Does the order of head noun and modifier explain response times in conceptual combination?

GERT STORMS

University of Leuven, Leuven, Belgium

and

EDWARD J. WISNIEWSKI

University of North Carolina, Greensboro, North Carolina

We describe a speeded sensibility judgment experiment in which noun–noun combinations in the Indonesian language were used that parallels Gagné and Shoben's (1997) study of combinations in the English language. Like English, Indonesian is read from left to right and contains noun–noun combinations that are formed by juxtaposing the nouns. However, unlike in English, the order of the modifier and the head noun is reversed. This difference between English and Indonesian combinations allowed us to assess whether sensibility judgments of combinations are affected primarily by the left–right order of the nouns or by different functional roles of the nouns (i.e., modifier vs. head noun). As in Gagné and Shoben's study, the modifier's relation frequency contributed significantly to predicting sensibility judgment times in a regression analysis, but the head noun's relation frequency did not. We discuss the implications of this finding for models of conceptual combination.

A very important aspect of human cognition is the ability to combine familiar concepts in novel ways to express new concepts and to refer to new situations. The ubiquitous nature of this ability is illustrated by language production. Virtually every sentence produced by a person reflects a novel combination of concepts. In the past 20 years, a large amount of research has focused on one type of conceptual combination—namely, novel noun–noun combinations consisting of a modifier (i.e., the leftmost noun) and a head noun (the rightmost noun; see, e.g., Costello & Keane, 2000; Estes & Glucksberg, 2000; Gagné & Shoben, 1997; Hampton, 1987; Medin & Shoben, 1988; Murphy, 1988; Wisniewski, 1996, 1997; Wisniewski & Middleton, 2002). Examples of novel combinations of familiar concepts that have appeared recently in print include *ostrich steak* (a steak made out of ostrich meat), *bait car* (a car that is used to catch car thieves),

and *bird flu* (a type of flu caught by eating chickens). Once initially produced, a novel combination may become familiar to a language community so that it becomes part of that language, as in *space shuttle* and *mad cow disease*. In many languages, the creation of such combinations is the primary mechanism that speakers use to expand their language (E. Clark, personal communication, 1995; Downing, 1977).

The present article addresses how people interpret novel combinations. The meanings of most novel combinations involve a relation between the constituents (see Wisniewski, 1997, and Wisniewski & Love, 1998, for evidence). For example, *mountain flower* could mean a “flower located on a mountain,” and *chocolate elephant* could mean “an elephant made of chocolate.” In understanding a novel combination, how do people determine the relation that links its constituents? One possibility is that the discourse context specifies this relation. For example, a story about a candy shop might mention that the shop sold elephants made of chocolate. Later in the story, a person might read that “chocolate elephants are very popular with the kids.” To understand *chocolate elephants*, the reader might recall the relation between elephant and chocolate (i.e., *made of*) that was presented earlier in the story.

In a number of studies, the role that such contexts play in the understanding of novel combinations has been examined (e.g., Gagné & Spalding, 2004; Gerrig & Bortfeld, 1999; Gerrig & Murphy, 1992; Murphy, 1990). For instance, Gerrig and Bortfeld identified novel combinations that had highly accessible out-of-discourse-context meanings (e.g., “a smile on a doll” for *doll smile*) and embedded the combinations in discourse contexts that

This research was partly supported by NSF Grant BCS-9975198, given to the second author. The authors thank Daniel De Amicis for programming the experiment, Lukas Lukas and Nelden Maria Djakababa for finding the Indonesian participants in Belgium, running the experiment, and crucially helping in the translations, N. Nurrachman for coordinating the noun–noun generation study at the University of Indonesia, Jakarta, and Francis Tuerlinckx for his help with the bootstrap test. We also thank Mark Keane and, especially, Arthur Glenberg for their trenchant comments on a previous version of the manuscript. The second author also thanks Bob Dylan for some of the inspiration. G.S. and E.J.W. contributed equally to this article; the order of authorship was determined alphabetically. Correspondence concerning this article should be addressed either to G. Storms, Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: gert.storms@psy.kuleuven.ac.be) or to E. J. Wisniewski, Department of Psychology, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402-6170 (e-mail: edw@uncg.edu).

suggested different, innovative meanings (e.g., “a smile on the face of a child that received a doll as a gift”). They found that highly accessible out-of-discourse meanings did not interfere with the interpretation of these combinations (as compared with combinations without such highly accessible meanings). Evidently, people relied primarily on the discourse context to determine the meaning of a novel combination and did not entertain a highly plausible out-of-context meaning as a possible interpretation.

However, in examining the role of discourse context in the understanding of novel combinations, researchers have used contrived discourse contexts (e.g., Gerrig & Bortfeld, 1999; Gerrig & Murphy, 1992; Murphy, 1990). In all of these studies, contexts have been used that *explicitly* mentioned the constituents of a novel combination and the relation between them *before* the reader encountered the combination (as in the example of *chocolate elephant* described above). However, Wisniewski and Clancy (2004) have examined a large number of novel combinations (744) in naturally occurring discourse contexts from magazine and newspaper articles. In contrast to the contrived contexts used by researchers, a large percentage of these contexts (86%) did not mention both the modifier and the head noun of the novel combination prior to its occurrence. Furthermore, an analysis of a random sample of the novel combinations (296) revealed that their discourse contexts specified their meanings in only 11% of the cases.

These results suggest that writers often coin novel combinations that they believe readers will understand and, thus, do not provide a context that explicitly spells out their meanings. Indeed, Wisniewski and Clancy (2004) also found that the meanings given to out-of-context novel combinations were often the same as the meanings given to those combinations when they appeared in their naturally occurring contexts. In contrast, subjects rarely gave meanings to out-of-context novel combinations that were the same as the meanings given to those combinations when they appeared in experimenter-contrived contexts.

