

# The Role of Ambiguity in the Interpretation of Noun-Noun Compounds

**Phil Maguire (phil.maguire@ucd.ie)**  
**Arthur W.S. Cater (arthur.cater@ucd.ie)**  
School of Computer Science and Informatics  
University College Dublin, Ireland

**Edward J. Wisniewski (edw@uncg.edu)**  
Department of Psychology,  
University of North Carolina at Greensboro, USA

## Abstract

Gagné and Shoben (1997) proposed that people are aware of how concepts are typically used to modify other concepts and that this modifier relation frequency influences the ease with which concepts can be combined. However, Wisniewski and Murphy (2005) suggested that Gagné and Shoben's conditions of relation frequency were confounded with plausibility. We conducted an experiment which investigated whether ratings of plausibility respond to relation frequency, a possibility that has been suggested by Gagné and Spalding (in press). We found that, when participants were provided with full descriptions of the combinations' referents, differences in plausibility between conditions of modifier relation frequency disappeared. A second experiment investigated why this should be the case. We found that many combinations were interpreted with a variety of relations, contrary to assumptions made by Gagné and Shoben (1997). Correlation analyses revealed strong associations between combination plausibility and measures of objective and subjective ambiguity, suggesting that the differences observed in Experiment 1 were most likely due to differences in combination ambiguity. We discuss the influence of ambiguity and the implications for Gagné and Shoben's (1997) CARIN theory.

**Keywords:** Conceptual combination; noun-noun compounds; plausibility ratings; ambiguity; CARIN theory.

## Introduction

Concept combination is a ubiquitous phenomenon, as evidenced by the volume of lexicalized compounds entering the English language (e.g. *soccer mom*, *beer gut*) and the frequent spontaneous use of such phrases in reference to novel concepts or ideas (e.g. *peasant dance*, *penguin film*). Not only can people understand concepts like *chocolate* or *rabbit* in isolation, they can also understand them in combination (i.e. *chocolate rabbit*). Recently, the area of conceptual combination has attracted much attention in cognitive science because of its potential to further our understanding of conceptual representation as well as language production and comprehension in general.

In their Competition Among Relations in Nominals (CARIN) theory, Gagné and Shoben (1997) proposed that the interpretation of a noun-noun combination occurs when a relation linking the two concepts is identified. According to this theory, the ease of interpretation of a combination is principally influenced by the relative frequency with which

the appropriate relation has been previously used with the modifier noun. For example, since combinations involving the modifier *mountain* frequently use the <LOCATED> relation, then it should be easy to interpret a combination like *mountain goat* while *mountain magazine* (a magazine <ABOUT> mountains) should be more difficult. As support for their theory, Gagné and Shoben (1997) demonstrated that when participants made speeded sensicality judgments for a set of combinations, those involving low modifier frequency relations took longer to interpret than those involving high modifier frequency relations, while the head noun's relation frequency had no discernible effect.

Wisniewski and Murphy (2005) challenged the reliability of these results by suggesting that the observed patterns in response time were primarily due to differences in the familiarity and plausibility of the stimuli. When they asked participants to rate the plausibility and familiarity of Gagné and Shoben's combinations, they found that low relation frequency items were rated as reliably less plausible and less familiar than the high frequency ones. Furthermore, these variables accounted for a far greater portion of the variance in sensicality judgment response time than did modifier relation frequency.

Gagné and Spalding (in press) countered this argument. They demonstrated that subjective familiarity can be influenced by relation frequency and that relation strength emerges as a significant predictor variable in multiple regressions involving measures of *objective* familiarity. Similarly, they defended the confounding of relation frequency and plausibility by interpreting the latter as a dependent variable. According to their view, plausibility is what theories of conceptual combination should seek to explain and therefore, on the pain of circularity, it should not be conceived of as an explanatory variable. As an explanation for why low frequency items were rated as less plausible, Gagné and Spalding suggested that plausibility might be responding to modifier relation frequency.

