

The role of experience in the interpretation of noun–noun combinations

Phil Maguire · Arthur W. S. Cater · Rebecca Maguire

Published online: 25 August 2007
© Springer Science+Business Media B.V. 2007

Abstract Gagné and Shoben's (J Exp Psychol Learn Mem Cogn 23:71–87, 1997) Competition Among Relations In Nominals (CARIN) theory maintains that the interpretation of modifier-noun combinations is influenced primarily by how the modifying noun has been used in the past. As support for this theory, they found that modifiers typically associated with the instantiated relation are interpreted reliably faster than those whose modifiers are less frequently associated with the relation. The CARIN theory explains this phenomenon by proposing that people store statistical distributions regarding the frequency with which modifying nouns have combined with each relation in the past. However, we maintain that an association between relation frequency and response time does not imply a causal influence. In this study we explore whether the effects observed by Gagné and Shoben were caused by the influence of relation frequency per se. Two experiments were conducted in which experiential knowledge about the modifier was controlled. The first experiment involved combinations whose modifiers were relatively rare and the second involved the presentation of nouns without a modifier-head syntax. In both of these experiments, knowledge about historical modifier usage was irrelevant. Our results show that correlations between modifier preference and response time persist even in situations where a knowledge of the modifier's history is not available. These findings provide converging evidence that the relationship between relation frequency and response time is not a causal one. Instead, an understanding of the relationship between modifier properties and usage, as appropriate to the given context, may be the dominant influence on interpretation in many circumstances. In light of this, we propose an alternative account of the factors influencing ease of interpretation.

P. Maguire (✉)
Department of Computer Science, NUI Maynooth, Co., Kildare, Ireland
e-mail: pmaguire@cs.nuim.ie

A. W. S. Cater
School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland

R. Maguire
Department of Psychology, Dublin Business School, 34/35 South William Street, Dublin 2, Ireland

Keywords Conceptual combination · Noun–noun compounds · CARIN theory · Relation frequency

1 Introduction

In everyday language, noun–noun compounds (also termed conceptual combinations) are frequently used in order to express new ideas and to encapsulate novel concepts and experiences (e.g. *vodka headache*, *phone table*). Compounding is a useful practice in that it greatly enhances the flexibility of language as well as offering an efficient means of communicating concepts for which there is no simple or suitable one-word expression. Such phrases can be exploited when the intended meaning can be derived given the surrounding context as well as that provided by the constituent nouns. However, the interpretation process is often not trivial, requiring an understanding of the addresser's communicative goals and a representation of the situation at hand as well as detailed world knowledge. People have an exceedingly well-developed means of understanding such novel compounds and the phenomenon is potentially revealing since it is intimately associated with language production and comprehension as well as conceptual representation. Indeed, the study of conceptual combination has been referred to as an exemplar of cognitive science (Murphy 1988), converging as it does on many areas of research including linguistics, psychology and artificial intelligence to name but a few.

In English, a language in which compounding is particularly productive, combinations consist of a modifier followed by a head noun. Usually, the head noun denotes the main category while the modifier implies a relevant subcategory or a modification of this set's typical members. In this way, a *mountain flower* is interpreted as a type of flower, and more particularly as one that is located in the mountains. In order to understand a combination like this, people must figure out the most likely way in which the noun *mountain* could be intended as a modification of the noun *flower*. Following this, the combined concept can then be elaborated upon and further inferences can be made (e.g. *mountain flowers* are likely to be wild).

1.1 Models of conceptual combination

Although several theories of conceptual combination have come to the fore in recent years, none of these has managed to encapsulate the phenomenon in a comprehensive manner. The Concept Specialisation model (Murphy 1988) and subsequently the Dual-Process theory (Wisniewski 1997) outline several psychological processes that allow people to successfully interpret novel noun combinations. These theories are centred on a two-stage interpretation process. The first involves a slot-filling mechanism where the modifier is inserted into the head concept to form an interpretation (e.g. in *kitchen chair*, the concept *kitchen* is inserted into the LOCATED slot of the concept *chair*). The second process constitutes an elaborative mechanism whereby world knowledge is used to expand these interpretations. In addition, Wisniewski's (1997) Dual-Process theory suggests a further alignment and comparison mechanism which can account for property-based and hybrid interpretations (e.g. a *robin snake* is a snake with a red breast). While these theories are informative, the description of the interpretation process they provide is quite vague.

The CARIN model (Gagné and Shoben 1997) differs from other theories in that it predicts the ease of the interpretation process rather than predicting the most likely interpretation.

According to CARIN, the interpretation of a novel noun–noun combination occurs when a person identifies a relation between the modifier and the head noun. The underlying principle of the theory is that the ease of interpretation of a combination is influenced by the availability of the relevant relation and that this availability is a function of how the modifier noun has been experienced in the past. For example, *mountain goat* should 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. In order to rationalise why this should be so, the CARIN theory proposes that people store statistical distributions representing how often every modifying concept has been used with each of a limited set of relation types. The theory proposes the existence of 16 possible relations that can be used to link concepts, including relations such as CAUSES, DURING, FOR and ABOUT. Gagné and Shoben (1997) posit that when people interpret a novel combination, the availability of each of the 16 relations is determined by the relation frequency distribution for that modifier. Although previous theories have stressed the importance of both constituents, CARIN clearly emphasises the importance of the modifier over that of the head, with the relation frequency of the head supposedly having little influence on ease of interpretation.

As evidence for their view, Gagné and Shoben (1997) demonstrated that the more frequently a relation is associated with the modifier noun of a combination, the easier it is to judge whether a combination involving that relation is sensible or not. At first blush, this effect seems intuitive. Consider the modifier *chocolate*: combinations in the form *chocolate x* are most easily interpreted using the structure *x MADE OF chocolate*. This reflects the most common instantiation of chocolate as a modifier, namely using the consistency relation. Illustrating this bias, a combination like *chocolate train* is more likely to be interpreted as a train made of chocolate than as a train containing chocolate, despite the latter being perhaps the more plausible. In addition, one would expect compounds in which the modifier is used in an atypical fashion to prove more difficult to interpret: for a compound such as *chocolate magazine*, the consistency sense of *chocolate* must be consciously suppressed, thereby lengthening the interpretation process. Following this rationale, Gagné and Shoben attributed differences in response times observed in a relevant sensibility judgement task to differences in modifier relation frequency.

1.2 Confounding association with causation

Other accounts of Gagné and Shoben's (1997) findings have been proposed which do not necessitate the maintenance of statistical distributions (e.g. Murphy 2002; Wisniewski 1997). These accounts suggest that differences in ease of interpretation arise due to the nature of the modifier's semantic representation and point out that combinational history may be an effect rather than a cause. Although Gagné and Shoben's (1997) results demonstrate an association between modifier history and response time, this finding alone does not require that one variable has a direct influence on the other. A correlation between ice-cream sales and drownings does not imply causation; instead, both are influenced by a more fundamental factor, namely the weather. Similarly, one might expect to find a correlation between the number of times a compound appears in an encyclopaedia and its ease of interpretation. Once again, this is not evidence that the encyclopaedia is influencing people's interpretation process. Instead, the reason for the association is that the overall frequency of the compound itself influences both the number of times it occurs in the encyclopaedia and people's familiarity with it. Thus, a lone correlation between response time and statistical frequency does not necessarily indicate a causal association. In order to establish such a relationship, alternative explana-

tions must be definitively ruled out and a clear mechanism of influence must be established. Specifically, there should exist unambiguous evidence that people are indeed influenced by statistical distributions based on the modifier.

