Modelling the Interpretation and Interpretation Ease of Noun-Noun Compounds Using a Relation Space Approach to Compound Meaning

Barry Devereux (Barry.Devereux@ucd.ie)
Fintan Costello (Fintan.Costello@ucd.ie)
School of Computer Science and Informatics, University College Dublin
Dublin, Ireland

Abstract

In this paper, we present a computational model of conceptual combination that introduces a new representation for the meaning of compounds: the relations used to interpret compounds are represented as points or vectors in a high-dimensional relation space. Such a representational framework has many advantages over other approaches. Firstly, the high-dimensionality of the space provides a detailed description of the compound's meaning; each of the space's dimensions represents a semantically distinct way in which compound meanings can differ from each other. Secondly, the spatial representation allows for a distance metric to measure how similar of different pairs of compound meanings are to each other. We conducted a corpus study, generating vectors in this relation space representing the meanings of a large, representative set of familiar compounds. A computational model of compound interpretation that uses these vectors as a database from which to derive new relation vectors for new compounds is presented. Also presented is a model of interpretation ease: that is, the ease or rapidity with which people can comprehend compounds. Our model uses ideas from the CARIN theory of conceptual combination about the modifier noun's role in the comprehension process; the model correlates as well as the traditional CARIN model with people's reaction times.

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

Introduction

Conceptual combination, the process that people employ when interpreting novel noun-noun compounds such as *volcano science*, *gas crisis* or *penguin movie*, is a non-trivial cognitive task, often requiring people to access complex knowledge about the two constituent concepts and about the world in general. For example, people can quickly and efficiently determine that the compound *penguin movie* refers to a movie *about* penguins, and not a movie *by* penguins (which would be the correct way to interpret the compound *penguin journey*), nor a movie *for* penguins (the correct way to interpret the compound *penguin enclosure*), nor any of the infinitely many other possible but implausible ways of interpreting that compound. Perhaps because of the complexity of the phenomenon, previous theories of conceptual combination have tended to focus on only some aspects of conceptual combination. For example, in Gagné and Shoben's (1997) CARIN model, the focus is on modelling the ease and rapidity with which people interpret noun-noun compounds (as measured by reaction time), but not other features of the conceptual combination process, such as how the correct relationship between the two constituent concepts is found or constructed. In this paper, our aim is to present evidence for a more comprehensive approach to conceptual combination, allowing us to model both the interpretation and interpretation ease aspects of noun-noun comprehension.

Conceptual combination can be regarded as a process which instantiates the most plausible or most appropriate relationship between the two constituent words in a compound (termed the modifier word and the head word, respectively). An important issue therefore for any model of conceptual combination is the manner in which the relationship between the modifier and head of a compound is represented; indeed, previous models of conceptual combination can be classified as belonging to two types, distinguished by how they represent relations. The first type, the concept specialization approach, assumes that instantiating a relation for a compound involves modifying a slot in the representation of the head word concept (for example, see Smith, Osherson, Rips & Keane, 1988). In the second type, the relationship between the two nouns is specified by means of a taxonomy of general relation categories. For example Levi (1978) describes a set of recoverably deletable predicates such as CAUSE, HAVE & FROM which are used to specify the meaning of compounds. The idea that the relationship between the constituents in a compound can be specified by a taxonomy of semantic primitives forms the basis for representing compound meaning in an important cognitive theory about conceptual combination, namely the Competition Among Relations In Nominals (CARIN) model (Gagné & Shoben, 1997).

The concept specialization approach and the taxonomic approach both assume that the meaning of a compound can be adequately captured by a simple label (either as a slot in the head concept or as a stand-alone relation category). One of the primary theses of this paper is that such a simple representation of compounds is inadequate; the relations instantiated during conceptual combination are semantically detailed entities and as such require a more complex mode of representation.

Our approach assumes that relations are as complex and as semantically non-trivial as the constituent concepts that they link are. We therefore represent relations in a way similar to how concepts have been represented in the classification literature (e.g. Nosofsky, 1984; Kruschke, 1992), using exemplars which are defined as sets of values on a set of dimensions. We generate relation exemplars using a corpus study where a large, represen-

Table 1: The dimensions used in the relation-rating task

No.	Relation	Example
1	H causes M	flu virus
2	M causes H	job tension
3	H has M	college town
4	M has H	lemon peel
5	H made of M	chocolate bar
6	H makes M	honey bee
7	H location is M	office friendships
8	H for M	plant food
9	H is M	guard dog
10	H uses M	machine translation
11	H derived from M	peanut butter
12	H about M	budget speech
13	H during M	summer clouds
14	H used by M	servant language
15	M location is H	murder town
16	H by M	student vote
17	M makes H	monastery beer
18	M is H	girl guide
19	H resembles M	mushroom cloud

tative sample of familiar compounds are rated on various semantic dimensions, derived from the taxonomy of relations used in the CARIN model (Table 1). The values on these dimensions for a compound then allow us to define the meaning of that compound as a vector or point in a high-dimensional semantic relation space.

