Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 2, 288–297

0278-7393/10/$12.00 DOI: 10.1037/a0018687

The Influence of Interactional Semantic Patterns on the Interpretation of Noun–Noun Compounds

Phil Maguire
National University of Ireland, Maynooth

Rebecca Maguire
Dublin Business School

Arthur W. S. Cater
University College Dublin

The CARIN theory (C. L. Gagné & E. J. Shoben, 1997) proposes that people use statistical knowledge about the relations with which modifiers are typically used to facilitate the interpretation of modifier–noun combinations. However, research on semantic patterns in compounding has suggested that regularities tend to be associated with pairings of semantic categories, rather than individual concepts (e.g., P. Maguire, E. J. Wisniewski, & G. Storms, in press; B. Warren, 1978). In the present study, the authors investigated whether people are sensitive to interactional semantic patterns in compounding. Experiment 1 demonstrated that the influence of a given modifier on ease of interpretation varies depending on the semantic category of the head. Experiment 2 demonstrated that the relation preference of the head noun influences ease of interpretation when the semantic category of the modifier is compatible with that preference. In light of these findings, the authors suggest that people are sensitive to how different semantic categories tend to be paired in combination and that this information is used to facilitate the interpretation process.

Keywords: conceptual combination, language comprehension, statistical linguistic knowledge, in-context interpretation, concept activation, CARIN theory

Conceptual combination is a productive strategy used by people to refer to entities for which no suitable one-word expression exists. In such cases, the combination of two or more existing concepts often provides a succinct means of referencing a novel concept or idea. In addition to the large number of compounds continually entering the vernacular (e.g., *soccer mom*, *tiger economy*), spontaneous combinations are often favored in everyday language because of their concision and flexibility (e.g., *penguin film*, *handbag dog*). In English, compounds consist of a modifier followed by a head noun. The head noun typically denotes the main category, whereas the modifier indicates a specialization of that category. For example, *mountain bird* refers to a type of bird, but, more specifically, it refers to a bird which <lives in> the mountains. Identifying the intended referent of a combination can often be a complex task, requiring an appreciation of the context as well as a detailed representation of the constituent concepts. Nevertheless, people have a natural propensity for deciphering the intended meaning of novel combinations, as evidenced by the fact that children as young as 3 years can produce and understand such phrases (Nicoladis, 2003). The study of how people interpret combinations efficiently and reliably holds the potential to inform theories of conceptual representation and language comprehension.

Phil Maguire, Department of Computer Science, National University of Ireland, Maynooth, Co. Kildare, Ireland; Rebecca Maguire, Department of Psychology, Dublin Business School, Dublin, Ireland; Arthur W. S. Cater, School of Computer Science and Informatics, University College Dublin, Dublin, Ireland.

Correspondence concerning this article should be addressed to Phil Maguire, Department of Computer Science, National University of Ireland, Maynooth, Co. Kildare, Ireland. E-mail: pmaguire@cs.nuim.ie

Theories of Conceptual Combination

A variety of different models of conceptual combination have been proposed (e.g., Costello & Keane, 2000; Estes & Glucksberg, 2000; Gagné & Shoben, 1997; Murphy, 1988; Wisniewski, 1997). 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. The concept specialization model (Murphy, 1988) and dual process theory (Wisniewski, 1997) are centered on a two-stage interpretation process. The first stage involves a slot-filling mechanism where the modifier is inserted into the head concept's schema to form an interpretation (e.g., in the case of *mountain bird*, the concept *mountain* is inserted into the <lives in> slot of the schema for *bird*). The second stage constitutes an elaborative mechanism whereby world knowledge is used to expand these interpretations (e.g., mountain birds prefer cooler climes). Wisniewski's (1997) dual process theory suggests a further alignment and comparison mechanism which can account for property-based and hybrid interpretations (e.g., interpreting a *robin snake* as a snake with a red breast).

Although these schema-based theories make accurate predictions about the type of interpretations that are produced for combinations, it is not clear how people initially identify the correct slot to be modified. According to Murphy (2002), "people use their general background knowledge to choose the slot that seems best" (p. 453). However, the amount of background knowledge associated with any concept is likely to be considerable, and most of it will be irrelevant to interpreting a particular combination. For example, to select the <lives in> relation for *mountain bird*, one does not need to know anything about the formation of mountains or the appearance of birds. Simply knowing that *mountain* is a geographical location and that *bird* is a land-dwelling creature is a sufficient basis for prioritizing this interpretation. This raises the question of whether people can somehow guide the activation of conceptual knowledge so that only the most relevant information is brought to mind.

Studies of noun compounds have indicated that they tend to follow regular patterns, leading linguists to propose that most can be satisfactorily ascribed to a limited set of forms (e.g., Downing, 1977; Levi, 1978). Furthermore, it has been suggested that people might exploit their knowledge of such regularities to streamline the interpretation process and activate conceptual knowledge selectively. For example, Warren (1978) posited that people are able to constrain the range of interpretation of a novel combination by applying their knowledge of how concepts tend to be related, thus facilitating the process of identifying a semantic relationship between a modifier and a head.

Developing this idea, Gagné and Shoben (1997) examined whether statistics describing the frequency of association between nouns and relations could account for differences in the ease with which combinations are interpreted. In their study, they identified a set of 16 possible relations that can be used to connect a modifier and a head (e.g., <made of>, <during>, <for>, <about>). Three sets of combinations were created, whose modifiers and heads were either strongly associated or weakly associated with the appropriate relation. HH combinations involved a relation that was high frequency for both the modifier and head concepts (e.g., *plastic toy*, <made of>). HL combinations involved a relation that was high frequency for the modifier and low frequency for the head (e.g., *plastic equipment*, <made of>), while LH combinations involved a relation that was low frequency for the modifier and high frequency for the head (e.g., *plastic crisis*, <about>). In a speeded sensicality task, Gagné and Shoben (1997) found that high modifier relation frequency combinations were interpreted more quickly and more accurately than low modifier relation frequency combinations. In contrast, the relation frequency of the head had no discernible effect on ease of interpretation. Modifier relation frequency was significantly correlated with response time ($r = .34$), whereas head relation frequency was not.

