Chapter 6

Production of multiword referential phrases: Inclusion of over-specifying information and a preference for modifier-noun phrases

Christina L. Gagné
Department of Psychology, University of Alberta

Thomas L. Spalding
Department of Psychology, University of Alberta

J. Claire Burry
Department of Psychology, University of Alberta

Jessica Tellis Adams
KidsAbility Centre for Child Development, Ontario

We examined the underlying psycholinguistic and cognitive factors that give rise to the production of multiword expressions. For example, if a story describes a woman buying a dog with blue fur, will people include the color of the dog when referring to the animal and, if so, in what syntactic form? In the experiment, participants read short stories that contained a concept that was presented as either a modifier-noun phrase (e.g., *the blue dog*) or full phrase (e.g., *the dog that was blue*). We also varied whether the property being highlighted was normal (e.g., *brown*) or distinctive (e.g., *blue*) for the head noun concept (e.g., *dog*). We found that participants are more likely to include distinctive properties than normal properties when referring to the concept. Although the selection of a syntactic form was partially influenced by the form of the information in the story, there was a strong overall bias toward using a modifier-noun phrase structure.

Christina L. Gagné, Thomas L. Spalding, J. Claire Burry & Jessica Tellis Adams. 2020. Production of multiword referential phrases: Inclusion of over-specifying information and a preference for modifier-noun phrases. In Sabine Schulte im Walde & Eva Smolka (eds.), *The role of constituents in multiword expressions: An interdisciplinary, cross-lingual perspective*, 155–178. Berlin: Language Science Press. DOI:10.5281/zenodo.3598564

1 Introduction

1.1 Aim and background

The aim of this chapter is to explore when and how multiword expressions are used within a referential context. In particular, we focus on production of referential expressions and examine what drives the inclusion of modifying information and the syntactic form of the expression. When referring to an object, person, or event, a speaker/writer is faced with the challenge of assigning linguistic labels to conceptual entities; often, several linguistic expressions can be used. For example, the same object can be referred to as *cup*, *ceramic cup*, or *cup that is made of ceramic*. What influences this decision? Two aspects of forming a referential expression are particularly relevant and will be the focus of our investigation. First, the speaker/writer might or might not include modifying information in the referential expression. Second, if modifying information is included, the expression might be a compound (e.g., *ceramic cup*) or a full noun phrase (e.g., *cup that is made of ceramic*). Although it is tempting to think of these as two separate ordered decisions (first decide whether or not to modify, then decide the form of the modification) we should note that these two aspects are not necessarily deliberate, conscious choices, nor need they be, strictly speaking, independent or sequential. Rather, the ultimate form of the expression may reflect underlying cognitive processes carried out within the language system that, working together, give rise to the form of the expression, and hence to both the syntax and the presence (or not) of modifying information.

Much of the existing work on compounds and modifier-noun phrases has focused on compound access and interpretation. The current study takes a different approach to this problem. Rather than focusing on the interpretation per se, we examine production to identify some of the expectations and biases that human users have about the use of modifying information during referential communication. When using referential expressions, speakers/writers attempt to establish both semantic co-ordination and lexical co-ordination with the addressee (e.g., Clark & Wilkes-Gibbs 1986; Garrod & Anderson 1987; Clark & Schaefer 1989). An attempt is made to synchronize the underlying mental model of the current situation as well as the specific expressions that are applied to particular entities within that model. In doing so, the speaker/writer draws on many different types of knowledge, including world knowledge, knowledge about information expressed in the conversation/discourse, and knowledge about linguistic conventions. Identifying the expectations that people have about the use of multiword expressions provides insight into how people are conceptualizing both the entities denoted by these constructions and the scenarios in which the constructions

are or should be used. Consequently, this area of research has implications for a variety of areas within the psycholinguistic and linguistic literature. In particular, the current project contributes to research that examines the contribution of the individual constituents to the understanding of the meaning of the whole expression, and the appropriateness of the use of the whole construction in a given situation.

The semantic transparency of the constituents of a compound has been a widely studied aspect of compound processing (Libben 1998; Jarema et al. 1999; Gagné & Spalding 2016; Smolka & Libben 2017). In general, compounds with opaque constituents (e.g., *humbug*) are more difficult to process than compounds with transparent constituents (e.g., *schoolyard*). Of course, in creating a multiword referential phrase that is new (as opposed to a known compound word, for example), the constituents will need to be relatively transparent in order to provide the information that would allow the communicative task to be successfully completed. However, at any given level of transparency there are other aspects that will influence whether a head noun is modified. In the current chapter, we will consider one of these factors, namely the distinctiveness of the property denoted by the modifier, which, like semantic transparency, is a semantic factor. Both *blue dog* and *brown dog* are semantically transparent expressions in that the meaning of the constituents contribute to the meaning of the whole. However, *blue dogs* are more distinctive compared to the concept *dog* than are *brown dogs*. We explore whether people are sensitive to the distinctiveness of a property during the formation of multiword expressions.

1.2 Overview of the chapter

In this chapter, we begin by providing an overview of the theoretical issues concerning the inclusion of modifying information and the use of either full phrases or modifier-noun phrases. Next, we present an experiment in which we manipulated two factors that might influence the production of referential expressions. In particular, we examined whether the distinctiveness of the modifying information influences whether that information is used when referring to the antecedent. In addition, we examined whether the syntactic form in which the modifying information is presented influences the form in which modifying information is conveyed. Finally, we discuss the relevance of the empirical data within a psycholinguistic context and highlight the implications of the data for multiword expressions and for modifier-noun phrases in particular.

