Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 4, 811–821

0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.4.811

COMMENTARY

A Reanalysis of the CARIN Theory of Conceptual Combination

Phil Maguire
National University of Ireland, Maynooth

Barry Devereux, Fintan Costello, and Arthur Cater
University College Dublin

The competition among relations in nominals (CARIN) theory of conceptual combination (C. L. Gagné & E. J. Shoben, 1997) proposes that people interpret nominal compounds by selecting a relation from a pool of competing alternatives and that relation availability is influenced by the frequency with which relations have been previously associated with the modifying concept. The current authors derived relation frequencies by using a sample of compounds occurring in the British National Corpus and compared them with those derived by Gagné and Shoben. The authors demonstrated that the original relation frequencies are unrepresentative and that Gagné and Shoben's technique for dichotomizing them into high and low is unreliable. In addition, the authors revealed anomalies in the mathematical instantiation of the CARIN model and showed that it does not provide evidence for competition among relations.

Keywords: conceptual combination, nominal compounds, thematic relations, CARIN model

When people use language to communicate, they often need to name concepts for which there is no simple or suitable one-word expression. In such cases, a compound formed from two nouns will frequently suffice, allowing the speaker to succinctly describe a complex concept in a way that can be reliably deciphered. In English, compounds consist of a modifier followed by a head noun, where the head noun typically denotes the main category and the modifier indicates a contrast or specialization of this category. In addition to the wealth of compounds continually entering the vernacular (e.g., *wedding tourist*, *coffee culture*), spontaneous combinations are ubiquitous in everyday language because of the way they provide a concise index to a novel instantiation of a concept (e.g., *apple-juice seat* could refer to the seat with apple juice opposite it; Downing, 1977).

The study of conceptual combination has the potential to reveal much about conceptual representation as well as the processes involved in language production and comprehension. As a result, there has been considerable interest in this phenomenon within cognitive psychology. In recent years, a variety of models of conceptual combination have been proposed (e.g., Costello & Keane, 2000; Murphy, 1988; Wisniewski, 1997). Most of these have tended to converge on the view that during the interpretation process, the basic head noun category is somehow refined or specialized by the modifier concept. For example, when the modifier *kitchen* is paired with the head noun *chair*, it serves to refine the concept *chair* by delineating a particular subset of that category, namely the type of chair found in kitchens.

One prominent theory that has adopted a different view is competition among relations in nominals (CARIN; Gagné & Shoben, 1997). According to this account, the interpretation of a novel noun–noun compound occurs when people identify the relation that exists between the modifier and the head noun. The CARIN theory uses a set of 16 potential relation types, including such relations as LOCATED, DURING, and FOR (e.g., *chair* LOCATED *kitchen*). Gagné and Shoben (1997) proposed that these basic thematic relations compete for selection and that the relative availability of the appropriate relation determines the ease with which the combination can be interpreted.

However, the CARIN theory diverges from other theories of conceptual combination regarding its proposed mechanism for how constituent nouns affect relation availability. Gagné and Shoben (1997) did not consider that a noun's influence on relation availability might be a function of its conceptual content. Rather, they suggested that this influence is a function of how a noun has been experienced in previous combinations. In other words, "people possess distributional knowledge about how often particular relations are used" (p. 74), and this knowledge affects the ease with which two constituents are combined. For example, Gagné and Shoben predicted that *mountain goat* would be easier to interpret than *mountain range* by virtue of the fact that *mountain* is more frequently used with the LOCATED relation than it is with the MADE OF relation. As a means of encapsulating relational

Phil Maguire, Department of Computer Science, National University of Ireland, Maynooth, County Kildare, Ireland; Barry Devereux, Fintan Costello, and Arthur Cater, School of Computer Science and Informatics, University College Dublin, Dublin, Ireland.

This research was funded in part by Irish Research Council for Science, Engineering and Technology Grant RS/2002/758-2 to Barry Devereux. We sincerely thank Christina Gagné for making available the data from Gagné and Shoben (1997). We also thank Edward Wisniewski and Rebecca Maguire for valuable comments and advice. Phil Maguire and Barry Devereux contributed equally to this article.

Correspondence concerning this article should be addressed to either Phil Maguire, Department of Computer Science, National University of Ireland, Maynooth, County Kildare, Ireland, or Barry Devereux, School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: pmaguire@cs.nuim.ie, phil.maguire@ucd.ie, or barry.devereux@ucd.ie

preference, Gagné and Shoben introduced the variable of relation frequency, that is, the relative frequency with which a noun is observed to combine with one particular relation. In their study, they found that the combinational behavior of the modifier *mountain* was encapsulated by a relation frequency of .82 for the LOCATED relation and a frequency .01 for the MADE OF relation.

Gagné and Shoben's (1997) approach of deriving individual context-insensitive relation frequencies for each constituent assumes that the influence of the modifier is separate and independent from that of the head noun. According to this view, both modifiers and heads have a separate influence on relation availability, one that is not affected by the opposite constituent. At first blush this is surprising, given that one would expect the significance of a modifier to vary depending on the head with which it is paired. Other theories of conceptual combination (e.g., Estes & Glucksberg, 2000) have proposed that the influence of both constituents is a joint interactive one. For example, *feather luggage* can be interpreted as light luggage because feathers have the salient property of being light and luggage has weight as a relevant dimension. On the other hand, the use of the modifier *feather* in a combination such as *feather storage* is unlikely to have the same effect, because storage does not have weight as a relevant dimension. If the significance of the modifier depends on how it is used, then its influence on the interpretative process will not be a consistent one.

In addition to partitioning the contribution of the modifier and head, Gagné and Shoben (1997) also drew a distinction between the relative influences of both constituents. After analyzing response times for a sample of combinations presented in a sensibility judgment task, they found that the modifier's relation strength (i.e., a value representing the relative availability of the appropriate relation) was significantly correlated with response time but that this was not the case for the head noun. This contrast is surprising given that other theories of conceptual combination have stressed the importance of both constituents. Intuitively, relation selection requires the consideration of both nouns so that the same relation is not always selected for a given modifier (e.g., *mountain height* and *mountain range* would be interpreted using the dominant LOCATED relation were it not for some form of head noun influence).

In summary, several aspects of the CARIN theory distinguish it from other theories of conceptual combination. In particular, it does not incorporate those elements that other models have consistently emphasized, namely conceptual content, a combinational process, and a contribution by the head noun. Studies supporting the influence of modifier relation frequency have often overlooked other potential confounding variables known to affect relation availability, such as familiarity and plausibility (see Wisniewski & Murphy, 2005). It has also been suggested that relation frequency might simply be an epiphenomenon of conceptual content (e.g., Maguire, Cater, & Maguire, in press; Maguire, Wisniewski, & Storms, 2007; Murphy, 2002). The empirical evidence in support of the CARIN theory rests primarily on the construal of a causal relationship between modifier relation frequency and ease of interpretation. This construal in turn rests on the assumptions that Gagné and Shoben's (1997) frequencies accurately reflect experiential knowledge and that relation frequency is a reasonable index of the availability of a thematic relation for a constituent. In our opinion, these assumptions have not been adequately justified. In the remainder of this article, we evaluate both the representativeness of Gagné and Shoben's frequencies and the reliability of the evidence supporting the influence of this variable on ease of interpretation. We also discuss the appropriateness of relation frequency as an operational definition of relation availability, and we consider other possible reasons for its observed association with response time.