On the basis of these findings, Gagné and Shoben (1997) developed the competition among relations in nominals (CARIN) theory. The theory proposes that people retain information about how often individual concepts have been used with each relation in the past. These statistics are then used to streamline the interpretation process by allowing a relation to be identified without the need to activate and manipulate detailed representations of the individual constituent concepts. In particular, the CARIN theory emphasizes the importance of the modifier for suggesting an appropriate relation: "Using the modifier's relational distribution to determine a suitable relation may be a means of constraining the amount of elaboration that is needed to obtain a more detailed interpretation of a phrase. . . people can identify that a *mountain bird* is a sensible phrase that uses the relation "noun *located modifier*" (a bird located in the mountains) before knowing in detail what a mountain bird is like" (p. 83, 84).

Evidently, the head must play some role in the interpretation process, or else the same relation would always be selected for a given modifier. Gagné (2002) elaborated on the original CARIN model, proposing that the head noun is used to assess the plausibility of the relations that are suggested by the modifier. For example, the <located> relation should be the first to be evaluated for *mountain planet*, because it is the relation with which *mountain* is most frequently associated as a modifier. This relation is subsequently rejected because the properties of *planet* do not support such an interpretation (planets are too large to be located in mountains), and the next most frequent relation is evaluated. According to this sequential access view, relations continue to be evaluated on the basis of their frequency of association with the modifier, until eventually one is accepted. Gagné and Shoben (2002) and Spalding and Gagné (2007) have described similar sequential models.

Additional Factors Influencing Interpretation

At first blush, the effect the CARIN theory describes seems intuitive. One would expect modifiers associated with a dominant relation (e.g., *plastic*) to be most easily interpreted when used with that relation (e.g., *plastic chair*, *plastic watch*, *plastic knife*). In addition, if people are predisposed to selecting dominant relations, then combinations involving less typical relations should be more difficult to interpret (e.g., *plastic factory*). However, statistics pertaining to the modifier alone may fail to adequately capture many of the salient patterns evident in compounding. For example, even though *plastic* is used with the <made of> relation up to 99% of the time (Maguire, Devereux, Costello, & Cater, 2007), this rule applies only when the head noun denotes a concrete object. Combinations like *plastic texture* or *plastic recycling* cannot be interpreted using the <made of> relation because neither of their modifiers is associated with a material constitution. Statistics pertaining to the modifier alone are too general to capture such details.

The CARIN theory's assumption that people activate the same set of invariant relation frequencies for a given modifier concept, regardless of the nature of the head, may therefore be overly simplistic. Other theories of conceptual combination have stressed that the influence of both modifier and head is a joint interactive one. For example, Estes and Glucksberg (2000) provided evidence that people are sensitive to the interaction between modifier and head features. A combination such as *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 the combination *feather storage* is unlikely to have the same effect, because *storage* does not have weight as a relevant dimension. According to this view, the influence that a given modifier or head has on the interpretation of a combination is dependent on its interaction with the opposite constituent, implying that their influences should not be modeled separately.

Maguire, Wisniewski, and Storms (in press) conducted an extensive corpus study involving combinations taken from the British National Corpus (BNC) and from the World Wide Web. They found that the relation frequency for a given modifier or head varies considerably depending on the semantic category of the opposite constituent. Even modifiers with highly skewed relation frequency distributions can exhibit considerable context-dependent variations in relation use. For example, *chocolate* combines most frequently using the <made of> relation, yet when this modifier is paired with a head of type [attribute] (e.g., *chocolate taste*), the <made of> relation is never used. Because Gagné and Shoben's (1997) measure of relation frequency averages relation incidence over a varied set of combinations, it fails to preserve this kind of information. Consequently, Maguire et al. (in press) found that the use of aggregated relation frequencies for individual nouns was a poor heuristic for predicting relation use.

Corpus studies of combination use have suggested that predictable regularities in compounding can be abstracted to the level of semantic categories, as opposed to being stored for individual concepts. For example, Warren (1978) detailed an extensive list of productive patterns involving different pairings of semantic categories (e.g., [substance–object]). Maguire et al. (in press) also observed that particular pairings of semantic categories tend to exhibit consistent relation preferences. For instance, they found that [time period–event] combinations were predominantly associated with the <during> relation (89%), whereas [area–animal] combinations were predominantly associated with the <located> relation (91%). These relation preferences are consistent because combinations involving the same pairing of semantic categories tend to share similar interactions of modifier and head features. For example, many animal concepts share the potential to be associated with some habitat, whereas many geographical areas share the potential to denote a habitat. As a result, combinations like *sea otter*, *desert rodent*, or *mountain bird* all share the potential to be interpreted using the <lives in> relation. For such combinations, the relation likelihood can be estimated based on the pairing of semantic categories; statistics tailored to the individual constituent concepts are less informative.

In sum, the CARIN theory proposes that the influences of the modifier and head should be modeled separately; it assumes that the influence exerted by a given modifier is a constant one which is not affected by the nature of the opposite constituent. In contrast, corpus studies have indicated that regularities in combination use often center on an interaction between the modifier and head, a view we henceforth refer to as the interactional hypothesis. In the remainder of this article, we present several experiments which contrast the predictions of the CARIN theory with those of the interactional hypothesis.

Experiment 1

Experiment 1 investigated whether a modifier's influence varies depending on the semantic category of the head noun. We contrasted three different conditions involving repeated modifiers with different heads. Each modifier had a dominant relation preference which was either appropriate or inappropriate to the combination. In the high condition, the modifiers were used with high-frequency relations (e.g., *plastic clock*), whereas in the other two conditions, they were used with low-frequency relations. The low-incompatible condition involved combinations with head nouns whose semantic category ruled out the inappropriate relation preference of the modifier (e.g., *plastic texture*; *texture* cannot be <made of> *plastic* because attributes are not associated with a material constitution). In contrast, the low-compatible condition involved combinations whose head nouns did not rule out the inappropriate relation preference of the modifier (e.g., *plastic dye*; being a concrete physical entity, a dye is associated with a material constitution and could therefore potentially be associated with the <made of> relation).