However, we believe that plausibility is very much an independent variable, and one which should intuitively influence ease of interpretation. The combination *plastic toy* is plausible by virtue of the fact that many toys are made of plastic. In contrast, *cooking treatment* is less plausible since the action of cooking does not suggest how it might be used as a treatment. Because this combination is unsupported by world knowledge it is more elusive and consequently, less plausible.

Since the plausibility of any concept, combined or otherwise, is determined by its semantic content, then this measure should not be influenced by the manner in which the concept is expressed. Given that modifier relation frequency is a factor which relates solely to interpretation and not to meaning, then logically it should not influence the subjective plausibility of a combined concept. Wisniewski and Murphy (2005) asked participants to rate the plausibility of “the thing referred to by the combination” as opposed to the plausibility of the phrase itself, and hence these ratings should not have been contaminated by modifier relation frequency as suggested by Gagné and Spalding (in press).

### Experiment 1

In the following experiment we investigated whether subjective plausibility is influenced by modifier relation frequency. In order to evaluate this hypothesis, we compared plausibility ratings for Gagné and Shoben’s (1997) stimuli in combination form with plausibility ratings for fully expanded paraphrases indexing the same referents. Since the paraphrases included an overt reference to the appropriate relation, this mitigated the influence of any factors relating to relation availability. Our hypothesis was that the differences in plausibility observed by Wisniewski and Murphy (2005) would remain after the influence of modifier relation frequency had been controlled for, thereby ruling it out as a possible influence.

### Method

**Participants** Forty-six first-year undergraduate students from University College Dublin participated in the study. All were native English speakers.

**Materials** In a stimulus pre-test we asked six participants to generate interpretations for the 57 experimental items from Gagné and Shoben’s (1997) Experiment 1. These comprised 19 combinations for three separate conditions of relation frequency, namely HH, HL and LH. Here, H and L refer to the frequency of the appropriate relation (i.e. High and Low), with the first and second letters denoting how frequently that relation is associated with the modifier and head noun respectively. Given that all of these combinations were designed to be sensible, we also included 19 implausible filler items so as to ensure that the full range of the plausibility scale would be used.

The dominant interpretation generated for each combination was used to generate its equivalent paraphrase. Each of these consisted of the original constituent nouns linked by the appropriate relation. In this way, the combination *water money* was reconstituted as “money for buying water” and *college magazine* became “a magazine about college”.

**Procedure** Participants were randomly assigned to either the combination condition or the paraphrase condition. All were presented with a random ordering of the 76 concepts in the relevant form and asked to rate the plausibility of each concept on a scale of 1 to 7. Following Wisniewski and

Murphy’s (2005) experimental paradigm, a rating of 1 indicated that the concept was “very weird” and 7 indicated that it was “very plausible”.

### Results and Discussion

For the combination condition, the average plausibility ratings for the HH, HL, LH and filler conditions were 5.1, 4.8, 4.3 and 2.2 respectively. For the paraphrase condition, the corresponding values were 5.8, 5.6, 5.6 and 2.0. These results are illustrated in Figure 1. Our combination plausibility ratings were consistent with those of Wisniewski and Murphy (2005),  $r = 0.88, p < .01$ . The most plausible combination was *student magazine* (6.91) and the least plausible was *water money* (1.74). The most plausible paraphrase was “a toy made of plastic” (6.78) and the least plausible was “a crisis involving plastic” (3.96).

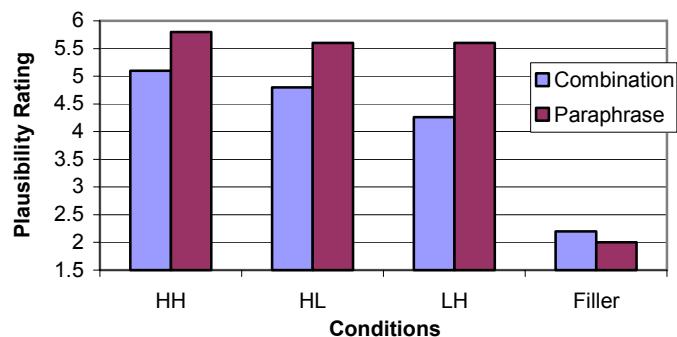


Figure 1. Average plausibility ratings for combinations and phrases in each condition of modifier relation frequency

