Propane Stoves and Gas Lamps: How the Concept Hierarchy Influences the Interpretation of Noun-Noun Compounds

Barry Devereux and Fintan Costello ({Barry.Devereux, Fintan.Costello}@ucd.ie)
Department of Computer Science, University College Dublin
Dublin, Ireland

Abstract

When people understand noun-noun compounds such as *kitchen mirror* they generate a relationship between the two constituent nouns, combining them together in a new concept. How do people determine or construct the correct relational link between the constituents? In our first experiment, we present evidence against one approach to noun-noun compounds, namely that of specifying the meaning of a compound with a single relation from a small taxonomy of general semantic relations. We found that people often select not one but several relations in the taxonomy for each compound meaning; for example, people classify *coffee stain* (meaning "a stain caused by the spilling of coffee") as *stain* MADE OF *coffee*, *coffee* MAKES *stain*, *stain* BY *coffee*, *coffee* CAUSES *stain* and *stain* DERIVED FROM *coffee*. We also found that compounds which had similar constituent concepts tended to be classified into similar relations. Our second experiment examines the role of the constituent concepts in determining the correct relation directly: again similar constituent concepts give rise to similar interpretations. Also, the results show that the constituent concepts' position in the conceptual hierarchy tends to influence interpretation: for example the compounds *propane stove* and *gas lamp*, which have conceptually similar constituents, tend to be interpreted in very similar ways.

Keywords: Conceptual combination; noun-noun compounds; CARIN; thematic relations.

Introduction

In English and other languages, novel noun-noun compound phrases such as *kitchen mirror* and *pear bowl* are commonly used. When people encounter such compounds they interpret them quickly and easily, determining that *kitchen mirror* means a mirror located in a kitchen and *pear bowl* means a bowl that is used for holding pears. How do people do this? *Conceptual combination*, the process that people use to determine the meaning of compounds, is non-trivial: to determine the meaning of a compound such as *kitchen mirror* people must instantiate the correct relational link between the head noun (H) and modifier (M) of the compound (yielding, for example, a mirror *located in* a kitchen) and discount the potentially infinitely many relational links that do not lead to plausible interpretations (for example, a mirror *used for holding* kitchens or a mirror *covered with* kitchens). In this paper our aim is to understand the nature of the relational links and examine the factors that influence what relational link is selected for a particular compound.

One of the most successful theories of conceptual combination in recent years has been the Competition Among Relations In Nominals (CARIN) model, introduced by Gagné and Shoben (1997). One of the key assertions of this theory is that the speed and ease with which a compound is interpreted depends on how often the type of relation used in the interpretation of that compound has been used with the modifier (i.e. the first word) in previously interpreted compounds. For example, the theory predicts that the compound *mountain stream* is easier to interpret than *mountain magazine* because the H LOCATION IS M relation used with *mountain stream* (i.e. *stream* LOCATION IS *mountain*) is also used in a great many other compounds which have *mountain* as their modifier (*mountain cabin*, *mountain goat*, and so on) whilst the H ABOUT M relation (i.e. *magazine* ABOUT *mountain*) is not commonly used with other compounds which have *mountain* as their modifier.

This approach assumes that there is a small number of general relations and that the same general relation can be used with a great many compounds. Gagné and Shoben therefore adopt a taxonomic approach to representing relations, defining a set of semantic primitives called *thematic relations*, following linguistic theories of compounds (e.g. Levi, 1978). The meaning of a compound is then specified simply by a classification of that compound into one of these 16 thematic relations (see Table 1).

The first aim of this paper is to investigate this taxonomic representation of relations. We ask whether the specific, concrete relations[1] used in the interpretation of compounds fall naturally into broad semantic relation categories. A finding that the specific relations used in the interpretations of compounds do not fall by similarity into distinct groupings would be problematic for taxonomic approaches as it would suggest relations have an internal structure that interacts in a non-trivial way with the two constituent concepts.

[1]Throughout this paper, we use the word *relation* in two contexts. A *thematic relation* is a broad semantic category that compounds can be classified as belonging to. A *specific relation* is the actual, concrete, fully-elaborated relationship between the two concepts in a particular interpretation of a noun-noun compound.