The CARIN theory asserts that the modifier's relation frequency has a direct effect on ease of interpretation, whereas that of the head noun has no effect. According to this view, high combinations (e.g., *leather scroll*) should be associated with the lowest response times and the highest level of accuracy. The CARIN theory does not predict any differences between the low-incompatible (e.g., *leather smell*) and low-compatible (e.g., *leather needle*) conditions, because both involve equally low-frequency modifier relations. According to Gagné's (2002) sequential access model, the same sequence of inappropriate relations will be evaluated in both cases. In contrast, the interactional hypothesis maintains that people are sensitive to the relation preference of the modifier–head pairing as a whole; the relation frequency of the modifier per se is less relevant. According to this view, the low-incompatible combinations should be easier to interpret than the low-compatible combinations because they are less suggestive of the misleading high-frequency modifier relation.

Method

Materials. A set of 20 combinations was generated for each of the three conditions: high, low-incompatible, and low-compatible (see Appendix). To obtain a reliable sample of high and low modifier frequency combinations, statistics were derived from the BNC, a representative sample of English containing over 100 million words of both spoken and written language (see Burnard, 1995). A set of 20 modifier nouns was selected which met the criterion that a 100-combination sample taken from the BNC contained a majority of combinations involving the dominant relation. High combinations used the modifier's dominant relation, whereas the low-incompatible and low-compatible combinations used an alternative relation which opposed the modifier's preference. The head nouns in the low-incompatible condition were selected so that their semantic category precluded the modifier's misleading relation bias (e.g., *paper thickness*), whereas those in the low-compatible condition were selected so that their semantic category was potentially compatible with the modifier's misleading relation bias (e.g., *paper equipment*). The 20 modifiers in each condition were repeated (e.g., *leather saddle*, *leather smell*, *leather needle*). Each participant saw the same set of 120 stimuli, comprising the three conditions of 20 items each and 60 nonsensical filler items.

Wisniewski and Murphy (2005) reported high correlations between participant-generated plausibility ratings and Web frequencies ($r > .90$). Also, Lapata, McDonald, and Keller (1999) found that log-transformed corpus co-occurrence frequencies were the most reliable statistical predictor of a combination's rated plausibility from among a range of predictors. Accordingly, we used Web counts to control for plausibility, taking the log of the Google hit counts returned for each combination. There was no significant difference in frequency, $F(2, 38) = 0.09$, $p = .91$, or in head length, $F(2, 38) = 0.26$, between conditions.

Procedure. The procedure for this experiment was similar to that used by Gagné and Shoben (1997). 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 judgments, pressing *J* for sense and *F* for nonsense. Each trial was separated by a blank screen lasting for 1 s. 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 given a short practice session where feedback was given regarding their judgments. The aim of this practice session was to familiarize them with the process of making quick judgments and also to set a reliable threshold for sensicality. After completing the practice session, participants were informed that the experiment was about to begin. The stimuli were then presented in a random order to each participant.

Participants. Twenty-four undergraduate students participated voluntarily in the experiment. All were native English speakers.

Results and Discussion

In calculating mean response times, a total of 21.8% of the data were eliminated from the analysis: 19.2% of responses were incorrect, and correct responses deemed unreasonably fast ($<$400 ms; 0.3%) or unreasonably slow ($>$6,000 ms; 1.3%) were excluded. Subsequently, any remaining response times which were more than three standard deviations outside each participant's mean for that condition were also eliminated (1.0%).

The mean response times were 1,436, 1,365, and 1,568 ms for the high, low-incompatible, and low-compatible conditions, respectively, and the mean accuracy rates were .79, .90, and .65, respectively. A series of repeated measure analyses of variance (ANOVAs) were conducted to examine the differences between conditions, using both participants and items as random factors. There was a main effect of condition on response time, $F_1(2, 46) = 5.00$, $p = .01$, $MS_e = 39{,}818.44$; $F_2(2, 38) = 3.52$, $p = .04$, $MS_e = 68{,}676.10$. As predicted by the interactional hypothesis, planned pairwise comparisons revealed a reliable difference in response time between the low-incompatible and low-compatible conditions, $F_1(1, 23) = 7.68$, $p = .01$, $MS_e = 51{,}836.67$; $F_2(1, 19) = 5.25$, $p = .03$, $MS_e = 91{,}803.52$. In contrast to the predictions of the CARIN theory, the difference between the high and low-compatible conditions was not significant, $F_1(1, 23) = 1.96$, $p = .18$, $MS_e = 44{,}776.06$; $F_2(1, 19) = 1.36$, $p = .26$, $MS_e = 74{,}528.11$. The difference between the high and low-incompatible conditions was reliable by participants but not by items, $F_1(1, 23) = 4.91$, $p = .04$, $MS_e = 22{,}842.58$; $F_2(1, 19) = 3.61$, $p = .07$, $MS_e = 39{,}696.68$. However, this difference was in the opposite direction to that predicted by the CARIN theory, with the low modifier frequency combinations being judged more quickly.

There was a main effect of condition on accuracy rate, $F_1(2, 46) = 55.53$, $p < .001$, $MS_e = 2.85$; $F_2(2, 38) = 9.55$, $p < .001$, $MS_e = 19.87$. As predicted by the interactional hypothesis, planned pairwise comparisons revealed a reliable difference between the low-incompatible and low-compatible conditions, $F_1(1, 23) = 87.06$, $p < .001$, $MS_e = 3.56$; $F_2(1, 19) = 16.80$, $p < .001$, $MS_e = 22.15$. In line with the predictions of the CARIN theory, there was a reliable difference in accuracy between the high and low-compatible conditions, $F_1(1, 23) = 26.05$, $p < .001$, $MS_e = 1.69$; $F_2(1, 19) = 4.83$, $p = .04$, $MS_e = 10.95$. However, there was also a reliable difference between the high and low-incompatible conditions, $F_1(1, 23) = 36.58$, $p < .001$, $MS_e = 3.29$; $F_2(1, 19) = 5.45$, $p = .03$, $MS_e = 26.51$; this effect was in the opposite direction to that predicted by the CARIN theory, with the low modifier frequency combinations being judged more accurately.