ESTIMATING REAL-WORLD RELATION FREQUENCY

Gagné and Shoben (1997) derived relation frequencies by pairing 91 head nouns and 91 modifiers from 100 familiar combinations in the Levi (1978) appendix (duplicate nouns were removed). This procedure resulted in 8,281 noun–noun phrases, of which 3,239 were judged as sensible. The interpretations of these 3,239 sensible combinations were then classified into the 16 relation categories provided for by the CARIN theory, and relation frequencies were determined by calculating the percentage of combinations involving each noun and relation. For example, the LOCATED relation was found to occur in 82% of the combinations involving *mountain* as a modifier.

Wisniewski and Murphy (2005) claimed that the accuracy of this method for computing frequency is unclear. According to the principles of the CARIN theory, relation frequencies should reflect the relative frequency with which a particular relation and noun have been paired in the written and spoken language to which people have been exposed. However, Gagné and Shoben's (1997) combinations were generated arbitrarily, and hence there is no guarantee that the relation frequencies derived in this manner represent the distributions present in natural language. As Wisniewski and Murphy pointed out, different modifiers and heads tend to combine with different sets of nouns, suggesting that one cannot accurately determine their relation frequency by pairing all of them with the same set. For example, *chocolate* tends to modify words such as *bar*, *cake*, and *biscuit*. On the other hand, *mountain* tends to modify words such as *goat*, *bird*, and *forest*. Although combinations like *chocolate goat* or *mountain cake* can be interpreted, such phrases are often quite bizarre. In addition, combinations in Gagné and Shoben's sample were weighted equally, meaning that very unusual phrases exerted as great an influence as those that are far more common (e.g., *gas scandal* was given the same weight as the more familiar *gas crisis*). It is not certain that frequencies derived in the above manner approximate frequencies observed in natural language.

Despite these issues, the same relation frequencies have been reused in several studies investigating the CARIN theory (e.g., Gagné & Spalding, 2004; Ramey, 2005). Storms and Wisniewski (2006) attempted to derive more representative statistics by asking participants to generate combinations for modifiers and head nouns. In this way, they avoided the problems of unnatural context and type-token frequency. However, this technique is still prone to inaccuracy and cannot be relied on to deliver representative relation frequencies. For example, Maguire et al. (in press) found that participants were predisposed to select the most available combinations first, leading to much repetition. Samples generated in this manner therefore failed to reflect the full range of combinational possibilities, with the more common relations being overrepresented and the rarer ones underrepresented.

Gagné and Shoben (1997) posited that relation frequencies reflect people's experiential knowledge about how often relations have been used with different nouns ("people have distributional knowledge about how often particular relations are used with modifiers," p. 74; also "knowledge about how likely particular relations are to be used with a given concept reflects a person's experience with the language," Gagné, 2002, p. 726). Consequently, the best method for estimating such frequencies is one that considers combinations occurring in a representative sample of language. A corpus analysis therefore provides the best and perhaps the only reliable method for deriving relation frequencies. Much of the support for the CARIN model depends critically on the representativeness of Gagné and Shoben's frequencies, making it important to establish their reliability.

In the following corpus study we addressed this issue by sampling combinations from a representative corpus of naturally occurring language. Deriving a more representative set of relation frequencies allowed us to better assess the rigor of both the CARIN theory and the evidence supporting it.

Corpus Study

In order to carry out the analysis, we required a representative corpus capable of yielding a sufficient number of combinations involving nouns in both the modifier and head positions. For this purpose we availed of the British National Corpus (BNC) World Edition (Burnard, 2000), a tagged annotated corpus containing over 100 million words. This contains samples of written and spoken language from a range of sources designed to represent a wide cross-section of current English. Noun–noun sequences were identified by means of Gsearch (Corley, Corley, Keller, Crocker, & Trewin, 2001), a chart parser that detects syntactic patterns in a tagged corpus via a user-specified, context-free grammar and a syntactic query.

Method

We derived relation frequencies for the heads and modifiers used in Gagné and Shoben's (1997) Experiment 1. In this experiment, they selected 19 of their 91 modifiers and 19 of their 91 heads in order to analyze the influence of modifier relation frequency alongside the influence of head relation frequency. The 38 nouns were paired so as to generate three separate conditions of modifier–head frequency for each noun, labeled HH, HL, and LH. Here H and L refer to the frequency of the appropriate relation (i.e., high and low), with the first letter denoting how frequently the relation was associated with the modifier and the second denoting how frequently that same relation was associated with the head noun. For example, they included the combination *mountain cloud* in the HL condition because the LOCATED relation was found to be high frequency for *mountain* as a modifier but low frequency for *cloud* as a head. All three conditions contained 19 combinations, with each of the 19 modifiers and 19 heads being used once. Using the BNC corpus, we reevaluated the modifier and head relation frequency distributions for these 38 nouns.

Materials

In extracting suitable combinations, we were forced to confront the issue of noun ambiguity, as some nouns (e.g., *plant*) combined using different senses. According to Gagné and Shoben (1997), "information about thematic relations is stored directly with the modifier concept" (p. 83). Therefore, if an ambiguous noun is used in reference to two distinct concepts, two separate relation frequencies should be applied. For example, the relation frequency brought to bear in interpreting the combination *water plant* (*plant* FOR *water*) should be based on those combinations where *plant* has been previously used in reference to the "factory" concept, as opposed to the "organism" concept. However, the experimental items and frequencies used by Gagné and Shoben suggest that noun ambiguity was overlooked. The item *water plant* was included in the HL condition, whereas *office plant* (*plant* LOCATED *office*) was used in the HH condition. In both cases, the same relation frequency was applied. Furthermore, in computing relation frequencies, Gagné and Shoben included all of the sensible combinations involving a particular noun, regardless of noun sense. As a result, *industrial plant* and *winter plant* were both included in computing the relation frequency of *plant*. Given that a distinction between words and concepts was not applied, the frequencies and stimuli used in evaluating the CARIN theory are inconsistent with the assertion that relation frequency distributions are stored at the conceptual level. However, for the sake of comparing our BNC frequencies with the original frequencies, we too ignored noun ambiguity and followed Gagné and Shoben's procedure.

