A Computational Model of Conceptual Combination

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

Rebecca Maguire (rebecca.maguire@dbs.ie)
Department of Psychology, Dublin Business School
34/35 South William Street, Dublin 2, Ireland

Arthur W.S. Cater (arthur.cater@ucd.ie)
School of Computer Science and Informatics,
University College Dublin, Dublin 4, Ireland

Abstract

We describe the Interactional-Constraint (ICON) model of conceptual combination. This model is based on the idea that combinations are interpreted by incrementally constraining the range of interpretation according to the interacting influence of both constituent nouns. ICON consists of a series of discrete stages, combining data from the British National Corpus, the WordNet lexicon and the Web to predict the dominant interpretation of a combination and a range of factors relating to ease of interpretation. One of the major advantages of the model is that it does not require a tailored knowledge base, thus broadening its scope and utility. We evaluate ICON's reliability and find that it is accurate in predicting word senses and relations for a wide variety of combinations. However, its ability to predict ease of interpretation is poor. The implications for models of conceptual combination are discussed.

Keywords: Conceptual combination; noun-noun compounds; paraphrase frequencies; WordNet; language comprehension.

Introduction

People using language to communicate often need to identify concepts for which there is no simple or suitable one-word expression. In such cases, a combination formed from two nouns will frequently suffice, allowing the speaker to succinctly describe a complex concept in a way that can be reliably deciphered (e.g. *kitchen sink*, *car magazine*). In English, a language in which compounding is particularly productive, combinations consist of a modifier followed by a head noun. Usually, the head noun denotes the main category while the modifier implies a relevant subcategory or a modification of that set's typical members. In this way, a *penguin film* is interpreted as a particular type of film, and more precisely as one that is about penguins.

Although conceptual combination has been the focus of much research, modelling the interpretation process has met with limited success to date. The various psychological theories of the phenomenon that have been proposed have tended to suffer from a lack of specificity regarding how commonsense knowledge is filtered, activated and applied (e.g. the Concept Specialization model, Murphy, 1988; the Dual-Process Theory, Wisniewski, 1997). In addition, the accuracy of computational models has been limited by the extent of the conceptual knowledge required to generate appropriate interpretations.

Outline of Theory

ICON is based on the findings of a series of studies investigating the cognitive processes involved in interpreting conceptual combinations (e.g. Maguire, Maguire & Cater, 2007; Maguire, Maguire & Cater, 2008). These studies have suggested that the influence of both noun constituents is an interactional one and that the range of interpretation is incrementally constrained until an appropriate interpretation is identified, at which point a modality-specific representation is instantiated. Maguire et al. (2007) proposed that conceptual knowledge is activated dynamically rather than 'all-at-once' and that concepts are dynamic and context-sensitive as opposed to being associated with a fixed set of features. For example, in the case of *plastic knife*, there is no need to activate the image of the canonical metal knife prior to combination. The knowledge that *plastic* is a substance and *knife* is an object is activated first and this is sufficient for indicating the <made of> relation. As a result, the conceptual content retrieved for the word *knife* remains appropriate to the context. This idea contrasts with other accounts such as Murphy's (1988) Concept Specialization model insofar as it does not require that both constituent concepts are fully activated prior to their combination. Instead, the abstract properties of the constituents are used to 'home-in' on the correct interpretation, avoiding the activation of inappropriate conceptual knowledge.

An important implication of our theory is that the contribution of a combination's constituents to the interpretation cannot be separated from each other. Instead, the interaction of noun properties must be taken into account. Our Interactional-Constraint (ICON) model views the identification of a referent as the main objective of combination interpretation and the dynamic strengthening of constraints as the process by which this is carried out. The model consists of a series of stages, each of which relates to a different component of the interpretation process. In order to facilitate the modelling process, these stages are consecutive and unidirectional.

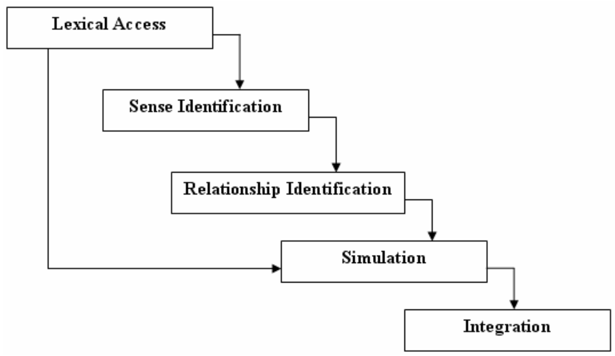

Figure 1: Stages involved in combination interpretation

While the output of our model is impoverished relative to the rich modality-specific representations that people can generate, it improves on previous models by providing both an interpretation and an estimation of the ease of that interpretation. ICON also obtains the senses of both constituents and is thus capable of interpreting combinations appearing in open text.