In sum, the low-incompatible combinations were verified more quickly and more accurately than the low-compatible combinations. Indeed, the low-incompatible combinations were interpreted more accurately than the high combinations and just as quickly. These results indicate that a modifier's influence on ease of interpretation varies depending on the nature of the head with which it is paired. For example, although *chocolate* as a modifier usually suggests the <made of> relation, this statistic is irrelevant in the case of *chocolate taste*, because combinations of the type [substance–attribute] are typically associated with the <modifier has head> relation. As a result, *chocolate taste* (low-incompatible) was interpreted more quickly than *chocolate factory* (low-compatible) and *chocolate rabbit* (high; 1,149, 2,370, and 1,658 ms, respectively). Because the CARIN theory assumes that the influence of a modifier on ease of interpretation is constant, it cannot explain the observed differences between the low-incompatible and low-compatible conditions. These findings support the interactional hypothesis, because they show that the influence of the modifier on ease of interpretation is altered by the semantic category of the head.

Modifier Primacy

From the results of Experiment 1, it is clear that the head noun can have a strong influence on the ease of interpretation of a combination. However, Gagné and Shoben's (1997) study did not detect any such influence. They advanced two possible explanations for the lack of a head noun effect. First, they suggested that people might rely more heavily on the modifier to guide relation selection because it is the first constituent to be encountered. However, Storms and Wisniewski (2005) replicated Gagné and Shoben's study in Indonesian, a language in which the ordering of the modifier and head is reversed relative to English. They also failed to observe a head noun influence, thus countering the ordering hypothesis.

A second possibility suggested by Gagné and Shoben (1997) is that the modifier is somehow privileged in determining the relation. Several priming studies have indicated that the function of the modifier noun is more closely associated with the relation than that of the head (e.g., Gagné, 2002; Jones, Estes, & Marsh, 2008; Maguire & Cater, 2004). Although a given head consistently denotes the same category for a wide range of combinations, the form of modification denoted by a modifier is more dependent on the type of relation involved. Nevertheless, this distinction in functional role does not imply that modifiers should be better predictors of relation use. The findings of several psycholinguistic studies have failed to support the modifier primacy hypothesis (e.g., Gagné & Shoben, 2002; Raffray, Pickering, & Branigan, 2007). In addition, the performances of various computational models have argued against the idea that modifiers should have a greater influence on relation selection. For example, studies by Kim and Baldwin (2005) and Rosario and Hearst (2004) both found that the head noun is more reliable than the modifier in predicting relation use.

Thus, neither the ordering hypothesis nor the modifier primacy hypothesis provides a satisfactory explanation for why previous studies did not observe a head noun influence. On the basis of the results of Experiment 1, we suggest that Gagné and Shoben's (1997) failure to anticipate an interaction between modifier and head may have hindered the detection of a head noun effect. Because their study was predicated on the assumption that relation frequency is stored for each concept individually, the manner in which semantic categories were paired was not controlled: Modifiers and heads were matched with no regard to the relational preference of the modifier–head pairing as a whole. The low head relation frequency condition included many high relation frequency modifiers, with the result that, in many cases, the pairing of semantic categories turned out to be supportive of the intended relation.

As an example, we consider the item *chocolate bird*. Gagné and Shoben (1997) found that *bird* as a head noun rarely combines using the <made of> relation (4%). Because the <made of> relation is high strength for *chocolate* as a modifier but low strength for *bird* as a head, the combination *chocolate bird* was included among Gagné and Shoben's HL combinations. However, according to the interactional hypothesis, the relevant statistic is not how *chocolate* or *bird* tends to combine overall. Knowing that *bird* is used with the <located> relation 43% of the time (e.g., *mountain bird*) is not relevant in this case, because *chocolate* is not a geographical location and is thus unlikely to denote the habitat of *bird*. Instead, the most relevant and useful statistic is knowing how combinations of type [substance–object] are usually interpreted. Because such combinations are strongly associated with the <made of> relation, it is not surprising that *chocolate bird* was interpreted more quickly than the HH combination *chocolate utensils* (1,129 ms and 1,267 ms, respectively). Rather than providing evidence of modifier primacy, this observation may instead be indicative of an underlying interactional influence between the modifier and head's semantic categories.

Experiment 2

In the following experiment we investigated whether the relation preference of the head noun can be shown to influence ease of interpretation when the semantic category of the modifier is appropriately controlled. Gagné and Shoben's (1997) choice of stimuli may have been unsuited to demonstrating a head noun influence. First, few of the head nouns they included in their study were associated with a dominant relation (e.g., nouns such as *bird*, *toy*, or *cloud* do not suggest a strong relation preference when considered in isolation). Second, these nouns were paired with modifiers which strongly distorted the relation preference of the phrase as a whole. For example, the interactional relation preferences of combinations such as *chocolate bird*, *mountain cloud*, and *plastic equipment* differ markedly from those of the head nouns per se (which, according to Gagné and Shoben, favor the <located>, <made of>, and <for> relations, respectively).