For each of the 38 modifiers and heads under investigation, we sought a random sample of 100 BNC combinations in which that word appeared in the relevant position. The modifier *floral*, which is an adjective, appeared only 3 times as a modifier in the BNC. However, because the CARIN theory proposes that relation frequencies are stored at the conceptual level, we were able to increase the sample by including the modifier's noun form (i.e., *flower*). The term *musical* is potentially a noun, although we inferred that Gagné and Shoben (1997) intended it as an adjective, given their interpretation of *musical light* as *light* MAKES *music* (and not a light that makes musicals). Accordingly, we sampled 100 combinations involving *music* rather than *musical*. The only two modifiers to occur fewer than 100 times were *urban* (56) and *servant* (76). Regarding Gagné and Shoben's head nouns, two were plural, and the rest were singular. Assuming that different lexical forms of a noun would be associated with the same relation frequency distribution, we enhanced the representativeness of our sample by including combinations with both singular and plural heads. Only 4 of the 19 heads did not yield the desired 100 samples, namely *antique* (9), *headache* (32), *utensil* (44), and *remedy* (84).

In total, we were able to extract a random sample of 100 valid compounds from the BNC for 32 of the 38 nouns involved in the study. In the 6 remaining cases, where the full 100-item sample could not be obtained, all those available were used, giving a total of 3,501 compounds. The sampling process considered all those noun–noun co-occurrences separated by a space, regardless of semantic transparency. In addition, we manually rejected all noun–noun sequences that did not constitute legitimate combinations. Among these noun–noun sequences were those where verbs had been incorrectly tagged as nouns (e.g., "the *mountain rose* up") as well as noun–noun co-occurrences (e.g., "last *year albums* were cheaper"). In total, 1,832 modifier and 1,669 head compounds were used to derive the frequencies, which compared favorably with the 1,038 modifier and 879 head compounds used by Gagné and Shoben (1997) for the same nouns. Interestingly, our corpus analysis revealed that Gagné and Shoben's modifier nouns occurred significantly more often than the head nouns, with averages of 2,399 and 739 occurrences, respectively, in the BNC, $t(36) = 3.03$, $p < .01$. This asymmetry is potentially problematic for an experiment contrasting the influences of modifier and head relation frequency.

Procedure

In order to derive relation frequencies for the nouns in question, Phil Maguire and Barry Devereux each classified half of the compounds into the most appropriate thematic relation category. We considered each token in the context of the sentence in which it occurred, selecting the relation appropriate to that particular use. However, in many cases, the task of classifying compounds into thematic relation categories did not prove to be straightforward.

We found that rarely was there just one clear thematic relation for a given compound. In many cases, there were at least two relations that could be judged as being appropriate. For instance, it was difficult to decide whether *family activities* was best classified as belonging to the thematic relation HAS (activities that a family has), LOCATED (activities within a family), or indeed FOR (activities for families), CAUSES (activities caused by a family), or BY (activities by a family). Similarly, *storm cloud* could reasonably be classified under CAUSES, LOCATED, DURING, or HAS. This issue highlights a significant limitation of the relation-based approach, namely that thematic relations are often too vague and ill-defined to serve as adequate representations of the specific, concrete relationships that are instantiated during conceptual combination (see Downing, 1977; Devereux & Costello, 2005). For example, the FOR relation is used both in interpreting *flu pills* and *fertility pills*, even though a clear distinction in meaning is apparent (*flu pills* are pills for attacking flu, but *fertility pills* are pills for improving fertility; Downing, 1977). Consequently, our emulation of Gagné and Shoben's (1997) procedure, whereby compounds were classified into precisely one thematic relation category, often necessitated selecting a single relation in a way that was quite unsatisfactory.

We also encountered the opposite problem, where compounds did not realistically fit any of the 16 thematic relations provided by CARIN (e.g., *chocolate eater*, *water supply*, *family commitments*, *music journalist*). Altogether, we were unable to identify any satisfactory relation for 7.5% of modifier compounds and 4.3% of head compounds. One might propose that including more relations in the taxonomy would allow these relations to be classified. However, doing so would exacerbate the earlier issue of most compounds having more than one reasonable thematic relation. The difficulties we experienced in applying Gagné and Shoben's (1997) taxonomy to a wide range of naturally occurring combinations suggest that adhering to a limited set of relation types is sometimes impractical.

We computed interrater reliability using a sample of 10 combinations for each of the 38 nouns, although we did not attempt to resolve disagreements given the fact that multiple relations were often equally appropriate. The level of agreement was 68%. Using the κ coefficient of agreement as a measure of reliability, we found that the relation categories were distinguished with a reproducibility of $\kappa = .65$, $p < .001$, where a value of .7 to 1 indicates the degree of reliability to be from good to perfect (Cohen, 1960).

Having classified all 3,501 compounds, we derived relation frequencies by computing the percentage of combinations using each relation, this being the same technique used by Gagné and Shoben (1997). Our relation frequencies were then compared with the original frequencies. It is worth noting that the values derived in this manner actually refer to relation proportions and not frequencies. However, for the sake of consistency with Gagné and Shoben's terminology we continue to refer to the percent proportion of a relation's occurrence as its relation frequency.

Results

Our analysis revealed that those combinations included in Gagné and Shoben's (1997) sample were not representative of the sample of combinations taken from the BNC corpus. Of the 1,038 sensible combinations they identified involving the modifier nouns in Experiment 1, only 51 of these types appeared in our sample (5%). Of the 879 combinations they used to derive head frequencies, only 62 of these types appeared (7%). Therefore, 94.1% of their dataset consisted of combinations that did not emerge given a representative sample of 100 tokens. These statistics demonstrate that the technique Gagné and Shoben used to generate their combinations did not result in a representative sample of combinations.

We used Spearman's ρ to compute the correlation between the BNC frequencies and Gagné and Shoben's (1997) frequencies. Correlations were carried out for each noun individually, and then these values were averaged.[1] The average ρ value for the modifiers was .64, whereas the average ρ value for the heads was .63. In total, 10 of the modifier distributions and 9 of the head distributions were significantly correlated at the $p < .01$ level. Figure 1 presents a sample comparison of both sets of frequencies for the modifier *student* and the head noun *cloud*.

In order to assess the degree to which these results support the accuracy of Gagné and Shoben's (1997) frequencies, we needed to compare them against some baseline. Accordingly, we split our BNC corpus data into two random halves and analyzed the level of agreement between them. This time the average ρ value for the modifiers was .83, whereas the average ρ value for the heads was .88. Furthermore, 17 of the modifier distributions and 18 of the head distributions were significantly correlated at the $p < .01$ level. Because the two halves of our data were strongly correlated, this suggests that the expected level of variability in relation frequency between two representative sources should be quite low and that even a sample size of 50 is capable of providing a reliable distribution. People are likely to have encountered many millions of combinations (as opposed to just 50), and thus we would expect the variability in experiential relation frequencies to be even lower still. The fact that the average ρ values for Gagné and Shoben's data were considerably lower than those for our split-corpus analysis shows that their method for deriving frequencies can be outperformed by using even a small sample of representative combinations.