The correlation between relation frequency and ease of interpretation can be construed as a spurious one. Evidently, modifying nouns will have an affinity for certain slots based foremost on the properties of that noun (e.g. *mountain*, being a geographical location, will have an affinity for modifying the locative slot in a wide variety of head nouns). Modifiers will combine more often using relations for which they have an affinity, meaning that similar nouns will combine in similar ways. Therefore, the tendency for *mountain* to combine using the LOCATED relation will reveal itself in any corpus analysis of combinations involving *mountain* as a modifier. In addition, modifiers with a preference for modifying particular slots should be easier to interpret when availing of such preferences, as opposed to when they use a relation that is not as clearly supported (e.g. *mountain x* should be easiest to interpret using the LOCATED relation since the noun *mountain* is clearly supportive of this role). From this perspective, a modifier's relation frequency can be viewed as an epiphenomenon of its semantic representation rather than as an independent variable. This intuitive account offers an alternative explanation for the association between relation frequency and ease of interpretation: when a certain relation is supported by the properties of a modifier it not only occurs more frequently with that modifier but is also easier to interpret.

Gagné and Shoben (1997) make it clear that relation frequencies reflect people's experiential knowledge about how often relations have been used with different nouns (see p. 74, "we suggest that people possess distributional knowledge about how often particular relations are used with modifiers"). However, this notion of stored distributions is problematic. The reliance on historical knowledge alone fails to explain how newly encountered modifiers can be interpreted or how combinations involving idiosyncratic relations can be understood (e.g. *party girl*, *fruit seller*). The theory also fails to acknowledge the possibility that relation distributions are far from random: if a modifier's history is closely linked to its semantic representation, then similar modifiers will display similar relation frequencies. Indeed, Devereux and Costello (2005) have shown that compounds with similar modifiers tend to be interpreted using similar relations, suggesting that relation preferences can be inferred from the properties of a modifier and hence that the storage of individual distributions may be unnecessary. For instance, people will realise that because the modifier *summer* is a time period, the DURING relation is the one most likely to be instantiated. Likewise, modifiers denoting locations or substances will convey similar biases, whether or not they have been previously encountered as a modifier.

Given these issues, we believe it is important to establish how much of the CARIN effect is due to the type of experiential knowledge described by Gagné and Shoben (1997) and how much is due to other factors. The focus of our study is thus on contrasting the CARIN theory with the alternative view that those effects previously attributed to modifier relation frequency should instead be attributed to the influence of modifier properties. The implication of this hypothesis is therefore that the correlation between relation frequency and response time is a spurious one. One important condition for the identification of a causal relationship is that the investigated effect should occur only when the causal factor is present. Therefore, if the correlation reported by Gagné and Shoben (1997) is indeed causal in nature, modifier influence should only be observed when statistical knowledge regarding modifier history is available. In the following experiments we investigate several cases in which modifier frequency information is not available. The purpose of these experiments is to allow us to determine how much of the effect observed by

Gagné and Shoben (1997) is due to statistical knowledge and how much is due to other factors.

1.3 Overview of rationale

The following experiments were designed in order to investigate the influence of the modifier noun in situations where knowledge about its historical usage is not available. The experimental designs were based on the original sensicality judgement paradigm used by Gagné and Shoben (1997). In Experiment 1, we replicated the design of the original study using uncommon modifiers which are rarely encountered. In Experiment 2 we re-used a selection of Gagné and Shoben's original materials and presented these without a combinational syntax. Here, participants were asked to relate individual noun concepts, and with no modifier indicated, knowledge regarding modifier history was again irrelevant. Results from both of these experiments supported the view that ease of interpretation is not directly influenced by modifier relation frequency.

2 Experiment 1

In this experiment we mitigated the influence of modifier experience by examining a set of relatively rare modifiers for which it could be assumed that participants had little probabilistic knowledge. In doing so, we were able to examine the influence exerted by factors other than the modifier's relation frequency distribution. Although CARIN makes no assumptions about the frequency of the modifier per se, the theory is based on the assumption that people maintain statistical knowledge about the modifiers in question. Consequently, in the case of particularly rare modifiers, the theory implies that modifier preference should exert less of an influence.

The experiment involved two conditions, one in which rare modifiers were combined using low frequency relations and the other condition in which the same modifiers were combined using a higher frequency relation. We hypothesised that a comparison of response times in the two conditions would be revealing: a lack of a difference would highlight the importance of experiential knowledge whereas a significant difference would indicate alternative causes for the observed phenomenon. It is worth noting that the high and low frequency conditions for this experiment refer to relative high and low frequencies: since we deliberately selected infrequently used modifiers, the absolute frequency of all relations was inevitably low. Accordingly, the high and low frequency labels used in this experiment are more akin to measures of preference rather than frequency. That is, they reflect how a modifier would behave in an infinitely large corpus. Since the CARIN theory supposes that knowledge about modifier use is obtained through previous exposure, it therefore implies that the relation preference of rare modifiers should not affect ease of interpretation.

2.1 Method

Participants judged whether noun–noun compounds were sensical or not. Half were sensical and half were nonsensical, according to the authors' prior judgements. The modifier relation frequency for the sensical compounds was divided between low and high frequency stimuli. Responses for sensicality judgements were timed.

2.1.1 Participants

Thirty-seven first-year undergraduate students from University College Dublin participated in the study for partial course credit. All were native English speakers.

2.1.2 Materials

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

When selecting materials, we attempted to incorporate a broad variety of concepts. Hence we included a rare time-period, *dusk*, as well as several locations (e.g. *swamp*, *pier*, *tavern*) and a substance modifier, *fudge*. Because modifiers of these types usually combine easily and hence are not rare, such examples proved difficult to find. Other candidates such as *ramadan* or *butane* were considered overly unusual and were not included because they might not have been known to all participants.

Following our selection of 19 modifiers, it was necessary to estimate the relation frequency distributions of each. Although all absolute relation frequencies were deliberately low, we wished to determine which relations were relatively lower or higher. Previous studies have generated relation distributions using several different techniques, including arbitrary pairings (Gagné and Shoben 1997) and corpus analysis (Maguire and Cater 2005). Because our materials were deliberately selected to have a low BNC frequency, a corpus analysis was not viable. Instead we adopted a technique used by Storms and Wisniewski (2006) which involved participant generation. In a stimulus pre-test we presented 18 participants with each of the modifiers and asked them to generate three possible combinations involving that modifier. The relation distributions were derived by ascribing the combinations garnered for each modifier to one of CARIN's 16 relation categories and determining the relative proportion in each. For example, when generating combinations involving *dusk* as a modifier, participants used the DURING relation 87% of the time (see Fig. 1).