Table 1: The taxonomy used in the CARIN model

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	canine companion
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

The second aim of the paper is to investigate the influence of the constituent concepts on how people construct the correct specific relation for a given compound. In particular, we aimed to investigate how the head and modifier's position in the conceptual hierarchy might lead to particular types of relations being selected for the compound. For example, compounds that have a type of artifact as their head noun (e.g. *juvenile underwear*) may tend to be interpreted with a FOR type relation, whereas compounds that have a type of substance as their modifier (e.g. *chocolate rabbit*) may tend to be interpreted with a MADE OF type relation. This question has implications for the CARIN model of conceptual combination, as the key prediction of the model (which is supported by the empirical evidence) is that it is the modifier and not the head noun that determines the speed with which compounds are interpreted. Evidence that the modifier tends to be more influential than the head in determining how a compound is interpreted would support this asymmetry between the roles of the head and the modifier. However, evidence that the head noun is as influential as the modifier in determining an interpretation would suggest that the situation is more complicated, with the modifier alone being the important factor in determining the *ease* of a particular interpretation but with both the head noun and the modifier being important in determining *what* that interpretation should actually be.

The two experiments described here use the same experimental procedure: participants were presented with a compound and an interpretation of it and were asked to select which thematic relations in the taxonomy they felt were appropriate paraphrases for the interpretation. In the context of such a relation selection task, it is important to make the distinction between classifying compounds into categories in the taxonomy and classifying the interpretations of compounds into categories in the taxonomy. Compounds can potentially be ambiguous; a compound can have two or more distinct meanings that fall into different taxonomic categories. (For example, the compound *cat rash* may be interpreted as either "a rash found on a cat" or "a rash caused by cats"). Previous research (e.g. Gagné & Shoben 2002) has assumed that each interpretation is specified by a particular relation; only in the case when a compound is ambiguous can people select more than one relation for it. What is of interest to us is what thematic relations are selected for a single interpretation of a compound. In our experiments, therefore, to avoid the possibility that participants select more than one relation for a compound only because the compound has more than one possible meaning, the participants are presented on each trial with both a compound and a single possible interpretation of it and are asked to make their relation selections on the interpretation of the compound.

Experiment 1

If a taxonomy is a true description of how relations are organized in the world, then compounds should tend to fall into one and only one taxonomic category, namely the one that forms the basis of the interpretation. However, if a taxonomy is an insufficient representation for relations, we would expect to find that people classify the interpretations of compounds into various different relation categories, that there is overlap between the relations selected for certain types of compounds, and that there is an interaction between what the constituent concepts are and what the chosen relations is, all of which would indicate that relations have a complex representational structure that interacts in a non-trivial way with the structure of the constituent concepts. Though our interest is in the validity of taxonomic approaches in general, we focused on the taxonomy of 16 thematic relations used with the CARIN model. To this taxonomy we added the relations M IS H and M MAKES H (plausible relations that were not included in that taxonomy). The primary question our experiment is designed to address is whether there is simply one correct relation for each compound meaning, or rather would the participants select more than one relation as being appropriate for a given compound interpretation.

Method

Materials Both authors generated interpretations for each of the 300 compounds presented in the appendices of Gagné (2001), yielding a total of 600 interpretations. Compounds that are very common (such as "air power") and therefore possibly lexicalized were removed from this potential set of materials. Compounds for which we could not find or deduce the correct thematic relations were also removed[2]. The materials were also subject

[2]Indeed, our experiment was in part motivated by our inability to confidently deduce the correct thematic relation for all 300 compounds. For example, it is difficult to decide whether *vapour humidifier* is properly classified as *humidifier* FOR *vapour* or *humidifier* CAUSES *vapour*.

to the constraint that, though each compound had two interpretations (one from each author), each compound could only appear once in the materials. The 60 experimental materials were selected from the remaining set of 382 compound and interpretation pairs. There were three factors in the experiment, which we omit here for brevity. Each participant saw half of the experimental items.

Participants Participants were 34 students of University College Dublin.

Procedure The 30 items were presented to participants sequentially on a Web page. For each experimental item, the Web page displayed a compound (for example *dog bed*) and an interpretation of that compound (for example "a bed where a dog sleeps"), followed by a scrollable list of the 18 relations (presented in the format *bed* ABOUT *dog*, *bed* FOR *dog*, and so on) which participants could move through by clicking the provided buttons. Participants could select relations by checking the checkbox that accompanied each relation. At the top of the page participants were instructed to select any of the 18 relations they felt were an appropriate paraphrase of the interpretation. They were instructed to select only one relation if they felt only one relation was appropriate or to select more than one relation if they felt that more than one relation was appropriate. The order of the 18 relations and of the 30 experimental were randomized for each participant. Each trial was self-paced.