Addressing these issues, we identified a set of head nouns with strong dominant relation preferences. In the high condition, combinations used the head's preferred high-frequency relation whereas in the low-compatible condition, combinations used an alternative low-frequency relation. Head nouns in the low-compatible condition were paired with a modifier whose semantic category was compatible with the head's dominant relation, so as to avoid altering the relation preference of the modifier–head pairing as a whole. For example, the head noun *shop* has a strong association with the <sells> relation, with 62% of *shop* combinations in the BNC using this relation. However, for this preference to apply, the modifier must denote an entity which could potentially be sold in a shop. Combinations such as *Christmas shop* or *gambling shop* cannot be interpreted using the <sells> relation, as the semantic categories of their modifiers rule out such an interpretation. Accordingly, to maintain the relevance of the head's context-free relation preference, we ensured that the semantic category of the modifier was compatible with that preference (e.g., *coat shop*, *baby shop*; both *coat* and *baby* denote physical objects which could potentially be sold). In addition, stimuli were selected so that the modifier relation frequency of the low-compatible combinations was higher than that of the high combinations. Thus, whereas the interactional hypothesis predicts faster response times and greater accuracy for the high stimuli than for the low-compatible stimuli, the CARIN theory predicts the opposite.

Method

Materials. A set of 20 combined concepts was generated for the two conditions (see Appendix). The 20 biased heads met the criterion that a random 100-combination sample of the BNC contained a majority of combinations involving the dominant relation for that head. The modifiers in both conditions were chosen so that their semantic category was compatible with the head's relation preference. Compatibility was defined on the basis of the range of semantic categories of the modifiers used with the dominant relation in the BNC sample. Specifically, any concept subsumed by the most specific common abstraction of these modifiers' hypernym trees was considered compatible. The WordNet lexical database was used to identify hypernym trees for nouns (see Miller, 1995). For example, the BNC includes the combinations *fish shop* and *plant shop*, both of which involve the dominant <sells> relation for *shop*. The hypernym tree for *fish* is [object] → [organism] → [animal], whereas that for *plant* is [object] → [organism] → [plant]. In this case, the most specific common abstraction for the two concepts is [organism]. Any modifier whose hypernym tree involves [organism] was therefore considered as being compatible with *shop*'s dominant <sells> relation (e.g., *baby shop*).

The log Google frequency of the combinations did not vary significantly between conditions, $F(1, 19) = 0.73$, $p = .40$, nor did the length of the modifiers, $F(1, 19) = 0.10$, $p = .75$. However, the modifier relation frequency of the low-compatible stimuli was significantly higher than that of the high stimuli, $F(1, 19) = 15.17$, $p = .001$.

Participants saw each head only once. Two lists of 20 stimuli were created such that there were an equal number of high and low-compatible items in each list. Participants were presented with the items from one of the two lists, along with 20 nonsensical filler items.

Procedure. The procedure for this experiment was the same as for Experiment 1.

Participants. Sixty-two undergraduate students participated voluntarily in the experiment. All were native English speakers.

Results and Discussion

Incorrect responses (13.9%) and unreasonably fast responses (<400 ms; 0.1%) were omitted from the analyses. There were no remaining slow responses (>6,000 ms). The mean response times were 1,262 and 1,557 ms for the high and low-compatible conditions, respectively. The mean accuracy rates were .95 and .77. Repeated measures ANOVAs were conducted to examine the influence of the head's relation preference on response time and accuracy rate, using both participants and items as random factors. There was a reliable difference in response time between the high and low-compatible conditions, $F_1(1, 61) = 30.74$, $p < .001$, $MS_e = 63{,}593.47$; $F_2(1, 19) = 18.21$, $p < .001$, $MS_e = 47{,}810.41$. The difference in accuracy rate was also reliable, $F_1(1, 61) = 68.19$, $p < .001$, $MS_e = .01$; $F_2(1, 19) = 19.27$, $p < .001$, $MS_e = 14.86$. A stepwise regression model was fitted using the variables combination frequency, modifier length, modifier relation frequency, and head relation frequency. Only head relation frequency, $F(1, 38) = 22.28$, $p < .001$, and modifier length, $F(2, 37) = 14.79$, $p < .001$, entered into the model. We also computed correlations between relation frequency and response time. In contrast to Gagné and Shoben's (1997) findings, head relation frequency was correlated with response time ($r = -.61$, $p < .001$), whereas modifier relation frequency was not ($r = .16$, $p = .32$).

In sum, these results demonstrate that the context-free relation preference of the head noun is associated with differences in ease of interpretation when the semantic category of the modifier is compatible with that preference. Specifically, combinations are more difficult to interpret when the relation preference of the head is misleading and the semantic category of the modifier fails to rule out that misleading bias (e.g., *baby shop*, *ship container*). Indeed, the correlation between head relation frequency and response time ($r = -.61$) was stronger than the correlation reported by Gagné and Shoben (1997) between modifier relation frequency and response time ($r = -.34$). These findings argue in favor of the interactional hypothesis because they reveal that, although the relation preference of the head can exert a strong influence on ease of interpretation, the nature of that influence depends on the semantic category of the opposite constituent. In contrast to the predictions of the CARIN theory, differences in modifier relation frequency were not correlated with differences in ease of interpretation.

Given that Gagné and Shoben (1997) did not control for interactional effects, the question remains as to why they observed a modifier influence but not a head influence. The same observation has been reported in subsequent studies investigating influences on ease of interpretation (e.g., Ramey, 2005; Storms & Wisniewski, 2005). We suggest that this phenomenon may reflect an asymmetry in the extent to which the modifier and head contribute to the interactional relation preference of a combination as a whole. Specifically, the influence of the head may be more context-sensitive than that of the modifier. Because concrete concepts exist in the four dimensions of physical reality, they are associated with a position in space and time and have some material constitution. Modifier concepts that can describe such features are thus capable of modifying the same slot for many different heads. For example, the modifier *plastic* can be used to denote the constitution of a wide variety of concepts (e.g., *knife*, *car*, *chair*). Equally, a modifier like *winter* will consistently suggest the <during> relation, whether paired with an artifact (e.g., *winter coat*), an event (e.g., *winter festival*), a time period (e.g., *winter evening*), a feeling (e.g., *winter depression*), or an act (e.g., *winter shopping*). Consequently, these kinds of modifiers exhibit a very consistent relation preference, one which is usually unaffected by the semantic category of the head noun, except in certain cases where that category happens to preclude the favored relation (e.g., *plastic texture*, *mountain height*, as in Experiment 1).