Identifying a Relation Taxonomy

Given that our model provides a specific relation as output, we are therefore confronted with the problem faced by previous models, namely specifying a limited range of relations. We do not maintain that people explicitly select from among a set of possible relations. Rather, they attempt to determine the referent, with a relation often emerging as an epiphenomenon of this process. Due to regularities in how entities can be related in the real world, many of the relationships between combinations will happen to fall into a number of discrete and coherent categories. While there will be many exceptions, the general interpretative form of many combinations can be reliably and informatively described using a limited taxonomy of relations. The generalisation of combination interpretations into a number of discrete categories is therefore a justifiable measure which can simplify the modelling process while retaining an acceptable level of informativeness.

We developed a concise taxonomy of relation categories approximating the relational gists that people form when interpreting combinations. This taxonomy was designed to balance coverage and parsimony, subsuming a significant proportion of combinations in a robust and consistent manner. In designing this taxonomy, we took into account previous efforts at categorisation (e.g. Gagné & Shoben, 1997), corpus statistics providing accurate relation frequencies (e.g. Cater & McLoughlin, 2000), and the real-world factors underlying epiphenomenal relation categories. Based on these considerations, a hierarchical taxonomy was identified as providing the most intuitive system for labelling combinations. We identified seven relation super-categories, dividing into 13 categories and 21 categories at successive depths of the hierarchy. The seven super-categories were as follows: <position>, <constitution>, <origin>, <effect>, <meronymy>, <predicative> and <topic>. According to Cater and McLoughlin's (2000) corpus study, these categories can account for 83% of compounds. A further <idiosyncratic> category was included as a catch-all for any remaining combinations. The distribution between categories was reasonably balanced, varying from a maximum of 24% for <effect> to 4% for <origin>.

Implementing the ICON Model

The stages in the ICON model are arranged as a cascade of discrete units, with the output of one stage being used as the input for the next. These stages exploit readily available sources of information with broad scope, namely WordNet, the British National Corpus (BNC) and the Google search engine. In designing the model, we sought to balance scope and utility with cognitive plausibility. However, while the latter might have been enhanced by providing a hand-crafted knowledge base for a limited set of concepts, this would have severely comprised the model's scope.

Stage 1: Lexical Access

Lexicalised phrases are more likely to be idiosyncratic than other combinations as the constituents do not need to be related in order for the referent to be retrieved (e.g. *passion fruit*). The combination can have a prior agreed meaning which is not reflected by any deducible relationship between the modifier and the head. Therefore, it is important for a model of conceptual combination to be aware of the degree of lexicalisation of a phrase. The first stage of the ICON model checks whether a combination is lexicalised and to what degree. A Google search is carried out for the combination preceded by the phrase "define: ". ICON returns a measure of the availability of the lexicalised phrase based on the number of definitions returned by the Google search.

Stage 2: Sense Identification

The second stage of the ICON model aims to simulate the processes by which people derive a semantic gist for the constituent nouns prior to integration. In the case of a combination, the opposite constituent represents the strongest constraint on interpretation. Accordingly, the input for the second stage of ICON is a pair of nouns, and the output is a pair of senses representing their semantic gists.

ICON uses the WordNet lexical database to assign word senses. Although WordNet information might not be adequate for assigning relations to combinations (cf. Cater and McLoughlin, 2000), it is better suited to discriminating between word senses, since dissimilar senses tend to combine with dissimilar sets of words (e.g. *bat cave*, *bat handle*). Senses in WordNet are numbered, generally according to frequency. This means that the sense number of a word has no semantic significance. For example, the artefact sense for *bat* is 3 while the artefact sense for *racket* is 4. These sense numbers have no significance outside the synset to which they relate and cannot be generalised in any way. Our solution to this problem is to generate all the possible sense permutations for a noun-noun compound and then compare these possibilities with a training set in order to ascertain which has the most merit.