---

[1] We thank Randi Martin for this suggestion.

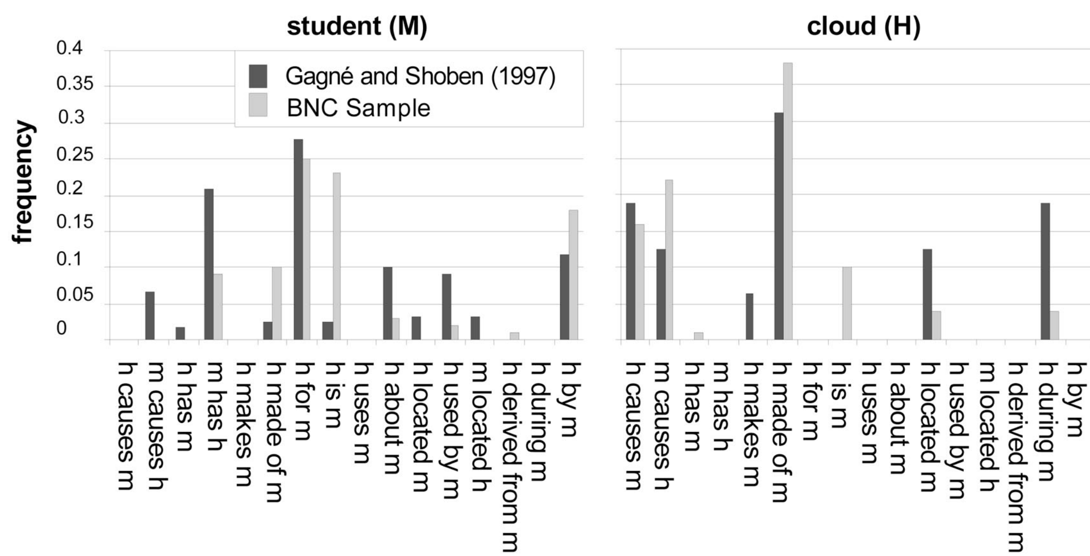

Figure 1. Comparison of the British National Corpus (BNC) relation frequency distribution and Gagné and Shoben's (1997) relation frequency distribution for the modifier (M) noun *student* and the head (H) noun *cloud*.

The correlation between Gagné and Shoben's (1997) data and the BNC data provides a potentially misleading measure of the appropriateness of their technique for deriving frequencies. Many relations cannot be used with certain nouns (e.g., the modifier *student* was never used with the DURING relation, because *student* is not a time period). Any method for deriving relation frequencies will presumably be capable of revealing this kind of information, no matter how unrepresentative it is. The inclusion of null data points in comparing relation frequency distributions is therefore uninformative for the purpose of assessing the appropriateness of Gagné and Shoben's method. Accordingly, we recomputed the Spearman correlations between both sets of frequencies, this time excluding those relations with no associated data. Two nouns had to be removed from the analysis because of lack of data (*plastic* and *antiques*). Correlating Gagné and Shoben's frequencies with the BNC frequencies, we found that the average ρ value for the modifiers was .43, whereas the average ρ value for the heads was .29. When we computed the same correlations between the two halves of our corpus data, the average ρ value for the modifiers was .77, whereas the average ρ value for the heads was .82. Although Gagné and Shoben's frequencies somewhat resemble the BNC data, they exhibit far greater variance than would be expected in a corpus sample and are thus unlikely to accurately reflect people's exposure to combinations.

Discussion

The BNC frequencies and Gagné and Shoben's (1997) frequencies generally agreed in some cases, although there was systematic divergence in others. We believe that this pattern can be attributed to fundamental differences in what was being measured. Gagné and Shoben's frequencies reflect the capacity of a set of words to combine with an arbitrary set of 91 nouns. In contrast, the BNC frequencies reflect how those words combine in everyday use. Relations that rely on the pairing of specific types of modifiers and heads cannot be accurately reflected by a set of combinations involving a limited set of nouns. As a result, differences in relation frequency often arose for relations relying on domain-specific pairings. For instance, the modifier whose relation distribution varied to the greatest extent was *servant*. In our BNC sample, this modifier occurred using the IS relation 76% of the time, as compared with the 5% obtained by Gagné and Shoben. The reason for this discrepancy is that *servant* can use only the IS relation when paired with a head noun that can act in a servant capacity (e.g., *servant girl*, *servant monkey*, *servant maid*). Because Gagné and Shoben's 91 heads did not include many nouns that could act as servants, their set of frequencies inevitably failed to reveal the true incidence of this relation for *servant* in natural language.

In the reverse scenario, Gagné and Shoben's (1997) relation frequencies were higher than the corresponding BNC frequencies in cases where a noun's capacity to combine using a particular relation exceeded the incidence of that relation in natural language. Hence, universal donor relations (e.g., LOCATED for *mountain*) and universal acceptor relations (e.g., ABOUT for *book*) had higher CARIN frequencies than BNC frequencies. A universal donor like *mountain* has the potential to specify the location of a wide selection of head nouns, resulting in a very high LOCATED frequency in Gagné and Shoben's dataset (82%). Yet despite this, the BNC revealed that *mountain* frequently combines using other relations (e.g., *mountain peak*, *mountain chain*), yielding a much lower figure for LOCATED (51%). Similarly, a universal acceptor like *book* has the potential to be modified by a wide range of modifier nouns using the ABOUT relation (75% in Gagné and Shoben's dataset) because books can have a wide variety of topics. Despite this, we found that in the BNC, *book* frequently combined with modifier nouns using relations other than the ABOUT relation (e.g., *guide book*, *prayer book*, *picture book*), therefore yielding only 22% for this relation.

In summary, the accuracy of Gagné and Shoben's (1997) method for deriving relation frequencies was compromised by nouns whose capacity for combining using certain relations did not reflect the actual frequency with which they combined using those relations in a representative sample of language. Arbitrarily pairing nouns using a finite set meant that some relations were overrepresented, whereas other relations requiring domain-specific terms were underrepresented.

Reanalysis of Gagné and Shoben's (1997) Experimental Paradigm, Using BNC Frequencies

In order to evaluate their hypotheses regarding relation frequency, Gagné and Shoben (1997) compared sensicality judgment times for combinations of high and low relation frequencies. They found that items in the HH and HL conditions were judged as sensible reliably faster than those in the LH condition, suggesting to them that the time taken to interpret a combination is influenced by modifier relation frequency. On the other hand, they found no reliable difference between the average time taken to judge HH and HL combinations. This they interpreted as evidence that head relation frequency does not influence ease of interpretation to the same extent.

Because our BNC frequencies deviated considerably from Gagné and Shoben's (1997) frequencies in some cases, we investigated the reliability of their classifications. In their study, all relations were partitioned into either high- or low-frequency 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 first identifying the highest frequency relation. 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 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.