Results

Across all 34 subjects and all 60 compound-interpretation items, a total of 3,296 relations were selected as being appropriate paraphrases, yielding an average of 3.23 relations being selected for each item by each participant. In only 288 (28.2%) of the 1,020 trials is just one relation selected, as assumed by the taxonomic approach. Figure 1 presents a histogram detailing the distribution of relation selections over all subjects for *coffee stain* = "a stain caused by the spilling of coffee" (classified as M CAUSES H in Gagne, 2001). Following the taxonomic approach, which assumes that one relation in the taxonomy is used to interpret a compound, we would expect participants to select one and only one relation for each compound meaning. Instead, we see a spread of relation selections over several different categories: for example, *stain* MADE OF *coffee*, *coffee* MAKES *stain*, *stain* BY *coffee*, *coffee* CAUSES *stain* and *stain* DERIVED FROM *coffee* are consistently selected by participants.

Why does more than one relation get selected? A possible explanation for why many different relations are being selected might be that the taxonomy is too large or possesses too much redundancy. For example, we might hypothesize that the H BY M and M CAUSES H relation categories are really two versions of the same relation: a *law* is BY a *government* if and only if the *government* CAUSES the *law*, and therefore both of these relations, rather than just one of them, tend to be selected for particular items. There are indeed correlations between

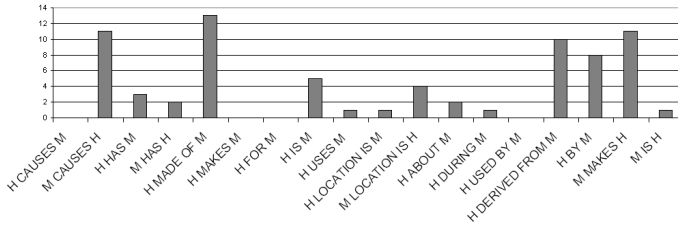

Figure 1: Distribution of relation selections for *coffee stain* = "a stain caused by the spilling of coffee"

pairs of response relations (e.g. for M CAUSES H and M BY H, $r = 0.688$; for M CAUSES H and H DERIVED FROM M, $r = 0.633$; all p's < 0.01) that would support this hypothesis. However, if two or more relations in the taxonomy really had the same or similar meanings, we would expect these relations to co-occur consistently across items: that is, for every compound meaning people should not select one of the relations without also selecting the other. This is not observed in many instances. For the interpretation of *job anxiety*, for example, people select the M CAUSES H relation but not the H BY M relation, in spite of the fact that there is a high correlation between these two relations. If M CAUSES H and H BY B mean the same or nearly the same thing, then it would be as natural or nearly as natural to say "anxiety that is by a job" as it is to say "anxiety that is caused by a job". Our correlation data support the hypothesis that the thematic relations intersect, in the sense that sometimes many relations are valid for a single compound meaning. But our data also suggest that all the relations in the taxonomy are distinct in the sense that it is not possible to replace one relation by another or merge two relations into one in a way that is meaningful for all compound interpretations. Both conclusions are problematic for the taxonomic approach. Furthermore, because we present people with a single concrete interpretation for a compound as well as the compound itself, it cannot be the possibility that the vagueness, ambiguity or the poorly-defined nature of compounds is resulting in more than one relation being selected; rather, it must be the result of the vagueness, ambiguity or the poorly defined nature of the thematic relations.

So, given these results, what are we to make of the relational link between the two constituent concepts of a compound? A possible shortcoming of the taxonomic view of relations is that there is no real interaction between the relations in the taxonomy and the semantic content of the concepts that they link. The CARIN model, for example, uses only distributional information about how frequently different thematic relations occur with the modifier concepts of previously encountered compounds. We wished to address how the meaning of the constituent concepts influenced our participant's relation selections. In particular, we were interested in whether our participants made similar relation selections for compound interpretations which had sim-

ilar constituent concepts. For example, for items with compounds such as *coil lock* and *electric brake*, which have similar head concepts (*lock* and *break*), would participants make similar relation selection judgements (and therefore produce similar relation frequency distributions)?