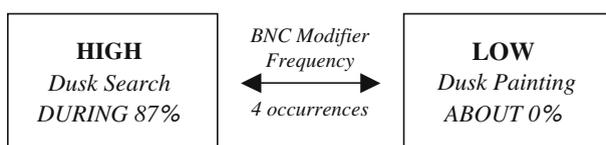


Fig. 1 Example of test stimulus

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

A problem with Gagné and Shoben's classification technique is that it allows for some relatively low relation frequencies to be classified as high frequency (see Maguire et al. 2007 for a detailed discussion). In order to compensate for this, we chose relations that would maintain a substantial gap in frequency between the high and low conditions: the average relation frequency for the high frequency stimuli was 0.53 whereas that for the low frequency stimuli was only 0.07.¹

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

Nineteen pairs of combined concepts were generated, as well as 38 nonsensical filler items (see Appendix A.1). Each pair of materials was controlled for length, plausibility and familiarity. The average number of letters in the high frequency ($M = 12.1$) and low frequency ($M = 12.3$) conditions was not reliably different, $t(18) = 0.38$, $p = 0.71$. In a stimulus pre-test two independent judges rated the plausibility of the 38 sensible materials on a scale of 1–5, where 5 was the most plausible. The judges were explicitly instructed only to evaluate the concept referred to by the combination and not the manner of its expression. The plausibility ratings did not differ reliably between the high frequency ($M = 3.8$) and low frequency ($M = 3.8$) conditions, $t(18) = 0.08$, $p = 0.94$.

Tagalakis and Keane (2006) have demonstrated that the familiarity of combinations has a large influence on response times in sensicality judgements. Although their familiarity ratings were generated by participants, we adopted Maguire and Cater's (2005) method, and assumed that the familiarity of a combination is related to its frequency. This value was then gauged by taking the number of hits generated by a Google search for that combination (see Keller and Lapata 2003). Using log Google hits, the average combination frequency of the high frequency ($M = 2.2$) and low frequency ($M = 2.2$) conditions did not vary, $t(18) = -0.12$, $p = 0.91$.

¹ It is worth noting that these values are more accurately described as relation proportions rather than frequencies. However, for the sake of consistency with Gagné and Shoben's (1997) study we will continue to refer to the percent proportion of a relation's occurrence as its 'relation frequency.'

2.1.3 Design

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

2.1.4 Procedure

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

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

2.2 Results

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

The mean response times were 1435 ms and 1554 ms for the high frequency and low frequency conditions respectively, and the mean accuracy rates were 0.91 and 0.82. A one-way repeated measures ANOVA was conducted to examine the modifier's influence on response times using both participants and items as random factors. The difference in response times between the high frequency and low frequency conditions was reliable across participants and across items, $F_1(1,36)=10.74$, $p < 0.01$; $F_2(1,18)=6.12$, $p = 0.024$. The difference in accuracy rates between these two conditions was also reliable, $F_1(1,36)=21.48$, $p < 0.01$; $F_2(1,18)=11.32$, $p < 0.01$.

These results show significant differences between the high frequency and low frequency conditions. This indicates that both rare and common modifiers exhibit the same phenomenon, in that combinations are interpreted reliably faster when the modifier is more typically associated with the instantiated relation.

2.3 Correlation and regression analyses

We obtained correlations between response time and the following measures: familiarity ($r = -0.42$, $p < 0.01$), relation frequency ($r = -0.38$, $p = 0.02$), plausibility ($r = -0.35$,

$p = 0.03$) and word length ($r = 0.10$, $p = 0.55$). We fitted a stepwise regression model using the data from the 38 experimental items and the three significant predictor variables. Both familiarity and relation frequency entered into the model and the resulting multiple correlation was 0.54. The standardised regression weights for familiarity and relation frequency were -0.38 ($p = 0.01$) and -0.33 ($p = 0.03$) respectively.

Although these results are consistent with the CARIN theory in that they support its premise regarding the association between modifier relation preference and ease of interpretation, they are simultaneously anomalous in that they contradict CARIN's explanation for the alleged influence of this entity. Gagné and Shoben's (1997) correlation between sensicality-judgement response times and relation strength was interpreted as a correlation between an *effect* (speedy judgement) and its *cause* (facilitated interpretation based on knowledge of the modifier's combinational history). However, in this case the modifiers were so unusual that participants should not have been able to rely on experiential knowledge to the same extent. Hence, if this variable were the sole contributor to the phenomenon we would have expected the effect to be significantly reduced. As this was not the case it appears that, in certain cases at least, the association between modifier preference and response time is due to factors other than previous experience of combinations involving the same modifier. The effect must be an effect of a different cause.

One might suggest that perhaps the rarity of the modifiers in question did not preclude the use of experiential knowledge, especially given that our results have reinforced the association between relation preference and response time. We cannot rule out the possibility that unusual modifiers are still associated with some (albeit less developed) level of statistical knowledge. It might also be the case that people can infer the relational frequency of rare modifiers based on their experience of other more common instantiations of the same concept, as suggested by Gagné (2002) (although it is debatable whether more common instantiations of concepts such as *pier*, *swamp* or *kettle* exist). On the basis of the present data, we cannot rule out these arguments conclusively, even if they appear unlikely. This motivated our second experiment in which we presented experimental stimuli in a manner so as to definitively rule out any influence of modifier history.

3 Experiment 2

The most unambiguous method by which to address the role of prior experience in conceptual combination would be to investigate cases in which no such experience is available. If it were practical, one would recruit speakers of a combination-free dialect as participants: individuals encountering combinations for the first time would be required to interpret them in the absence of any statistical knowledge, thereby providing an ideal condition for investigating CARIN's premises. Unfortunately, given the ubiquity of conceptual combination, virtually every language contains combinations of some form, rendering this paradigm infeasible. Another approach would be to present concepts as images, thus avoiding the activation of any linguistic knowledge. Unfortunately, the use of images is unreliable as picture recognition is greatly influenced by canonicity and view specificity. Furthermore, some concepts cannot be represented pictorially: a representation of fruit would invariably refer to some particular types of fruit as opposed to the general concept. Similarly, conveying abstract concepts like *dilemma* or *justice* would be problematic.

In order to sidestep these problems while simultaneously suppressing the activation of experiential knowledge, an alternative method of presentation was adopted. This paradigm was based on the reasoning that statistical knowledge regarding modifiers is irrelevant when

a noun is not acting as a modifier. Since noun concepts are not interpreted as modifiers in the absence of a head, CARIN-style statistical knowledge can have no influence on the interpretation of individual nouns. For example, in reading the noun *chocolate* in isolation, only the regular noun sense is activated. Modifier interpretations such as MADE OF *chocolate* or CONTAINING *chocolate* are considered only when the noun is used in combination with an appropriate head. As a result, *chocolate*'s relation frequency as a modifier does not affect the interpretation of its independent form.