In the remainder of the paper, we present models which use this relation space approach to representing compound meaning. The first model is a relation selection model which predicts the interpretation of a given compound (in the context of our representation system, it finds the point or vector in the relation space that best corresponds to the compound's meaning). The second model predicts reaction time or interpretation ease; given a compound and an interpretation for it, it estimates how long it would take a person to produce that interpretation for that compound. In both cases, the models use information about how the constituents of known compounds (i.e. those from our corpus study) tend to combine. To foreshadow the results, we found that both the relations generated by our approach and the predicted reaction times correspond closely to those generated by people.

Corpus Study

To test these ideas, a large, representative sample of noun-noun compounds was required. In a previous study (Keane & Costello, 1997) participants were presented with some common examples of noun-noun compounds and were instructed to generate as many noun-noun compounds as they could think of. They were instructed that the compounds should be "commonplace ones – that is, ones used in everyday life that are familiar to you." The eight participants generated a total of 680 compounds. Several steps were then taken to ensure that this sample was a representative set of familiar compounds. Very infrequent compounds (where the log of the number of hits in a Google search was more than 2 standard deviations below the average) were removed from the sample. A priming effect was evident in the generated compound lists; often a participant would generate a sequence of compounds that used the same modifier and similar relations (for example one participant generated "*train ticket*, *cinema ticket*, *bus ticket*" in sequence) or used the same head and similar relations (for example another participant generated "*rugby ball*, *soccer ball*, *tennis ball*" in sequence). Therefore, to counteract the effects of this priming, compounds were removed from the set of 680 available compounds so that each modifier word and each head word appeared only once in the items generated by each participant. Compounds were removed in a way that aimed to maximize the total number of compounds remaining. In cases where a choice between compounds did not affect this aim, the most frequently occurring compound was retained. Examples such as *electric razor*, *bob cat*, and *mini skirt* which could not reasonably be judged noun-noun compounds were also removed. After these considerations, a sample of 384 representative, familiar compounds remained.

For each of these 384 compounds, the first author rated the appropriateness (on a scale from 0 to 10) of each of the 19 relation categories for that compound sequentially. The order in which the relation categories were rated was randomized. (A new random ordering of the relation categories was used after every 15 compounds). As a test of the robustness and sensibility of these ratings, the second author rated a random sample of 15% of the compounds; inter-rater reliability was high (across all items and relation categories, $r = 0.75$, $p < 0.001$, $N = 1140$).

Results

Following previous studies (e.g. Devereux & Costello, 2005), the relation category ratings produced in this way were regarded as being histograms representing the relationships between the heads and the modifiers of compounds. Compounds which have very similar relationships instantiated between the head and the modifier (as *toilet seat*, *watch strap*, and *door handle* do, for example) tend to have very similar relation rating histograms (for example, with relatively high values on the H FOR M and H HAS M dimensions). Compounds which have very different relationships instantiated between the head and the modifier (for example, *watch strap* and *air rifle*) tend to have very different relation rating histograms. Furthermore, the relation rating histograms can be thought of as points or vectors in a 19-dimensional Euclidean space. This allows us to measure the distance (or, inversely, similarity) between pairs of compound meanings using the standard Euclidean metric.

Across the 384 items, there is a large degree of correlation between many pairs of relation categories (Table

Table 2: Relation categories with high rating correlations

Relation 1	Relation 2	r
H makes M	H causes M	0.85
M makes H	H by M	0.80
M location is H	H has M	0.76
H makes M	H derived from M	0.67
H is M	H made of M	0.63
H location is M	M has H	0.60
H is M	M is H	0.58
H for M	H derived From M	-0.58