Results were analyzed with participants as a random variable and not with items as a random variable. The reason for this was because we wished to determine whether modifier relation frequency affected the subjective plausibility of the items in question rather than drawing an inference from this set to a larger population. For the combination condition, a within-subjects analysis of variance revealed statistically reliable differences among the three conditions of relation frequency,  $F(2, 44) = 45.82, p < .01, MSE = .10$ . A planned comparison revealed that combinations in the low modifier relation frequency condition (LH) were rated as significantly less plausible than those in the high modifier relation frequency conditions (HH and HL),  $F(1, 44) = 78.92, p < .01, MSE = .09$ . For the paraphrase condition, a separate within-subjects analysis of variance also revealed statistically reliable differences among the three conditions of relation frequency,  $F(2, 44) = 4.43, p = .02, MSE = .068$ . Contrary to hypothesis, a planned comparison revealed that the difference between the low and high modifier relation frequency conditions was not significant,  $F(1, 44) = 2.07, p = .21, MSE = .08$ . This challenges the view that Gagné and Shoben’s (1997) conditions of modifier relation frequency were confounded with plausibility, since differences for the paraphrases were not evident.

The two sets of plausibility ratings were closely correlated,  $r = .64, p < .01$ . We compared how well combination plausibility versus paraphrase plausibility

predicted Gagné and Shoben's (1997) sensicality judgment response times for the stimuli. The correlation between combination plausibility and response time was significant,  $r = -.67, p < .01$ . However, a two-tailed z-test revealed that the corresponding correlation involving paraphrase plausibility,  $r = -.31, p = .02$ , was significantly weaker,  $z = 2.55, p = .01$ . This implies that ratings of combination plausibility were influenced by factors relating to interpretation and not just the plausibility of the referent concept per se.

Wisniewski and Murphy (2005) found that when they entered plausibility and modifier relation frequency into a multiple regression, the latter did not emerge as a significant predictor variable of response time. While their ratings of plausibility were potentially contaminated by relation frequency, our paraphrase ratings could not have been. Accordingly, we performed two separate multiple regressions, one involving combination plausibility and the other involving paraphrase plausibility, both having response time as the criterion variable. For the regression involving combination plausibility, the multiple correlation was .69 ( $R^2$  of .47) and the standardized regression weights were -.62 for plausibility ( $p < .01$ ) and -.19 for modifier relation frequency ( $p = .07$ ). The fact that relation frequency did not emerge as a significant predictor variable replicates Wisniewski and Murphy's finding. However, for the regression involving paraphrase plausibility, the multiple correlation was .46 ( $R^2$  of .21) and the standardized regression weights were -.31 for plausibility ( $p = .02$ ) and -.34 for modifier relation frequency ( $p < .01$ ). This finding supports the possibility that combination plausibility ratings were affected by modifier relation frequency.

If differences between combination and paraphrase plausibility are due in part to the mitigation of modifier relation frequency in the latter condition, then we would expect such differences to be correlated with relation frequency: low modifier relation frequency combinations should benefit the most from having their relation made explicit. Indeed, this correlation proved to be significant,  $r = -.32, p = .02$ . The average increase in plausibility for the low modifier relation frequency items was 1.3 whereas that for the high frequency items was only 0.7.

These results lend credence to Gagné and Spalding's (in press) assertion that plausibility ratings for combinations are affected by factors relating to interpretation, one of which includes modifier relation frequency. Combinations with low modifier relation frequency were rated as significantly less plausible than those with high modifier relation frequency. Yet, when the same concepts were presented along with the appropriate relation, both types were rated equally plausible. Thus, it would appear that the influence of modifier relation frequency on relation availability may have influenced plausibility ratings in the combination condition.