We implemented the same procedure as above using the BNC frequencies. This revealed a considerable level of disagreement between the relation classifications based on the BNC frequencies and those based on Gagné and Shoben's (1997) frequencies. Of the 45 high-frequency modifier relations we identified, 38% had been categorized as low frequency in Gagné and Shoben's analysis. Of the 40 high-frequency head relations we identified, 48% had previously been categorized as low frequency.

Subsequently, we examined the reliability of the classification of relations involved in Gagné and Shoben's (1997) Experiment 1. According to the BNC, 47%, 47%, and 58% of combinations were misclassified as in the HH, HL, and LH conditions, respectively. In several cases, both the modifier and the head frequency obtained by Gagné and Shoben disagreed with the BNC classification. For example, *floral language* (USES) and *servant book* (FOR) both appeared in the HL condition but should have been LH materials, according to the BNC data. This level of discrepancy raises concern about whether differences between Gagné and Shoben's experimental conditions can realistically be attributed to differences in relation frequency.

On average, differences between Gagné and Shoben's (1997) frequencies and the BNC frequencies led to a disagreement in categorization for 51% of the combinations in Experiment 1. The extent of this disagreement is primarily due to the sensitivity to noise of Gagné and Shoben's method for dichotomizing relation frequencies. The use of a precise 60% dichotomization threshold (or indeed any other precise percentage) is unreliable because, theoretically, an infinitesimal variation in the frequency of a relation can make the difference between that relation being categorized as high or low. The susceptibility of Gagné and Shoben's technique to frequency variance is also exacerbated by the fact that it depends not just on the relation in question but also on its competitors, meaning that a tiny variation in any relation frequency can affect classification. Such effects are not realistic because, first, relation frequencies are only approximations based on a small sample of combinations, and second, combinational experience will inevitably vary from person to person. We encountered several cases where the classification of a combination into one particular condition was dependent on less than 1% of a difference in its relation distribution. For example, Gagné and Shoben found that the head noun *headache* combined based on the CAUSES relation 60.3% of the time and the LOCATED relation 12% of the time. Because the CAUSES relation occurred more than 60% of the time, it alone was the only high-frequency relation. However, had this figure been only 0.4% lower, then the LOCATED relation would also have been high frequency. Given that it is unreasonable to expect relation frequencies to be accurate to within 1%, relations should not be dichotomized based on a paradigm that is sensitive to such small differences. As a result, our use of a similar yet slightly different set of relation frequencies resulted in a substantial level of disagreement in classification.

An alternative approach to relation frequency categorization would have been to simply use relation frequency directly and to sample at the extremes of the spectrum. Had the high-frequency relations consisted of only those most frequent, and the low-frequency relations consisted of only those most rare, then the distinction between high and low would have been more rigorous. Maintaining a tangible gap between high and low relation frequencies would have rendered categorization insensitive to small variations as well as exposing more clearly the potential differences between the processing of both types of combination. As it was, any relation with frequency up to 40% could potentially have been classified as low frequency, whereas relations with frequency as low as 3% could have been classified as high frequency.

We computed the level of correlation between relation frequency and sensicality judgment response times (collected in Gagné & Shoben, 1997, Experiment 1) using the BNC frequencies. If people derive relation frequencies based on combinational experience, we would expect the BNC frequencies to reflect this experience more closely. Given the CARIN theory's claim that relation frequency influences ease of interpretation, we would thus expect corpus-derived frequencies to be more closely correlated with sensicality judgment response times than Gagné and Shoben's frequencies, which are not representative of natural language. However, the correlation between BNC modifier relation frequency and response time ($r = -.33$, $p < .01$) was no stronger than that obtained using Gagné and Shoben's frequencies ($r = -.34$, $p < .01$). Because the more representative relation frequencies were no better at predicting response time, this suggests that the association between relation frequency and response time is not experientially grounded. Instead, it may be the case that relation frequency is simply an epiphenomenon of conceptual content, insofar as both interpretation and combination use are influenced by conceptual features (see Maguire et al., in press; Maguire, et al., 2007; Murphy, 2002).

The correlation between BNC head frequency and response time was not significant ($r = .02$, $p = .88$). In this regard, the BNC data replicated Gagné and Shoben's (1997) original finding, namely that modifier relation frequency was correlated with response time but head relation frequency was not. Although this might appear to support the CARIN theory, there are reasons to suspect that the effect is not robust. First, as we discovered, the chosen modifiers occurred significantly more often than the head nouns in the BNC, which could potentially explain any difference in influence. Furthermore, Wisniewski and Murphy (2005) have claimed that Gagné and Shoben's stimuli were confounded with plausibility and familiarity, these being more accurate predictors of response time than relation frequency. Although Gagné and Spalding (2006) demonstrated that subjective familiarity can be influenced by relation frequency, the effect was relatively small. Therefore, it seems unlikely that all of the difference in familiarity between Gagné and Shoben's conditions of relation frequency can be explained in this way. Maguire, Cater, and Wisniewski (2006) also found that participants were more likely to interpret the low-frequency combinations with a variety of relations, undermining the CARIN approach of modeling conceptual combination with a single relation frequency. When they reran Gagné and Shoben's analyses and included only the less ambiguous stimuli, the correlation between modifier relation frequency and response time was not significant. Thus, it may be the case that the CARIN effect can be attributed to Gagné and Shoben's choice of materials.

REANALYSIS OF THE CARIN MODEL

Gagné and Shoben (1997) provided two sources of evidence to support the view that relation frequency influences ease of interpretation. In the previous section, we examined the reliability of one source of evidence, namely, the findings from a series of experiments comparing response times for high and low relation frequency items. The second source of evidence they provided involved the formulation of a model expressing relation strength as a function of relation frequency. They found that the output of this model was significantly correlated with response time and interpreted this as evidence supporting the influence of relation frequency. Subsequent studies (e.g., Gagné & Spalding, 2006) have argued specifically for the strength measure over raw frequencies because of its supposed encapsulation of competition. In this section, we examine whether the model provides for competition among relations.