2). For example, there is a high degree of correlation between the M MAKES H and H BY M relation categories – if one of these relations is highly appropriate for a compound, then the other one is also likely to be appropriate. There are several potential problems with such interdependencies between relations. Firstly, relation categories with a high correlation tend to be semantically similar: for example, M MAKES H tends to mean the same thing as H BY M in many contexts. If a relatively large subset of the 19 relation categories used in the corpus study tend to be semantically similar, then the result is that this aspect of relation meaning will be amplified in the relation space. Secondly, as one of the aims of this work is to describe a model that predicts appropriate relation points in the space for new compounds, the possibility that interdependencies between relations can potentially lead to misleading evidence in support of a model must be addressed. For example, if a model correctly predicts that the meaning of a particular compound is modelled by a point in the relation space that has high values on the M MAKES H and H BY M dimensions, then statistically this counts as two points of agreement for the model. However, in reality the model has only captured one aspect of the meaning of the compound; the second point of agreement is simply a result of the arbitrary choice of dimensions for the relation space.

We therefore used Principal Components Analysis to reduce the dimensionality of the relation space and to produce a set of dimensions which have no significant intercorrelations across the compound ratings in our corpus. A principal components analysis with varimax rotation was applied, which reduced the relation space to six dimensions. There were no significant correlations between these six new dimensions ($r < 0.001$, $p > 0.999$ for every pair of dimensions).

It is worth noting that these six new dimensions appear to have qualitative interpretations. For example, the first dimension can be paraphrased as M GENERATES H; *brain child*, *lamb wool*, *body heat* and *horse manure* are the compounds with the highest values on this dimension. The second dimension can be paraphrased as H CONTAINS M (*mine field*, *boat race*, *book store* and *disk drive* are the highest on this dimension). The third dimension can be paraphrased as M IS H (*boy scout*, *girl guide*, *sofa bed*, *guard dog*). Similarly, the fourth, fifth and sixth dimensions can be paraphrased as H USED BY M, H GENERATES M, and H ON THE OCCASION OF M, respectively.

Modelling Conceptual Combination

In this section, we propose models of several aspects of conceptual combination that utilize our relation space representation of relations. Firstly, we present a model of how people interpret or comprehend a compound: given a novel compound, how do people find an appropriate or sensible way of linking the modifier and head of the compound? Secondly, we present a model examining the factors that influence the speed and ease with which people interpret different compounds.

Finding the correct relation for compounds

Our model of conceptual combination presupposes that people have experiential knowledge about how concepts that they are familiar with have combined with other concepts in the past (i.e. people have knowledge about how semantic relationships have been instantiated between pairs of concepts). For example, people have experiential knowledge about the concepts *nest* and *tree* and how they tend to combine in the world by having a "located in the branches of" relationship instantiated between them. So for example, if a relationship is not known and needs to be constructed between for the concepts *cocoon* and *bush*, people might construct a relationship that is similar to the relationships they know can exist between concepts similar to *cocoon* (e.g. *nest*) and concepts similar to *bush* (e.g. *tree*).

In our model, we assume that our large corpus of familiar, commonplace noun-noun compounds is representative of how concepts combine in the world; it therefore uses the set of 384 familiar compounds and their vector-represented relations as its knowledge base. In constructing the correct relation vector for a novel compound, the model uses information about the relation vectors that exist for compounds in the database which have a head or a modifier that is similar to the head or modifier of the novel compound. More specifically, the new relation vector is constructed as the weighted average over the relation vectors already plotted in the space, where the weight parameter on each term of the average is derived from the total head and modifier similarity of the known compounds. Formally, the relation vector $\vec{r}_{m,h}$ for a novel compound "m h" is defined to be

$$\vec{r}_{m,h} = \frac{\sum_{c \in S} (sim(m, m_c) + sim(h, h_c))^{\alpha} \cdot \vec{r}_c}{\sum_{c \in S} (sim(m, m_c) + sim(h, h_c))^{\alpha}}$$

where $sim(m, m_c)$ and $sim(h, h_c)$ denote the semantic similarity between constituent concepts m & m_c and h & h_c respectively. S is the set of known compounds plotted in the relation space; the sum is taken over all compounds c represented as vectors in this space. α is a