We noticed from our stimulus pretest that the interpretations provided for some combinations were very consistent, whereas others varied widely. For example, *plastic toy* was always interpreted as "a toy made of plastic" while *wood money* was interpreted as "money derived from the wood trade", "money for buying wood" and "money

made of wood". Crucially, the LH combinations appeared to be more ambiguous than those in the other conditions. This suggested to us that plausibility ratings might somehow have been influenced by combination ambiguity.

Gagné and Shoben's (1997) measure of relation frequency assumes that combinations will always be interpreted with one particular relation. If combinations can be interpreted using several different relations then this invalidates the basis for relation frequency. For example, "money <DERIVED FROM> wood" uses a low frequency relation but "money <MADE OF> wood" uses a high frequency relation; if participants generate both interpretations, then the combination cannot be successfully categorized.

We suspected that differences in plausibility may have arisen because participants rating combinations did not generate the intended interpretation. The following experiment examined this possibility.

## Experiment 2

This experiment was designed to investigate whether the plausibility ratings in Experiment 1 might have been influenced by the inherent ambiguity of the combined concepts in combination form. For instance, the referent of *wood money* intended by Gagné and Shoben (1997) was "money derived from wood". Although ratings for our paraphrases reflected the plausibility of this specific concept, participants viewing the combination might not necessarily have generated the same interpretation. Some participants may have interpreted *wood money* as meaning "money made out of wood" or "money for buying wood" and therefore rated these as very implausible. Others may have been unable to generate any satisfactory interpretation. In short, plausibility ratings for combinations could not be guaranteed to reflect the plausibility of the intended referent if not all participants interpreted them in the intended manner. In the following experiment we investigated the consistency with which Gagné and Shoben's experimental stimuli were interpreted. The same 57 experimental items used in Experiment 1 were presented as combinations and participants were asked to provide the interpretation which they felt was most probable, and to give a rating of how likely it was to be the intended one.

## Method

**Participants** Sixty first-year undergraduate students participated in the experiment. All were native English speakers. The data from 12 participants were excluded for either not following instructions or for failing to complete the task.

**Materials** The 57 experimental stimuli used in Gagné and Shoben's (1997) Experiment 1 were presented in combination form.

**Procedure and Design** The 57 items were counterbalanced across participants, with each participant viewing 19 items. One combination was presented for each of the 19 modifier

nouns, this set constituting a distribution of HH, HL and LH items. Participants were informed that the combinations were genuine and were asked to provide an interpretation for each one. After writing down this interpretation, they were also required to provide a rating from 1 to 7 which reflected how likely they thought it was that their interpretation was the intended one. A rating of 1 indicated that this was “very unlikely” while 7 indicated that it was “very likely”. For demonstration purposes, several examples were provided along with interpretations and likelihood ratings. These included very specific combinations (e.g. *kitchen door*) and very ambiguous combinations (e.g. *shovel bird*).

## Results and Discussion

For each combination, we obtained a set of 16 interpretations and 16 associated likelihood ratings. We clustered interpretations garnered for each combination into distinct categories which we felt reflected the different types of interpretation. Any disagreement in classification was resolved through discussion. A total of 183 interpretations were identified for the 57 stimuli, with an average of 3.2 different interpretations per item.

We conceived of the two measures in our experiment as constituting approximate measures of objective ambiguity (number of different interpretations produced) and also of subjective ambiguity (participants’ confidence that their interpretation was the intended one). Both measures varied considerably by-item. Fourteen combinations resulted in a single interpretation, whereas *cooking treatment* produced 10. Similarly, confidence in interpretations ranged from a low of 2.2 for *gas antiques* to a high of 6.7 for *plastic toy*. Out of the total of 16 participants that viewed each combination, the number that rated their confidence above the midpoint varied from a low of 1 for *gas antiques* to a full complement of 16 for *gas crisis*, *student magazine* and *plastic toy*.