Given that discourse context often does not specify the relation between the constituents of a novel combination, how do people determine that relation? Gagné and Shoben (1997) have provided part of the answer to this question. They suggested that people use distributional knowledge of relations associated with the modifier noun in order to interpret the combination (Gagné & Shoben, 1997). This distributional knowledge reflects the frequency with which the modifier has participated in various relations in previously encountered combinations. For example, when *mountain* occurs as a modifier in familiar combinations, it typically instantiates a location relation (as in *mountain stream*, *mountain resort*, *mountain goat*, etc.) and is only rarely involved in other types of relations (cf. *mountain range*). This idea forms the basis for Gagné and Shoben’s competition among relations in nominals (CARIN) model.

The most important evidence for the CARIN model is the finding that the more frequently a relation is associ-

ated with the modifier noun of a combination, the easier it is to judge that the combination is sensible. In two experiments, Gagné and Shoben (1997) had subjects decide as quickly as possible whether a noun-noun combination was sensible. Various relations linked the modifier and the head nouns together to form sensible interpretations. Gagné and Shoben manipulated the frequency with which these relations were associated with the modifier and head nouns. They found shorter response times for combinations involving a relation frequently associated with the modifier noun than for combinations involving a relation less frequently associated with the modifier. For example, people more quickly judged *mountain bird* as sensible than they judged *mountain magazine* as sensible. The meaning of the former combination involves a locative relation (i.e., “bird *in* the mountains”), whereas the latter combination involves an *about* relation (“magazine *about* mountains”). According to Gagné and Shoben, the former relation is more frequently associated with familiar combinations having *mountain* as a modifier than is the latter relation.

Surprisingly, the frequency with which relations were associated with the head noun did not influence judgment times. For example, the *for* relation is more frequently associated with utensils (when used as a head noun) than is the *made of* relation. However, this difference in frequency did not predict why people were faster to judge *chocolate utensils* (“utensils *for* chocolate”) as sensible than *wood utensils* (“utensils *made of* wood”).

Gagné and Shoben’s (1997) results raise the question of why the frequencies of the relations associated with the modifier, but not those of the head noun, affect the comprehension time for combinations. Gagné and Shoben suggested that: “the modifier may have more of an influence than the head noun because it is encountered first, and therefore, highly frequent relations for the modifier become activated prior to frequent relations for the head noun. If so, then the influence of the modifier may simply be due to syntactic ordering of the nouns: for compounds in English, the modifier always precedes the head noun” (p. 83). However, they also noted that they could not “rule out the possibility that the modifier has some semantic privilege as well” (p. 83). In particular, it is the modifier that specifies how a subtype of a head noun category contrasts with other subtypes. For instance, the modifiers of *shoebox*, *lunchbox*, *toolbox*, *tackle box*, and *sandbox* distinguish subtypes of boxes on the basis of their contents. Likewise, the modifiers of *mud snake*, *pine snake*, *corn snake*, and *water snake* distinguish subtypes of snakes on the basis of habitat. Thus, people may initially focus on the modifier because it is *relevant* to distinguishing the referent of a novel combination from other subtypes of the head noun category (see also Estes & Glucksberg, 1999).

As Gagné and Shoben (1997, p. 83) suggested, an obvious way to test the word order and semantic privilege hypotheses is to examine the interpretation of combinations in a language in which the order of the modifier

and the head noun is the reverse of that in English. The present article adopts this approach by examining the interpretation of combinations in the Indonesian language. Like English, Indonesian is read from left to right, and the language has similar noun–noun combinations that are formed by juxtaposing nouns. However, contrary to English, the head noun precedes the modifier. For instance, bookshop is translated as *toko buku* in Indonesian (literally, “shop book” in English). Consequently, using Gagné and Shoben’s procedure with Indonesian combinations allows us to alter the activation order of the head noun and the modifier and their respective relation distributions, as compared with English. If the activation order of the relation distributions of the head noun and the modifier is crucial for Gagné and Shoben’s finding, response times in Indonesian should be shorter for high-frequency relations of the head noun, with the distribution of relations associated with the modifier having no influence. This finding would support the word order view. If, however, response times are still determined by the relation distribution of the modifier and are independent of the distribution of relations associated with the head noun, the semantic privilege view will be supported.

A second purpose of the article concerns how Gagné and Shoben (1997) determined the relation frequencies of the modifier and head nouns. These frequencies are intended to reflect how often relations have appeared in familiar combinations that people have read or heard. But, as will be detailed below, Gagné and Shoben determined these relation frequencies by examining arbitrarily created combinations, rather than familiar combinations. In Part 1 of the experiment that follows, we will argue that this approach does not have high face validity and will detail an alternative method for computing relation frequencies that does. We used this method in determining the relation frequencies of Indonesian combinations. In Part 2 of our experiment, we will examine the word order and semantic privilege hypotheses. Native Indonesian speakers judged the sensibility of combinations involving relations that varied in how frequently they were associated with the head and modifier nouns.

In the General Discussion section, we will assess the implications of our findings for the CARIN model. In addition, the model does not explain how the relation that forms the basis of a combination’s meaning is selected from among the relations associated with the modifier (e.g., why the *uses* relation would be selected, rather than the *made-of* relation, in determining the meaning of *chocolate chef*). We also will describe plausible selection mechanisms and how they might be integrated into the CARIN model, drawing on schema-based models of conceptual combination (Murphy, 1988; Wisniewski, 1997) and recent work on perceptual symbol systems (Barsalou, 1999; Stanfield & Zwaan, 2001).

EXPERIMENT

In the first part of the experiment, we determined relation frequency, using a measure that had greater face

validity than the one developed by Gagné and Shoben (1997). We first will describe Gagné and Shoben’s measure and its limitations and then will derive our alternative measure. In Part 2 of the experiment, we repeated Gagné and Shoben’s study, using our alternative measure with Indonesian subjects.