Combinations in the training set were obtained from two sources. From the BNC we selected a random set of 300 combinations with between 10 and 100 occurrences. This frequency criterion was observed in order to ensure that combinations would be relatively familiar (and thus context-independent) but not to the point of being lexicalised. In order to compensate for the abstractness of these terms, we also included a selection of 100 non-lexicalised, non-idiosyncratic participant-generated combinations. An independent judge selected the most appropriate sense for each of the nouns in the 400 combinations. The sense selection algorithm follows the same principles as Kim and Baldwin's (2005) model for relation selection. The modifier and head of the input combination are compared against the 400 combinations in the training set. A similarity rating is calculated using Seco, Veale and Hayes's (2004) WordNet similarity metric, which takes into account the most specific common abstraction between two WordNet synsets. The similarity value is calculated for the first words and the second words separately and then multiplied, in recognition of the fact that the semantic significance of a combination is interactional as opposed to additive. We also included an additional component which considers the frequency and dominance of the potential word senses based on the Senseval frequencies provided in WordNet. All sense permutations are ranked according to these measures and the one with the highest overall confidence value is selected as the most likely. These senses are then passed to the third stage of the model.

Relationship Identification

The third stage of ICON reflects the idea that people's awareness of productive combinational patterns allows them to constrain their interpretation so that irrelevant features of the constituent nouns are not activated. This stage represents the initial integration of both constituents, taking into account the constraints imposed by the combinational syntax. For example, people will realise that a combination of type [substance – artefact] (e.g. *plastic chair*) is likely to involve a <made of> type relation before retrieving the precise features of the constituent concepts. ICON's third stage takes in two words and their pre-selected WordNet senses as input and outputs a relational label representing how the interaction of the gist of the constituent nouns initially constrains the overall interpretation of the phrase.

Diagnostic WordNet Patterns

WordNet contains information which is useful for identifying certain patterns of combinations. For example, the position of a concept in the lexical hierarchy can allow accurate inferences regarding animacy, concreteness, abstractness or, for instance, membership of location, time period and substance categories. Consequently, WordNet data is successful in identifying relations associated with such concepts (e.g. <located>, <during>). Machine learning techniques work optimally when the noise in the data is minimal. It is therefore important to ensure that all of the variables included in a model are diagnostic of the output. Accordingly, we sought to identify the subset of WordNet information that is most diagnostic of relation use. Maguire, Maguire and Cater (2008) demonstrated that the influence of a combination's constituents on the interpretation process is interactional: the effect of a particular modifier or head is very much dependent on the opposite constituent. In addition, Maguire, Wisniewski and Storms (2007) found that, taken together, the general categories of the modifier and head nouns are often diagnostic of a relation. Based on a thorough analysis of the 400 combinations in our training set, we identified 24 diagnostic modifier-head WordNet patterns exhibiting broad coverage (e.g. [time – event], [solid – object], [agent – object]).

Paraphrase Data

Although WordNet data is useful for predicting some relations, others involve aspects of conceptual content which are not reflected in the organisation of the hierarchy (e.g. size, shape, appearance etc.). The limited success of models based solely on hierarchical data emphasises that other sources of information are required in order to accurately model the interpretation process (Cater & McLoughlin, 2000). In light of this, ICON supplements statistical WordNet-based data with combination paraphrase frequencies harvested from the Web. The Web as a corpus has the benefit of being broad, extensive and easily available, and thus represents a very practical and useful source of information for minimally supervised linguistic models. Rather than needing to specify exactly what knowledge people are sensitive to during the interpretation process, paraphrase frequency data represents the cumulative influence of such knowledge. For example, in order to ascertain the probability of the <about> relation for *penguin film*, ICON takes into account the number of hits garnered for the paraphrase "a film about penguins".

The accuracy of the paraphrasing technique depends on using paraphrases that introduce as little noise as possible. Unfortunately, Web searches are inherently noisy. Punctuation is ignored and part of speech information is not available. While it is straightforward to generate paraphrases with a high true positive rate, it is more difficult to reduce the number of false positives. Even if a paraphrase only produces inappropriate high frequencies very occasionally, this can still impact the reliability of the data when the intended relation is itself infrequent. For example, paraphrases involving the preposition *with* provide a high number of hits for combinations involving the <has feature> relation (e.g. "table with a drawer" – 652, "clock with a pendulum" – 4,820). Intuitively, such paraphrases should provide a lower number of hits for combinations that do not involve this relation (e.g. “clock with a kitchen” – 3). However, when considering a large sample of combinations, false positives become apparent (e.g. “table with a garden” – 36,400, “clock with a metal” – 11,900). While the design of paraphrases in previous studies has been guided by the ratio of false negatives to true positives (i.e. the ability of the paraphrase to detect a relation), this ratio is of far less concern as it reflects the scope of a paraphrase, not its reliability. In contrast, even a relatively low number of false positives can reduce the informativeness of paraphrase frequencies to the level of random noise. In light of this, we subjected our candidate paraphrases to rigorous testing in order to establish their diagnosticity. This process revealed four particularly salient problems affecting the reliability of paraphrase frequencies.