In order to assess the implications of ambiguity for measures of relation frequency, we divided interpretations into those that involved Gagné and Shoben’s (1997) intended relation versus those that did not. Although they reported no statistics of this nature for their first experiment, Gagné and Shoben conducted a similar analysis for the interpretations given in Experiment 3 of their study. For this analysis, eight participants provided interpretations of the combinations involved. Gagné and Shoben (1997) found that the interpretations generated included the intended relation 82% of the time. They also reported that on a participant-by-participant basis, the percentage of agreement ranged from 71% to 90% and described this level of agreement as “encouraging” (p. 79). We conducted the same analysis using the interpretations given in the present study, and obtained a similar figure: the overall proportion of combinations which were interpreted using the relation intended by Gagné and Shoben was 73.6%. However, closer inspection revealed that the by-participant analysis was

extremely misleading: variance in agreement was primarily manifested at the item and not at the participant level.

For 20 of the 57 stimuli, 100% of participants interpreted the combination using Gagné and Shoben’s intended relation (e.g. *financial crisis*, *plastic toy*). On the other hand, only a single participant interpreted the combinations *floral language* and *floral toy* using the relations <USES> and <IN> respectively; all other participants used alternative relations. For 15 of the stimuli, the number of participants using the correct relation were in the minority. Furthermore, the average level of agreement varied across conditions: the figures for the HH, HL and LH conditions were 85%, 69% and 67% respectively.

A correlation between *intended* relation frequency and use of the intended relation proved significant,  $r = .31$ ,  $p = .02$ . Thus, the lower the intended relation frequency, the more likely it was that participants used a different relation which was potentially not be a low frequency one. Interestingly, the correlation between Gagné and Shoben’s (1997) response times and use of the intended relation was also significant,  $r = .50$ ,  $p < .01$ . The longer the average response time, the more likely participants were not to use the intended relation.

The substantial by-item variance in ambiguity deals a serious blow to the CARIN perspective on combination interpretation. Gagné and Shoben (1997) derived values for their hypothetical variable of relation strength by assuming that participants were largely using the same relation, the one they themselves selected. However, we have shown that while some materials were reliably interpreted using this relation, others were rarely interpreted so. As a result, the relation frequencies used by Gagné and Shoben were in many cases inaccurate and furthermore, the extent of this inaccuracy was confounded with the variable under investigation: low frequency items were more likely *not* to be interpreted using that low frequency relation. In many cases, those combinations intended as low frequency were actually interpreted with high frequency relations, often the same ones as those used in the high modifier relation frequency conditions. For example, Gagné and Shoben intended *winter book* as “a book about winter”, the <ABOUT> relation representing a low frequency relation for *winter*. However, half of our participants interpreted this combination as “a book one reads during winter”, with the <DURING> relation being the intended relation for *winter* in both the HH and the HL conditions (i.e. *winter cloud*, *winter town*). In another case, *wood money* was intended as “money derived from wood”. However, 38% of participants interpreted it with the high frequency <MADE OF> relation from the HL condition (i.e. “money made of wood”) and another 13% interpreted it with the high frequency <FOR> relation from the HH condition (i.e. “money for buying wood”). Because of this, many of Gagné and Shoben’s (1997) relation strength values would have been inappropriate.

We attempted to compensate for this inaccuracy by only considering the 42 items for which the majority of

participants used the intended relation. Gagné and Shoben reported a significant correlation of  $r = .44$  between relation strength and response time when all 57 combinations were included. We performed the same analysis but used an appropriate nonparametric correlation and included only the 42 less ambiguous combinations. The correlation was not significant, Spearman's  $\rho = .187, p = .24$ .

Not only does this undermine the empirical evidence supporting the CARIN theory, it also brings into question the validity of that theory's approach to conceptual combination in general. Modifier relation frequency cannot be used to model the interpretation process when no agreement on the appropriate relation exists. We have shown that multiple relations can be used with many of Gagné and Shoben's (1997) combinations, with some being interpreted using both high frequency and low frequency relations. Since CARIN is founded on the assumption that relations compete based on their availability, it therefore cannot explain why relations of varying availability should be selected for the same combination.