In determining the relation frequencies of modifier and head nouns, Gagné and Shoben (1997) selected 91 head nouns and 91 modifiers. The nouns were taken from 100 familiar combinations in the appendix of Levi (1978). (Duplicate nouns were removed.) They then created 8,281 noun–noun combinations by pairing every modifier with every head noun and judged which phrases had sensible meanings. This procedure yielded a corpus of 3,239 sensible combinations, which Gagné and Shoben then classified into the 15 relation categories shown in Table 1. They then determined the frequency with which a relation was associated with a modifier or a head noun by calculating the percentage of sensible combinations involving that noun and relation. For example, for the head noun *cloud*, the *in* relation had a frequency of 13%, since 13% of the sensible combinations with *cloud* as a head noun had interpretations involving the *in* relation.

Gagné and Shoben (1997) then selected sensible combinations with interpretations based on relations that were highly associated with both the modifier and the head noun (HH), highly associated with the modifier but infrequently associated with the head noun (HL), or infrequently associated with the modifier and highly associated with the head noun (LH). They dichotomized a relation (as high or low) for the modifier and for the head noun by using an arbitrary criterion.¹

However, there are reasons to question the validity of the relation frequencies determined by Gagné and Shoben (1997). First, the frequency with which a relation is associated with a noun should reflect how often that relation and noun have appeared in familiar combinations that people have read or heard. But because Gagné and Shoben arbitrarily paired nouns to construct combinations, it is likely that their sample contained some non-negligible proportion of unfamiliar combinations (per-

Table 1
Examples of CARIN’s Basic Relations

Relation	Example
Noun <i>causes</i> modifier	flu virus
Modifier <i>causes</i> noun	financial crisis
Noun <i>has</i> modifier	coal town
Modifier <i>has</i> noun	plant shape
Noun <i>makes</i> modifier	musical light
Noun <i>made of</i> modifier	vegetable cube
Noun <i>for</i> modifier	financial remedy
Modifier <i>is</i> noun	fruit food
Noun <i>uses</i> modifier	floral language
Noun <i>about</i> modifier	plastic crisis
Noun <i>located</i> modifier	urban scandal
Modifier <i>located</i> noun	murder town
Noun <i>used by</i> modifier	servant language
Noun <i>derived from</i> modifier	wood money
Noun <i>during</i> modifier	winter cloud

haps even a significant majority). Thus, one cannot be confident that the relation frequencies derived from such an analysis necessarily mirror the frequencies based on familiar combinations that speakers themselves produce and read in everyday discourse settings.

Second, it is unclear whether Gagné and Shoben's (1997) sample of combinations is broad enough to accurately reflect the relation frequencies associated with a noun. In particular, the same set of 91 head nouns was used to determine the relation frequencies of each modifier noun. Likewise, the same set of 91 modifier nouns was used to determine the relation frequencies of each head noun. However, our intuition is that different modifier nouns often combine with quite different head nouns (and vice versa). For example, head nouns of combinations involving the modifier *chocolate* include *bar*, *cake*, *candy*, and *chip*; the modifier *flu* tends to modify words like *epidemic*, *medicine*, *shot*, and *symptom*. But we are not familiar with phrases such as *flu bar*, *flu cake*, *flu candy*, *chocolate epidemic*, *chocolate medicine*, or *chocolate symptom*. Thus, if modifier nouns tend to combine with different head nouns, one may not be able to accurately determine their relation frequencies from the same set of head nouns. A similar argument applies to the relation frequencies of head nouns.

Third, Gagné and Shoben's (1997) sampling procedure did not address type–token differences between relations (Wisniewski & Murphy, 2005). In particular, some combinations are extremely common, with many occurrences or tokens (e.g., *dog house*, *newspaper headline*). They may have a greater effect on relation frequencies than do less common phrases with very few tokens. At least, frequency of tokens in categorization experiments strongly affects typicality ratings and categorization (Barsalou, Huttenlocher, & Lamberts, 1998; Spalding & Murphy, 1999). However, Gagné and Shoben's analysis gave equal weight to phrases in their corpus that were common and uncommon (e.g., in determining the relation frequencies for the modifier *gas*, the familiar phrase *gas crisis* was given the same weight as the less familiar phrase *gas scandal*). In our study, we will address whether the frequency of tokens affects response times.

Finally, as was noted, Gagné and Shoben (1997) dichotomized the relation frequencies as high or low. However, as a result of this dichotomization, information about differences in frequencies is lost. Therefore, we also conducted regression analyses of response times in Part 2, using nondichotomized frequencies.

Given these concerns, Part 1 of the present experiment determined relation frequencies by using familiar phrases that were generated by the subjects, rather than by using combinations formed by arbitrarily pairing nouns. Such a measure has more face validity than the one developed by Gagné and Shoben (1997). In particular, the subjects generated familiar noun–noun combinations, given either the modifier or the head noun. After classifying all the generated combinations in terms of the relation cat-

egories noted above, we straightforwardly derived relation frequencies through counting.

In Part 2 of the experiment, we used this alternative measure of relation frequency to determine whether Gagné and Shoben's (1997) effect of modifier relation frequency on response time was due to word order or to the semantic privilege of the modifier. Consequently, we used a procedure identical to that in Gagné and Shoben's, which involved the subjects' making sensibility judgments.²

Part 1: Determining Relation Frequencies

Method

Subjects. Fifty-seven students from the psychology faculty of the University of Indonesia, Campus Depok, Jakarta, 19 to 24 years of age, participated in the noun–noun generation task. They were native speakers of Indonesian. The subjects were paid the equivalent of about \$2 for their participation.