In light of this, we presented noun concept pairs as individual nouns. Our method of presentation ensured that the nouns would not be interpreted as constituents of a combination, and hence that statistical knowledge regarding modifier relation frequency would not be activated. By controlling for the influence of experiential knowledge in this way, we were able to investigate the nature of relation selection when that process is based solely on the conceptual content of the constituent nouns.

3.1 Experimental design

As a reference to the interpretation of noun–noun compounds, the term conceptual combination is potentially misleading since constituent concepts are not necessarily activated in their prototypical form. Compound interpretation is typically guided by the existence of a combinational syntax: in English this requires that the first noun in the pair denotes a relevant modification of the second noun. However, this structure serves to alter the significance of both nouns. Rather than being interpreted as references to individual concepts, paired nouns are interpreted in the context of their contribution to indexing a combined concept. For example, when interpreting a combination such as *gas lamp*, people do not need to simulate the concepts *gas* and *lamp* separately before attempting to relate the two. Instead, people are guided by selective knowledge of the constituent nouns in the context of their unique pairing to single out the relevant combined concept.

In the following experiment we investigated a more basic form of conceptual combination, in that participants were required to simulate the prototypical forms of the constituent nouns before attempting to relate the two in some way. This primitive and undirected combinational process is how we would expect individuals speaking a combination-free dialect to approach such phrases for the first time: without heuristics for deducing compound meaning and without knowledge about how nouns in combination are typically used. The following experiment allowed us to investigate the differences between basic conceptual combination and syntax-guided combination, therefore controlling for the influence exerted by experiential knowledge.

The experimental design was based on the rationale that presenting nouns without a combinational syntax would compel participants to relate the noun pairs by virtue of their semantic representation as opposed to using experiential knowledge. According to the CARIN theory, the differences in ease of interpretation observed by Gagné and Shoben (1997) should only be apparent in cases where statistical knowledge regarding modifier relation frequency is activated. Hence, these differences should be eliminated given the novel method of presentation. In testing this hypothesis we intended to evaluate the claim that such differences arise due to the influence of relation frequency.

Two separate conditions were identified, both involving the same experimental materials. In the combined condition, participants were presented with noun pairs as regular compounds and were required to judge the sensibility of these combinations. In the independent condition, a different group of participants was presented with the same pair of nouns as independent words and they had to decide whether the two could be related in a sensible way. According

to CARIN, the interpretation of a novel noun–noun combination occurs when people identify a relation between the modifier and the head noun. As participants in the independent condition were asked to search for a sensible relation linking the two concepts, the CARIN theory would maintain that participants in both conditions were effectively carrying out the same ultimate task, albeit in different ways. In the combined condition, participants could exploit experiential knowledge as CARIN supposes; in the independent condition, they could not.

3.2 Method

This experiment involved a similar method to that of Experiment 1, in that participants were required to judge whether combinations were sensible or otherwise and responses were timed. Experimental stimuli were divided into three conditions of modifier relation frequency according to the classification used by Gagné and Shoben (1997). One half of participants viewed compounds as combinational phrases; the other half viewed them as stand alone nouns.

3.2.1 Participants

Forty first-year undergraduate students from University College Dublin participated in the study for partial course credit. All were native English speakers.

3.2.2 Materials

For the sake of comparison, we selected a subset of the materials used by Gagné and Shoben (1997) for our experiment. In their study, materials were divided into three categories, namely HH, HL and LH conditions. For these labels, the letters H and L refer to the frequency of the instantiated relation (High or Low), with the first letter denoting how frequently that relation is associated with the modifier and the second denoting how frequently it is associated with the head. Thus *mountain magazine* (a magazine ABOUT mountains) is classified as an LH combination because the ABOUT relation is low frequency for the modifier *mountain* but high frequency for the head *magazine*.

We selected a sample of 10 materials from each of the HH, HL and LH conditions used in Gagné and Shoben's (1997) study. These materials were selected based on several criteria. Any of Gagné and Shoben's modifiers that were adjectives (e.g. *thermal*, *historical*) were excluded from consideration. We also ignored combinations involving ambiguous nouns used in a subdominant sense (e.g. *water plant*). In a re-analysis of the original CARIN study, Wisniewski and Murphy (2005) suggested that the plausibility and familiarity of Gagné and Shoben's materials may not have been controlled, thus confounding response times for the various conditions. Indeed, many of Gagné and Shoben's materials are quite bizarre (e.g. *olive area*, *cooking hole*). In order to control for this, two independent judges divided the original stimuli into ambiguous and unambiguous categories. Any of the combinations selected as ambiguous by either of the judges was excluded. Following this, 10 materials were randomly selected from the remaining items in each condition (see Appendix A.2).

Several measures were taken to ensure that noun pairs presented in the independent condition would not be interpreted as a combination. Firstly, the presentation of nouns in both conditions was separated by a 1,000 ms visual mask, intended to prevent participants inadvertently reading noun pairs as phrases. As a further measure, we took advantage of the fact that modifiers in the English language are nearly always singular: having a plural in the modifier slot violates the weak constraints for a combinational syntax. Thus when two nouns

are presented together and the first is plural, the phrase can typically only be interpreted as two separate nouns and not as a combination (e.g. *mountains+cloud* cannot be read as a single unit). We therefore pluralised a large portion of the filler modifiers in the independent condition (e.g. *dogs+vet*, *sandwiches+filling*). The presence of these pluralised fillers was intended to prevent participants from adopting the strategy of viewing the two nouns as a combination, which might have supported their search for a linking relation.

The experimental materials used in both the combined and the independent conditions were identical. Aside from the instructions, the only other difference between the two conditions was the pluralisation of the filler modifiers. In the independent condition, we used 30 sensical combinations with pluralised modifying nouns, 10 nonsensical combinations with singular modifying nouns and 10 nonsensical combinations with pluralised modifying nouns. In the combined condition, all of the corresponding fillers had singular modifiers.

3.2.3 Design

A 2-way mixed model design was employed, with the two conditions of presentation (independent and combined) as a between-participants measure, and Gagné and Shoben's three conditions of relation frequency as a within-participants measure (HH, HL and LH). The dependent variables were response time and accuracy rate and these were used to infer ease of interpretation. Participants were randomly assigned to one of the between-participants conditions, with a total of 20 in each. Each participant was presented with 80 stimuli, comprising the same set of 30 experimental stimuli for both between-subject conditions and the 50 filler items.

3.2.4 Procedure

As in Experiment 1, participants sat in front of a computer screen and placed the index finger of their left hand on the F key of the computer keyboard and the index finger of their right hand on the J key. The participants in the combined condition were instructed that they would be presented with a series of concept combinations for which they had to make sensicality judgements and that the constituent nouns of the combinations would be displayed one after the other. In contrast, participants in the independent condition were informed that they would be shown two concepts, and that they had to decide whether a sensible relation between the two concepts existed. Both sets of participants were instructed to press J for sense and F for nonsense and emphasis was again placed on the fact that they should only press F if the item was truly incomprehensible. In both conditions, each word was displayed by itself for one second in the centre of the screen, separated by the mask screen lasting for one second. After viewing the second word in the trial, participants had to make a sensicality judgement by pressing the appropriate key. The same materials were presented in both conditions and so the modifying noun was always presented first, although this was not made explicit.