### **Influence of Ambiguity**

Given that many of Gagné and Shoben's (1997) stimuli can be interpreted using different relations, the use of modifier relation frequency is not appropriate for explaining differences in the ease of interpretation or plausibility of those combinations. We investigated whether combination ambiguity might constitute a more appropriate predictor variable. The correlation between Gagné and Shoben's (1997) sensality judgment response times and the number of different interpretations generated was significant,  $r = .43, p < .01$ . The correlation between response time and participants' average confidence rating was also significant,  $r = -.62, p < .01$ . A further set of correlations examined whether ambiguity was associated with plausibility. We found that combination plausibility was significantly correlated with both number of interpretations,  $r = -.52, p < .01$ , and with confidence ratings,  $r = .82, p < .01$ . Furthermore, the average differences between combination and paraphrase plausibility were significantly correlated with both number of interpretations,  $r = .27, p = .04$ , and confidence ratings,  $r = .54, p < .01$ . These results suggest that the differences in plausibility observed in Experiment 1 can be better explained by differences in ambiguity rather than differences in modifier relation frequency.

The fact that differences between combination and paraphrase plausibility ratings were associated with differences in ambiguity suggests two possible explanations. Firstly, participants may have been less likely to find the intended referent of ambiguous combinations and consequently, may have generated less plausible interpretations. A second possibility is that ratings were influenced by the ambiguity of the combination as well as the plausibility of the referent concept. Thus, ambiguous compounds may have been rated as less plausible even when participants were able to generate the intended interpretation.

The principles of pragmatics maintain that words, whether in combination or otherwise, are used to convey meaning, and this meaning is what renders their use sensible. Since the ambiguity of a combination is inextricably linked to its success as a communicative expression, this factor is likely to influence its subjective plausibility: any combination which inadequately constrains the interpretation process will be viewed as unacceptable. A significant correlation between our subjective and objective measures of ambiguity ( $r = -.64, p < .01$ ) reveals that participants were able to make reliable estimates regarding combination ambiguity. Hence, participants may have rated ambiguous combinations as less plausible because they seemed less meaningful.

Differences in ambiguity can also explain differences in sensality judgment response time. Firstly, the ambiguity of a combination will no doubt influence the length of time needed to interpret it. When high levels of agreement exist, it is usually because the range of interpretation is highly constrained. Such overt constraints are likely to guide people quickly to the intended referent. In contrast, ambiguous combinations will be less constrained, necessitating a broader search in order to identify an appropriate interpretation. People may also be tempted to continue searching for a potentially superior interpretation in cases where an initial interpretation seems unconvincing.

In addition, the interpretation process must be followed by some form of decision process, as interpretation alone is not sufficient evidence for sensality. Gagné and Shoben's (1997) view of sensality as being equivalent to successful interpretation is unrealistic. Given sufficient motivation and imagination, virtually any combination of nouns can be connected in some way and the challenge therefore lies in observing some threshold of acceptability (e.g. *metal cloud*, *fish legs*, and *mountain whale* can all be interpreted, yet are clearly not sensible). With combinations located along a continuum of sensality, judgments will no doubt be influenced by their proximity to the threshold of acceptability. Combinations that are clearly sensible or clearly non-sensible will be easy to judge. Conversely, those that are not so clear will be more difficult to judge and the process will take longer. Since ambiguity is likely to lower the perceived sensality of a combination, Gagné and Shoben's (1997) ambiguous stimuli were inevitably more difficult to judge than those that were overtly meaningful.

## **General Discussion**

In Experiment 1 we investigated whether subjective plausibility ratings for combinations are influenced by modifier relation frequency. Our results appeared to suggest that this was the case: differences in plausibility ratings between conditions of modifier relation frequency disappeared when the relation was made explicit. However, in Experiment 2 we showed that modifier relation frequency could not explain differences in plausibility. Given that some combinations were interpreted with both low and high frequency relations, a single frequency categorization could

not be applied. On the other hand, we found that measures of objective and subjective ambiguity were closely correlated with both plausibility and ease of interpretation. Therefore, although the referents of Gagné and Shoben's (1997) combinations were not confounded with plausibility, many of their stimuli were ambiguous, thereby invalidating the relevance of relation frequency.