Materials. The stimuli were phrases of the form *modifier X* (e.g., *chocolate X*) and *X head noun* (e.g., *X crisis*). The modifiers and head nouns were taken from the 57 test combinations used in Gagné and Shoben's (1997, Experiment 1) speeded sensibility judgment task. These combinations consisted of equal numbers (19) of HH, HL, and LH combinations (based on the English classifications). They were created so that the same modifiers and head nouns appear once in each of the conditions (HH, HL, and LH), making the modifiers and the head nouns equivalent across conditions (see Table 2 for a list of these combinations). Thus, in the present study there were 19 *modifier X* phrases and 19 *X head noun* phrases. Prior to forming these phrases, the combinations were translated by a native Indonesian speaker into the Indonesian language. Another native speaker conducted a cross-translation (from Indonesian back into English) to double check the quality of the translations. Only 3 of the 57 English noun–noun compounds differed from the original set used for the Indonesian translation, and these differences were due to the use of synonyms.

Procedure. The subjects received a booklet with 19 versions of *head noun X* and 19 versions of *X modifier*. (Recall that the order of the modifier and the head noun is reversed, as compared with English.) The subjects were instructed to replace the *X* with a noun that made a familiar combination and to write down as many completions as they could, with a maximum of 10, for any given head noun or modifier. The stimuli were presented in two different random orders (each distributed to half of the subjects) with *X modifier* and *head noun X* items intermixed. The task lasted about 40 min.

Results

Our intuitions were confirmed about how modifier nouns combine with head nouns (and vice versa). That is, a given modifier tended to combine with a large majority of head nouns with which other modifiers did not combine. Specifically, 74.7% of the head nouns that the subjects used to complete the *X modifier* phrases were unique to a given modifier. Likewise, a given head noun tended to combine with a large majority of modifier nouns with which other head nouns did not combine. Specifically, 81.7% of the modifier nouns that the subjects used to complete the *head noun X* phrases were unique to a given head noun. These results contrast with the method that Gagné and Shoben (1997) used to determine relation frequencies. As has been noted, they determined the relation frequencies of each modifier from

Table 2
Noun–Noun Combinations Used in the Experiment in Indonesian
and in English and Their Classifications

album musik	musical album	noun <i>about</i> modifier
antik kertas	paper antiques	noun <i>made of</i> modifier
bahasa rumah	home language	noun <i>located</i> modifier
buku bunga	floral book	noun <i>about</i> modifier
burung gunung	mountain bird	noun <i>located</i> modifier
cahaya kota	urban light	noun <i>located</i> modifier
kota keluarga	family town	noun <i>for</i> modifier
krisis gas	gas crisis	noun <i>about</i> modifier
mainan plastik	plastic toy	noun <i>made of</i> modifier
majalah akademi	college magazine	noun <i>about</i> modifier
peralatan siswa	student equipment	noun <i>for</i> modifier
perbaikan memasak	cooking remedy	noun <i>for</i> modifier
perlakuan kayu	wood treatment	noun <i>for</i> modifier
skandal pembantu	servant scandal	noun <i>about</i> modifier
uang air	water money	noun <i>for</i> modifier
album kertas	paper album	noun <i>made of</i> modifier
antik gas	gas antiques	noun <i>uses</i> modifier
buku pembantu	servant book	noun <i>for</i> modifier
cahaya musik	musical light	noun <i>makes</i> modifier
krisis keuangan	financial crisis	noun <i>caused by</i> modifier
mainan memasak	cooking toy	noun <i>for</i> modifier
majalah siswa	student magazine	noun <i>for</i> modifier
peralatan kayu	wood utensils	noun <i>made of</i> modifier
peralatan plastik	plastic equipment	noun <i>made of</i> modifier
perbaikan rumah	home remedy	noun <i>located</i> modifier
perlakuan akademi	college treatment	noun <i>located</i> modifier
skandal kota	urban scandal	noun <i>located</i> modifier
uang keluarga	family money	modifier <i>has</i> noun
album kota	urban album	noun <i>about</i> modifier
antik kantor	office antiques	modifier <i>has</i> noun
bahasa pembantu	servant language	noun <i>used by</i> modifier
burung air	water bird	noun <i>located</i> modifier
cahaya rumah	home light	noun <i>for</i> modifier
kota music	musical town	noun <i>has</i> modifier
krisis plastik	plastic crisis	noun <i>about</i> modifier
mainan bunga	floral toy	noun <i>located</i> modifier
majalah gunung	mountain magazine	noun <i>about</i> modifier
peralatan keluarga	family utensils	noun <i>used by</i> modifier
peralatan kertas	paper equipment	noun <i>for</i> modifier
perbaikan keuangan	financial remedy	noun <i>for</i> modifier
perlakuan memasak	cooking treatment	noun <i>uses</i> modifier
skandal siswa	student scandal	noun <i>about</i> modifier
uang kayu	wood money	noun <i>derived from</i> modifier

the same set of head nouns. Likewise, they determined the relation frequencies of each head noun from the same set of modifier nouns. However, our findings suggest that the relation frequencies of any given modifier are based on head nouns that differ from those associated with other modifiers. Likewise, the relation frequencies of any given head noun are based on modifiers that differ from those associated with other head nouns.

An unanticipated finding was that the subjects produced a greater number of types for *X modifier* phrases than for *head noun X* phrases ($M = 42.7$ per modifier vs. $M = 30.8$ per head noun). This difference was statistically significant [$t(36) = 3.15, p < .01$]. They also listed a greater number of tokens for *X modifier* phrases than for *head noun X* phrases ($M = 314.2$ per modifier vs. $M = 260.1$ per head noun). This difference approached statistical significance [$t(36) = 1.57, p < .13$]. Evidently, for Indonesian subjects, knowledge about the nouns that combine with modifiers is more readily accessible than

knowledge about the nouns that combine with head nouns. Later, we will discuss the implications of this finding for the word order and semantic privilege hypotheses.

