Interpreting Noun-Noun Compounds with Rare Modifiers

Phil Maguire (phil.maguire@ucd.ie)
Arthur Cater (arthur.cater@ucd.ie)
Department of Computer Science, University College Dublin
Belfield, Dublin 4, Ireland

Abstract

Previous studies in conceptual combination (Gagné & Shoben, 1997; Storms & Wisniewski, in press) have shown that a noun-noun combination is easier to interpret when it instantiates a thematic relation that is frequently associated with the modifier and more difficult to interpret when it instantiates a relation that is less typically associated with the modifier. In this experiment we investigated the extent to which the effects observed in these studies can be attributed to prior experience of the modifier as opposed to other sources of information that might influence interpretation. This was achieved by examining the influence of relatively rare modifiers which, because of their rarity, were unlikely to have a substantial combinational history. Gagné and Shoben's (1997) CARIN model suggests that the influence of the modifier is due to statistical knowledge that people store about its typical usage. As a result, the model implies that the relation type frequencies of rare modifiers should exert a lesser influence due to speakers' lack of statistical knowledge about them. However, our results showed that rare modifiers behave in a similar fashion despite the fact that they might never have been encountered in a modifying capacity. We consider possible explanations for this finding including Gagné's (2002) proposal of a secondary process as well as the possibility of naturally arising differences in ease of interpretation.

Introduction

In everyday conversation, noun-noun compounds such as *garden chair* or *coffee cup* are frequently used. The existence of such conceptual combinations greatly enhances the flexibility of language as well as making communication more efficient. Upon encountering such combinations, people are able to interpret their meanings quickly and efficiently, determining that *garden chair* is a chair found in a garden and that *coffee cup* is a cup for holding coffee. This process, although intuitive, is not straightforward and requires a sophisticated combination of world knowledge and common sense. However, it remains unclear how much of the information used to relate two concepts is based on prior experience at the lexical level and how much depends on the higher conceptual level.

Gagné and Shoben (1997) demonstrated that the more frequently a relation is associated with the modifier noun of a combination, the easier it is to judge whether a combination involving that relation is sensible or not. Similar effects have also been observed in other languages including Indonesian and French (Storms & Wisniewski, in press; Maguire & Cater, 2004). This phenomenon has been encapsulated in the form of the Competition Among Relations in Nominals (CARIN) model (Gagné & Shoben, 1997). According to the model, *mountain stream* is relatively straightforward to interpret because *mountain* is frequently associated with the <*located*> relation. Conversely, *mountain magazine* is more difficult because it involves the <*about*> relation, a more unusual relation for *mountain*. CARIN seeks to explain these differences through its assumption that people use statistical knowledge about the modifier in order to interpret a combination. This knowledge comprises a distribution of the relative frequencies with which the modifier has combined using each of 16 relation types outlined by Gagné and Shoben (1997). The CARIN model therefore necessitates that people store distributions for every noun in the lexicon representing how often that noun has been associated with the various relation types in the past.

However, this idea of stored distributions does not explain how newly encountered modifiers can be interpreted, nor does it take account of the fact that certain predictions about a modifier's relation type preference can be reliably made based on its most basic features. For example, speakers will realize that because the modifier *summer* is a time period, the <*during*> relation is the one most likely to be instantiated. Likewise, modifiers denoting locations or substances will convey similar biases, whether or not they have been previously encountered as a modifier. A study by Devereux and Costello (2005) revealed that compounds that have similar modifiers tend to be interpreted using similar relations, suggesting that relational preferences can be inferred from the properties of the modifier without having to be stored.

The following experiment investigated the possibility that modifier properties might be responsible for its influence on the ease of interpretation. This was achieved by differentiating between the influence exerted by statistical knowledge about the modifier per se and alternative sources of influence.

Experiment

This experiment was designed in order to investigate the extent to which the effects observed by Gagné and Shoben (1997) can be attributed to statistical knowledge about how a modifier has combined in the past. Accordingly, we restricted the influence of experience by examining a set of relatively rare modifiers for which it could be assumed that speakers had little statistical knowledge. In doing so, we were able to examine the influence exerted by factors other than the modifier's relation type frequency distribution. Although CARIN makes no assumptions about the frequency of the modifier per se, the model is based on the assumption that people maintain statistical knowledge about the modifiers in question. Consequently, in the case of particularly rare modifiers, the model implies that there will be little or no influence.