Participants were initially given a short practice session where feedback was given regarding their judgements. The aim of this practice was to set a reliable threshold for sensicality and also to familiarise them with the nature of the task. Participants in the independent condition were shown pairs of concepts with pluralised modifying nouns (e.g. *tomatoes+ sandwich*, *dogs+allergy*). After making a sensicality judgement participants were then informed whether the two nouns could be related and if so, the nature of the relation (e.g. "an allergy caused by dogs"). Similarly, in the combined condition, participants were shown concept

pairs and after making a sensibility judgement, were shown how the combination could be interpreted. Upon completing this practice session, participants were informed that they were beginning the main part of the experiment. The materials were then presented in a random order to each participant.

3.3 Results

A total of 20.5% of trials were omitted from the analysis of the results, 15% in the combined condition and 26% in the independent condition. In 11.9% of combined trials and 23% of independent trials the incorrect response was given and hence these data were excluded. Responses were also eliminated if they were deemed unreasonably fast (0.1% of trials <400ms in the combined condition) or slow (1.7% > 4,000ms in the combined condition, 2.5% > 8,000ms in the independent condition)². After this initial elimination process, any remaining response times which were more than three standard deviations outside each participant's mean were also excluded. This removed another 1.3% of trials in the combined condition and 0.5% in the independent condition. In the combined condition, the mean response times were 1093, 1145 and 1254ms for the HH, HL and LH conditions respectively while in the independent condition, the mean response times were 1,853, 1,873 and 2,213ms. Two ANOVAs were conducted; the first was a 2-way mixed model ANOVA, with presentation method as a between participants measure and relation frequency as a within participants measure. The second was mixed model ANOVA with presentation as a within items measure and relation frequency as a between items measure. Contrary to predictions made by the CARIN theory, there was no significant interaction between method of presentation and relation frequency, indicating that the influence of relation frequency was not affected by method of presentation, $F_1(2, 76) = 1.89, p = 0.16$; $F_2(2, 27) = 0.50, p = 0.61$. The analyses revealed a main effect of relation frequency by participants, $F_1(2, 37) = 7.89, p < 0.01$, but not by items, $F_2(2, 27) = 2.26, p = 0.12$. The lack of significance in the latter case may be attributable to the relatively low number of items in each condition. There was a significant main effect of presentation method, with response times being significantly longer in the independent condition than in the combined condition, $F_1(1, 38) = 15.81, p < 0.01$; $F_2(1, 27) = 92.13, p < 0.01$. Post-hoc analyses carried out using Bonferroni adjustments revealed significant differences between the HH and LH conditions and between the HL and LH conditions by participants (both $ps < 0.01$). The mean accuracy rate for HH, HL and LH in the combined condition was 0.92, 0.87 and 0.75 respectively while in the independent condition the mean accuracy rate was 0.83, 0.76 and 0.64. A second set of ANOVAs revealed no significant interaction between accuracy and method of presentation, $F_1(2, 76) = 0.08, p = 0.93$; $F_2(2, 27) = 0.04, p = 0.96$, indicating that the pattern of accuracy rates was similar for both methods of presentation. As with response time, there was a significant main effect of relation frequency by participants, $F_1(2, 76) = 25.38, p < 0.01$, but not by items, $F_2(2, 27) = 2.83, p = 0.08$. There was also a significant main effect of presentation method, $F_1(1, 38) = 52.39, p < 0.01$; $F_2(1, 27) = 14.07, p < 0.01$. Pairwise comparisons revealed significant differences between the HH and LH conditions and between the HL and LH conditions by participants (both $ps < 0.01$).

These results show significant differences in response time and accuracy rates between different conditions of modifier frequency for both the combined and independent conditions. The differences follow the same pattern as the one revealed by Gagné and Shoben's

² The threshold for identifying extreme outliers in the independent condition was set higher due to the fact that responses in this condition were longer. Hence, the use of the same threshold was not appropriate.

(1997) orthogonal contrasts: the HH and HL conditions were reliably different to the LH condition but there was no reliable difference between HH and HL. The absence of an interaction between presentation method and modifier frequency contradicts the hypothesis that eliminating the alleged influence of statistical knowledge would mitigate differences in ease of interpretation. Given that a similar pattern was observed for both methods of presentation, the implication is that the factors influencing response time and accuracy rate are not related to experiential knowledge of the modifier. Since relation selection in the independent condition could only have been carried out based on the semantic representation of the constituent concepts, it appears that differences in the ease of relating a pair of noun concepts, equally evident in both conditions, are accounted for by an explanation other than that provided by CARIN.

3.4 Correlation analyses

In order to determine the contribution of modifier history towards the overall variance in response time, we obtained correlations between relation frequency and the dependent measures. Each of the experimental stimuli was assigned a value corresponding to the relative frequency with which its modifier was associated with the instantiated relation. We used the same frequencies as Gagné and Shoben (1997), which they derived by pairing 91 heads with 91 modifiers. Considering the 3,239 sensible combinations which emerged from this process, they analysed the relative frequencies with which each modifier combined using the 16 possible CARIN relations. For example, compounds with the *mountain* modifier are typically interpreted using the LOCATED relation (e.g. *mountain cloud*) and according to Gagné and Shoben's frequencies, the relative frequency of this relation for *mountain* is 82%. Using these values, we obtained correlations between response time and relation frequency and between accuracy rate and relation frequency.

An important aim of the experiment was to investigate, as in Experiment 1, whether the correlation between modifier history and ease of interpretation remains when the alleged causal mechanism for that influence is absent. To find correlations in both conditions would provide compelling evidence against a causal relationship of this nature. On the other hand, to find no significant correlation in either would undermine the influence of this factor. We found that neither the correlations between relation frequency and response time or those between relation frequency and accuracy proved significant for either the combined ($r = -0.11$, $p = 0.58$; $r = 0.31$, $p = 0.09$) or the independent conditions ($r = -0.24$, $p = 0.21$; $r = 0.28$, $p = 0.14$). The lack of any significant correlation challenges the importance of modifier relation frequency as a factor in the interpretation process.

Wisniewski and Murphy (2005) have claimed that Gagné and Shoben's (1997) findings may have been due to the confounding of relation frequency with plausibility and familiarity. They demonstrated that Gagné and Shoben's low frequency stimuli happened also to be less plausible and less familiar (e.g. *wood money*) than materials of higher relation frequency (e.g. *plastic toy*). In light of this, it may be that our inclusion of only the more sensible of Gagné and Shoben's (1997) materials may have served to reduce the association between relation frequency and response time. Investigating further, we conducted an analysis of how well plausibility and familiarity predicted response times in our experiment, making use of the ratings obtained by Wisniewski and Murphy (2005) using a group of 30 participants. The correlations between mean response time and familiarity and plausibility were significant for both the combined ($r = -0.61$; $r = -0.56$) and the independent conditions ($r = -0.57$; $r = -0.58$). The correlations between accuracy rate and familiarity and plausibility were also highly significant for the combined ($r = 0.69$; $r = 0.66$) and the independent conditions

($r = 0.72$; $r = 0.73$; all $ps < 0.01$). These findings are consistent with those of Wisniewski and Murphy (2005), revealing that familiarity and plausibility are strong predictors of response time. Since these variables relate to the referent concept rather than its method of communication, they offer an explanation as to why the same pattern of differences was observed in both the combined and independent conditions: stimuli referring to more plausible and more familiar concepts were interpreted reliably faster and more accurately regardless of being presented as a combination or otherwise.