The CARIN model conceives of relation strength as a function of one relation's frequency relative to its competitors. Although relation frequency is already an expression of the relative frequency of a relation, Gagné and Shoben (1997) justified the utility of their model through their assertion that it reflects the level of competition among relations vying for selection ("the CARIN model claims that the ease with which the appropriate relation can be found depends on both the strength of the to-be-selected relation and on the strength of the alternatives," p. 81). Furthermore, "the availability of other high-strength [relations] should slow the interpretation of conceptual combinations" (p. 81).[2] Therefore, if there is any other relation close to or exceeding the frequency of the appropriate relation, the interpretation of the combination should be slowed because of competition between the two. However, if there is no such other relation, then interpretation should be faster because the correct relation can be selected without effective competition. In order for a formula to reflect this effect, its output must therefore be sensitive to the distribution of relation frequency among the alternative relations. In a series of worked examples introducing their model, Gagné and Shoben made use of the following initial formula:

$$\text{Relation strength} = \frac{P_{\text{selected}}}{P_{\text{selected}} + P_1 + P_2 + P_3}. \quad (1)$$

Here, P_{selected} is the frequency, or proportion of times, that the selected (i.e., appropriate) relation has been used with compounds that have the same modifier, and P_1, P_2, and P_3 are the corresponding proportions for the three highest frequency alternative relations for that modifier. One significant limitation of this formula's capacity to reflect competition among relations concerns its insensitivity to the distribution of frequencies in the denominator. Whether or not one of these alternative relations has a frequency close to that of the selected relation has no particular influence on the output. For example, suppose the proportional frequency of the selected relation is .50. If one of the alternative relations also has a frequency of .50 and the other two have frequencies of .00, then the formula will yield a strength of $.50 / (.50 + .50 + .00 + .00) = .50$. However, if the three most frequent alternative relations have evenly distributed frequencies of .167 (which together sum to .50), then again this yields a strength of .50. According to the formula, these two situations are indistinguishable, although the first case clearly involves competition between two relations with similar frequencies, whereas the second does not.

There is a sense in which this initial formula indirectly reflects information about competition among relations. For example, in cases where there are only four nonzero relation frequencies (i.e., the three leading competitors to P_{selected} are relatively strong), the denominator of the equation will be 1 and the computed strength will simply be P_{selected}. However, in cases where there are more than four nonzero frequencies in the relation distribution (i.e., the leading competitors are weaker), then the denominator of the equation will be slightly less than 1 and, therefore, the computed strength will be marginally greater than P_{selected}. Although this pattern reflects the general idea that interpretation will be easier in cases where the competitors are weaker, the effect is limited because the output is dominated by P_{selected}. Indeed, the measures of P_{selected} and strength are nearly identical ($r > .99$, for both Gagné and Shoben's [1997] frequencies and the BNC-derived frequencies). Given that the formula's output varies so little from P_{selected} and that the distribution of frequencies in the denominator has no effect, its success in reflecting competition among relations is minimal. Moreover, Gagné and Shoben presented their formula as an application of Luce's (1959) choice rule (p. 81), which means that, ideally, all possible alternatives should be included in the denominator. However, if all of the relation frequencies are included, then the denominator sums to one and the output of

[2] In referring to relation strength here, Gagné and Shoben (1997) meant to refer to relation availability and hence relation frequency. Although the term *relation strength* was also used to refer to the output of the CARIN model, such a reference would produce a circular definition in the given context.

Equation 1 is identical to $P_{selected}$. In this case, the formula cannot reflect competition.

Following the worked examples of their model, Gagné and Shoben (1997) introduced a refinement whereby each of the variables is transformed using the negative exponential function and a free parameter α:

$$\text{Relation strength} = \frac{e^{-\alpha P_{selected}}}{e^{-\alpha P_{selected}} + e^{-\alpha P_1} + e^{-\alpha P_2} + e^{-\alpha P_3}}. \quad (2)$$

This was again presented as an application of Luce's (1959) choice rule, which formulates the probability of selecting a particular item from a pool of alternatives. In the context of the CARIN model, the particular item is the selected relation and the pool of alternatives constitutes all thematic relations in the CARIN taxonomy. In adopting the negative exponential transformation, Gagné and Shoben stated that they were following previous applications of Luce's choice rule, citing that by Sadler and Shoben (1993). However, its use in Equation 2 is problematic because the phenomenon being modeled is of a fundamentally different nature. For example, the probability that an item i is selected from a pool of n alternatives in an analogical reasoning task was defined by Sadler and Shoben as

$$p(i) = \frac{e^{-\alpha d_i}}{\sum_{j=1}^{n} e^{-\alpha d_j}}, \quad (3)$$

where d_i is the distance between item i and an ideal solution to the analogy problem. The probability that item i is selected is therefore a decreasing function of distance, in that it decreases as the distance between i and the ideal solution point increases. In this and similar models, the negative exponential transformation of distance is used to generate a measure of psychological similarity. Viewed from this perspective, Luce's choice rule models the probability of item i being selected as an increasing function of the similarity of i to the ideal solution.

However, these transformations are not appropriate in CARIN's case, because the measure of relation frequency is not analogous to that of distance. According to Gagné and Shoben's (1997) theory, when the frequency of a relation is high, the probability that that relation will be selected is also high. This happens to be the opposite pattern to that modeled by Sadler and Shoben (1993). If the CARIN model is to be specified as an application of Luce's (1959) choice rule, then the frequency measures for each relation should not be transformed using a negative exponential. Instead, a transformation that is an increasing function of frequency is required.

The consequences of the use of a negative exponential in Equation 2 are anomalous in two ways. First, the description of the model's output as denoting relation "strength" contradicts the reality of the situation: A high $P_{selected}$ will result in a low output value and vice versa, yielding a positive correlation between response time and relation strength. Second, the use of negative exponentials means that the largest term in the denominator, and therefore the one with the greatest influence, will always come from the lowest relation frequency. Examining these issues, we calculated the linear correlation between each of the exponentially transformed variables and the output of the formula for the items in Gagné and Shoben's (1997) Experiment 1 (using the original α value of 36). The strongest correlation was found for $P_{selected}$, $r = .59$ (a consequence of its presence in the numerator of the formula). However, of the competing relations, P_3 was found to be the most strongly correlated, $r = -.41$, with P_2 the second, $r = -.36$, and finally P_1 the least, $r = -.18$. The fact that the lowest term in the denominator has the greatest influence on the output is incompatible with the theory described by Gagné and Shoben. It is also contrary to reason: P_3 is the smallest relation frequency and should therefore exert the least competitive influence. The anomaly emerges because the use of the negative exponential coupled with a high α value makes the higher relation frequency terms in the denominator so small that the lowest frequency term becomes the most significant one. For the purpose of elucidating the behavior of Gagné and Shoben's model, we provide a simplified formula in Equation 4 that approximates the output of Equation 2, given these observations. Here, P_{lowest} refers to the lowest relation frequency in the denominator, that being either $P_{selected}$ or P_3:

$$\text{Relation strength} \approx \frac{e^{-\alpha P_{selected}}}{0+0+0+e^{-\alpha P_{lowest}}} = e^{-\alpha(P_{selected}-P_{lowest})}. \quad (4)$$

Note that the $P_{selected}$ term in the numerator will also be close to zero but that variance in this term is directly correlated with the output. From Equation 4, we can see that the output of the CARIN model is approximated by $e^{-\alpha(P_{selected}-P_{lowest})}$. An analysis of the data confirmed this relationship, yielding a correlation between ln(*relation strength*) and $P_{selected} - P_{lowest}$ of $r = -.999$. In other words, the output of the CARIN model is dominated by only two values, namely $P_{selected}$ and P_{lowest}. Because no other relation frequencies have a significant influence on the output of the formula, it clearly cannot support competition among relations.