The experiment involved two conditions, one in which rare modifiers were combined using low frequency relations and the other condition in which the same modifier was combined using a higher frequency relation. We hypothesized that a comparison of response times in the two conditions would prove to be revealing: a lack of difference would emphasize the importance of statistical knowledge whereas the opposite case would indicate the existence of alternative sources of modifier influence.

Method

Participants Thirty-seven first-year undergraduate students from University College Dublin participated in the study for partial course credit.

Materials In order to carry out the experiment we needed to find nouns whose frequency as modifiers we could determine as being relatively rare. For this purpose we required a measure of how frequently a noun was likely to occur as a modifier and we therefore consulted the British National Corpus (BNC), a tagged, annotated corpus containing over 100 million words. Using the Gsearch chart-parser (Corley et al., 2001), we were able to extract the one million noun-noun phrases contained within the corpus. By examining the number of times any given noun occurred as a modifier within this subcorpus, we were able to select materials of sufficient rarity. An upper limit of 25 occurrences was set as a selection criterion, which is a frequency of not more than once in every 4 million words. The average number of occurrences of the selected modifiers was only 8.8. To put this number into perspective, the modifier *family* occurs over 8,800 times in the BNC, while the modifier *water* occurs over 6,800 times. Moreover, the average number of occurrences of the modifiers used in Gagné and Shoben's (1997) first experiment is over 2,500, which is a frequency of once in every 39,000 words. Thus for every one of our modifiers encountered in a piece of text, we would expect to find nearly 300 of Gagné and Shoben's modifiers. The rarity of our modifiers guaranteed that their relation type frequency distributions would be significantly less developed than those of modifiers used in previous studies.

When selecting modifiers, we attempted to incorporate a broad variety of concepts. Hence we included a rare time-period, *dusk*, as well as several locations (e.g. *swamp*, *pier*, *tavern*) and a substance modifier, *fudge*. Because such modifier types usually combine easily, these particular examples proved difficult to find. Other candidates such as *ramadan* or *butane* were overly unusual and were not included because their unfamiliarity had the potential to confuse participants.

Following our selection of 19 modifiers, we needed to estimate the relation type frequency distributions of each. Previous studies have generated relation distributions using several different techniques, including arbitrary pairings (Gagné & Shoben, 1997) and corpus analysis (Maguire & Cater, 2005). Because our materials were deliberately selected to have a low BNC frequency, a corpus study was not viable. Instead we adopted a technique used by Storms and Wisniewski (in press) which involved participant generation. In a stimulus pre-test we presented 18 participants with each of the modifiers and asked them to generate three possible combinations for each. The relation distributions were derived by ascribing the combinations garnered for each modifier to one of CARIN's 16 relation categories and determining the relative proportion in each.

For the purposes of comparison we decided to follow Gagné and Shoben's (1997) technique for dichotomizing the relation type frequencies into "high" and "low" categories. The high frequency relations for any given modifier denoted those relations with the highest relative frequencies for that modifier. This group was determined by firstly identifying the highest frequency relation based on the participant-generated distributions. If that relation accounted for 60% or more of the sensible combinations for that modifier, then that one relation was the only high frequency relation. If not, the relation with the next highest frequency relation was added to the high frequency group, until the selected relations accounted for 60% or more of the sensible combinations for that modifier. All other relations were considered low frequency. Following this, we were able to generate two combinations for each modifier, one which used a high frequency relation and the other which used a low frequency relation (see Figure 1).

Although we followed Gagné and Shoben's classification technique in this regard, the vagaries of this arbitrary paradigm allow for some relatively low relation frequencies to be classed as high frequency. As a result, we ensured that the difference in relation frequencies between the high and low conditions was real and exaggerated: the average frequency for the high frequency materials was .53 whereas that for the low frequency materials was only .07.

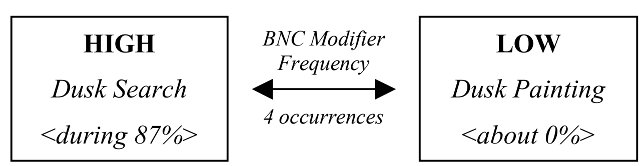

Figure 1: Example of experimental materials.