On the other hand, few heads are associated with a consistent form of modification for a wide range of different modifiers. Head nouns with a dominant relation preference tend to require a particular type of modifier for that preference to be manifested (e.g., *soup* as a head noun combines frequently using the <made of> relation but only involving [food] type modifiers). Unlike modifiers, head nouns rarely maintain the same relation preference across a wide variety of different combinations. This asymmetry in consistency of influence means that for some concepts, the context-insensitive measure of relation frequency (as calculated by Gagné & Shoben, 1997) provides a closer reflection of modifier use than head use, an effect particularly marked for substances, time periods and locations. As it happens, 11 of Gagné and Shoben's (1997) 19 LH combinations involve modifiers of this type. This might explain why they observed a significant correlation between ease of interpretation and modifier relation frequency.

Although a greater proportion of modifiers than heads may be associated with a consistent relation preference, the vast majority of nouns are unlikely to have any consistent preference either as a modifier or as a head. For such concepts, modifier or head relation frequency as generalized over all previous occurrences is uninformative. The results of Experiments 1 and 2 have shown that in such cases, the interaction between the modifier and head must be taken into account to make a reliable prediction regarding the appropriate relation.

Influence of Discourse Context

Most studies investigating conceptual combination have examined out-of-context interpretations, and consequently, the influence of discourse context has tended to be overlooked. It seems intuitive that discourse context must play an important role in how a combination is interpreted. For example, if one were to use the combination *plastic bin* while pointing toward a metal bin full of plastic bottles, this would no doubt enhance the availability of the <for> relation while diminishing the perceived likelihood of the <made of> relation. Gerrig and Bortfeld (1999) argued that the interpretation of combinations in discourse contexts is not affected by their out-of-context interpretations. According to this view, the influence exerted by a given modifier and head depends entirely on the situation in which they are used, as opposed to being fixed. They demonstrated that combinations with highly accessible out-of-context meanings (e.g., *doll smile*) did not experience interference when endowed with different, innovative meanings in discourse contexts (e.g., the smile on the face of a child that receives a doll as a gift). This finding suggests that people might be sensitive to how context affects relation likelihood.

Examining this possibility, we conducted an experiment in which 40 participants were asked to complete a compound phrase given the modifier. Half were presented with the extract "a mountain *x*," whereas the other half were presented with the extract "John was going on a hike. He went down to his local hardware store to buy some mountain *x*." Both groups were told that the extract had been taken from a narrative and that the final word had been replaced with an *x*. They were allowed three guesses as to what the hidden word might be. Previous studies investigating combination interpretation have used the analysis of participant-generated combinations to derive relation frequencies (e.g., Maguire, Cater, & Maguire, 2006; Storms & Wisniewski, 2005). We applied the same technique, ascribing each of the responses to one of Gagné and Shoben's (1997) relation categories to derive a relation probability distribution for the modifier *mountain* as perceived by each of the two groups. These relation distributions are illustrated in Figure 1.

On the basis of the combinations generated by the context-free group, the relation with the highest frequency was <located> with 47%, followed by the <for> relation with 17%. In contrast, every combination generated by the context group involved the <for> relation. These results indicate that people have an implicit understanding of how relation likelihood is affected by discourse context. Thus, if people do rely on statistical knowledge to guide the relation-selection process, it seems likely that such knowledge reflects the interaction between modifier, head and context, as opposed to being represented by a single context-insensitive statistic.

A study by Gagné and Spalding (2004) found that combinations were interpreted more quickly in a supportive context but that high modifier relation frequency stimuli were easier to interpret regardless of context. However, this observation does not support the argument that the modifier should be modeled independent of other contributing factors. It merely demonstrates that representing the modifier's influence using a simplified context-independent statistic may be adequate for observing significant, albeit weakened, effects in certain cases. In light of the current findings, it seems likely that the use of a more detailed context-specific measure would prove a more reliable predictor of ease of interpretation.

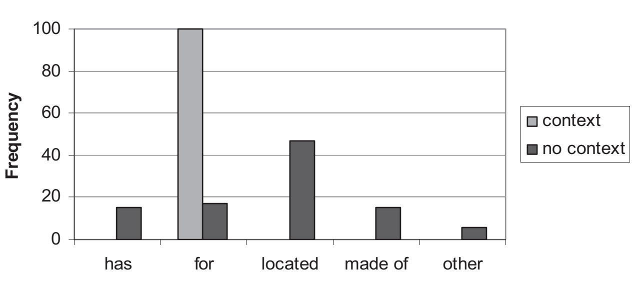

Figure 1. Proportional relation frequencies for combinations generated in context and no-context conditions.

General Discussion

Although schema-based theories (e.g., Murphy, 1988; Wisniewski, 1997) make accurate predictions about the types of interpretations that are produced for combinations, they assume that a full conceptual schema must be activated each time a noun is encountered in the head role. However, in any given case, only a small subset of this information is needed to interpret a combination; activating context-inappropriate features is more likely to impair than to aid comprehension (McElree, Murphy, & Ochoa, 2006). Statistic-based theories, such as Gagné and Shoben's (1997) CARIN theory, are based on the idea that people can activate conceptual knowledge selectively by exploiting regular patterns that exist in compounding, thus streamlining the interpretation process and avoiding the consideration of irrelevant information. There is considerable empirical support for the idea that language users represent probabilistic information about words, phrases, and other linguistic structures and that this information plays a role in language comprehension (see Jurafsky, 1996; Landauer, 2002; MacDonald, 1994; Simpson & Burgess, 1985). Event-related brain potential recordings have shown that people use the content of a sentence to estimate relative likelihoods for upcoming words and that the preactivation of these words reflects the probability of their occurrence (Delong, Urbach, & Kutas, 2005). Given that patterns of combination use exhibit consistent regularities, it seems reasonable that people should represent and exploit these regularities to facilitate the interpretation of combinations.