Long-range Dependencies Often the nouns contained in a paraphrase are not the arguments of the relationship being described. For example, the sentence “a lack of money *caused* the family to beg” provides an inappropriate hit for the paraphrase “money *caused* the family”.

Compound Truncations Even when a paraphrase represents a genuine relationship between two noun concepts, the first and last nouns may be part of compounds which have been truncated. For example, a search for the paraphrase “college *has* a treatment” might be carried out to obtain information regarding the likelihood of the <has> relation. Unfortunately, sentences such as “college *has* a treatment room” or “college *has* a treatment facility” provide inappropriate hits.

Ambiguous Connectives Individual prepositions often suggest different relations in different circumstances. For example, the preposition *about* is strongly associated with the <topic> relation (e.g. “magazine *about* sports”). However, in some cases, the same preposition can be used to denote dispersal in a general area (e.g. “cloud *about* the mountain”).

Context-Specific Relationships Even when a paraphrase hit is genuine and error free, the relationship expressed between the two concepts might be a context-specific one which does not apply to any other context. For example, the sentence “he put the magazine *in* the car” does not indicate the existence of a particular type of magazine found in cars.

In order to mitigate the sources of noise highlighted above and improve the reliability of the data, we sought an optimal set of paraphrases through a process of trial and error. Potential paraphrase templates were identified and used to obtain frequency information. Subsequently, this was analysed for reliability and the paraphrases were refined so as to reduce the influence of noise. In order to boost the diagnosticity of our paraphrases, we included verbs in as many paraphrases as possible (e.g. *located in* and *found in*, as opposed to the preposition *in* by itself) and made use of delimiting words such as *the* and *that*. In total, we developed 14 paraphrase templates to provide accurate information on the appropriateness of the various relations in the taxonomy for a given combination.

Another problem facing paraphrase models is that of data sparseness. Paraphrases for uncommon combinations (e.g. *banana phone*, *giraffe race*) yield fewer genuine hits, thus increasing the influence of false positives. In addition, the more obvious the relationship between two concepts, the less likely it is to be explicitly paraphrased (e.g. “jar made of glass” has a hit count of 811 while “lamp made of glass” has a hit count of 3,730). In light of this, an extra generalisation component is incorporated into ICON to compensate for combinations of low frequency.

This generalisation component works by identifying common combinations in the BNC which are as similar as possible to the input combination. It assumes that combinations above a critical level of similarity are likely to use the same relation (e.g. *plastic cup*, *metal spoon*). Paraphrase frequencies are then obtained for the similar BNC examples, augmenting the data set and mitigating the effect of noise. In filtering the BNC’s combinations, the generalisation component initially considers any nouns contained in the first level of WordNet hyponyms for both modifier and head and then continues to extend the depth of the search into the hyponym tree at the same rate for both constituents until all such possibilities have been considered. A match is returned whenever a BNC combination is identified which involves a modifier and head within the limits of the current search space. If the required number of examples is not found then the search space is extended to include the subtrees of the modifier and head’s direct hypernyms and finally the subtree of the grandparent hypernyms. If this still fails to yield the required examples then the search is terminated at this point. Any combinations exhibiting a greater dissimilarity are less likely to use the same relation as the original combination.

Combining the Data

The paraphrase data constitute 14 separate numerical variables, each representing the ratio between the log of the paraphrase hit count and the log of the combination hit count. This information, together with the single nominal WordNet variable, forms a 15-dimensional vector for each combination. These data are intended to represent the experiential knowledge which people use to constrain the interpretation of a combination based on the activation of generalised properties of the modifier and head.