Generating combinations for the rare modifiers in the low frequency condition proved particularly challenging. The most obvious candidates for this condition usually involved heads that were extremely biased towards the instantiated relation (e.g. *seller*, *maker* etc.). For example, incorporating *tavern* into a combination that avoids the <*located*> relation is difficult, and the most obvious candidates are those involving biased heads such as *tavern owner* or *tavern cleaner*. Unfortunately such combinations could not be included as Maguire and Cater (2005) have demonstrated that biased heads facilitate interpretation: the inclusion of such materials would thus have confounded response times for the low frequency condition. In light of this, we observed the same threshold for bias as Maguire and Cater and only accepted head nouns that combined using the same relation less than 60% of the time. This level of bias was determined by a random sample of 100 combinations from the BNC.

Nineteen pairs of combined concepts were generated, as well as 38 nonsensical filler items (see Appendix). Each pair of materials was controlled for length, plausibility, familiarity and head frequency. The average number of letters in the high frequency (M = 12.1) and low frequency (M = 12.3) conditions was not reliably different, $t(18) = .38$, $p = .71$. In a stimulus pre-test two independent judges rated the plausibility of the 38 sensible materials on a scale of 1 to 5, where 5 was the most plausible. The judges were explicitly instructed only to evaluate the concept referred to by the combination and not the manner of its expression. The plausibility ratings did not differ reliably between the high frequency (M = 3.8) and low frequency (M = 3.8) conditions, $t(18) = .08$, $p = .94$.

Tagalakis and Keane (2003) demonstrated that the familiarity of combinations has a large influence on response times in sensibility judgments. Although their familiarity ratings were generated by participants, we followed Maguire and Cater's (2005) experimental technique, and assumed that the familiarity of a combination is related to the frequency of its occurrence. This frequency was then gauged by taking the number of hits generated by a Google search for that combination. Using log Google hits, the average combination frequency of the high frequency (M = 2.2) and low frequency (M = 2.2) conditions did not vary $t(18) = -.12$, $p = .91$. Finally, we also controlled for the frequency of the head by taking the log of the number of times it occurred as a head in the BNC. The average head frequency of the high frequency (M = 1.9) and low frequency (M = 2.2) conditions showed no significant difference, $t(18) = 2.11$, $p = .05$.

Design A within-participants design was used for the experimental manipulation of condition. Each participant saw the same set of 76 stimuli, comprising the high and low frequency conditions of 19 materials each and the 38 nonsensical filler items.

Procedure Participants sat in front of a computer screen and placed the index finger of their left hand on the F key of the computer keyboard and the index finger of their right hand on the J key. They were informed that a series of noun-noun compounds would be displayed on the screen for which they would have to make sensibility judgments, pressing J for sense and F for nonsense. Emphasis was placed on the fact that they should only press F if the combination was truly incomprehensible. Trials were preceded by a blank screen lasting for one second. The combination then appeared in the middle of the screen and participants had to make a decision by pressing the appropriate key.

Participants were initially asked to carry out a short practice session in which they received feedback regarding their judgments. The aim of this practice was to familiarize them with the process of making quick sensibility judgments and also to set a reliable threshold for sensibility. Without such a measure, participants would have been liable to disregard unusual but potentially sensible combinations as nonsense. After completing this practice session, participants were informed that they were now beginning the experiment. The materials were then presented in a random order to each participant.

Results and Discussion

A total of 13.4% of trials were omitted from the analysis. In 9.0% of the trials, the incorrect response was given and hence these were not considered. Additionally, responses deemed unreasonably fast (< 400ms, 0.3%) or unreasonably slow (> 4000ms, 2.8%) were also excluded. After this initial elimination process, any remaining response times which were more than three standard deviations outside each participant's mean were also excluded. This eliminated another 1.4% of responses.

The mean response times were 1,435 ms and 1,554 ms for the high frequency and low frequency conditions respectively, and the mean accuracy rates were .91 and .82. A one-way ANOVA was conducted to examine the modifier's influence on response times using both participants and items as random factors. The difference in response times between the high frequency and low frequency conditions was reliable across subjects and across items, $F_1(1,36) = 10.74$, $p < .01$; $F_2(1,18) = 6.12$, $p = .024$. The difference in accuracy rates between these two conditions was also reliable, $F_1(1,36) = 21.48$, $p < .01$; $F_2(1,18) = 11.32$, $p < .01$.

These results show significant differences between the high frequency and low frequency conditions. This indicates that both rare and common modifiers exert a similar influence on interpretation: combinations are interpreted reliably faster when the modifier is more typically associated with the instantiated relation, whether that modifier is rare or not.