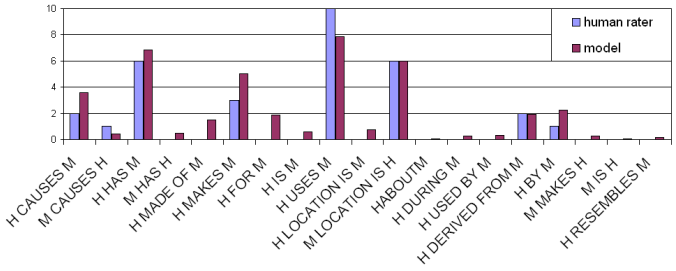

Figure 1: The model's performance for the compound *steam boat*

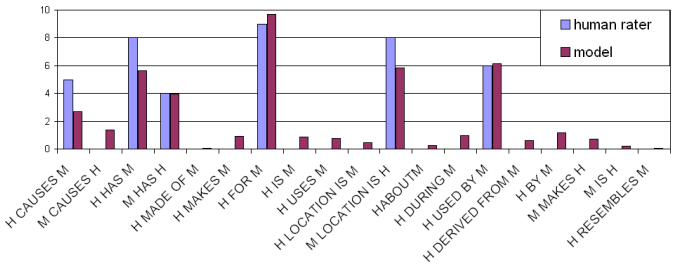

Figure 2: The model's performance for the compound *billiards room*

free parameter which describes the relative importance of constituent semantic similarity.

In estimating the similarity between concepts, we use the Intrinsic Information Content metric for semantic similarity in WordNet (Seco, Veale & Hayes, 2004). In this approach, similarity between concepts is calculated using the hierarchical structure in WordNet: the semantic similarity between nouns is taken to correspond to the ratio of the amount of intrinsic information in the two nouns' most specific common abstraction in the WordNet ontology to the amount of intrinsic information in the two nouns themselves. For example, *film* and *magazine* have similarity of 0.74 with this metric by virtue of the fact that that they both have *medium* as a superordinate category, while *newspaper* and *magazine* have similarity of 0.92 by virtue of the fact that that they both have *public press* as a superordinate category. In the latter case the similarity score is higher because *public press* is a more specific category than *medium* is.

Performance of the model To test the performance of the model, we used the model to compute new relation vectors for each of the 384 items in our experiment. In computing the relation vector for each compound, the compound in question was excluded from S, the database of known compounds already plotted in the relation space. (there were therefore 383 terms in the numerator and denominator of the model's formula). For the full 19 dimensional space, there was a high agreement between the computed vector and the human ratings (across all 384 items $\times$ 19 relation categories, $r = 0.58$, $p < 0.001$, $N = 7296$, optimal value of $\alpha = 10$). Using the 6 dimensional, PCA-generated relation space, the agreement was also respectable ($r = 0.44$, $p < 0.001$, $N = 2304$, optimal value of $\alpha = 8.5$). An example of the model's predictions for the items *steam boat* and *billiards room* are presented in Figures 1 and 2, respectively.

Ease and rapidity of interpretation

Previous studies (e.g. Gagné & Shoben, 1997; Wisniewski & Murphy, 2005) have investigated what influences the ease and rapidity with which noun-noun compounds can be interpreted by experimental participants. These experiments typically consist of participants making sense-nonsense judgements about a set of compounds (and an equal-sized set of nonsense fillers) and reaction times being recorded. The CARIN model proposed by Gagné and Shoben (1997) posits that the ease and rapidity with which a compound can be interpreted, as measured by reaction time, depends on the availability of the various thematic relations for the modifier of that compound. In particular, the CARIN model predicts that reaction times should be quickest when the compound is interpreted using a thematic relation that is often associated with that compound's modifier. For example, the CARIN theory predicts that the compound *mountain stream*, which is interpreted with the *location* thematic relation, should be interpreted easily, as the modifier *mountain* is often associated with the *location* relation (in compounds such as *mountain cabin*, *mountain trail*, *mountain goat*, etc.). Conversely, the theory predicts that *mountain magazine* will be more difficult to interpret as the *about* thematic relation is rarely associated with the modifier *mountain*.

The framework that we have proposed for representing the meaning of compounds differs considerably from the representations assumed by the CARIN model. In that model, compound meaning is represented by the simple assignment of one of 16 thematic relation categories to the compound. The approach we have proposed in this paper allows for potentially infinite variation in compound meaning (corresponding to the infinite number of points in the relation space). Furthermore, our representational framework allows for a graded measure of relation similarity: that two compounds such as *propane stove* and *gas lamp* are interpreted in very similar ways is captured in our approach by the fact that the points for these compounds are relatively close to each other in the relation space.