1.3 What motivates the inclusion of modifying information?

The expressions used to denote referents reflect how the speaker/writer is conceptualizing the object and, in particular, how he/she chooses to distinguish it from other items (Brown 1958; Olson 1970). Indeed, speakers are sensitive to both nonlinguistic- and linguistic-ambiguity during referential communication and attempt to avoid producing ambiguous expressions (Ferreira et al. 2005). A key issue for the current research concerns the factors that lead people to include modifying information rather than using an unmodified noun when producing a referential expression. The inclusion of modifying information serves several linguistic and psychological functions. Most often, modifying information is used to distinguish among potential referents (Downing 1977; Brekle 1986). There are often situations in which using the category label alone would not be sufficient. Consider a situation in which there are several cups on a table. To refer to a particular cup, for example, a speaker might specify its material and use either a full noun phrase (e.g., *May I have the cup that is ceramic?*) or a compound (e.g., *May I have the ceramic cup?*). Both utterances involve combining information about the head noun concept (e.g., *cup*) with information about a modifying concept (e.g., *ceramic*). This combination of information, in turn, allows the unambiguous identification of the referent within the available set of potential referents.

Several experiments on referential communication that used a visual display of objects (Tanenhaus et al. 1995, see also Frank & Goodman 2012) have found that speakers use a pre-nominal adjective (e.g., *tall glass*) in a context in which there are contrasting members (e.g., *a short glass*), which is consistent with the hypothesis that speakers try to make their utterances as informative yet as economical as possible (Grice 1975). The pre-nominal adjective is used to uniquely identify one object among several objects. However, the motivation for using modifying information appears to go beyond merely disambiguating among multiple possible referents because it is often included even when there is no need to provide additional information. This phenomenon of providing modifying information even in cases where such information is not needed to identify the referent is known as over-specification. Indeed, there are a number of studies showing that participants include adjectives during referential communication even though this additional specification is not required to identify the referent (e.g. Pechmann 1989; Sedivy 2003; Maes et al. 2004; Koolen et al. 2013).

Over-specification performs various functions in addition to identifying referents. For example, modifying information (e.g., *the cup that is on the shelf near the plate*) is used to shift the addressee's focus of attention (Ariel 1990; Prince 1992; Gundel et al. 1993; Chafe 1994). Another reason for using modifying information is to conform to pre-established conversational pacts (Brennan & Clark

1996; Ibarra & Tanenhaus 2016). Conversational partners often converge on an expression and will persist in using that expression even when there is no longer a need to include the additional information. To use an example from Brennan & Clark (1996), the term *pennyloafer* was initially used to denote a particular shoe among other possible shoes. However, the speaker continued to use this term rather than switching to using the simpler term *shoe* even when no other shoes were present in the display.

From a cognitive processing perspective, over-specification appears beneficial to both the speaker and the listener. For example, it aids in the identification of objects in a visual array and, consequently, speakers are more likely to produce over-specified expressions when they were asked to imagine that the task was very important (i.e., when told to imagine that the control panel is being used for long-distance surgery) than when they were not given such a scenario (Arts et al. 2011). Over-specification also benefits production (Pechmann 1989). Consistent with this idea, redundant information is more likely to be included when the speaker is under time pressure. Koolen et al. (2016) conducted a study in which participants referred to target objects in a visual array of objects. Participants were more likely to use over-specifying information when they were under a time constraint (e.g., they had to respond within 1000 ms) than when they had an unlimited amount of time to refer to the target object. Koolen et al. (2016) concluded that when individuals are under pressure, they are more likely to use quick heuristics and therefore select properties of an object based on their perceptual salience rather than discriminatory power.

Overall, there appear to be many reasons for why speakers might choose to include modifying information in referential expressions. In the current experiment, we focus on additional usage of modifying information that has not been fully explored in the literature. In particular, we propose that modifying information might be used to mark a conceptual distinction among category members and, in particular, to make explicit note of particularly distinctive information.

Studies on referential expressions within a visual context (i.e., situations in which objects are presented visually) indicate that the distinctiveness of visual properties within the display influences referential expressions. Participants were more likely to provide modifying information (i.e., to produce over-specified expressions) when the property of an object is atypical (e.g., Westerbeek et al. 2015). For instance, Rubio-Fernández (2016) used a referential communication task in which participants asked the researcher to click on objects that were presented in an array on the computer screen. In the first experiment, participants saw pictures of paper dolls and a display of paper clothes that were either all the same color (e.g., *brown purse*, *brown shirt*, *brown dress*, and *brown shoes*) or different colors (e.g., *yellow purse*, *pink shoes*, *blue dress*, and *red pants*). In the second ex-

periment, participants saw arrays of animals, fruits, vegetables, and artifacts that either had typical colors (e.g., *brown camel*) or atypical colors (e.g., *blue camel*). Participants tended to use a redundant color adjective in instances where such modifying information would be unnecessary (e.g., *the blue dress*, where only one dress could be a possible referent) more often when the object was an atypical color than when the object was a stereotypical color. These results suggest that modifying information is used when the concept has been modified with a distinctive property. Furthermore, participants provided modifying information more often when the color was a central property of the object category (e.g., referents such as clothing yielded a higher usage of redundant color adjectives than did geometrical figures). Taken together, these results suggest that a key characteristic in terms of determining whether modifying information is provided is conceptual distinctiveness rather than perceptual/visual distinctiveness. That is, the distinctiveness of the information relative to the category itself, rather than just within the visual display.