Correlation and Regression Analysis We obtained correlations between response time and familiarity ($r = -.42$, $p < .01$), relation frequency ($r = -.38$, $p = .02$), plausibility ($r = -.35$, $p = .03$), modifier word frequency ($r = -.18$, $p = .27$), head frequency as a head ($r = -.083$, $p = .62$) and word length ($r = .101$, $p = .55$). The correlations between response time and familiarity, relation frequency and plausibility were significant. We then fitted a stepwise regression model using the data from the 38 experimental items and the three significant predictor variables. Both familiarity and relation frequency entered into the model and the resulting multiple correlation was .54. The standardized regression weights for familiarity and relation frequency were -.38 (p = .01) and -.33 (p = .03) respectively.

The correlations between response time and relation frequency support CARIN's premise that the relational preference of the modifier affects the ease of interpretation. However, CARIN claims that speakers are aware of these preferences through statistical knowledge based on the modifier's combinational history. The modifiers used in our experiment were much rarer than those used in previous studies and consequently would have been encountered hundreds of times less often. If prior experience were the only factor involved we would have expected the influence of the modifier to be significantly reduced. As this was not the case we conclude that, in certain cases at least, the influence of modifier preference is due to factors other than past experience of combinations in which it has been involved.

General Discussion

Although our results are consistent with the CARIN theory in that they support its claim regarding modifier influence, they are simultaneously anomalous in that they contradict CARIN's explanation for this influence. Gagné and Shoben's (1997) correlation between sensibility-judgment response times and relation strength was interpreted as a correlation between an *effect* (speedy judgment) and its *cause* (facilitated interpretation based on knowledge of the modifier's combinational history). However, in this case the modifiers were so rare that participants could not have had the necessary experiential knowledge. The effect therefore must be an effect of a different cause. We attempt to reconcile these findings with the CARIN view and consider the implications for other theories of conceptual combination.

The Experiential View

Although the lack of statistical knowledge available for the rare modifiers in our experiment would seem to rule out such knowledge as the cause of the modifiers' influence, Gagné (2002) proposed the possibility of a secondary process with the potential to account for such an anomaly. In this study relation priming was observed between modifiers that were semantically related, suggesting that speakers can be influenced by how modifiers sharing common features have combined in the past. For example, relation priming was found to occur between combinations such as *tea stain* and *coffee tension* due to the similarity of their modifiers. As a result, Gagné (2002) proposed the existence of two separate processes, one involving the relation distribution of the relevant modifier and a separate process involving knowledge about how other semantically related modifiers have combined in the past. Similarly, Tagalakis and Keane (2005) have suggested that speakers can interpret novel combinations by relying on analogous well-known examples.

Following this logic, one might propose that participants were able to infer how rare modifiers were likely to combine based on their experience of other similar modifiers. However, the existence of a process that relies on similar modifiers inevitably raises the question of how a selection of such modifiers could be invoked. The majority of the rare modifiers used in our study do not have a readily identifiable correspondent that is simultaneously common enough to offer a more detailed frequency distribution (e.g. *pier*, *swamp*, *kettle*). Because of this limitation, the existence of such a process would not be able to account for our results.

In order to fully reconcile the behavior of rare modifiers with the CARIN theory, its dependence on statistical knowledge must be reassessed. Considering a hybrid approach, one could speculate that rather than storing relation distributions linked to individual words, speakers might encode distributions relating to particular properties that have a large bearing on how a modifier is likely to be used. The potential effectiveness of such a strategy is underlined by Devereux and Costello's (2005) finding that the properties of a modifier are a reliable predictor of how it will combine. For example, the modifiers *mountain*, *forest* and *river* are relatively common and all display very high instances of the <*located*> relation. Thus speakers might infer that any geographical entities denoting a location will display a similar bias. Upon encountering a combination involving a rare modifier of this type (e.g. *gorge snake*), interpretation would be facilitated by this knowledge even without having previously encountered the modifier itself. In this way, a modifier's relation availability could be gauged based on its ontological category, thereby explaining how the combinational preferences of our rare modifiers could have influenced interpretation.

It seems likely that speakers are indeed aware of correlations of this nature between modifier properties and relational preferences. Speakers of any language will realize that modifiers denoting a location will be strongly predisposed towards the <*located*> relation, substance modifiers will be strongly biased towards the <*made of*> relation and similarly, time period modifiers will be biased towards the <*during*> relation. This realization is possible independently of any statistical knowledge and does not require similar instances to be recalled individually from memory as proposed by Gagné (2002).