We next calculated relation frequencies for the modifier and the head noun of every test combination used by Gagné and Shoben (1997, Experiment 1). For each familiar noun–noun phrase that was produced in the generation task, two judges independently classified its meaning as one of the relations used by Gagné and Shoben (see Table 1). The judges disagreed on 10.3% of their classifications. These differences were resolved through discussion. For each test combination and its relation, we then computed the relative frequency with which that relation was associated with the modifier noun and with the head noun. For example, the test combination *mountain bird* had the *in* relation. According to the judges, 59.5% of the familiar combinations in which *mountain* was a modifier had meanings that involved the *in* relation. Likewise, 44.4% of the familiar combinations in

Table 3
Average Median Response Times (RTs, in Milliseconds)
for Conditions (Speeded Sensibility Judgment Task),
With Proportions of Trials on Which
a Test Combination Was Judged as Nonsensical

Condition	RT	P
HH	1,188	.25
HL	1,152	.33
LH	1,304	.25
LL	1,369	.37

Note—HH, relations highly associated with both the modifier and the head noun; HL, relations highly associated with the modifier but infrequently with the head noun; LH, relations infrequently associated with the modifier but highly associated with the head noun; LL, relations infrequently associated with the modifier and the head noun.

which *bird* was the head noun had meanings that involved the *in* relation.

We calculated relation frequencies on the basis of both tokens and types. That is, type relation frequencies for the modifier and head noun of a combination were based only on the number of different familiar combinations produced by the subjects. But token relation frequencies for the modifier and head noun were based on the number of different familiar combinations, as well as their frequencies.

We then used the type relation frequencies to classify the test combinations into HH, HL, LH, and LL types, using Gagné and Shoben's (1997) procedure for classifying a relation as H or L (see note 1). The frequencies for the test combinations, based on familiar combinations generated by the Indonesian subjects, differed considerably from the classification obtained by Gagné and Shoben for the same combinations. As a result, only 25 of the 57 combinations were classified as identical types in both classifications. For example, the relation frequencies of the modifier *family* and the head noun *town* in the combination *family town* were .27 and .10, respectively. However, Gagné and Shoben obtained relation frequencies of .24 and .27 for the modifier and the head noun, respectively, of this same combination. Thus, they classified this combination as HH, but we classified it as HL. Furthermore, our classification yielded some LL combinations, although none of Gagné and Shoben's combinations was classified as LL. In our classification, the four types of combinations—that is, HH, HL, LH, and LL—occurred with frequencies equal to 16, 13, 15, and 13, respectively.

There are at least two possible reasons why test combinations had different classifications in Indonesian, as compared with English. First, our method for deriving relation frequencies was different from the one used by Gagné and Shoben (1997). Second, it seems unlikely that two different languages and cultures would have similar distributions of familiar combinations. Thus, relation frequencies would be expected to differ for the same modifiers and head nouns.

Part 2: Speeded Sensibility Judgment Task

Method

Subjects. Thirty Indonesian citizens living in Leuven participated in the speeded sensibility judgment task. They were mostly students from the University of Leuven and their relatives. All were native speakers of the Indonesian language. The subjects were paid the equivalent of about \$7 for their participation.

Materials. The stimuli were the 57 test combinations and 57 filler combinations used in Gagné and Shoben's (1997, Experiment 1) speeded sensibility judgment task. The test combinations consisted of 16 HH, 13 HL, 15 LH, and 13 LL combinations, determined from the data generated in Part 1. The filler combinations did not have sensible meanings. As in Gagné and Shoben, each combination appeared in a sentence at its end. We used the same sentences as Gagné and Shoben (translated into Indonesian).

Procedure. The subjects in the speeded sensibility judgment task sat in a quiet room in front of a Dell portable computer. They placed the index finger of the left hand on the F key of the keyboard and placed the index finger of the right hand on the J key. The keys were labeled so that the key marked *sense* corresponded to the subject's dominant hand and the other key was labeled *nonsense*.

Before the beginning of the experiment, the subjects read the instructions on the screen, and the experimenter answered questions about the procedure. Trial presentation in the experiment was self-paced. The subjects received 24 sensible and 24 nonsensical practice trials, none of which included words from within the experimental stimulus set. There was a short break after the practice trials and halfway through the experimental stimulus set.

Each trial began with the message "Ready?*" on the computer screen. The asterisk marked the spot where the first word of the sentence would appear. The subjects initiated each trial by pressing the space bar. As in Gagné and Shoben (1997), each sentence was displayed in two parts. First, the sentence frame was displayed (e.g., *On Wednesday, Jack thought about the*, translated into Indonesian). The subjects pressed the space bar once they had read and understood the beginning of the sentence. Pressing the space bar removed the sentence frame and displayed an asterisk at the left side of the screen for 250 msec. The asterisk indicated the location of the beginning of the first word of the noun–noun combination. Finally, the asterisk was replaced by a word pair (e.g., *album kertas*, the Indonesian translation of *paper album*), and the subjects indicated whether the word pair made sense by pressing the appropriate key.

Results

We omitted trials of the speeded sensibility judgment task from the response time analysis if either the reading time for the frame or the response time to the noun–noun combination was less than 250 msec. This procedure resulted in 1.3% of the trials being deleted. We also omitted the 30.3% of the trials on which the subjects judged a test combination as nonsensical. (On average, the subjects judged 24.0% of the nonsensible filler combinations as sensible.)