The aim of the current study is to explore the role of conceptual distinctiveness by examining whether the tendency to mention distinctive properties extends to situations in which the objects are not physically present. In particular, we will focus on a situation in which the contrast with other category members is implied or based on conceptual knowledge within a story context, rather than presenting the objects in a visual display. For example, mentioning that flowers are either fresh or wilted implicitly contrasts the flowers with ones that are not fresh or not wilted. Moreover, in the context of buying flowers as a gift, it is more typical to buy ones that are fresh than ones that are wilted. Thus, from a conceptual perspective, the property *wilted* is more distinctive for flowers than is *fresh*.

Conceptual distinctiveness is related to the issue of contrast. The notion of contrast between categories and subcategories has long played an important role in linguistic and psycholinguistic theories. Indeed, the principles of contrast and mutual exclusivity (Clark 1983; Carstairs-McCarthy 2010) are well-known constraints on word learning. In terms of multiword expressions, previous research on conceptual combination suggests that the notion of contrast influences how people use noun phrases. For example, Gagné & Murphy (1996) found that when verifying whether a property is true of a modifier-phrase (e.g., *submarine door*), people took less time to verify a property that was true of the phrase but not generally true of the head noun (e.g., *made of metal*) than to verify a property that was true of both the phrase and the head noun (e.g., *solid*). This finding suggests that people are sensitive to the extent to which the modified concept (e.g., *submarine door*) is semantically/conceptually distinctive from other members of the head noun concept (e.g., *door*).

In terms of judgments about whether a concept has a particular property, several studies (Connolly et al. 2007; Gagné & Spalding 2011; 2014b; Hampton et al. 2011; Jönsson & Hampton 2012) have shown that properties that are true of the head noun (e.g., *kites have strings*) are viewed as being less true of the modified head (e.g., *silk kites have strings*). This effect (known as the modification effect) appears to be driven by the expected level of contrast between the combined concept (e.g., *silk kites*) and the head concept (e.g., *kites*); when making judgments about the likelihood that a property is true, participants are influenced by the meta-knowledge that modified concepts are used to signal that the subcategory is similar to the category (e.g., *silk kites have many properties in common with kites*) but also that the subcategory is somehow different than the category (Gagné & Spalding 2011; 2014b; Spalding & Gagné 2015). These two expectations account for why properties that are true of the head noun are judged as being less true of the modified concept, and that properties that are false of the head noun (e.g., *candles have teeth*) are judged to be more true (but still unlikely) of the modified concept (e.g., *purple candles have teeth*). Indeed, the effects of the expected contrast is so strong that the same effects are seen even when the modifier is a non-word (e.g., Gagné & Spalding 2015).

Thus, we conclude that conceptual contrast or conceptual distinctiveness is a critical factor in the use and understanding of multiword phrases and compound words in general and is therefore likely to contribute to the production of such phrases.

1.4 When modifying information is included, how is it expressed?

If modifying information is included, the syntactic form which expresses this information can still vary. In English, modifying information can be expressed as a full noun phrase (e.g., *a dog that is blue*) or as a modifier-noun phrase (e.g., *a blue dog*). Do people have a priori biases toward using one linguistic expression over another? The answer is not immediately obvious because intuitions based on ease of processing do not correspond with the tendency for expressions to become shortened over time.

In terms of ease of processing, there is an advantage to using a full phrase because noun compounds are particularly challenging to interpret (Lapata 2002; Copestake & Briscoe 2005; Libben 2014). Much of the difficulty lies in recovery of an implicit underlying relation between the modifier and head noun concept. A modifier-noun phrase is more ambiguous than a full phrase, in that the full noun phrase explicitly describes the exact nature of the modification that is being performed (e.g., *oil for babies*) whereas, for modifier-noun phrases (e.g.,

baby oil) the nature of the modification is implicit and must be reconstructed by the listener/reader (see Levi 1978; Gagné & Shoben 1997). The term "modifier-noun phrase" most often refers to constructions that are novel (e.g., *apple juice seat*; *mountain magazine*), but, can also refer to lexicalized open (unspaced) compounds (e.g., *hunting dog*; *paper bag*). Indeed there seems to be commonalities in the processing of novel noun phrases and lexicalized compounds (Gagné & Spalding 2006). Psycholinguistic research has shown that human language users actively make use of relations during the processing of both novel and established/lexicalized compounds (Gagné & Shoben 1997; Gagné 2002; Gagné & Spalding 2009; 2014a). This research indicates that, during the comprehension of noun compounds, the more available the required relation is, the easier it is to select the relation and, consequently, the less time it takes to interpret the compound. In other words, the more difficult it is to recover the implicit underlying relation, the more difficult it is to interpret a compound (see, for example, Gagné & Shoben 1997; Spalding & Gagné 2014; Schmidtke et al. 2018).

Given the difficulty inherent in recovering implicit semantic relations, one would presume that it would be advantageous to overtly express the relation and, consequently, to avoid the use of compounds. Yet, this is not what happens within the human language system. Over time, lexicalized phrases are often truncated and become compounds (e.g., *our lady's bug* became *ladybug*). Similarly, compounds can become non-compounds (e.g., *electronic mail* became *e-mail* and, more recently, *email*); the words *lord* and *lady* are derived from Old English compounds *half-weard* 'bread-keeper' and *halfdige* 'bread-kneader'. This truncation that occurs on a global (and more long-term) level within a language also occurs during local interactions. During referential communication, for example, linguistic expressions are often shortened (Garrod & Anderson 1987; Brennan & Clark 1996). For example, in one experiment, a geometric figure that was initially described as looking *… like a person who's ice skating, except they're sticking two arms out in front* became *the ice skater* (Clark & Wilkes-Gibbs 1986). Similarly, an object that was initially referred to as *the car that has like … blueprints painted on the side of it sorta* was later referred to as *the blueprint car* (Metzing & Brennan 2003). In sum, there appears to be a preference toward using syntactically simpler expressions such as compounds, even though such expressions are inherently more ambiguous than full expressions which specify the relation overtly.