This possibility would obviate the storage of frequency distributions for every noun in the lexicon and would be far more efficient: speakers would mainly be storing distributions for specific features rather than storing relation distributions for each modifier word independently, the majority of which are quite unusual. However, the further CARIN deviates from being a completely experientially-driven model and the more semantic components it incorporates, the more it converges with schema-based theories of conceptual combination (e.g. Murphy, 1990;

Wisniewski, 1996). If one accepts that modifier properties influence relation availability, then the importance of statistical knowledge is diminished. One must then accept the basic foundation of schema-based theories, namely that constituent noun properties constrain and dictate the linking relation. As further evidence challenging the experiential view, it is worth noting that no study so far has actually investigated the link between ease of interpretation and statistical knowledge per se: all relation frequencies derived in such studies were based on the modifier's *potential* to combine using the various relations (see Gagné & Shoben, 1997). No relation frequencies have yet been derived based on a measure of how often they would actually be expected to occur in everyday discourse. As a result, these studies do not offer conclusive evidence for the influence of statistical knowledge; on the contrary they merely highlight a correlation between ease of interpretation and how a modifier is likely to combine. On the basis of our current findings it would appear that this appreciation of a modifier's combinational preference is unlikely to be based on statistical knowledge.

Naturally Arising Differences

Other accounts of Gagné and Shoben's (1997) findings have been proposed which do not necessitate the maintenance of statistical knowledge (e.g. Murphy, 2002; Wisniewski, 1997). These accounts suggest that differences in ease of interpretation arise due to the nature of the modifier's semantic representation and point out that combinational history may be an effect rather than a cause. Given our findings, we suggest that the ease of interpretation is not directly influenced by relation frequencies or relation availability. Instead, the ease of interpreting a modifier with a certain relation varies depending on the modifier's properties and this ease of interpretation corresponds naturally with the proportion of combinations that happen to use that relation. For example, the properties of *summer* make it easiest to interpret using the <*during*> relation, and this preference is also reflected in the way that the majority of combinations involving *summer* do involve this relation.

From this perspective it is clear how the relation distribution per se does not necessarily affect how easily a combination is interpreted. Instead, it seems more reasonable that the ease of interpretation is influenced by the nature of the concepts being combined and that the distribution for any given modifier is merely a reflection of how natural it is for that modifier to combine using each relation type. Thus the relationship between the naturalness and the incidence of a certain relation is not a cause and effect relationship: relation frequency covaries with ease of interpretation not because one influences the other, but because both are influenced in the same way by the modifier's preference for combining with certain relations. In other words, not only will natural combinations be interpreted faster, they will also tend to occur more frequently.

Gagné and Shoben's (1997) finding that combinations like *mountain stream* are interpreted faster than *mountain magazine* can easily be explained using this rationale. *Mountain* names a concept which has the prominent feature of being a geographical location. It is easiest for a modifier to combine using a relation that engages a salient attribute, thus *mountain stream* is relatively straightforward to interpret. Furthermore, the majority of combinations involving *mountain* will instantiate the <*located*> relation, as *mountain* tends to modify the locative attribute of a wide range of head nouns. The fact that *mountain* prefers the <*located*> relation is not an arbitrary fact about the word that needs to be stored in the mental lexicon but rather it is a tendency that is reflected both by its high relation frequency and by faster response times. On the other hand, combining *mountain* and *magazine* is a more complicated process, as *magazine* names a concept for which location is not such a salient attribute and the potential of *mountain* to act as a subject matter is not obvious. Correspondingly, few combinations involving *mountain* will happen to instantiate the <*about*> relation. In this case it is clear that although the relation distribution for *mountain* can be used to predict the ease with which it will combine with the various relation types, the distribution per se does not influence the interpretation process.

This idea can also explain our findings regarding rare modifiers. Consider the example *fudge*: the nature of this rare modifier makes *fudge dessert* faster and easier to interpret than *fudge wrapper*. This is because *fudge*, being a substance, is most easily combined with a head noun that instantiates the <*made of*> relation. This relational preference is due to the obvious potential of *fudge* to indicate material composition. The same preference is also reflected in our participant-generated combinations, the vast majority of which also involve the <*made of*> relation.