To test this hypothesis we required both a measure of concept similarity and a measure of the similarity of relation selection distributions for items. To measure how similar or different the constituent words used in our compounds were, we applied the Intrinsic Information Content (IIC) metric for semantic similarity in WordNet (Seco, Veale & Hayes, 2004). This metric calculates semantic similarity between nouns using the hierarchical structure in WordNet: the similarity between nouns is taken to correspond to the amount of intrinsic information in the two nouns' most specific common abstraction; that is, their relative closeness in the conceptual hierarchy. Hence, *film* and *magazine* have similarity of 0.74 with this metric by virtue of the fact that that they both have the abstract concept MEDIUM as an abstraction.

For measuring the similarity of participants' relation selection distributions for items in our experiment, we regarded frequency distributions (like those in Figure 1) as points in an 18-dimensional space, where the 18 thematic relations correspond to the space's dimensions. Applying the standard Euclidean metric gives us a measure of all the pairwise distances between experimental items in this space. We applied correspondence analysis to produce a 2-dimensional projection of this space. This graph is intuitively sensible: items with similar frequency distributions (and which therefore use similar specific relations) tend to be close together, dissimilar ones do not (we omit the graph for brevity; a similar graph for different materials is presented in Figure 2). For example, the minimum distance between two distinct points in this space is 4.0, between the interpretations of *propane stove* = "a stove that uses propane" and *gas lamp* = "a lamp that uses gas as fuel" (for which the thematic relations H HAS M and H USES M tended to be selected), indicating that participants produced similar histograms of relation selections for these items. The distance between the points for these compound interpretations is small because the specific relations used to link the constituent concepts in the interpretations of these compounds are very similar.

The metric for relational distance described above was transformed into a measure of relation similarity by the transformation $1/e^{d(r_1,r_2)}$, where $d(r_1, r_2)$ is the distance between items r_1 and r_2 in the relation space. We computed all distinct pairwise relation similarities for the 60 items in our experiment. We also computed the pairwise similarity scores for the head concepts and the modifier concepts using the IIC metric. The more similar the head concepts in a pair of compounds were, the more similar the selected relation distributions were for those compounds (Spearman's $\rho = 0.214, n = 1770, p < 0.001$)[3]. How similar the modifier concepts were also had a significant, though weaker, influence on how similar the relations used in the corresponding compound interpretations were (Spearman's $\rho = 0.057, n = 1770, p = 0.017$). That the influence of the head noun is as strong as the influence of the modifier is surprising in light of previous evidence suggesting the importance of the modifier alone in determining the speed with which an interpretation can be produced. Therefore, the influence of the head noun and the modifier were investigated further in a second experiment.

[3]Spearman's ρ was used as the data are not normally distributed. Qualitatively similar results are obtained using Pearson's r.

Experiment 2

The aim of the second experiment is to examine in closer detail how the correct relationship between the two constituent concepts is formed, and investigate the possible factors that influence this process. In particular, we aim to separate out the influence of the head and the influence of the modifier and examine the role both these factors play in determining the correct specific relation. For example, the compound *summer rain* might be interpreted using a DURING type of relation because *summer* is an interval of time and this property might guide the interpretation process towards that kind of relation. Alternately, perhaps the head concept influences the type of interpretation constructed. For example, perhaps *party girl* is interpreted using a LIKES type of relation because *girl* is a person, and this property might guide the interpretation process towards a LIKES relation.

Many theoretical accounts of conceptual combination have examined the role of the properties of the two constituent concepts in a noun-noun compound in the interpretation process (e.g. Costello & Keane, 2001). These approaches typically take an explicit representational framework for a concept's properties and explicitly enumerate a set of attributes for the concept. We take a different approach: we examine how a concept's position in a conceptual hierarchy such as WordNet influences compound interpretation. We were interested in whether compounds with similar constituent concepts (i.e. concepts located close together in the hierarchy) would tend to have similar relations selected by participants (indicating that similar specific relations are used in their interpretations). We chose three groups of modifiers and three groups of head nouns from the hierarchy, and head group and modifier group were used as the two (three level) experimental factors of our experiment. The experiment is designed to examine whether either the head or modifier is dominant in influencing participants' relation selections.

Method

Materials Again, our materials were selected from the 300 compounds in the appendices of Gagné (2001) and their associated interpretations. Compounds that used an adjective in the modifier position (such as *underdeveloped gland*) were removed from the potential set of materials. We first performed a Ward's Method cluster analysis on both the set of head nouns and the set of

modifiers, using the semantic similarity metric, obtaining 18 clusters for each. These clusters were then used as the basis for our common abstraction groupings. For example, the modifiers *peasant*, *doctor*, *employee* and so on were grouped together in the same cluster; therefore the common abstraction PERSON from the WordNet hierarchy was used as one of the modifier groupings.