ICON analyses the data from the various sources in order to ascertain the extent to which an input combination is indicative of a particular relation. First, a Support Vector Machine (SVM) is used to train the model. The training set involves the same 400 representative combinations used in Stage 2, each with the correct relation provided. ICON then uses this data to make predictions for each input combination and its generalized examples retrieved from the BNC. A set of relation probabilities is generated for each combination. Finally, the relation probabilities for each of the related group of combinations are averaged to yield a set of generalised probabilities. The relation with the highest average probability is then outputted as the most likely relation for the input combination. The associated probability provides a measure of the model's confidence that the chosen relation is the correct one.

Results

In order to evaluate the performance of a computational model, its output must be compared against pre-defined correct outputs as well as with human performance at the same task. Because of the range of participant-derived data available to us from previous experiments, we used Gagné and Shoben's (1997) Experiment 1 stimuli in evaluating ICON's performance. For these combinations we were able to obtain average participant response times, plausibility and familiarity judgments, subjective and objective ambiguity ratings and also a set of 16 different interpretations for each combination. Gagné and Shoben's combinations exhibit considerable variability in plausibility, lexicalisation and ambiguity (e.g. *plastic toy* versus *cooking treatment*), allowing ICON's performance to be tested for a broad range of inputs of varying difficulty.

Based on preliminary analyses, the intermediate 13-relation taxonomy was adopted as the most reliable output of Stage 3. We found that this taxonomy provided the optimal compromise between coverage and specificity, maximizing the accuracy of the model. Subsequently, WordNet data and paraphrase frequencies were obtained for the 57 combinations in the test set. The generalisation component was used to obtain five similar examples (when possible) for each of the stimuli. In total, lexicalised definitions were obtained for seven of the combinations and the dominant word senses were identified for 46 of the 57 combinations. Based on this output from Stage 2, a total of 251 similar examples were retrieved from the BNC, yielding an average of 4.4 similar examples per input combination. Paraphrase hit count ratios were obtained for these combinations and these data were added to the test set. The SVM algorithm was then applied, yielding 13 relation probability values for each combination in the training set. These were then averaged between the input combinations and their similar examples to produce 57 sets of values. Finally, the relation obtaining the highest average probability was chosen in each case.

The ambiguity of Gagné and Shoben's (1997) materials meant that in many cases there was no single correct interpretation (e.g. *college treatment*). In order to appropriately assess ICON's performance, we compared the relations selected by the model with the interpretations produced by participants in Maguire, Cater and Wisniewski's (2006) experiment. For each combination, we identified the proportion of participant interpretations that involved ICON's choice of relation. On average, the baseline dominant interpretation was used by 74.7% of participants (*SD* = 22.4%). The model's output relation was on average used by 45.8% of participants (*SD* = 39.5%). The agreement between the model's selection and the participants' selection varied from between 100% for combinations like *mountain bird*, *office plant* and *student equipment* to 0% for incorrectly interpreted combinations such as *servant language* (<topic>), *music album* (<predicative>) and *flower toy* (<for>). Some of the relations outputted by ICON were intuitively plausible but were unsupported by participant interpretations. For instance, the combination *college headache* was plausibly interpreted by ICON as using the <located> relation, yet was never interpreted in this way by the participants in Maguire et al.'s (2006) study. Of the 57 combinations, 32 were interpreted by ICON using the dominant relation, 14 using a subdominant relation and 11 involved relations that were unsupported by any of the participants' interpretations.

In general, the similar examples retrieved for the 57 test combinations were appropriate. The main sources of error in our model were therefore due to over-generalisation of the WordNet data and to the inaccuracy and sparseness of the Web paraphrase frequencies. For example, the combination *water bird* was inaccurately interpreted by ICON as using the <for> relation. The five similar examples retrieved for *water bird* included *salt fish*, *water fish*, *water weed*, *water flower* and *water snake*. Intuitively these combinations are all suggestive of the <located> relation. However, not one of them was interpreted in this way by ICON. First, the modifier *water* does not fall into any of our WordNet categories (only solids are included as substances). Second, the paraphrase hit counts for these combinations are low, since these organisms <live in> rather than being <located in> water. As a result, spurious paraphrase hits resulted in misleading probabilities (e.g. "…fish used for salt rich diet feeding studies…"). This example demonstrates how ICON struggles to identify relations that are even slightly idiosyncratic, since these are not detectable with regular WordNet patterns or standard paraphrases. The inaccuracy of our model highlights the difficulty of using a limited number of discrete variables to represent the extensive range of knowledge which people bring to bear on the interpretation process. It also reveals the pitfalls associated with adopting a rigid relation selection process as the basis for interpretation. The fact that people do not experience the same difficulty in interpreting combinations with unusual relations indicates that they do not group relations into a limited number of discrete categories.

