are indistinguishable to 5 decimal places. However, competition amongst relations predicts the first scenario should be faster that second, as in the first scenario the selected relation only has to compete with relations of the same or less frequency, whereas in the second scenario it is competing with a relation over three times more frequent than it. Again, the problem is with the negative exponential function: as discussed above, the term e^{-36*P} is negligible unless P has a very small value.

Another way to assess the competition proposal is by looking directly at the experimental data. To do this we selected, out of the set of compounds used in Gagné & Shoben's Experiment 1, the set of compounds where the correct relation was also the most frequent relation for the modifier of that compound. We divided this set of compounds into two groups: the 'close' group (compounds for which there was another competing relation with a proportional frequency within 0.10 of $P_{selected}$), and the 'distant' group (those for which there was no other competing relation with a proportional frequency within 0.60 of $P_{selected}$). We selected the criteria values of 0.10 and 0.60 so that there would be an equal number of compounds in both the close and distant groups. The competition among relations approach would predict that, on average, people would have faster reaction times for the distant group (in which there is no other relation competing with the correct relation for the compound), but slower reaction times for the close group (in which there is another relation close in frequency to the correct relation, and competing with that relation). However, analysis showed that the average reaction time for the close group was $1030ms$ $(SD = 94ms)$ while the average reaction time for the distant group was $1073ms$ $(SD = 75ms)$: people reacted, on average, faster to compounds in which there was another relation close in frequency to the correct relation, and slower to compounds in which there was no such such competing relation. This is opposite to the pattern that would be expected by the competition model. Of course, this a post-hoc selection of data from the experimental results, using a relatively small set of data points, and for which there was no sig-

nificant difference between the two groups (t = 0.32 in an unpaired t-test). It does, however, along with our other findings, suggest an absence of explicit evidence for the CARIN model's competition proposal.

In light of these issues, how does the exponential-term model give a significant correlation of 0.45 with response times obtained in Gagné & Shoben's experiment? An analysis of the data shows that the strength score captures information about the number of relations greater than $P_{selected}$; this suggests that what this equation is actually modeling is the rank of $P_{selected}$ (the correlation between the rank of the selected relation and response strength computed by Equation 3 is 0.93). Gagné & Shoben found that the rank of the correct relation was significantly correlated with response time ($r = 0.45$, the same correlation as seen for the exponential equation). It seems that the reason the exponential equation gives a significant correlation with response time is not because it involves competition among relations, but simply because it reduces to the rank of the correct relation.

Discussion

In this paper we have presented both good and bad news for the CARIN model of conceptual combination. Our good news is that the relation frequencies used in the model are confirmed by the close match between a selected subset of those frequencies and a set of relation frequencies derived from a more realistic and more general source: the BNC corpus. This is encouraging for the thematic relations approach because it suggests that relation frequency is a fairly constant property of the words used in compound phrases, with similar relation frequencies being returned for the constituent words of compounds when they are sampled from a randomly selected dataset (as Gagné & Shoben did in their study) and when they are sampled from a large and representative corpus of written text (as was the case for our study). We have also, however, presented bad news for this approach, in showing that the equations used in the competition among relations in nominals (CARIN) model do not actually describe competition, and so the correlations between reaction times and response strengths computed by these equations cannot be taken as support for the CARIN model of conceptual combination.

In the face of this problem for the CARIN model, we propose an alternative model for reaction times in noun-noun compounds: one based simply on the rank of the correct relation for a compound, in order of modifier frequency. One advantage of this modifier rank measure is that it correlates significantly with reaction times in the experiments, as Gagné and Shoben report. A second advantage is that it has a suggestive relationship with ideas of memory access and recall. For example, it could be that when people encounter a compound they query their long-term memory for examples of relations occurring in previously-seen compounds with the same modifier. As each example of a possible relation is returned, its suitability for the compound in question is assessed, and the relation is either accepted or rejected. Given the associative nature of memory, relations are likely to be returned in rank order according to their association with the modifier. Thus the response time for a given compound would be proportional to number of relations that need to be considered and rejected before an acceptable relation is found; in other words, proportional to the rank of the correct relation for that compound.

This proposal is purely speculative, based solely on the observed relationship between relation rank and response time in Gagné and Shoben's (1997) data. There are many other ways in which the observed influence of relation rank on reaction time could be explained and modelled: an important area for future work is to investigate and distinguish between these possible accounts. We think that our results here provide a solid grounding for that work, and may lead to future understanding of the fascinating processes by which people understand nominal compounds.

References

Cohen J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Corley, S., Corley, M., Keller, F., Crocker, M. W. & Trewin, S. (2001). Finding syntactic structure in unparsed corpora: The Gsearch corpus query system. *Computers and the Humanities*, 35, 81-94.

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint guided conceptual combination. *Cognitive Science*, 24(2).

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.

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

Luce, R. D. (1959). *Individual Choice Behavior.* Wiley, New York.

Sadler, D. D., & Shoben, E. J. (1993). Context effects on semantic domains as seen in analogy solution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 128-147.

Shepard, R.N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space *Journal of Experimental Psychology*, 55, 509-523.

Storms, G. & Wisniewski, E. J. (in press) Does the order of head noun and modifier explain response times in conceptual combination? *Memory and Cognition.*

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, 4(2), 167-183.