In order to illustrate the problems with the model's output, we consider a case where a $P_{selected}$ of .40 competes with relation frequencies of .50, .10, and .00 and another case where the competing relation frequencies are .30, .30, and .00. We have claimed that the negative exponential transformation means that the lowest term in the denominator will be the only important one. Because both of these cases will have a zero frequency in their denominator, this suggests that both will produce virtually identical outputs, despite the fact that the selected relation is dominant in only the latter case. Indeed, this is the pattern that emerges. On the basis of Gagné and Shoben's (1997) α value of 36, the relation strength in the first scenario is 5.6×10^{-7} / [$(5.6 \times 10^{-7}) + (1.5 \times 10^{-8}) + (2.7 \times 10^{-2}) + 1$] = 5.4×10^{-7}. The relation strength in the second scenario is 5.6×10^{-7} / (($5.6 \times 10^{-7}) + (2.0 \times 10^{-5}) + (2.0 \times 10^{-5}) + 1) = 5.6 \times 10^{-7}$, a figure that is identical to the former to 7 decimal places, or one millionth of one standard deviation of the strength values obtained for the items in Gagné and Shoben's Experiment 1.

That the formula is dominated by $P_{selected}$ and P_{lowest} means that we can consider the output as falling into two distinct categories. If $P_{selected}$ is the lowest term in the denominator, then the output will approach one. If $P_{selected}$ is not the lowest term in the denominator, then the output will be closer to zero. Evidencing this trend, 53% of all the relation strength values derived by Gagné and Shoben (1997) are lower than .01, which is within 3% of one standard deviation from the zero edge of the range. In all of these cases, the near zero output is due to the fact that the P_3 term is lower than $P_{selected}$. This phenomenon leads to some surprising relation strengths for Gagné and Shoben's experimental items. For example, the relation strengths for both *plastic toy* (MADE OF) and *plastic crisis* (ABOUT) are lower than .01 (differing by less than 1% of the standard deviation of the strength values) despite the fact that the former uses a high frequency relation ($P_{\text{MADE OF}}$ = .68) and the latter uses a low frequency relation (P_{ABOUT} = .18). The reason for this anomaly is that in both cases P_{lowest} is only .02; with the negative exponential transformation, the term featuring P_{lowest} exerts by far the greatest influence on the computed strength. As well as *plastic crisis*, other combinations for which there is a similar effect include *urban album* (strength = .01) and *mountain magazine* (strength = .02). Therefore, some low modifier frequency combinations receive a very low value for relation strength, despite the fact that these values should be reserved for stimuli that are the easiest to interpret. These examples are evidence that the output produced by the CARIN model does not encapsulate the theory of conceptual combination outlined by Gagné and Shoben.

Given that the formula depends on the lowest frequency in the denominator, then the number of terms included is clearly crucial. Gagné and Shoben (1997) claimed that whereas "the decision to include only four relations in the denominator of the strength ratio used in the CARIN model is somewhat arbitrary, the ability to predict response time on the basis of our model is relatively insensitive to the number of relations that appear in the denominator" (p. 82). This statement does not address how the inclusion of lower frequencies affects the output of the formula. For example, including four competitors in the denominator instead of three means that the relation strength of *cooking treatment* changes considerably, from 0.85 to 0.39 (with α = 36). Here, P_4 is relatively small (0.02). However, because of the negative exponential transformation, it exerts the greatest influence on the formula, more than doubling the size of the denominator and thereby halving the output value. Thus, the CARIN model can be very sensitive to the number of terms included in the denominator. Indeed, if all possible relations are included, in accordance with Luce's (1959) choice rule, then all of the computed relation strengths fall below 0.07.

In their study, Gagné and Shoben (1997) obtained a significant correlation between response time and the output of their relation strength formula and interpreted this as support for the CARIN theory. The strength of this correlation, r = .44, is indeed higher than that between response time and relation frequency, r = –.34. How can this be the case if the behavior of the model does not reflect the theory? One explanation is that Gagné and Shoben's model includes a free parameter α, which was selected so as to minimize the difference between the output and response time. If we remove the free parameter α in Equation 2, we find that the correlation drops to r = .34, which is no better than the correlation using $P_{selected}$. It therefore appears that the CARIN model does not offer any advantage over $P_{selected}$ other than the natural advantage gained by introducing a free parameter to optimize the degree of fit to the data.

Alternative Sources of Evidence for Competition among Relations

The theory described by Gagné and Shoben (1997) implies that the output of the CARIN model should be interpretable as the posterior probability for a categorical target variable, namely the selected relation. This kind of multiple logistic function, sometimes known as the softmax activation function, is frequently used in training neural networks (e.g., Bridle, 1990). The output is given by the following equation, where $p(i)$ refers to the probability of item i being selected, and q_j refers to the activation of item j:

$$p(i) = \frac{e^{q_i}}{\sum_{j=1}^{n} e^{q_j}}. \quad (5)$$

Equation 5 resembles Equation 3 (used by Sadler & Shoben, 1993), with the exception that the exponential transforms are positive as opposed to negative. Because an increase in activation level q_i is associated with an increase in the probability of item i being selected, the role of q in Equation 5 is analogous to Gagné and Shoben's (1997) concept of relation frequency. In order to investigate the possibility of a competition-among-relations effect, we applied the activation function to both Gagné and Shoben's frequencies and the BNC frequencies. However, the output from this function was not more strongly correlated with response time than was relation frequency (r = –.30, p = .02, using the BNC frequencies; r = –.32, p = .01, with Gagné and Shoben's frequencies). In a separate analysis, we refined Equation 5 by introducing a scaling factor α to the exponential transforms in order to optimize the output. When we used both sets of frequencies, the scaling factor that maximized the correlation between output and response time was one that reduced the formula to a function of $P_{selected}$. In the case of Gagné and Shoben's frequencies, the optimal positive valued α was a value convergent on zero. When we used this value for α, it produced a relatively invariant denominator, mitigating the influence of every relation frequency other than the selected one. To confirm this relationship, we assumed an arbitrarily low value of .001 for α and obtained a correlation of $r >$ .99 between $P_{selected}$ and the activation output. When we used the BNC frequencies, the optimal value for α tended toward infinity. With such a high value, the output converges on one if $P_{selected}$ is the highest relation and zero if it is not. Clearly, such a dichotomy cannot account for competition among relations, because it fails to reflect any information regarding the proximity or number of competing relations. These results undermine Gagné and Shoben's claim that ease of interpretation should be modeled as a competition among relations.