One might counter that ambiguity itself is unlikely to be an independent variable and hence combinations in the LH condition might have appeared more ambiguous *because* they were associated with low frequency relations. However, this perspective is illogical. A combination is ambiguous by virtue of the fact that it can be interpreted using a variety of different relations, all of which have different relation frequencies. Clearly then, the assumption of a single interpretation cannot be applied to explain differences in ambiguity. As a result, modifier relation frequency cannot explain the differences in plausibility and ease of interpretation observed for the combinations in Gagné and Shoben's (1997) study.

The ambiguity of a combination is likely to be determined by multiple factors. Firstly, familiarity is likely to play a large role in guiding interpretation. For example, the lexicalized term *palm tree* is associated with one dominant interpretation which overrides the need for a combinatorial process. In other cases, context is the crucial element in constraining possible meanings. Given this clue, an otherwise ambiguous combination can be interpreted with certainty. Finally, world knowledge combined with personal experience will have a large effect on ambiguity: combinations that can be interpreted in a way that is consistent with world knowledge will appear more convincing and hence less ambiguous.

Although it may be the case that ambiguity is affected by some factor akin to relation bias, we stress that this bias is likely to be based on the interaction of both the modifier and the head noun. For example, the modifier *mountain* has a preference for the <LOCATED> relation, yet this bias is only relevant when it can be supported by the head. Thus, a combination like *mountain height* does not suggest the <LOCATED> relation, since *height* cannot have a location. Similarly, the head noun *soup* is often biased toward the <MADE OF> relation, but only when the modifier denotes a food substance (see Maguire & Cater, 2005).

### Judgments of Plausibility

It may be misleading to consider the plausibility of a combination as absolute. Combined concepts are only generally used in situations where they make sense. Upon encountering a compound phrase, it is not logical to entertain the possibility that it might be implausible: either one understands it or one does not, but either way the combination is almost certainly intended to be plausible. Intuitively, combinations will only appear implausible in cases where they are interpreted in the absence of the context which motivated and justified their use. For instance, the intended referent of a combination like *energy headache* might prove elusive to someone for whom the

relevant context was not available. However, to state that this combination is implausible or even less plausible than other combinations is misleading because in its original context, it makes perfect sense (i.e. "Germany's *energy headache* will dominate next week's cabinet meeting", Irish Times, January 14<sup>th</sup> 2006). Therefore, a more accurate description would be to say that the interpretation of *energy headache* is context dependent.

The ambiguity revealed in Experiment 2 was undoubtedly due to the dependence of some combinations on a suitable context for meaningfulness. Combinations like *wood money* and *cooking treatment* were presented in isolation despite the fact that they could not be reliably or consistently interpreted under these circumstances. As noun-noun compounds are typically only used in situations where they can be reliably interpreted, this raises serious questions about the validity of analyzing response times for context-dependent combinations presented in isolation. A sensibility judgment task involving such stimuli is unlikely to reveal anything other than the extent to which they are dependent on context for their meaning.

### Conclusion

Although we have found that Gagné and Shoben's (1997) conditions of modifier relation frequency were not confounded with plausibility, we have shown that many combinations were ambiguous. This undermines the empirical support for the CARIN theory. Firstly, many of the relation frequencies used by Gagné and Shoben were not appropriate because participants interpreted combinations using many different relations. Secondly, when we eliminated those combinations which the majority interpreted using a different relation, we found no significant correlation between response time and relation strength. Future study investigating the influence of modifier relation frequency should ensure that combinations can be reliably interpreted given the level of context in which they are presented.

### References

- 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. & Spalding T. L. (in press). Relation availability was not confounded with familiarity in Gagné and Shoben (1997): Comment on Wisniewski and Murphy (2005). *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- 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*.
- Wisniewski, E. J., & Murphy, G. L. (2005). Frequency of Relation Type as a Determinant of Conceptual Combination: A Reanalysis *Journal of Exp. Psychology: Learning, Memory, and Cognition*, 31, 169-174.