We use the distribution of the relation vectors in the relation space database to predict reaction time in our model. For example, in our framework, cases such as *mountain stream* should be easy to interpret not because the same thematic relation is used most often for compounds with the modifier *mountain*, but because compounds with modifiers semantically similar to *mountain* tend to have similar relations (that is, relations that are close in the space). Cases such as *mountain magazine* should be difficult to interpret not because the same thematic relation is used very infrequently for compounds with *mountain*, but because compounds with modifiers with high semantic similarity to *mountain* tend to be

interpreted with relations that are very different (i.e. in a different region of the space) compared to the relation used to interpret *mountain magazine*. Following the evidence (e.g. Gagné & Shoben, 1997) that it is the modifier and not the head that primarily determines reaction time, we can hypothesize that the compound interpretation process begins with the retrieval of known compounds which have modifiers that are the same or very similar to the modifier of the target compound. Following our methodology for modelling relation selection, we can define a weighted-average of these retrieved modifier-similar vectors using the formula

$$\vec{r}_{msim} = \frac{\sum_{c \in S} sim(m, m_c)^\alpha \cdot \vec{r_c}}{\sum_{c \in S} sim(m, m_c)^\alpha}$$

where, as before, α is a free parameter that scales the relative importance of concept similarity.

In Gagné and Shoben's approach, the correct relation for a compound is known to the reaction time model. This makes sense: the time taken to interpret a compound depends on how the compound is actually interpreted; reaction time is then modelled as being a measure of how untypical or implausible the correct relation is for compounds with the given modifier. We follow the same approach, and assume that the correct relation vector for a compound is available to our model of interpretation ease. In essence, or model predicts that reaction time is proportional to the likelihood that the quickly retrieved, modifier-similar compounds (calculated using the equation above) will lead to the correct interpretation.

In a sense-nonsense judgement task, participants need only minimally interpret the compounds; therefore nuances of interpretation should not effect reaction times. In our representation, that a sensible compound has *some* meaningful interpretation is represented by the fact that its relation vector has a high value on at least one dimension. Our model therefore takes the highest component in the known, correct relation vector for the compound and examines the relative strength of that component in the weighed average of the highly modifier-similar relation vectors retrieved at the beginning of the interpretation process.

Performance of the model To test this model, we used the RT data from Gagné and Shoben (1997) Experiment 1. In that experiment, 45 of the 57 items have a noun in the modifier position; these items were rated on the 19 relation dimensions by the first author, in the manner described above for the main corpus study. The weighted average vector for the modifier-similar compounds was then calculated. For each compound, the highest rated component was extracted and the value of this component on the weighted average of the modifier-similar retrieved vectors was used as the statistic of interest. We found that the correlation of this statistic improved as α increased towards infinity (though the correlation was significant for all α's above 11). As α approaches infinity, the weighted average equation becomes

$$\vec{r}_{msim} = \frac{\sum_{c \in S_m} \vec{r_c}}{|S_m|}$$

where S_m denotes the small (typically one or two elements) subset of the relation space database which have maximal modifier-similarity to m.

This version of the model yields the best success in predicting the reaction times from Gagné & Shoben (1997), Experiment 1 ($r = -0.40$, $p < 0.01$, $N = 45$). If the predictor statistic is taken to be the *rank* of the highest component in the correct vector amongst the components of the modifier-similar average vector, then the correlation improves ($r = 0.47$, $p < 0.01$, $N = 45$), which is comparable to the CARIN model's performance over the same items ($r = 0.47$). Using the rank rather than the absolute strength can be motivated psychologically if one assumes that people use the retrieved modifier-similar relations to direct their search for plausible relations. For example, if the rank is three then this means that there are two more plausible regions in the space that need to be examined before the correct region is considered; if the rank is four there are three regions in the space that need to be considered first, and so on. In this view, retrieving the meanings of known compounds with similar modifiers at the start of the interpretation process is a kind of heuristic people employ to direct their search for plausible interpretations.

A version of the above model which uses the sum of head and modifier similarity (rather than modifier similarity alone) does not give significant correlations for any value of α. This is interesting as it supports one of the key tenets of the CARIN model: ease or rapidity of compound comprehension depends on how relations are associated with the modifier but not on how relations are associated with the head noun. The current work therefore suggests that this finding is robust with respect to both the choice of relation representation and choice of corpus database used.