We examined the data in more detail for other indications of a competitive process. Considering the set of compounds used in Gagné and Shoben's (1997) first experiment, we selected those compounds where $P_{selected}$ was the most frequent relation and correlated the average response time for these with the distance to the nearest competing relation. This correlation was not significant when using either Gagné and Shoben's frequencies (r = –.09, p = .66, N = 28) or the BNC frequencies (r = .08, p = .70, N = 26). In order to separate the influence of relation frequency from that of competition among relations, we performed a multiple regression using $P_{selected}$ and competitor distance as the predictor variables and response time as the criterion variable. Using Gagné and Shoben's frequencies, we found that the multiple correlation was .10 (R^2 of .01) and the standardized regression weights were –.32 (competitor distance, p = .73) and .24 ($P_{selected}$, p = .79). Using the BNC frequencies, we found that the multiple correlation was .10 (R^2 of .01) and the standardized regression weights were –.28 (competitor distance, $p = .86$) and .36 ($P_{selected}$, $p = .81$). The fact that neither variable emerged as a significant predictor of response time undermines Gagné and Shoben's view that relation frequency and competition are the principal factors influencing ease of interpretation.

Investigating the possibility of a more localized competitive effect, we again examined the cases where $P_{selected}$ was the most frequent relation. From this sample, we identified items comprising a "close" group (compounds for which there was another competing relation with a frequency within .10 of $P_{selected}$) and also a "distant" group (those for which there was no competing relation with a frequency within .60 of $P_{selected}$). The criterial values of .10 and .60 were chosen so that there would be an equal number of compounds in both the close and distant groups. According to the competition-among-relations approach, we would have expected average response times to be faster for the distant group and slower for the close group. However, analysis revealed that the average response time for the close group was 1,030 ms ($SD = 94$ ms), whereas that for the distant group was 1,073 ms ($SD = 75$ ms). In other words, participants did not react significantly faster when there was no competing relation (they were in fact slower, but not significantly so, $p = .32$). Although the items that were most distant from a competitor also happened to have higher relation frequencies, the effect of this trend on response time would have been in the same direction as any competitive effect. Thus, the fact that the observed difference was in the opposite direction provides converging evidence that the availability of competing relations is not a factor in interpretation.

Relation Frequency as a Measure of Relation Availability

The problems we have highlighted regarding the influence of relation frequency and the evidence provided in support of this influence do not undermine the relation-based approach to conceptual combination in general, nor do they challenge the possibility that relation availability influences ease of interpretation. On the contrary, several studies (e.g., Gagné, 2001) have shown that when all other factors are held constant, combinations are interpreted more easily when the availability of the intended relation is enhanced. Our main concern is that relation frequency per se does not constitute an accurate reflection of relation availability. The frequency of a given relation is a relatively naive measure because it is averaged over every previous encounter of a noun in combination. In contrast, the actual level of relation availability is likely to be far more complex, taking into account the multitude of factors that interact to influence the perceived likelihood of a relation. Such factors, although possibly including relation frequency, will certainly also include familiarity and plausibility as well as context (see Gagné & Spalding, 2004; Gerrig & Bortfeld, 1999; Wisniewski & Murphy, 2005). Knowledge of the situation at hand often supports the identification of an appropriate relation, and therefore, the effect that a certain modifier or head has on relation availability is likely to vary in different circumstances. Hence, the application of a single aggregate measure for a noun is unlikely to encapsulate the precise nature of its influence across every unique situation. In our opinion, the investigation of other influences in parallel with relation frequency has the potential to inform a more comprehensive theory of conceptual combination.

CONCLUDING REMARKS

Our finding that there is little evidence in support of the CARIN model is of considerable significance to the study of conceptual combination given the prominence of this theory in the area. We have highlighted several problems with the current version of the theory, in particular regarding its mathematical instantiation and the frequency dichotomization paradigm used in the experiments supporting it. As subsequent studies have used the same technique and the same formula, these problems have been propagated through many experiments investigating the influence of relation frequency (e.g., Gagné & Spalding, 2004; Storms & Wisniewski, 2006). We suggest that to avoid these problems, future studies of this nature should meet several criteria. First, relation frequencies should be based on a large representative sample such as the one used in this study. Second, only unambiguous nouns and combinations should be used in experiments to ensure that measures of relation frequency are appropriate and reliable. Third, high relation frequencies should refer to those frequencies that are significantly higher than average, and low frequencies should be those that are significantly lower than average. Fourth, relation frequency should not be confounded with other variables that are known to influence response time, such as familiarity or plausibility. Future studies should consider, in addition to these criteria, the possibility that relation frequency is merely an epiphenomenon of conceptual content and therefore not functional at all. If this is the case, it might prove more fruitful to directly examine the influence of concept properties on interpretation.

References

Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogleman Soulie & J. Herault (Eds.), *Neurocomputing: Algorithms, architectures and applications* (pp. 227–236). Berlin: Springer-Verlag.

Burnard, L. (2000). *Reference guide for the British National Corpus* (world edition). Oxford, United Kingdom: Oxford University Computing Services.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

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.

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint guided conceptual combination. *Cognitive Science, 24,* 299–349.

Devereux, B., & Costello, F. J. (2005). Investigating the relations used in conceptual combination. *Artificial Intelligence Review, 24,* 489–515.

Downing, P. (1977). On the creation and use of English compound nouns. *Language, 53,* 810–842.

Estes, Z., & Glucksberg, S. (2000). Interactive property attribution in concept combination. *Memory & Cognition, 28,* 28–34.

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

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

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. (2004). Effect of discourse context and modifier relation frequency on conceptual combination. *Journal of Memory and Language, 50,* 444–455.

Gagné, C. L., & Spalding, T. L. (2006). Relation availability was not confounded with familiarity or plausibility in Gagné and Shoben (1997): Comment on Wisniewski and Murphy (2005). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1431–1437.

Gerrig, R. J., & Bortfeld, H. (1999). Sense creation in and out of discourse contexts. *Journal of Memory and Language, 41,* 457–468.

Levi, J. (1978). *The syntax and semantics of complex nominals.* New York: Academic Press.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York: Wiley.

Maguire, P., Cater, A. W., & Maguire, R. (in press). The role of experiential knowledge in the interpretation of noun–noun compounds. *Artificial Intelligence Review.*

Maguire, P., Cater, A. W., & Wisniewski, E. J. (2006). The role of ambiguity in the comprehension of noun–noun combinations. In R. Sun & N. Miyake (Eds.), *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (pp. 537–543). Mahwah, NJ: Erlbaum.

Maguire, P., Wisniewski, E. J., & Storms, G. (2007). A corpus analysis of conceptual combination. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of the Second European Cognitive Science Conference* (pp. 419–424). London: Routledge.

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

Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Ramey, C. H. (2005). Is presentation order a confound for modifier–noun compounds? In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (p. 2544). Mahwah, NJ: Erlbaum.

Sadler, D. D., & Shoben, E. J. (1993). Context effects on semantic domains as seen in analogy solution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 128–147.

Storms, G., & Wisniewski, E. J. (2006). Does the order of head noun and modifier explain response times in conceptual combination? *Memory & Cognition, 33,* 852–861.

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

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

Received April 20, 2006
Revision received February 5, 2007
Accepted February 7, 2007 ■