In order to ascertain whether ICON provides any insight regarding the cognitive processes involved in combination interpretation, we correlated the output variables with the participant-derived measures relating to ease of interpretation. The confidence values outputted from Stage 2 and Stage 3 did not correlate significantly with any of the participant-derived variables. In other words, the cases which ICON found difficult to interpret did not correspond with the cases that people found more difficult to interpret. Many of Gagné and Shoben's (1997) unambiguous combinations were easily interpreted by ICON (e.g. *plastic toy*), but the model ran into difficulty with irregular interpretations (e.g. *water bird*). Because the information that people can avail of is far more extensive than that represented in our model, its failings are simply an artefact of its design and consequently are not reflected in the participant-derived variables.

We found that the overall similarity of the generalised BNC examples (a measure we term 'regularity') was significantly correlated with response time, $r = -.30$, $p < .01$ and plausibility, $r = .34$, $p < .05$. Because this regularity measure is determined by the range of combinations present in the BNC, it is unaffected by the choice of input data (i.e. diagnostic WordNet patterns and paraphrase templates). The finding that regularity is correlated with ease of interpretation supports our idea that generalised information is used to initially constrain the interpretation process and suggests that people are sensitive to the way in which general word categories tend to combine. For example, a combination like *frog tail* initially seems plausible and suggests the <is part of> relation because it conforms to a regular pattern (i.e. [animal – body part]) shared by many other combinations (e.g. *dog tail*, *frog leg* etc.).

General Discussion

A significant limitation of our model is that its fails to implement the simulation and integration stages of the interpretation process (cf. Figure 1). As a result, the interpretation that ICON produces is simply a propositional label. In reality, this is a very poor reflection of the detailed representations that combinations are intended to elicit. Most of the variability in ease of interpretation is likely to be manifested in these latter stages when a situated simulation must be instantiated (cf. Barsalou, 2003). Therefore, no matter how perfect the knowledge used in modelling the initial three stages, we would still not obtain a strong correlation with ease of interpretation.

Linguistic modelling has been slow to take into account the embodied approach. Since differences in ease of interpretation are most likely to be manifested at the stage when modality-specific information is invoked, any model which claims to accurately reflect such differences must be viewed sceptically. Current knowledge bases simply do not contain the kind of information which would allow these kinds of cognitive processes to be accounted for. The result is that much of the information that is brought to bear in interpreting a combination cannot easily be modelled computationally without resorting to a task-specific hand-tailored knowledge base. While heuristics such as paraphrase frequencies can be used to implicitly detect noun properties, this approach will inevitably fall short because it fails to appreciate the underlying cause: words evoke detailed mental representations.

Conclusion

We have provided a computational model which performs reasonably accurately in ascribing combinations to a limited taxonomy of relations. However, the performance of the model does not correspond with human performance. One of its most significant limitations is that it is based solely on word-level statistics and hence does not take into account modality-specific conceptual knowledge. In addition, ICON's performance highlights that rigid adherence to a limited relation taxonomy is unrealistic and unsatisfactory. Although this can simplify the modelling process, people do not select from among a limited set of relations, nor do they explicitly represent such relations. The extensive variety of rich interpretations that can be produced for a combination emphasises the fact that the processes involved in language interpretation are often not amenable to such simplifications. In conclusion, our research on conceptual combination has highlighted the fact that language cannot be divorced from the embodied cognitive processes which inspire it. Accordingly, the challenge for future research in this area is to investigate exactly how conceptual knowledge is represented. Only then will an accurate model of conceptual combination be possible.

References

Barsalou, L.W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*, 513-562.

Cater, A. & McLoughlin, D. (2000). Cluster strategy for choice of compound noun interpretation rules. *Proceedings of the 11th Conference on Artificial Intelligence and Cognitive Science*.

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.

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

Maguire, P., Cater, A.W.S. & Wisniewski E. J. (2006). The role of ambiguity in the comprehension of noun-noun combinations . *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*.

Maguire, P., Maguire, R., & Cater, A.W.S. (2007). In search of the frog's tail: Investigating the time course of conceptual knowledge activation. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Maguire, P., Maguire, R., & Cater, A.W.S. (2008). Factors affecting the ease of interpretation of noun-noun compounds. *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*.

Maguire, P., Wisniewski, E.J. & Storms, G. (2007). A corpus analysis of conceptual combination. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*.

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

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