Table 3 lists the average median response time for each test condition and the proportion of trials that were judged nonsensical for each condition (see Ratcliff, 1993, on the appropriateness of using medians in analyzing response times). We conducted a two-way analysis of variance (ANOVA) with median response time as the dependent measure and dichotomized modifier frequency (H or L) and head noun frequency (H or L), based on the Indonesian generation data (see Part 1), as the independent vari-

ables. Consistent with Gagné and Shoben's (1997, Experiment 1) findings, only the effect of the modifier relation frequency was statistically reliable [$F(1,53) = 8.97, p < .01$]. The head noun relation frequency and the interaction between modifier and head noun relation frequencies did not yield significance (both F s < 1). Thus, although the proportion of nonsensicality judgments for the test combinations was relatively high, the pattern of findings mirrored those in Gagné and Shoben. This finding suggests that our subjects may have used a higher threshold to evaluate a combination as sensible but also engaged in processing that was similar to that of Gagné and Shoben's subjects.

We also conducted an ANOVA with proportion of nonsensical judgments as the dependent measure and dichotomized modifier and head noun frequencies as the independent variables. None of the effects was statistically reliable (all F s < 1). Thus, although the proportion of nonsensicality judgments for the test combinations was relatively high, it did not vary systematically with the independent variables.

Correlations and regressions. The *token* relation frequency of the modifier showed a correlation of $-.46$ with the response times, which was significant ($p < .01$). However, the token relation frequency of the head noun showed a correlation of only $-.23$, with the response times, which was not significant. The token relation frequencies of the modifier and the head noun showed a correlation of $-.03$ with each other. The corresponding correlations based on types, instead of tokens, were $-.38$ ($p < .01$), $-.23$ (n.s.), and $-.10$ (n.s.), respectively. These findings are consistent with Gagné and Shoben's (1997) model, which predicts that the relation frequency of the modifying noun, but not that of the head noun, predicts response times in judging the sensibility of noun–noun combinations.

To examine how well our modifier relation frequency measure compared with Gagné and Shoben's (1997) measure, we performed two multiple linear regressions. One used our modifier and head noun type relation frequencies as predictor variables and Indonesian response times as the criterion variable. The other used Gagné and Shoben's modifier and head noun type relation frequencies as predictor variables and their American response times as the criterion variable.³ For the Indonesian analysis, the multiple correlation was $.44$ (with an R^2 of $.19$). The standardized regression weights were $-.40$ ($p < .01$) and $-.23$ (n.s.), respectively. For the American analysis, the resulting multiple correlation was $.34$. The standardized regression weights for the type relation frequency of the modifier and for the head noun were $-.37$ ($p < .02$) and $-.06$ (n.s.), respectively. Thus, our modifier frequency measure accounted for almost twice as much variance in the Indonesian response times (19.27%), as did Gagné and Shoben's modifier relation frequency measure for their American response times (11.97%). However, a bootstrap test (Efron & Tibshirani, 1993) showed that the difference between the R^2 s in the two analyses was not significant ($p = .27$).

A final analysis examined how well token modifier relation frequency compared with type modifier relation frequency in predicting the Indonesian response times. We performed a multiple linear regression, using the modifier and the head noun token relation frequencies as predictor variables and response time as the criterion variable. The resulting multiple correlation was $.48$ (with an R^2 of $.23$). The standardized regression weights for the token relation frequency of the modifier and for the head noun were $-.44$ ($p < .01$) and $-.17$ (n.s.), respectively. Note that the token-based analysis yielded an R^2 of $.23$, which was higher than the R^2 of $.19$, obtained in the type-based analysis. This finding implies that token-based measures explain 3.7% more of the variance in the response times than do type-based measures. However, a bootstrap test (Efron & Tibshirani, 1993) showed that the difference between the R^2 s in the two analyses was not significant ($p = .20$).

GENERAL DISCUSSION

Our findings are consistent with the CARIN model, which suggests that the relation frequency of the modifier, but not that of the head noun, predicts response times in judging the sensibility of noun–noun combinations. Importantly, our results directly address the reason why the modifier has this influence. Gagné and Shoben (1997) provided two possible explanations for the modifier's importance. One explanation, favored by Gagné and Shoben (p. 83), is based on word order: Because English speakers encounter the modifier first in processing a combination but the head noun second, relations associated with the modifier are accessed first and are used as the basis of a combination's interpretation. However, Gagné and Shoben also left open the possibility that the modifier may have certain semantic privileges in determining the meaning of a combination. Our results, when taken together with those of Gagné and Shoben, suggest that modifiers have greater access to relations than do head nouns, regardless of word order. In showing the same effect for modifier frequency with combinations in a language in which the head and the modifier nouns are reversed, our findings support the view that the modifier has semantic privileges over the head noun in determining the interpretation of a combination.

Another piece of evidence for the semantic privilege hypothesis is our finding that Indonesian subjects more readily produced familiar noun–noun combinations when given a modifier (i.e., an *X modifier* phrase to complete) than when given a head noun (i.e., a *head noun X* phrase to complete). This result suggests that the modifier more readily accesses familiar combinations involving that modifier. Hence, it has greater access to the relations that form the basis of interpretations of combinations involving that modifier.

This finding also raises the question of whether such an effect occurs for modifiers in English. To address this question, we examined data from an unpublished study

by the second author. Analogous to our Indonesian study, 56 American subjects were given *modifier X* and *X head noun* phrases and were asked to generate up to four familiar combinations for each phrase.⁴ As in the Indonesian study, the phrases were constructed from the modifiers and head nouns used in Gagné and Shoben (1997, Experiment 1). Despite having only to complete four phrases per noun, the subjects produced a greater number of types for *modifier X* phrases than for *head noun X* phrases ($M = 20.0$ per modifier vs. $M = 17.2$ per head noun). This difference was statistically significant [$t(36) = 2.16, p < .04$]. They also listed a greater number of tokens for *X modifier* phrases than for *head noun X* phrases ($M = 29.6$ per modifier vs. $M = 25.7$ per head noun). This difference was also statistically significant [$t(36) = 3.16, p < .0004$]. Thus, in both Indonesian and English, the modifier more readily accesses nouns with which it has combined than does the head noun.