In a further analysis, we investigated how well the ease of finding a relation between two separate noun concepts could predict the ease of interpreting the same concepts presented as a combination. Accordingly, we correlated the dependent variables in the independent and combined conditions. The correlation between response times in both conditions was significant, $r = 0.44$ as was that between accuracy rates, $r = 0.78$ (both $ps < 0.01$). In other words, concept pairs that were easy to interpret as a combination were also easy to relate when presented independently. Also, concept pairs that were often misjudged as nonsense when presented as a combination were also frequently misjudged when presented independently. The above correlations imply that the overtness of the relationship between a pair of concepts is a good predictor of the time taken to interpret those concepts in combination, thereby emphasising the importance of noun conceptual content.

3.5 The influence of familiarity

The influence of familiarity can be manifested in two different ways, namely phrase familiarity or referent familiarity. If the phrase itself is a familiar lexicalised one then the concept it indexes can be activated directly, thus avoiding the combinational process and facilitating interpretation (e.g. *interest rate*). Phrases that have been encountered many times previously will also benefit, even if not fully lexicalised (e.g. *dog basket*). However, our results suggest that the influence of familiarity was not related to phrase familiarity but to referent familiarity. A pair of one-tailed z-tests revealed no significant differences between the familiarity correlation coefficients in the combined condition and those in the independent condition ($z = -0.23$, $p = 0.41$, $z = -0.22$, $p = 0.41$ for response time and accuracy correlations respectively). Since materials in the independent condition were not presented as combinational phrases, this suggests that phrase familiarity was not responsible for the high correlations observed in that condition or by deduction, the combined condition. Instead, the familiarity ratings we used are likely to have reflected the familiarity of the referent concept. For example, a significant facilitating factor in the interpretation of *gas lamp* is the fact that the concept it refers to is highly available. This factor should also facilitate the interpretation of alternative expressions of the same concept (e.g. *fuel lamp*). The influence of referent familiarity is not dependent on a combinational syntax and consequently it can explain why the correlations between familiarity and the dependent measures were similar for both methods of presentation. Given that the combinations used in our experiment were supposedly novel, it seems intuitive that the influence of familiarity should be attributed to conceptual rather than to phrasal effects.

Ratings of concept familiarity, phrase familiarity and concept plausibility are likely to be highly correlated since familiar concepts will undoubtedly be very plausible and by the same token, the combinations that index them will be very familiar. This hypothesis is supported by a high correlation of 0.94 between the plausibility and familiarity ratings for our materials and the high correlations of 0.90 and 0.92 reported by Wisniewski and Murphy (2005) for Gagné and Shoben's (1997) pair of experiments. Given this close relationship, it may ordinarily be very difficult to separate out the relative contributions of these different factors.

However, our experiment has succeeded in demonstrating that for relatively novel stimuli, the main influence of familiarity can be linked to experience of the referent rather than of the phrase.

3.6 Method of interpretation

Given our results, one might object that, despite our efforts, participants in the independent condition were somehow interpreting the modifying nouns in a modifier sense. However, in a post-hoc manipulation check for the independent condition, no participants reported having realised that concept pairs could be viewed as comprising a compound phrase. This position is also supported by the empirical evidence. Firstly, a main effect of presentation condition was observed: response times in the independent condition were significantly longer than those in the combined condition. Moreover, accuracy rates were significantly lower. If participants had been processing the word pairs as combinations then overall differences in response times or accuracy rates would be expected.

One might argue that statistical knowledge about how a noun concept can be used as a modifier is still activated even when that noun is not being interpreted in a modifying role. In order to investigate this possibility we analysed the differences between response times to stimuli in the high modifier frequency conditions (HH and HL) and the sensical pluralised fillers. According to CARIN, combinations in the HH and HL conditions should benefit from statistical knowledge whereas pluralised modifiers can have no associated history since nouns like *tomatoes* are not typically used as modifiers.³ Nevertheless, a post-hoc analysis revealed that participants interpreted the pluralised fillers in the independent condition significantly faster than the high modifier relation frequency stimuli, $t(19) = 2.89$, $p < 0.01$. Setting aside the multitude of variables one would expect to influence the interpretation of different combinations, this is the opposite pattern to that which would be expected had participants been benefiting in any way from the availability of statistical knowledge. It therefore suggests that such knowledge was not a factor for the independent materials.

3.7 Conceptual combination versus syntactical compounding

This experiment has offered a conclusive demonstration that people are capable of relating concepts without those concepts being presented in a combinational form. This should not come as a surprise, since the ability to relate concepts is not one that exists solely for the purpose of interpreting noun–noun compounds. Instead, the capacity of understanding how concepts relate is a key element of cognition and is continually exercised in routine linguistic and reasoning tasks. In order to successfully interact with the environment, people avail of an in-depth appreciation of how different entities interact (e.g. *milk COMES FROM cow*, *fish LIVES IN river*). Accordingly, it seems unreasonable that the cognitive processes involved in conceptual combination would have emerged solely for the purposes of interpreting a narrow range of relatively infrequent linguistic phrases. Instead, we believe that many of these processes are likely to be manifested in diverse areas of human cognition, hence our finding that relation selection is not critically dependent on a combinational syntax.

Given that concepts can be related by virtue of their semantic representation alone, the investigation of conceptual combination raises two separate questions: firstly, how do people relate concepts at the conceptual level and secondly, how does the phenomenon of noun–noun compounding streamline this process at the linguistic level? Although method of presenta-

³ There are rare cases where plural modifiers in compounds are accepted, for example *sports columnist* or *war crimes tribunal*. These are however very rare indeed.

tion did not affect the relative differences in response times between the relation frequency conditions, we found that the materials in the independent condition were interpreted reliably slower than those in the combined condition (1980 ms as opposed to 1164 ms). This disparity suggests that the presence of a combinational syntax greatly enhances the fluency of interpretation. Although the effect may have been partially due to the unnaturalness of the task, the longer response times in the independent condition suggest that the syntactical constraints imposed by having a designated modifier and a designated head are important for facilitating interpretation: knowing which noun is acting as a modifier greatly speeds up the process. Without the clue afforded by syntax, participants in the independent condition may have felt the need to select from among a considerably greater number of possible relationships, some of which may have involved the first of the two concepts acting like a head instead of as a modifier. Furthermore, the constraints imposed by syntax may have streamlined the interpretation process by allowing both nouns to be evaluated in the context of their pairing. Participants in the combined condition may have been able to identify the more relevant features relating to the combination, thereby avoiding the activation of inappropriate information: this process may not have been possible in the independent condition. Considering as an example the experimental item *gas lamp*, participants in the combined condition would have evaluated the concept *lamp* in the context of it being associated with *gas*, thereby tailoring their activation of both concepts. On the other hand, participants in the independent condition had to activate both concepts fully before searching for an appropriate linking relation, presumably instantiating the prototypical electric lamp before being forced to suppress this interpretation. The difference in response times observed between both conditions provides us with an insight into conceptual combination and suggests that this process does not involve the straightforward combination of two individual concepts. Instead, the use of a combinational syntax allows people to constrain the activation of both constituents.