Each compound in the set of available experimental stimuli belonged to one modifier grouping and one head noun grouping. As it was impossible to use all 18 levels of both modifier grouping and head grouping in a factorial design, we chose three of the most populous groupings for both the modifier and the head as these groupings allowed us to maximize the number of experimental items used in the experiment. The three modifier groups corresponded to the abstractions ACT (e.g. *murder*), PERSON (e.g. *infant*) and SUBSTANCE (e.g. *grain*) in WordNet. The three head groups corresponded to the abstractions ACT (e.g. *attempt*), ARTIFACT (e.g. *pill*) and RELATION (e.g. *law*) in WordNet. The noun-noun compounds used as our experimental stimulus fell into both one of the modifier groupings and one of the head groupings (e.g. *murder attempt*, *infant pill* and *grain law*). The use of these particular three modifier abstractions and these particular three head abstractions allow us to select three experimental items for each modifier grouping × head noun grouping cell, producing a 3×3 factorial design with a total of 27 items. This design maximizes the number of items in the experiment: no other symmetrical factorial design could be constructed from the available materials that used more than 27 items.

Participants Participants were 44 current undergraduate Computer Science students of University College Dublin.

Procedure Each participant saw all 27 experimental items, presented in the manner described for Experiment 1. The original set of 16 thematic relations used in the CARIN model were used as the relations participants could select.

Results

Across all 44 subjects and 27 items, 3,035 relations were selected as being appropriate paraphrases, yielding an average of 2.55 relations being selected for each item by each participant. In 40.9% of trials exactly one relation is selected. (The primary reason this is higher than for Experiment 1 is that six of the 44 participants always selected no more than one relation.) As was the case for the first experiment participants do not consistently select one appropriate relation for each compound interpretation, as is implied in the methodology of the CARIN model. Again, there were examples of significant correlations between pairs of response relations (e.g. for M LOCATED H and H HAS M, $r = 0.862$; for H CAUSES M and H HAS M, $r = 0.846$). The experiment therefore corroborates the evidence from the first experiment suggesting that a taxonomic approach to representing relations is insufficient.

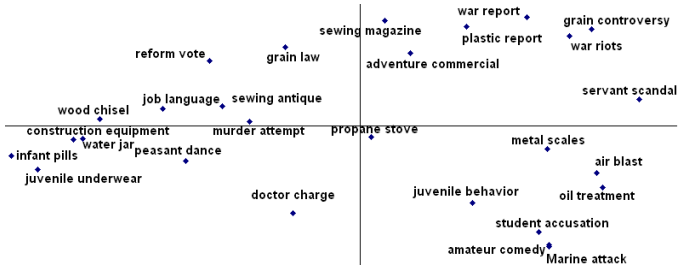

Figure 2: Correspondence analysis graph of the relation space.

The primary question investigated by the experiment is whether the specific relation generated between the modifier and the head is influenced by the abstractions that the head and modifier belong to. Also, the experiment is designed to address whether there is any difference between how much influence is exerted by the head and how much influence is exerted by the modifier. Following the methodology described in the results of Experiment 1, we used the frequency with which relations were selected for items to plot the specific relation used with every compound interpretation as a point in a 16 dimensional space. We applied correspondence analysis to explore this relation space (Figure 2); again the 2-dimensional graph was intuitively sensible: items using similar specific relations fall very close together in the graph, whereas items using dissimilar specific relations do not. For example, the two points that are closest in the 16-dimensional space are *sewing magazine* = "a magazine about sewing" and *plastic report* = "a report about the plastic industry", which are both also close to *war report* and *adventure commercial*, and other items that use similar specific relations in their interpretations. Other groupings by specific relation similarity can also be discerned in the graph.

As before, the pairwise similarity scores for compound interpretations were calculated, as were the pairwise similarities between all modifiers and all head nouns. As in the first experiment, how similar the head concepts were had a significant influence on how similar the specific relations used in the corresponding compound interpretations were (Spearman's $\rho = 0.133, n = 351, p = 0.013$). How similar the modifier concepts were also had a significant, though slightly weaker, influence (Spearman's $\rho = 0.119, n = 351, p = 0.026$). The results therefore corroborate the findings of the first experiment, and again suggest that both the abstract features of the head noun and of the modifier are important in determining what specific relation is instantiated for compounds.