We also obtained findings that support the CARIN model, using a measure of relation frequency that we suggested has greater face validity than the one used by Gagné and Shoben (1997). In particular, our measure was derived from phrases that were familiar to subjects, rather than from combinations formed by arbitrarily pairing nouns (as in Gagné & Shoben, 1997). As was detailed earlier, this measure reflects a more plausible way of approximating the actual relation frequencies of nouns, since it is based on a sample of the phrases from which these frequencies are hypothesized to arise. As suggestive support for the greater validity of our measure, multiple regression analyses revealed that it accounted for more of the variance in the Indonesian response times than did Gagné and Shoben's measure for their American response times. However, this difference was not statistically reliable.

We also addressed the distinction between relations based on the types of combinations versus the tokens of a combination (i.e., its frequency of occurrence). CARIN's measure of relation frequency is based only on the types of combinations. We found that token relation frequencies of modifiers better predicted response times than did type relation frequencies. However, this difference in prediction was not statistically reliable.

Extending the CARIN model: The importance of a selection mechanism. The CARIN model proposes that in understanding a novel combination, people consider the ways in which the modifier noun has previously combined with other nouns (represented as different relations). Furthermore, the time it takes to interpret a combination by using a particular relation depends on the frequency with which that relation has occurred in previous combinations involving that modifier. Thus, people will interpret *mountain bird* more quickly than *mountain magazine*. The former interpretation involves a *location* relation that has occurred in many previous combinations involving *mountain* as a modifier. In contrast, the latter interpretation involves an *about* relation that has occurred in only a few combinations involving *mountain* as a modifier. Gagné and Shoben's (1997) find-

ings, as well as our own, support these aspects of the model.

However, the CARIN model does not specify how a relation is selected. For example, CARIN does not explain why people typically interpret *mountain magazine* as "a magazine *about* mountains" but not as "a magazine *located on* the mountains." Thus, the model requires a mechanism that decides which potential relation should form the basis for the interpretation of a novel combination.

Schema-based models of conceptual combination provide one candidate selection mechanism. These models suggest that one interprets a noun–noun combination by filling a slot of the head noun with the modifier noun, using a constraint-based process (Murphy, 1988; Wisniewski, 1997). Thus, one might interpret *robin snake* by filling a slot in *snake* (e.g., the slot *eats*) with the modifier concept *robin* to produce the meaning "a snake that eats robins." Importantly, to explain which slot is selected to be filled by the modifier, these models emphasize the importance of constraints on the fillers of a slot. That is, a slot specifies preconditions that a potential filler must meet. For example, the filler of the *eats* slot of *snake* would have to be edible (cf. *rock snake*).

A simple way to integrate the relation frequency aspect of CARIN and the constraint-based aspect of the schema-based views is to assume that these approaches work together to derive an interpretation of a combination. When people encounter a novel combination, the modifier noun activates information about how it has previously combined with other nouns. This information forms the basis for candidate meanings of the combination. (As our findings and those of Gagné & Shoben [1997] suggest, this kind of information is activated much less, or perhaps not at all, by the head noun.) However, the actual meaning of the combination must be selected via a constraint-based process akin to the one proposed by schema-based views.

To illustrate the interaction of these processes, consider how a person might interpret *chocolate chef*. Consistent with the CARIN model, reading the phrase *chocolate chef* may activate the *made of* relation associated with many familiar phrases that have chocolate as a modifier (e.g., *chocolate chips*, *chocolate cake*, *chocolate bar*, and *chocolate ice cream*). However, consistent with schema-based views, a person's schema of *chef* might indicate that a chef is a human being, and knowledge about human beings suggests that they usually cannot be made of chocolate. Thus, the person would not normally prefer the *made of* relation as the basis for the interpretation of *chocolate chef*.⁵ As a result, the process would consider another relation associated with the modifier *chocolate* (e.g., *uses*). It would then determine that this relation yields the plausible interpretation "chef that *uses* chocolate for cooking" for *chocolate chef*. In particular, a person's schema of *chef* provides constraints on what chefs use in cooking (e.g., food, utensils, etc.). A person can then infer from a schema of *food* that chocolate is a type of food and that, therefore, chefs could use chocolate in cooking.

There are other candidate selection mechanisms besides the one based on the traditional schema-based view. Much recent research suggests that our knowledge of the world may not be represented as schemas (in the traditional sense) but as perceptual symbols (for a review, see Barsalou, 1999). In brief, perceptual symbols are neural representations of our multimodal perceptual experiences. Groups of such symbols are organized into a simulator that allows a cognitive agent to construct a simulation of an entity or event in its absence (analogous to simulations involved in mental imagery). For instance, reading about a nail pounded into a wall may activate a simulator with perceptual symbols that allows one to simulate a horizontal nail (see Stanfield & Zwaan, 2001, for evidence).

How might the relation frequency aspect of CARIN be integrated with the perceptual symbols approach? Consider again how a person might interpret *chocolate chef*. Reading *chocolate* may not activate a *uses* relation per se. Instead, it may lead a person to simulate uses of chocolate on the basis of the person's past perceptual experiences, either directly or indirectly, with chocolate. If one of these experiences involved a chef cooking with chocolate, the person could plausibly interpret *chocolate chef* as a "chef that uses chocolate." Alternatively, the person might alter a simulation involving chocolate to include a chef. As indirect evidence for these possibilities, people appear to simulate the referents of adjective-noun combinations (Barsalou, Solomon, & Wu, 1999; Wu, 1995).