On the basis of these findings, one would expect an overall bias towards using a truncated expression (e.g., using *wilted flowers* or *even flowers*, rather than *flowers that are wilted*). However, this bias must also be considered in light of another bias reported in the literature – namely, the tendency for people to reuse recently encountered syntactic structures. For example, Bock (1986) demon-

strated that speakers tend to re-use a syntactic structure from the priming sentence when describing a scene. This effect has been examined in a variety of context including examinations of whether it can be driven by a single word as in the case of featural accounts of syntactic priming. For example, Melinger & Dobel (2005) found that production preferences for dative alternation can be biased by prior exposure to a single verb. However, most relevant for the current project concerns studies that focus on the creation of referential expressions. Syntactic convergence occurs during referential communication. For example, participants were more likely to describe a picture of a red sheep as *The sheep that's red* when the confederate recently described a picture of a red door as *The door that's red* than when it was described as *a red door* (Cleland & Pickering 2003). This result suggests that participants tend to re-use syntactic structures, especially when the prime and target sentences share lexical items such as *red* (see also Chang et al. 2003). Similarly, Tarenskeen et al. (2015) found that when participants use modifying information to describe a target item from a visual array of six drawings of clothing, there is a tendency to continue to re-use the same syntactic structures.

These studies all demonstrate that participants have a tendency to re-apply the same syntactic structure that was used with one object/entity (e.g., *sheep*) when subsequently referring to a separate object/entity (e.g., *door*). However, an unresolved question concerns whether syntactic priming will occur in a task in which participants are introduced to a concept (e.g., *apples that are rotten*) and then are asked a question requiring them to refer to that same concept. This situation directly pits the bias towards truncation against the bias towards re-using syntactic expressions. The current experiment will investigate this issue.

2 Experiment

2.1 Overview and rationale

We examine the types of referring expressions that people produce when referring to a concept that has been encountered in a short description of a scenario. The experiment was designed to address two key issues:

1. whether the distinctiveness of the modifying information being conveyed about a target entity in a story influences whether that information is included when the participant is asked to refer back to the entity and

2. whether the syntactic form in which the modifying information is presented influences the form in which modifying information is conveyed.

Participants read short stories and then answered a comprehension question that would require them to refer to something in the story. For example, one story described a woman buying a pet. The target antecedent was the dog that she purchased. We varied the type of modifying information that was presented with the target antecedent. The information was either normal or typical for the object or was distinctive. To illustrate, all participants read a version of the story in which the color of the dog was mentioned. For half of the participants, the dog was described as having brown fur (a normal feature for dogs), and for the other half, the dog was described as having blue fur (a distinctive feature for dogs). We were interested in what the participants would produce when they were asked *What kind of pet did Sally buy?*

We predict that distinctiveness will influence whether participants choose to include modifying information in their linguistic expression. Properties that are unusual or distinctive for the head noun will be seen as especially relevant and, consequently, will be more likely to be included in the description provided by the participants. However, properties that are not unusual will be deemed less relevant (because the majority of members of the head noun category have the same property) and therefore less likely to be included. Thus, when referring to a dog that was previously mentioned in a short story, participants will be more likely to include modifying information when the dog was described as having an atypical color such as *blue* relative to when the dog was described as having a typical color such as *brown*, because the resulting subcategory is more distinctive and therefore will tend to more readily identify the appropriate referent. In short, there are lots of brown dogs, but relatively few blue dogs in the world, and, consequently, it should be more informative to refer to the subcategory of *blue dogs* than to the subcategory of *brown dogs*. Note, however, that in no case is the modifying information required to uniquely identify the referent.

In terms of the syntactic form that is used to convey the modifying information, the existing literature points to two conflicting predictions. On one hand, people might show a tendency toward using a modifier-noun phrase even when the information is presented as a full noun phrase. Two considerations arise here. First, the modifier-noun phrase is shorter and syntactically simpler and, thus, might generally be preferred. Second, a modifier-noun phrase is more ambiguous than a full phrase, in that the full noun phrase explicitly describes the exact nature of the modification that is being performed (e.g., *a dog that is blue*) whereas, for modifier-noun phrases the nature of the modification is implicit and must be reconstructed by the listener/reader (Downing 1977; Levi 1978). Having the relation directly specified (e.g., *crayon that is made of plastic*, or *sunshine in the morning*) removes uncertainty about relation selection (Gagné & Spalding 2014a;

2015). Thus, there could be some trade-off, in which speakers or writers generally prefer to use the shorter modifier-noun phrase, as long as they have reason to believe that the recipient will understand the implied connection between the modifier and the head noun concepts. Gagné & Spalding (2004) found that the presence of a referent in a discourse made modifier-noun phrases easier to comprehend, even though the phrase itself had not been presented. In the present study, all of the stories include information (either in the form of the full noun phrase or the modifier-noun phrase) that should make it easy for a recipient to understand the modifier-noun phrase. Therefore, the participants, in responding to the question about the target antecedent, might show a general preference for the modifier-noun phrase.

On the other hand, the form in which the information was initially presented in the preceding discourse might influence the manner in which the information is later conveyed due to syntactic priming. That is, when information is presented as a modifier-noun phrase, then people should be more likely to produce a modifier-noun phrase than when the information is presented as a full noun phrase. This prediction is derived from research on the activation of syntactic structure during speech production that demonstrates that speakers tended to reuse a syntactic structure from the priming sentence when describing a scene (Bock 1986; Bock & Loebell 1990).