Nevertheless, the particular form of statistical knowledge suggested by the CARIN theory appears too simple to capture many of the regularities evident in compounding. The experiments described in this article have provided converging evidence that the influence of the modifier and head on ease of interpretation is an interactional one. Specifically, it appears that people rely on the pairing of semantic categories to guide the early stages of processing. In both experiments, combinations were more easily interpreted if the pairing was one frequently associated with the appropriate relation. In contrast, combinations involving a misleading pairing of semantic categories were often misinterpreted. For example, in Experiment 1 participants were liable to misinterpret low-compatible combinations, such as *leather needle*, using the modifier's dominant relation, suggesting that they were responding to the interaction between a [substance] modifier and an [artifact] head, while overlooking the specific features of these concepts which rule out the inappropriate <made of> relation (i.e., leather is soft, whereas needles are sharp).

In light of the current experimental findings, we propose two refinements to the CARIN theory. First, we propose that statistics describing patterns of combination use should be modeled at the level of semantic categories rather than at the level of individual concepts. Concepts in the same semantic category will tend to share similar features and will thus tend to appear in combination using the same kind of relations. This allows relation preferences to be generalized by category as opposed to being maintained for each concept individually. Second, we propose that statistics for predicting relation likelihood must take into account the interaction between modifier and head to be useful. The relation that exists between a pair of concepts is a result of the interaction between those concepts' features. As a result, it cannot be reliably predicted by considering each constituent separately. Averaging relation frequency for a given concept over a wide range of different contexts fails to preserve information about how the use of that concept differs across contexts.

Guided Interpretation Using Statistical Knowledge

The results of a study by Maguire, Maguire, and Cater (2007) provide further evidence that people rely on generalized statistical knowledge to guide interpretation. They contrasted the amount of time taken to reject implausible combinations that belonged to productive and unproductive modifier–head pairings. For example, both *toboggan wheel* and *shirt wheel* are implausible. However, whereas the category [garment–mechanical device] is relatively unusual, the category [vehicle–mechanical device] is a productive one which is often associated with the <has> relation. Maguire et al. found that combinations belonging to productive modifier–head pairings took significantly longer to judge as nonsensical. For example, participants were slower to reject *toboggan wheel* than *shirt wheel* (1,942 ms vs. 1,511 ms), even though toboggans per se have no association with wheels.

Taken together with the findings of Experiments 1 and 2, these findings support the idea that people initially rely on statistical knowledge to guide the interpretation of a combination, as opposed to activating default schemas for the constituent concepts. We suggest that this involves a process of iterative specification, whereby the meaning of each constituent is constructed with respect to the context in which it has been used. As noted by Barsalou (2003), a concept is not a fixed unit of semantic knowledge but rather "a skill for constructing idiosyncratic representations tailored to the current needs" (p. 521). We posit that when people encounter a combination, they begin with generalized representations of the constituent concepts; these representations are gradually specified in a concurrent manner, taking into account the interaction between the words and the context. In this way, statistical knowledge can be used to guide the process of interpretation without needing to activate potentially inappropriate fine-grained features. For example, people will realize that *stone squirrel* matches the pattern [substance–object] and will be guided toward the <made of> relation. With this in mind, they can tailor their representations for *stone* and *squirrel* accordingly and avoid activating inappropriate features like "runs" or "is brown" (cf. McElree, Murphy, & Ochoa, 2006; Swinney, Love, Walenski, & Smith, 2007).

Modeling the Influence of Statistical Knowledge

The idea that people store statistical knowledge about how semantic categories tend to relate raises the question of whether such categories can be clearly delineated. However, ascribing concepts to a single category or identifying a limited set of relationships between categories would risk oversimplifying the specificity of people's linguistic awareness. For instance, the WordNet hierarchy provides the following hypernym tree for the concept *squirrel*: [object] → [organism] → [animal] →[mammal]. Accordingly, a combination like *stone squirrel* could be viewed as simultaneously belonging to both the category [substance–object], which is very productive, but also the category [substance–animal], which is relatively unusual. In addition there is no guarantee that people's statistical knowledge adheres to the organization typically found in a lexical hierarchy. Cater and McLoughlin (2000) developed a computational model for identifying relations based on the positions of modifier and head concepts within WordNet. This model performed relatively poorly because many forms of semantic category were not represented. For example, if the head noun *bag* is modified by a concept that can be contained in a bag, then the resulting combination can be interpreted using the <for> relation (e.g., *mail bag*, *coin bag*, *sweet bag*). However, WordNet does not contain a category for a "collection of small things that can be stored in a bag." Similarly, although *seat*, *mirror*, *chain*, and *brake* can all describe part of a motorbike, these concepts are not grouped together under a single category but are instead scattered throughout the WordNet tree. Cater and McLoughlin concluded that, in many cases, the type of information needed to interpret combinations is unlikely to be reflected by the arrangement of semantic categories in a lexical hierarchy. A comprehensive model would also need to tackle the challenging issue of discourse context. As we have shown, context has a strong influence on perceived relation likelihood, indicating that people are sensitive to how different situations affect the probability of relationships between concepts.

In sum, it should not be assumed that the statistical knowledge people use to facilitate combination interpretation is amenable to straightforward classification; on the contrary, such knowledge is likely to be highly elaborate. Further research is required to clarify the nature of this knowledge and the process by which it is applied.

Conclusion

The CARIN theory fills a significant gap in the understanding of conceptual combination in that it offers an explanation for how a relation can be efficiently identified without needing to consider a vast quantity of background knowledge. The results of our experiments have vindicated the CARIN theory's central premise, namely that people use statistical knowledge about how combinations are typically interpreted to facilitate relation selection. However, they also indicate that the knowledge that people bring to bear in interpreting a combination is more complex than the CARIN theory allows. Experiments 1 and 2 provide converging evidence that the influences of the modifier and head are intertwined and cannot be modeled independent of each other. Although it may be possible to detect an association between modifier relation frequency and ease of interpretation, we have shown that both constituents must be taken into account to provide a fuller picture. The failure of previous studies to detect a head noun effect suggests that the influence of the head may be more context-sensitive than that of the modifier.