4 General discussion

The results from our experiments have called into question the empirical support for CARIN by challenging the construal of a causal relationship between modifier experience and ease of interpretation. In order to investigate the association between these variables, we considered situations where no causal mechanism for a direct influence was possible. These included the use of extremely rare modifiers as well as presenting concept pairs without any combinational syntax. Although experiential knowledge about how modifiers combine was irrelevant in both these cases, an association between relation frequency and response time persisted. These findings provide converging evidence against a causal relationship and challenge the mechanism of influence proposed by *Gagné and Shoben (1997)*.

The CARIN theory clearly indicates that relation frequencies are derived from experience (e.g. “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). A pertinent question is whether relation frequency can be derived in any way other than through experience. If so, then people would be required to somehow evaluate every noun’s potential to combine. This would involve considerable cognitive effort and since assessments of combinational potential would necessarily be based on conceptual content, it is unclear why this would not be explicitly recognized. The possibility that relation frequency is based on existing conceptual knowledge invalidates the justification for storing such frequencies in the first place. It may be that *Gagné and Shoben (1997)* did not intend that relation frequency should repre-

sent a tangible cognitive entity but instead conceived of it as a hypothetical variable used to constrain and simplify the modelling process. In this case, relation frequency would simply be an epiphenomenon of conceptual content as opposed to constituting a genuine factor in the interpretation process.

Various probabilistic models of human language comprehension have been proposed in the past, based on the idea that probabilistic information about words, phrases and other linguistic structure is represented in the minds of language users and plays a role in language comprehension. Indeed, experiments related to general statistical language models show that humans are very good predictors of word usage (see [Jurafsky 2003](#)). This suggests that prior experience, as well as something akin to frequency distributions, could indeed be a factor in human language processing. Although the current findings challenge the CARIN theory, they certainly do not deny this possibility. In many circumstances, people may well be aware of how particular nouns are typically used in combination. However, the results of our experiments imply that experiential knowledge is unlikely to be manifested in the way that CARIN claims.

We believe that the CARIN theory is too simple to represent a realistic model of human language comprehension. For a start, there is little evidence to suggest that storing one basic statistical distribution per modifier is the best way to encapsulate experiential knowledge. Different contexts surrounding a modifier can alter its significance, and thus the experiential knowledge that is brought to bear in interpreting a combination should be context appropriate and not simply an aggregate of all previous encounters with that same modifier. For example, the modifier *mountain* is typically interpreted using the LOCATED relation but in the case of *mountain height*, this knowledge is irrelevant since *height* cannot have a location: the more appropriate information in this case is knowing how *mountain* is typically used in combination with a measurement. Although a study by [Gagné and Spalding \(2004\)](#) suggested that the CARIN effect occurs independently of context, they used the same head noun in each condition. Given that the head noun is an important contextual component, this is likely to have mitigated the intended variance of context between conditions.

It appears that specialised statistical knowledge is unnecessary in the majority of cases, since the preference of a modifier is usually a function of its properties (e.g. one can infer that *mountain* prefers the LOCATED relation as it refers to a place; [Devereux and Costello 2005](#)). In light of this, it appears unlikely that people would fail to exploit predictable patterns in how modifiers are used (e.g. the knowledge that time periods tend to combine using the DURING relation is not information that needs to be learned and stored separately with every time period modifier). Our study has demonstrated that people are able to relate concepts by virtue of their semantic representation alone, highlighting the importance of conceptual content. Consequently, any theory hoping to offer a comprehensive account of conceptual combination should incorporate this factor in some way. In pursuing an entirely statistical model, CARIN fails to recognise the possibility that relation frequency might simply be an epiphenomenon of conceptual content.

A significant oversight of the CARIN theory is that it fails to incorporate the influence of the head noun, an element that other theories have consistently emphasised (see [Maguire and Cater 2005](#)). Given that the interpretation of any combination inevitably requires the consideration of both constituents, Gagné and Shoben's theory therefore struggles to explain how a low frequency relation can ever be selected. Their model implies that the same relation, namely that most frequent, will always be the one considered for any modifier noun. Thus no matter what head noun *mountain* is paired with, the LOCATED relation should always be the one that comes to mind, even in cases where it is clearly not applicable (e.g. *mountain height*).

The theory fails to explain how any other relation can be selected or how rare modifiers or unique relations can be interpreted (e.g. *fruit seller*).

In summary, although people might avail of knowledge about how nouns have combined in the past, the type of information brought to bear is unlikely to resemble that proposed by the CARIN model. We have shown that the differences in response time interpreted by Gagné and Shoben (1997) as support for the influence of relation frequency are unlikely to have been caused by such.

4.1 Alternatives to CARIN

If the association between relation frequency and ease of interpretation observed by Gagné and Shoben (1997) is not causal, then how can it be explained? Wisniewski and Murphy (2005) suggested that the apparent influence of relation frequency was actually due to differences between the familiarity and plausibility of the stimuli: just as participants were quicker to judge HH and HL stimuli as sensible than the LH stimuli, they also rated the HH and HL phrases as more plausible and more familiar. Maguire et al. (2006) investigated these differences in plausibility and found that they were associated with differences in ambiguity. That is, low modifier relation frequency items were found to be more ambiguous and were interpreted with a variety of relations, therefore invalidating the notion of relation frequency. When Maguire et al. (2006) re-ran Gagné and Shoben's analyses while omitting the most ambiguous combinations, the correlation between modifier relation frequency and response time was no longer significant. As a result, it may be that the CARIN effect is simply an artefact of Gagné and Shoben's stimuli and can be attributed to their choice of combinations.

Despite the evidence supporting the influence of plausibility and familiarity on ease of interpretation, it is worth noting that for Experiment 1 herein, novel stimuli were created, and these unambiguous stimuli were controlled for the above variables. Although familiarity and plausibility emerged again as significant predictors, how can the longer response times for the low modifier frequency condition be explained in this case? We suggest that an additional factor is involved in ease of interpretation which is based on the level of 'relation constraint' exerted by the constituent nouns. We conceive of this measure as reflecting how strongly the pairing of nouns constrains the appropriate relation: the greater the extent to which the properties of the nouns suggest the relationship, the easier the compound phrase is to interpret. Our constraint-based view assumes that, for non-lexicalised phrases, interpretation will be guided by how well the context and the properties of the noun pair constrain the search for an interpretation. This factor integrates the influence of both noun constituents in the context of their pairing as well as incorporating the influence of familiarity and plausibility.