2.2 Method

2.2.1 Participants

Fifty-four introductory psychology students participated for partial course credit. All participants were native speakers of English. The data from two participants were not used because they did not follow instructions. Thus, data from 52 participants were included in the analyses.

2.2.2 Materials and procedure

Twenty-eight short stories were constructed. Each story was under 65 words long and contained a target antecedent (i.e., the antecedent that we will be eliciting) for which we provided modifying information. We varied whether the modifying information was distinctive (e.g., *blue fur*) or usual (e.g., *brown fur*) for the head noun (e.g., *dog*) in the context of the story. In addition, we varied the syntactic form in which the modifying information was presented: the information was presented as a modifier-noun phrase (e.g., *brown dog; blue dog*) or full noun

phrase (e.g., *a dog that is brown; a dog that is blue*). These two variables were crossed which yielded four experimental conditions. For example, one story was:

> Sally loves animals. She decided to get a pet. So she went to the pet store to see what was there. Sally immediately set her eyes on a [*blue dog/brown dog/dog that was blue/dog that was brown*]. She picked him up and knew instantly that he was going to be a great companion for her.

Only one of the expressions within the square brackets was presented to a particular participant. The items were counter-balanced such that each participant saw an equal number of stories in each of the four conditions and each item was seen only once by each participant. Order of presentation was randomized for each participant. The full list of target items (i.e., the unusual, normal, and head noun) is listed in the Appendix.

Participants viewed the stories one at a time on a computer screen. They were instructed to read each passage carefully and were allowed as much time as necessary to complete the task. After each story, participants answered two questions about the story. The first question required people to recall the referent of the target noun phrase from the story. It specifically required the participant to respond by describing the target concept. For example, a question might ask "What kind of pet did Sally get?" The participant typed in their answer. The second question was also associated with the passage, and asked about another aspect of the story.

2.3 Results

Two of the authors classified the responses into four categories based on how the participants referred to the antecedent: modifier-noun phrase (e.g., *blue dog* or *brown dog*), full phrase (*dog that is blue* or *dog that is brown*), and head noun only (*dog*). In addition, a fourth category was used for "other" responses. Three main types of responses fell under this category. The first were responses that did not provide a specific answer (e.g., "I don't know", "it doesn't say"). The second were responses that did not address the question (e.g., "What does Nathan cut quickly" was intended to elicit either green or yellow grass, but the participant responded "because his parents are coming home"). The third type of response did not directly refer to the target reference (e.g., "What does Katie wear to keep her feet warm" was intended to elicit snake slippers or soft slippers but the participant responded "fuzzy slippers").

Inter-rater agreement was 100%. Table 1 displays the number of responses (for each condition) in each category. Overall, participants generally did include modifying information; modifying information was provided in 962 out of 1456 responses, and the vast majority (84%) of these responses were in the form of a modifier-noun phrase. The responses that were coded as "other" were not included in further analyses and, thus, the percentage with which a category was used within each of the four experimental conditions was calculated based only on responses in the form of a modifier-noun phrase, full phrase, and head noun only.

Table 1: Number of responses and row percentages (in parentheses) for each condition that were modifier-noun phrase, full phrase, head noun only, or other. Each row sums to 364.

Experimental condition		Response type			
Property	Form	modifier NP	full phrase	noun	other
non-distinctive	modifier-noun phrase	220 (60.44)	5 (1.37)	91 (25.00)	48 (13.19)
non-distinctive	full phrase	111 (30.49)	48 (13.19)	157 (43.13)	48 (13.19)
distinctive	modifier-noun phrase	303 (83.24)	2 (0.55)	37 (10.16)	22 (6.04)
distinctive	full phrase	177 (48.63)	96 (26.37)	59 (16.21)	32 (8.79)
	Total	811 (55.70)	151 (10.37)	344 (23.63)	150 (10.30)

We conducted two separate analyses. The first analysis focused on whether Form and Distinctiveness affected the likelihood of including modifying information. The second analysis examined whether Form and Distinctiveness influenced the form (e.g., full phrase vs. modifier- noun phrase) in which the modifying information was conveyed. In both analyses, the dependent variable was binary (i.e., is modified vs. not modified for the first analysis, and compound vs. phrase for the second analysis) and, consequently, we used the *melogit* function in Stata 15 to fit a mixed-effects model for binary responses. The experimental variables, Form and Distinctiveness, were included as fixed effects, and subjects and items were included as crossed random effects. The estimates of the fixed effects are reported as log odds.

To examine whether the syntactic form (e.g., *wilted flowers* vs. *flowers that are wilted*) in which the information had been presented in the story and the distinctiveness of the property influenced the likelihood of including modifying information when referring to the antecedent, we fit a model in which the dependent variable was whether the participant's response included modifying information; modifier-noun phrase and full phrase responses were coded as 1 and the

head noun only responses were coded as 0. Both the distinctiveness of the property and the form in which the information was presented in the story influenced whether modifying information was included in the response. Participants were more likely to provide modifying information when the property presented in the story was distinctive (e.g., *wilted* as a property of *flowers*) rather than usual (e.g., *fresh* as a property of *flowers*), 86% vs. 61%, $b = 1.48, \text{SE} = 0.24, z = 6.22, p < 0.0001$, and when the property had been presented as a modifier-noun phrase rather than a full-phrase (81% vs. 67%), $b = -1.15, \text{SE} = 0.19, z = -5.96, p < 0.0001$. The two predictor variables (Form and Distinctiveness) did not interact with each other, $b = 0.46, \text{SE} = 0.31, z = 1.48, p = 0.14$.