We have proposed two refinements that would allow the CARIN theory to account for a greater range of semantic patterns in compounding. First, relation frequency statistics should be centered on semantic categories rather than individual words. Second, these statistics should take into account the interaction between the modifier and head, as opposed to representing the influence of each constituent separately. A more comprehensive model which incorporates such interactional influences may be more successful in accounting for differences in ease of interpretation.

References

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

Burnard, L. (1995). *User's reference guide for the British National Corpus.* Oxford, England: Oxford University Press.

Cater, A., & McLoughlin, D. (2000). Cluster strategy for choice of compound noun interpretation rules. In J. Griffith & C. O' Riordan (Eds.), *Proceedings of the 11th Conference on Artificial Intelligence and Cognitive Science,* 25–34.

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

Delong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117–1121.

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. (2002). Lexical and relational 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., & Shoben, E. J. (2002). Priming relations in ambiguous noun–noun combinations. *Memory & Cognition, 30,* 637–646.

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.

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

Jones, L. L., Estes, Z., & Marsh, R. L. (2008). An asymmetric effect of relational integration on recognition memory. *Quarterly Journal of Experimental Psychology, 61,* 1169–1176.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137–194.

Kim, N. S., & Baldwin, T. (2005). Automatic interpretation of noun compounds using WordNet similarity. In R. Dale & K. F. Wong (Eds.), *Proceedings of the 2nd International Joint Conference on Natural Language Processing* (Vol. 3651, pp. 945–956). Jeju, Korea: Springer.

Landauer, T. K. (2002). On the computational basis of learning and cognition: Arguments from LSA. *Psychology of Learning and Motivation, 41,* 43–84.

Lapata, M., McDonald, S., & Keller, F. (1999). Determinants of adjective–noun plausibility. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 30–36).

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

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9,* 157–201.

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

Maguire, P., Cater, A. W., & Maguire, R. (2006). The role of experience in the interpretation of noun–noun compounds. *Artificial Intelligence Review, 25,* 139–160.

Maguire, P., Devereux, B., Costello, F., & Cater, A. W. (2007). A reanalysis of the CARIN theory of conceptual combination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 811–821.

Maguire, P., Maguire, R., & Cater, A. W. (2007). In search of the frog's tail: Investigating the time course of conceptual knowledge activation. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of the European Cognitive Science Conference 2007* (pp. 728–733). Hove, East Sussex, England: Erlbaum.

Maguire, P., Wisniewski, E. J., & Storms, G. (in press). A corpus study of semantic patterns in compounding. *Corpus Linguistics and Linguistic Theory*.

McElree, B., Murphy, G. L., & Ochoa, T. (2006). Time-course of retrieving conceptual information: A speed-accuracy tradeoff study. *Psychonomic Bulletin & Review, 13,* 848–853.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38,* 39–41.

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.

Nicoladis, E. (2003). What compound nouns mean to preschool children. *Brain and Language, 84,* 38–49.

Raffray, C. N., Pickering, M. J., & Branigan, H. P. (2007). Priming the interpretation of noun–noun combinations. *Journal of Memory and Language, 57,* 380–395.

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.

Rosario, B., & Hearst, M. A. (2004). Classifying semantic relations in bioscience text. In *Proceedings of the ACL 2004* (pp. 430–437). Morristown, NJ: Association for Computational Linguistics.

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 28–39.

Spalding, T. L., & Gagné, C. L. (2007). Semantic property activation during the interpretation of combined concepts. *The Mental Lexicon, 2,* 25–47.

Storms, G., & Wisniewski, E. J. (2005). Does the order of head noun and modifier explain response times in conceptual combination? *Memory & Cognition, 33,* 852–861.

Swinney, D., Love, T., Walenski, M., & Smith, E. E. (2007). Conceptual combination during sentence comprehension: Evidence for compositional processes. *Psychological Science, 18,* 397–400.

Warren, B. (1978). Semantic patterns of noun–noun compounds. In *Gothenburg Studies in English 41, Goteburg: Acta Universtatis Gothoburgensis.*

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.

Appendix

Experiment 1 Stimuli

High		Low-Incompatible		Low-Compatible	
mountain	badger	mountain	height	mountain	range
chocolate	rabbit	chocolate	taste	chocolate	shelf
metal	column	metal	industry	metal	factory
steel	shovel	steel	mining	steel	polish
paper	bracelet	paper	thickness	paper	equipment
timber	ceiling	timber	exports	timber	saw
desert	road	desert	climate	desert	shoes
plastic	clock	plastic	texture	plastic	dye
city	port	city	size	city	book
gold	mask	gold	trade	gold	safe
office	fan	office	layout	office	report
leather	scroll	leather	smell	leather	needle
baby	crocodile	baby	vomit	baby	doctor
male	ostrich	male	tendency	male	razor
village	traffic	village	isolation	village	drawing
liquid	porridge	liquid	flow	liquid	basin
cardboard	package	cardboard	strength	cardboard	scissors
garden	spade	garden	width	garden	magazine
park	scenery	park	preservation	park	photo
winter	journey	winter	duration	winter	memories

Experiment 2 Stimuli

High		Low-Compatible	
tent	hire	party	hire
bracelet	box	shed	box
coat	shop	baby	shop
bleach	stain	curtain	stain
bottle	factory	turkey	factory
wallet	thief	celebrity	thief
wheat	shortage	famine	shortage
potato	stew	family	stew
guitar	collector	museum	collector
carpet	seller	street	seller
cinema	manager	apprentice	manager
drilling	noise	attic	noise
coal	supplier	wedding	supplier
clock	museum	tourist	museum
pheasant	hunter	hound	hunter
dust	allergy	infant	allergy
chin	injury	fist	injury
petrol	container	ship	container
biking	magazine	gift	magazine
pumpkin	sauce	pastry	sauce

Received August 7, 2008
Revision received May 21, 2009
Accepted June 9, 2009 ■