Summary and conclusions. Our findings provide further evidence for the view that the modifier has a special influence on how people interpret novel noun-noun combinations. In understanding a novel combination, people focus on how the modifier (rather than the head noun) has previously combined with other nouns, which guides them toward an interpretation. We also showed that this influence is independent of word order and results from the modifier's having semantic privileges over the head noun. At the same time, we argued that interpreting a novel combination also requires a mechanism for selecting one of the ways that the modifier has previously combined with other nouns. We described several candidate selection mechanisms. A hybrid model of this sort could provide a more complete account of conceptual combination.

REFERENCES

- BARSALOU, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, **22**, 577-660.
- BARSALOU, L. W., HUTTENLOCHER, J., & LAMBERTS, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology*, **36**, 203-272.
- BARSALOU, L. W., SOLOMON, K. O., & WU, L. L. (1999). Perceptual simulation in conceptual tasks. In M. K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, typological, and psychological perspectives in cognitive linguistics: The proceedings of the 4th conference of the International Cognitive Linguistics Association* (Vol. 3, pp. 209-228). Amsterdam: Benjamins.
- COSTELLO, F. J., & KEANE, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, **24**, 299-349.
- DOWNING, P. (1977). On the creation and use of English compound nouns. *Language*, **53**, 810-842.
- EFRON, B., & TIBSHIRANI, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- ESTES, Z., & GLUCKSBERG, S. (1999). Relevance and feature accessibility in combined concepts. In M. Hahn & S. C. Stones (Eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (pp. 149-154). Mahwah, NJ: Erlbaum.
- ESTES, Z., & GLUCKSBERG, S. (2000). Interactive property attribution in concept combination. *Memory & Cognition*, **28**, 28-34.
- GAGNÉ, C. L., & SHOEN, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 71-87.
- GAGNÉ, C. L., & SPALDING, T. L. (2004). Effect of discourse context and modifier relation frequency on conceptual combination. *Journal of Memory & Language*, **50**, 444-455.
- GERRIG, R. J., & BORTFELD, H. (1999). Sense creation in and out of discourse contexts. *Journal of Memory & Language*, **41**, 457-468.
- GERRIG, R. J., & MURPHY, G. L. (1992). Contextual influences on the comprehension of complex concepts. *Memory & Cognition*, **17**, 194-207.
- HAMPTON, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, **15**, 55-71.
- LEVI, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
- MEDIN, D. L., & SHOEN, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, **20**, 158-190.
- MURPHY, G. L. (1988). Comprehending complex concepts. *Cognitive Science*, **12**, 529-562.
- MURPHY, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory & Language*, **29**, 259-288.
- RATCLIFF, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**, 510-532.
- SPALDING, T. L., & MURPHY, G. L. (1999). What is learned in knowledge-related categories? Evidence from typicality and feature frequency judgments. *Memory & Cognition*, **27**, 856-867.
- STANFIELD, R. A., & ZWAAN, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, **12**, 153-156.
- WISNIEWSKI, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory & Language*, **35**, 434-453.
- WISNIEWSKI, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, **4**, 167-183.
- WISNIEWSKI, E. J., & CLANCY, E. J. (2004). *You don't need a weatherman to know which way the wind blows: The role of discourse context in conceptual combination*. Unpublished manuscript.
- WISNIEWSKI, E. J., & LOVE, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory & Language*, **38**, 177-202.
- WISNIEWSKI, E. J., & MIDDLETON, E. L. (2002). Of bucket bowls and coffee cup bowls: Spatial alignment in conceptual combination. *Journal of Memory & Language*, **46**, 1-23.
- 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, & Cognition*, **31**, 169-174.
- WU, L. L. (1995). *Perceptual representation in conceptual combination*. Unpublished doctoral dissertation, University of Chicago.

NOTES

1. First, the highest frequency relation was determined for a modifier (or head noun). If that relation was involved in at least 60% of the sensible combinations for that constituent, it was classified as a high-frequency relation for that constituent. (All other relations involving that constituent were considered low-frequency relations.) If not, the process was repeated with the next highest frequent relation until the se-

lected relations accounted for 60% of the interpretations for that constituent. These relations were classified as high frequency, and all remaining relations were considered low frequency.

2. The reader may wonder about the appropriateness of sensibility judgments as a measure of how people understand novel combinations. Evidence suggests that sensibility judgments reflect how subjects understand novel combinations. For example, Murphy (1990) found that response times for novel noun–noun combinations were slower than those for atypical adjective–noun combinations that, in turn, were slower than those for typical adjective–noun combinations. This pattern of findings occurred for sensibility judgments, as well as for times to read sentences containing these combinations. Gerrig and Bortfeld (1999) found that response times were faster for noun–noun combinations having highly accessible meanings than for combinations having less accessible meanings, regardless of whether subjects made sensibility judgments or generated interpretations. Also, in an unpublished study using American subjects and the combinations from Gagné and Shoben's (1997) first experiment, the subjects either made sensibility judgments or generated interpretations of the combinations. The correlation between these response time measures was .52.

3. We thank Christina Gagné for providing us with the modifier and head noun type relation frequencies and response times from Experiment 1 of Gagné and Shoben (1997).

4. The subjects were undergraduates from Northwestern University, who participated as part of a course requirement. The *modifier X* and *X head noun* phrases were constructed from the stimuli used in Experiments 1 and 3 of Gagné and Shoben (1997), with duplicate stimuli removed. The phrases were then randomly divided into seven lists with eight *modifier X* phrases and eight *X head noun* phrases (and one list containing the remaining seven *modifier X* phrases and five *X head noun* phrases). Each subject received one of the lists, and 8 subjects completed each type of list.

5. Given an appropriate discourse context, a noun such as *chef* could refer to a representation of a chef, rather than to a real chef. In this case, the context would suggest that the *made of* relation is the preferred relation.

(Manuscript received May 6, 2003;
revision accepted for publication August 25, 2004.)