The second analysis was conducted using only the responses that included modifying information (i.e., only the full phrase and modifier-noun responses) so that we could test whether the form in which the modifying information was presented in the story and the distinctiveness of the property influenced the way in which participants conveyed the modifying information in their response. The dependent variable corresponded to whether the response was a modifier-noun phrase (1 = modifier-noun phrase and 0 = full phrase). Participants were more likely to provide a modifier-noun response when the story used a modifier-noun form (predicted $M = 0.99, \text{SE} = 0.009$) than when the story used a full phrase form ($M = 0.64, \text{SE} = 0.05$), $\chi^2(1) = 31.47, p < 0.0001$. Note that because there are only two levels of the variable, the reverse is also true: namely, that participants are more likely to provide a full-phrase response when the story used a full-phrase form than when the story used a modifier-noun form. The type of property used in the story did not strongly influence whether participants used a modifier-noun form, $\chi^2(1) = 3.05, p < 0.08$.

Distinctiveness and Form interacted, $b = -1.99, \text{SE} = 0.58, z = -3.41, p = 0.001$, and, therefore, we examined the simple effects at each level of form. Distinctiveness of the property had no effect on whether the response was a full phrase or modifier-noun phrase when the modifying information was presented as a full phrase, $\chi^2(1) = 2.43, p < 0.12$. However, when the modifying information was presented as a modifier-noun phrase, the response was more likely to be a modifier-noun phrase when the property was unusual/distinctive than when the property was normal, $\chi^2(1) = 8.91, p < 0.003$.

3 Discussion

We explored two aspects of the production of multiword referential expressions: inclusion of modifying information and syntactic form, with a particular focus on

modifier-noun phrases (e.g., *blue dog* and *brown dog*) and full noun phrases (e.g., *dog that is blue* and *dog that is brown*). The experiment directly pitted the bias towards truncation against the bias towards re-using syntactic expressions. The findings make three primary contributions to the literature on multiword expressions. First, we demonstrate the influence of semantic/conceptual knowledge on the inclusion of modifying information. In particular, the degree of conceptual contrast seems to be critical in determining whether modifying information is included when the referential expression is produced. Second, our results reveal the primacy of modifier-noun phrase constructions (over full phrase constructions) as a means of conveying that information. Third, while it is possible that there are small effects of syntactic repetition, or a general bias to use shorter syntactic forms for a reference to an already identified object from the story, the bias towards the modifier-noun phrase appears to be the main driver of the syntactic form of the referential expression, at least in this particular communicative task.

3.1 Including modifying information

Previous research using visual displays of objects found that over-specification was more likely when a property was visually distinctive or salient such as when one object was a different color than other objects in the display (e.g., in a visual display in which one dog is blue and the others are orange). The current results extend this finding to a situation where the objects are not physically present and the distinctiveness of a property is based on conceptual knowledge about the modifier and head noun concepts. For example, *blue* is distinctive for *dogs* but not for *skies*. The knowledge needed to determine distinctiveness comes from past history and knowledge of the concepts involved rather than from visual information that is presented in the experiment. Therefore, our finding suggests that people are sensitive to conceptual distinctiveness in addition to (as shown in previous research) referential distinctiveness. To illustrate, in general language usage, a category name (e.g., *dog*) typically refers to a generic type (i.e., to the category of dogs). However, in our study, the referent was always a particular category member, not a generic category. Whether participants used a generic label or modified construction depended on the distinctiveness of the property (relative to the head noun category) used in the story. In this respect, our data highlights the role of a particular type of implicit information, namely knowledge about the nature of the category-subcategory similarity. In particular, the category label (i.e., *dog*) was used when the particular referent in the story was not unusual; that is, when the entity being described was similar to the generic

representative of the category. Note that the modifying information was not required to uniquely identify the referent (i.e., there was only one dog in the story), yet participants often opted to include this information, especially when it was distinctive. Thus, the inclusion of modifying information corresponded to a conceptual distinction rather than a purely referential one in that participants were sensitive to semantic and conceptual knowledge about the category to which the referent belonged.

There are several possibilities for why participants tended to provide overspecified expressions especially when the referent had a distinctive property than when it had a normal property. One possibility is that the distinctive properties are just much more salient. For example, work on memory has suggested that features that violate expectations are often noticed and remembered particularly well (e.g., a skull in an office setting, see Brewer & Treyens 1981). In general, people make note of properties that are not similar to those they have seen before and, when communicating, they might prefer to explain these differences to others in the simplest way possible (Garrod & Anderson 1987; Markman et al. 1997). In the current experiment, the distinctive properties might have been more noticeable than normal properties, and this difference might have prompted participants to include them in their response. Another possible explanation is that the distinctive features are more likely to be incorporated into the representation of the target referent because they tend not to be true of the head noun. This explanation is consistent with previous research on novel combined concepts that suggests that features that are true of the entire phrase but not of the head noun in general (e.g., *white* for *peeled apples*) are more available than features that are true of the head noun (e.g., *round*) (Springer & Murphy 1992; Gagné & Murphy 1996) and also with evidence suggesting that people strongly expect property differences between things named with modified and unmodified nouns (Gagné & Spalding 2011; 2014b; Spalding & Gagné 2015). In our experiment, the normal properties were ones that tended to be true of the head noun concept, whereas the distinctive properties were not generally true of the head noun concept. Thus, it is possible that the distinctive property was more likely to be integrated into the representation of the target referent than was the normal property. If so, then distinctive properties would be more likely to be included in the participant's response than would normal properties.