Intrinsic relation complexity Both the CARIN model and our model of reaction time assume that ease or rapidity of compound interpretation depends on how relations are distributed amongst previously experienced modifiers. This approach assumes that there is no intrinsic difference between different types of relations: different types of relations are equally easy to interpret if the relation distribution patterns are the same. Our relation space representation for relations allows us to investigate intrinsic differences in relations however; we can investigate whether compounds located in some regions of the relation space take longer to comprehend than others. To investigate this possibility, we calculated correlations between RT and the values on each of the 19 dimensions across the 45 items. The two lowest correlations with RT were for the M LOCATION IS H and H HAS M dimensions ($r = -0.31$ & $r = -0.20$, respectively). That compounds with high values on these dimensions are easy to interpret seems sensible; these relations are

indeed intuitively simple. The two highest correlations were for the H CAUSES M and H MAKES M dimensions ($r = 0.20$ & $r = 0.25$, respectively). Again, that compounds with high values on these dimensions tend to be difficult to interpret is sensible; these types of relations may require more causal links or the construction of more complex scenarios in order to be interpreted fully. We therefore hypothesize that intrinsic information about the meaning and complexity of relations, as well as experience about how relations are instantiated for already-known compounds, may be important information to the question of conceptual combination. In any case, these correlations between dimension scores and reaction times offer further evidence that relations have an intrinsic complexity that cannot be captured by simple slot-value or taxonomy views of relation meaning.

Conclusion

In this paper, we have presented a new methodology for representing the relations instantiated between the two concepts of a noun-noun compound. Hypothesizing that these relations have a complex internal semantic structure, we represent them in a manner analogous to how concepts themselves are represented in the classification learning literature. Following these approaches, we represent compound meanings as exemplars in a high dimensional semantic relation space.

In our model of compound interpretation, we have shown how knowledge of known, familiar compounds influence the interpretation of new compounds. Such an approach is intuitively sensible; an interpretation of a compound will only be plausible if it is consistent with how concepts are known to combine in the world (which is knowledge inherent in our database of known compound interpretations). The value of α is relatively low (8.5 for the 6-dimensional relation space) in this model therefore, as plausibility is a function of total world knowledge. Our model of ease and rapidity of interpretation assumes that people begin the process of interpreting a compound by first retrieving a relatively small set of known compounds which have modifiers with very high similarity to the modifier of the target compound. The value of α is relatively high in this model (and optimal at infinity) because retrieval begins with the most similar modifiers possible. It is important to point out that there is no contradiction in these two approaches. Our models suggest that people retrieve highly modifier-similar compounds as a kind of heuristic employed to find the region of the relation space that is most likely to contain the correct, most plausible relation for the compound. However, our model of compound interpretation suggests that the most plausible meaning of a compound is determined not by the contribution of the modifier alone but by modifier and head together; it is therefore possible that the initial modifier-similar retrieved relations do not lead to a plausible interpretation and in such cases the search for a plausible interpretation must continue by retrieving more, less similar, modifiers. The CARIN model, on the other hand, has no mechanism for modelling how the actual interpretation of a compound is found; for example, it predicts correctly that the reaction time for *mountain magazine* should be long (because the most likely relation is incorrect), but does not explain how the correct, unlikely *about* relation is actually reached. However, the CARIN model is not inconsistent with the principle that both head and modifier may be important in judging the plausibility of a candidate interpretation. In future work, we hope to develop a more detailed computational model of conceptual combination that searches for plausible interpretations of compounds using the similar modifier retrieval heuristic that we have posited in this paper.

Acknowledgements

We sincerely thank Christina Gagné for making available the data from Gagné and Shoben (1997), Experiment 1. We thank Georgios Tagalakis for a SWI-Prolog implementation of Seco et al.'s metric. This research was funded by a grant from the Irish Research Council for Science, Engineering and Technology to the first author.

References

Devereux, B., & Costello, F. (2005) Propane Stoves and Gas Lamps: How the Concept Hierarchy Influences the Interpretation of Noun-Noun Compounds *Proceedings of the 27th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23(1), 71–87.

Keane, M.T. & Costello, F. J. (1997). Where do "Soccer Moms" Come From?: Cognitive Constraints on Noun-Noun Compounding in English. *Proceedings of MIND II.* Dublin: Dublin City University.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.

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

Nosofsky, R. (1984). Choice, Similarity, and the Context Theory of Classification *Psychological Review*, 85(3), 207-238.

Seco, N., Veale, T. & Hayes, J. (2004). An Intrinsic Information Content Metric for Semantic Similarity in WordNet. In the *Proceedings of ECAI'2004, the 16th European Conference on Artificial Intelligence.* Valencia, Spain.

Smith, E., Osherson, D., Rips, L., and Keane, M. (1988) Combining prototypes: A selective modification model. *Cognitive Science*, 12, 485-527.

Wisniewski, E. J., & Murphy, G. L. (2005). Frequency of relation type as a determinant of conceptual combination: A reanalysis. *Journal of Experimental Psychology Learning, Memory, & Cognition*, 31, 169-174.