For further analysis, we found the number of times the most frequently selected thematic relation was selected for compound interpretations in each head category and the number of times the most frequently selected thematic relation was selected in each modifier category. The statistic of interest was whether the most frequently selected thematic relations for the three head categories were selected more frequently than the most frequently

selected thematic relations for the three modifier categories. For each subject, the average number of times the most frequently selected relations were selected for the head categories and the average number of times the most frequently selected relations were selected for the modifier categories was calculated. Again, the results suggest that the head is of slightly greater influence than the modifier (averaging across subjects as well as categories, $\mu_H = 3.713$, $\mu_M = 3.507$), but again the difference is not significant ($t(43) = 1.648$, $p = 0.107$).

Conclusion

In the experiments we investigated conceptual combination using a task where people were shown an interpretation for compounds (like *pumpkin sauce* = "a sauce which contains pumpkin") and were asked to indicate which relation or relations they thought were appropriate paraphrases for those interpretations. In the taxonomic approach to conceptual combination, each interpretation for a given compound is the result of selecting a single general taxonomic relation to use in interpreting that compound. The taxonomic approach thus expects that every compound interpretation belongs in one particular taxonomic relation category (the relation that was used to generate that interpretation). Our results, however, suggest that this is not the case: participants in our experiment typically felt that the interpretations we showed them belonged in more than one taxonomic category. For example, most participants felt that *job anxiety* = "anxiety about losing a job" was an example of the relations H ABOUT M, H DERIVED FROM M and M CAUSES H: most participants placed this interpretation into those three different taxonomic categories.

In light of these findings, the taxonomic theory does not appear to be correct. So how do we explain people's interpretation of noun-noun compounds? One possibility would be to retain the set of relations used in taxonomic theories of compounding, but allow that most interpretations are produced by a combination of multiple relations. In this approach, relations would not be seen as taxonomic categories into which compounds must be placed, but rather as atoms which could be used to construct interpretations for compound phrases.

This is the approach we took: we represented the specific relations used in compound interpretations as points in a relation space. Our results suggest that such a representation for relations is meaningful in the sense that similar specific relations are grouped close together in the relation space and dissimilar specific relations are not. We have shown that similar relations in this space tend to have similar heads and similar modifiers. This result is important because it suggests that a compound's constituent words play a significant role in determining which specific relation is constructed for that compound (more specifically, the position of a compound's words in WordNet's lexical hierarchy seems to be important in relation selection). Furthermore, we have shown that the head in a compound has at least as strong an influence on relational similarity as the modifier. This is interesting because it contrasts with (but does not contradict) previous empirical results (e.g. Gagné & Shoben, 1997) showing that the modifier in a compound has a stronger influence than the head in determining the speed with which people interpret compounds.

It may be possible to explain this apparent divergence between speed of interpretation (influenced only by modifier relation distribution and not by head relation distribution) and the selected relations (influenced by both head and modifier) by proposing a 'generate-and-test' model of conceptual combination. In this 'generate-and-test' approach people would interpret a compound by first using the modifier to select a set of candidate relations for that compound (explaining the modifier's influence on speed of interpretation) and then using both concepts to select the correct relation from that set of candidates (explaining why compounds with similar heads tend to use similar relations). Indeed, Gagné and Shoben (2002) discuss the possibility that the head noun is primarily used to assess the plausibility of the relation suggested by the modifier. The modifier may act as a heuristic identifying which candidate relations should be considered first, whilst both the head and modifier's position in the conceptual hierarchy may be used to evaluate which of these candidate relations are acceptable or plausible. Such a theory is attractive as it would unify the relation-focused approach of the CARIN model with the concept-focused approaches of the other main theories of conceptual combination.

Acknowledgements

This research is funded by a grant from the Irish Research Council for Science, Engineering and Technology. We also thank Georgios Tagalakis for providing a Prolog implementation of Seco *et al*'s semantic similarity metric.

References

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

Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory & Cognition, 27(1)*, 236–254.

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.

Gagné, C. L., & Shoben, E. J. (2002). Priming relations in ambiguous noun-noun combinations. *Memory & Cognition, 30(4)*, 637–646.

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

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.