3.2 Selection of syntactic form

There is some tendency to reproduce the syntactic form in which the information was first presented; responses using a modifier-noun phrase are common,

but are even more used when the story also uses a modifier-noun phrase then when the story uses a full-phrase. Furthermore, although responses using a full response were relatively rare, the vast majority of responses that used a full phrase ($n = 144$) were produced when the story also used a full phrase whereas only 7 responses using a full phrase were produced when the story did not use a full phrase. This finding is consistent with previous research on syntactic priming (Bock 1986; Bock & Loebell 1990) that found that people are more likely to produce passive constructions when describing a scene when previous sentences contained passive constructions than when previous sentences did not contain passive constructions. The current experiment examined part of a sentence, namely, the structure of a noun phrase, and also found support for syntactic priming.

However, the selection of syntactic form was not completely determined by the form presented in the story. Instead, there was a strong preference toward using a modifier-noun phrase (e.g., *wilted flowers*) rather than a full noun phrase (e.g., *flowers that are wilted*). Previous work on referential communication has indeed shown an overall trend towards the use of shortened expressions (Brennan & Clark 1996; Markman et al. 1997) and analyses of text corpora also show evidence of text compression (Marsh 1984). Thus, the preference for a modifier-noun phrase might reflect a tendency to select a syntactically simpler construction. Modifier-noun phrases are syntactically simpler than full noun phrases and yet still provide information that allows the reader/listener to identify a subcategory of head noun (e.g., *ceramic cup* refers to a particular subcategory of the category *cup*). Thus, modifier-noun phrase constructions offer a balance between syntactic simplicity and informativeness. At the same time, there was little evidence to suggest that participants selected a head noun only structure over a modifier-noun structure, even though head noun only structures are syntactically simpler than modifier-noun phrase structures. That is, rather than exhibiting an overall bias towards shortening, per se, our data indicate a bias towards modifier-noun phrase use, which suggests that modifier-noun phrase might have a special status in the language. Although full phrases (e.g., *flowers that are fresh*) were almost always shortened (to either a modifier-noun phrase or noun, e.g., *fresh flowers* or *flowers*), modifier-noun phrases were rarely shortened to noun-only. Thus, the use of a modifier-noun phrase rather than a full phrase might reflect something about the special status of modifier-noun phrases rather than a general bias toward syntactically simple constructions, per se. That is, it seems likely that modifier-noun phrases are particularly useful for conveying subcategory information. People are sensitive to overt cues that indicate the existence of a contrast set, such as the presence of the word *only*, and the inclusion of this

cue affects the relative ease of resolving main clause/reduced relative clause ambiguities (Sedivy 2002). Perhaps the inclusion of modifying information in the context implied the existence of a contrast set. This might have encouraged people to use a modifier-noun phrase when referring to the target referent because this construction indicates a contrast set (Markman 1991).

In sum, we see some evidence for syntactic priming in that the form of the presentation in the story could reduce the bias to producing modifier-noun phrases, but the influence of the prior form was relatively weak in that it was not able to overturn the strong preference for modifier-noun phrases constructions. Similarly, although we see some degree of shortening of the referring phrase, there still seems to be a preference for maintaining at least a modifier-noun construction, rather than just a generic noun, even though no modifying information was required in order to identify the referent in the story. This was particularly true when the modifying information was atypical.

4 Conclusion

Our data reveal that the context in which the linguistic expressions are used provides useful cues as to the form that the linguistic expression will take and provide insight into the expectations/biases that languages users use during referential communication. During conversation and referential communication, modifier-noun phrases (e.g., *rotten apple*) are produced for several reasons including distinguishing among potential referents and maintaining conversational pacts. The current experiment demonstrates that modifier-noun phrases also are produced in order to highlight conceptually distinctive properties. The finding that distinctiveness influenced the use of modifying information provides insight into how people use multiword expressions to convey information about how they are conceptualizing the various entities about which they are communicating. In particular, the form of the linguistic construction (e.g., noun versus modifier-noun phrase) provides useful cues as to the intended meaning. Furthermore, although the participants were somewhat sensitive to the syntactic form with which the target was presented, there was a strong bias for the modifier-noun phrase form. In sum, it appears that modifier-noun phrases have a privileged status among multiword expressions and provide a good compromise between competing principles of conveying sufficient information and using simple syntactic structures.

References

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: New York: Routledge.

Arts, Anja, Alfons Maes, Leo G M Noordman & Carel Jansen. 2011. Overspecification in written instruction. *Linguistics* 49(3). 555–574.

Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18. 355–387.

Bock, Kathryn & Helga Loebell. 1990. Framing sentences. *Cognition* 35(1). 1–39.

Brekle, Herbert E. 1986. The production and interpretation of ad hoc nominal compounds in German: A realistic approach. *Acta Linguistica Academiae Scientiarum Hungaricae* 36(1-4). 39–52.

Brennan, Susan E. & Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(6). 1482–1493.

Brewer, William F. & James C. Treyens. 1981. Role of schemata in memory for places. *Cognitive Psychology* 13(2). 207–230.

Brown, Roger. 1958. How shall a thing be called? *Psychological Review* 65(1). 14–21.

Carstairs-McCarthy, Andrew. 2010. *The evolution of morphology*. Oxford: Oxford University Press.

Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chang, Franklin, Kathryn Bock & Adele E Goldberg. 2003. Can thematic roles leave traces of their places? *Cognition* 90(1). 29–49.

Clark, Eve V. 1983. *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Clark, Herbert H. & Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13(2). 259–294.

Clark, Herbert H. & Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22. 1–39.

Cleland, Alexandra A. & Martin J. Pickering. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language* 49(2). 214–230.

Connolly, Andrew C., Jerry A. Fodor, Lila R. Gleitman & Henry Gleitman. 2007. Why stereotypes don't even make good defaults. *Cognition* 103(1). 1–22.