Not only does this constitute a more sensitive measure than the use of a single value for every occurrence of a particular noun, it also explains the association between relation frequency and response time without necessitating a causal mechanism: a modifying noun which strongly constrains the search for a relation because of its properties will also happen to combine more often using those relations which satisfy its constraints. For example, the modifier *summer* strongly suggests the DURING relation because it is a time period, a fact which will reduce the ambiguity and thereby speed up the interpretation of any phrases of the type *summer x* which use this relation. At the same time, *summer* will also happen to combine with this relation more often than with any other relation, since it is the one best supported by that concept's properties. Hence, it is not the statistical knowledge of how *summer* has combined in the past which influences interpretation, it is the knowledge of how

nouns exhibiting such properties are likely to be used as modifiers. For example, in the case of the rare modifier *dusk*, the reason that *dusk search* was interpreted more quickly than *dusk painting* was not because of experiential knowledge about how *dusk* had combined in the past. It was because *dusk*, being a time period, strongly suggested the DURING relation in conjunction with the activity *search*; the head noun *painting*, in its dominant sense as an artefact, did not support such a preference.

5 Conclusion

In recent years, the phenomenon of conceptual combination has attracted a great deal of attention in the area of cognitive science. One of the most prominent models of the combinational process has been the CARIN model, proposed by Gagné and Shoben (1997). In support of this model, they interpreted an observed association between modifier relation frequency and ease of interpretation as evidence that people store statistical distributions regarding the historical frequency with which modifier concepts have combined with different relations. Investigating the association between these variables, we conducted two separate experiments in which experiential knowledge was not available. In both, we observed the same differences in ease of interpretation, thereby challenging the notion of a causal relationship, as suggested in Gagné and Shoben's study.

Our findings undermine one of the central principles of the CARIN theory, namely that people rely on statistical knowledge regarding modifier relation frequency for interpreting combinations. Given that aggregate distributions are redundant in context and that a modifier's preference is often a function of its semantic representation, the processes described by CARIN are unnecessary in many cases. The theory does not recognise the intuitive link between relation frequency and noun properties, meaning that the statistics on which it bases itself may be measured at the wrong level of abstraction. Thus, although the influence of experiential knowledge is probable, it seems unlikely that it should manifest itself in the form advocated by Gagné and Shoben (1997).

In contrast, we have suggested that the influence of experiential knowledge is property specific rather than noun or concept specific and can therefore be generalised based on the ontology or properties of the noun (e.g. time period+event=DURING). Moreover, we believe that the knowledge that is brought to bear in interpreting a combination will be relevant to the context in which it is required, rather than constituting a single value aggregated over all previous exposures. This alternative account of experiential knowledge is capable of explaining the effects observed in our experiments and presents a clear progression from Gagné and Shoben's (1997) model.

We have shown that those effects interpreted as supporting the CARIN theory are unlikely to have been caused by experiential knowledge. Other factors such as conceptual content, familiarity and plausibility are more closely associated with ease of interpretation and, seeing as these have yet to be explored in detail, the emphasis that the CARIN theory places on relation frequency appears unwarranted. Indeed, given the difficulties associated with controlling for the multitude of other more influential variables inherent to conceptual combination, as well as the inevitable variance in participants' historical exposure, it may not be possible to ever satisfactorily demonstrate an unambiguous CARIN effect.

Acknowledgements The authors wish to thank Edward Wisniewski and Suzanne Guerin for valuable contributions and advice. We thank Edward Wisniewski for providing us with the familiarity and plausibility ratings used in Experiment 2.

Appendix A

Appendix A.1 Test stimuli for experiment 1

High frequency	Low frequency
Dusk search	Dusk painting
Fudge dessert	Fudge wrapper
Earwig nest	Earwig poison
Hoover cable	Hoover sound
Mackerel trawler	Mackerel can
Blueberry pancake	Blueberry fungus
Lasagne fork	Lasagne restaurant
Trombone parade	Trombone stand
Kettle switch	Kettle steam
Whiskey flask	Whiskey cellar
Swamp animal	Swamp disease
Magpie claw	Magpie forest
Turnip dinner	Turnip fibre
Pier stroll	Pier timber
Tornado devastation	Tornado shelter
Dentist glove	Dentist torture
Jeep window	Jeep sale
Tavern bench	Tavern road

Appendix A.2 Test stimuli for experiment 2

HH	HL	LH
Cream sauce	Chocolate bird	Summer money
Flu pills	Office headache	College headache
Gas lamp	Paper album	Gas cloud
Party dance	Family money	Sports tension
Sugar scales	Coffee sugar	Water bird
Grain alcohol	Plastic ladder	Fish gland
Murder report	Smoke problem	Mountain magazine
Gas crisis	Finger nerve	Milk virus
Tax pressure	Morning prayers	Plastic crisis
Wood shavings	Winter breeze	Paper equipment

References

- Devereux B, Costello F (2005) Investigating the relations used in conceptual combination. *Artif Intell Rev* 24:489–515
- Corley S, Corley M, Keller F, Crocker MW, Trewin S (2001) Finding syntactic structure in unparsed corpora: the Gsearch corpus query system. *Comput Hum* 35:81–94
- Gagné CL (2002) Lexical and relation influences on the processing of novel compounds. *Brain Lang* 81:723–735
- Gagné CL, Shoben EJ (1997) Influence of thematic relations on the comprehension of modifier-noun combinations. *J Exp Psychol: Learn, Mem Cogn* 23:71–87
- Gagné CL, Spalding TL (2004) Effect of discourse context and modifier relation frequency on conceptual combination. *J Mem Lang* 50:444–455
- Jurafsky D (2003) Probabilistic modeling in psycholinguistics: linguistic comprehension and production. In: Bod R, Hay J, Jannedy S (eds) *Probabilistic linguistics*. MIT Press, Cambridge, MA
- Keller F, Lapata M (2003) Using the web to obtain frequencies for unseen bigrams. *Comput Linguist* 29(3):459–484
- 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*. Erlbaum, Hillsdale, NJ
- Maguire P, Cater A, Wisniewski EJ (2006) The role of ambiguity in the comprehension of noun–noun combinations. *Proceedings of the twenty-eighth annual conference of the cognitive science Society*. Erlbaum, Hillsdale, NJ
- Maguire P, Devereux B, Costello F, Cater AW (2007) A reanalysis of the CARIN theory of conceptual combination. *J Exp Psychol: Learn, Mem Cogn* 33(4):811–821
- Murphy GL (1988) Comprehending complex concepts. *Cogn Sci* 12(4):529–562
- Murphy GL (2002) *The big book of concepts*. MIT Press, Cambridge MA, pp 463–464
- Storms G, Wisniewski EJ (2006) Does the order of head noun and modifier explain response times in conceptual combination?. *Mem Cogn* 33:852–861
- Tagalakis G, Keane MT (2006) Familiarity and relational preference in the understanding of noun–noun compounds. *Mem Cogn* 34(6):1285–1297
- Wisniewski EJ, Murphy GL (2005) Frequency of relation type as a determinant of conceptual combination: a reanalysis. *J Exp Psychol: Learn, Mem Cogn* 31:169–174
- Wisniewski EJ (1997) When concepts combine. *Psychon Bull Rev* 4:167–183