It is also worth noting that our participant-generated combinations are unlikely to be a reflection of statistical knowledge in the first place. Presumably, participants were not recalling previously encountered combinations involving the rare modifiers but were instead actively generating combinations. As a result, the emergent participant-generated frequencies reflect the ease of combining these modifiers with the various relations rather than reflecting the participants' knowledge of their relation distributions: the correlation we obtained between response time and relation frequency can be more accurately described as a correlation between response time and naturally arising relational preferences. In this case the relational preferences are reflected both in the ease of interpretation and in the range of combinations generated by participants.

This account of our findings is more satisfactory than Gagné and Shoben's (1997) statistically based model. The degree to which distributional knowledge would simplify the comprehension process would be minimal at best, yielding only a rough likelihood for the occurrence of each relation type. Such probabilities would be of no benefit in the absence of a marked bias in relation frequencies and crucially, such knowledge alone could never determine the correct relation: the interpretation process unavoidably requires a detailed consideration of both constituent concepts. As a result, CARIN is an inadequate model of conceptual combination: at best it accounts for only a small portion of response time variance and has nothing to say about how a relation is finally selected or elaborated. For these reasons, a theory that can explain the differences in ease of interpretation as a natural phenomenon is far more desirable than one which necessitates the maintenance of a large volume of statistical knowledge.

Conclusion

It seems likely that there are multiple factors that influence interpretation and hence it may be the case that both statistical knowledge and modifier properties affect this process. Although the greater effect might arise naturally due to a modifier's conceptual content, response times to combinations could still be influenced by linguistic elements based on prior experience. As suggested by Wisniewski (1997), a noun's combinational history may have the effect of suggesting candidate meanings and thereby narrowing down the search space. In effect, natural preference and modifier experience may be involved in a form of feedback loop: modifiers that, because of their properties, are commonly intended in one particular modifying sense may tend to evoke those senses regardless of other influences. For example, the word *chocolate* used as a modifier is nearly always used in the <*made of chocolate*> sense and this fact could make other uses of *chocolate* less available. In this way, a combination like *chocolate train* is more likely to be interpreted as "a train made of chocolate" rather than as "a train carrying chocolate".

Despite this, experiential biases are unable to account for the influence of rare modifiers and consequently, the results of our experiment require an alternative explanation. In conclusion, our findings are most satisfactorily explained by the existence of naturally arising differences in interpretation difficulty which are determined by the nature of a combination's constituent concepts.

Acknowledgements

The authors would like to thank Ed Wisniewski and Rebecca Maguire for useful comments on this work.

References

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

Devereux, B. & Costello, F. (2005). Propane stoves and gas lamps: how the concept hierarchy influences the interpretation of noun-noun compounds. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.*

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.

Gagné, C. L. (2002). Lexical and relation influences on the processing of novel compounds. *Brain and Language, 81*, 723-735.

Maguire, P. & Cater, A. (2004). Is conceptual combination influenced by word order? *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, 133-136.

Maguire, P. & Cater, A. (2005). Turnip soup: Head noun influence in the comprehension of novel compounds. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.*

Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language, 29,* 259–288.

Murphy, G. L. (2002). *The Big Book of Concepts*, 463-464. Cambridge, MA: MIT Press.

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

Tagalakis, G. & Keane, M. T. (2003). Modeling the understanding of noun-noun compounds: The role of familiarity. *Proceedings of the European Cognitive Science Conference*, 319-324.

Tagalakis, G. & Keane, M. T. (2005). How understanding novel compounds is facilitated by priming from similar known compounds. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.*

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language, 35,* 434–453.

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

Appendix: Experimental Items

High Frequency	Low Frequency
Dusk Search	Dusk Painting
Fudge Dessert	Fudge Wrapper
Earwig Nest	Earwig Poison
Hoover Cable	Hoover Sound
Mackerel Trawler	Mackerel Can
Blueberry Pancake	Blueberry Fungus
Lasagna Fork	Lasagna Restaurant
Trombone Parade	Trombone Stand
Kettle Switch	Kettle Steam
Whiskey Flask	Whiskey Cellar
Swamp Animal	Swamp Disease
Magpie Claw	Magpie Forest
Turnip Dinner	Turnip Fiber
Pier Stroll	Pier Timber
Tornado Devastation	Tornado Shelter
Dentist Glove	Dentist Torture
Jeep Window	Jeep Sale
Tavern Bench	Tavern Road