Copestake, Ann & E.J. Briscoe. 2005. Noun compounds revisited. In John I. Tait (ed.), *Charting a new course: Natural language processing and information retrieval*, vol. 16, 129–154. Dordrecht: Springer.

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

Ferreira, Victor S, L Robert Slevc & Erin S Rogers. 2005. How do speakers avoid ambiguous linguistic expressions? *Cognition* 96(3). 263–284.

Frank, Michael C & Noah D Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998.

Gagné, Christina L. 2002. Lexical and relational influences on the processing of novel compounds. *Brain and Language* 81(1-3). 723–735.

Gagné, Christina L. & Thomas L. Spalding. 2014a. Conceptual composition: The role of relational competition in the comprehension of modifier-noun phrases and noun–noun compounds. In Brian Ross (ed.), *Psychology of learning and motivation*, vol. 59, 97–130. New York: Elsevier.

Gagné, Christina L. & Thomas L. Spalding. 2015. Semantics, concepts, and metacognition: Attributing properties and meanings to complex concepts. In Laurie Bauer, Lívia Körtévlyessy & Pavol Štekauer (eds.), *Semantics of complex words*, 9–25. New York: Springer.

Gagné, Christina L. & Gregory L. Murphy. 1996. Influence of discourse context on feature availability in conceptual combination. *Discourse Processes* 22(1). 79–101.

Gagné, Christina L. & Edward J. Shoben. 1997. Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(1). 71–87.

Gagné, Christina L. & Thomas L. Spalding. 2004. Effect of discourse context and modifier relation frequency on conceptual combination. *Journal of Memory and Language* 50(4). 444–455.

Gagné, Christina L. & Thomas L. Spalding. 2006. Using conceptual combination research to better understand novel compound words. *SKASE Journal of Theoretical Linguistics* 3(2). 9–16.

Gagné, Christina L. & Thomas L. Spalding. 2009. Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language* 60(1). 20–35.

Gagné, Christina L. & Thomas L. Spalding. 2011. Inferential processing and metaknowledge as the bases for property inclusion in combined concepts. *Journal of Memory and Language* 65(2). 176–192.

Gagné, Christina L. & Thomas L. Spalding. 2014b. Subcategorisation, not uncertainty, drives the modification effect. *Language, Cognition and Neuroscience* 29(10). 1283–1294.

Gagné, Christina L. & Thomas L. Spalding. 2016. Effects of morphology and semantic transparency on typing latencies in English compound and pseudo-compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(9). 1489–1495.

Garrod, Simon & Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27(2). 181–218.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics 3: Speech acts*, vol. 3, 41–58. New York: Academic Press.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Hampton, James A., Alessia Passanisi & Martin L. Jönsson. 2011. The modifier effect and property mutability. *Journal of Memory and Language* 64(3). 233–248.

Ibarra, Alyssa & Michael K. Tanenhaus. 2016. The flexibility of conceptual pacts: Referring expressions dynamically shift to accommodate new conceptualizations. *Frontiers in Psychology* 7. 561.

Jarema, Gonia, Céline Busson, Rossitza Nikolova, Kyrana Tsapkini & Gary Libben. 1999. Processing compounds: A cross-linguistic study. *Brain and Language* 68(1-2). 362–369.

Jönsson, Martin L. & James A. Hampton. 2012. The modifier effect in within-category induction: Default inheritance in complex noun phrases. *Language and Cognitive Processes* 27(1). 90–116.

Koolen, Ruud, Martijn Goudbeek & Emiel Krahmer. 2013. The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science* 37(2). 395–411.

Koolen, Ruud, Emiel Krahmer & Marc Swerts. 2016. How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distractor distance. *Cognitive Science* 40(7). 1617–1647.

Lapata, Maria. 2002. The disambiguation of nominalizations. *Computational Linguistics* 28(3). 357–388.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Libben, Gary. 1998. Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language* 61(1). 30–44.

Libben, Gary. 2014. The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology* 31(1–2). 8–25.

Maes, Alfons, Anja Arts & Leo Noordman. 2004. Reference management in instructive discourse. *Discourse Processes* 37(2). 117–144.

Markman, Arthur B., Takashi Yamauchi & Valerie S. Makin. 1997. The creation of new concepts: A multifaceted approach to category learning. In TB Ward, SM Smith & J Vaid (eds.), *Conceptual structures and processes: Emergence, discovery, and change*, 174–208. Washington, DC: American Psychological Association.

Markman, Ellen M. 1991. *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Marsh, Elaine. 1984. A computational analysis of complex noun phrases in Navy messages. In *Proceedings of the 10th International Conference on Computational Linguistics and the 22nd Annual Meeting of the Association for Computational Linguistics*, 505–508.

Melinger, Alissa & Christian Dobel. 2005. Lexically-driven syntactic priming. *Cognition* 98(1). B11–B20.

Metzing, Charles & Susan E. Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* 49(2). 201–213.

Olson, David R. 1970. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review* 77(4). 257–258.

Pechmann, Thomas. 1989. Incremental speech production and referential overspecification. *Linguistics* 27(1). 89–110.

Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In William C. Mann & Sandra A. Thompson (eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text*, 295–325. Amsterdam: John Benjamins.

Rubio-Fernández, Paula. 2016. How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology* 7. 153.

Schmidtke, Daniel, Christina L. Gagné, Victor Kuperman, Thomas L. Spalding & Benjamin V. Tucker. 2018. Conceptual relations compete during auditory and visual compound word recognition. *Language, Cognition and Neuroscience* 33(7). 923–943.

Sedivy, Julie C. 2002. Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language* 46(2). 341–370.

Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32(1). 3–23.

Smolka, Eva & Gary Libben. 2017. ‘‘Can you wash off the hogwash?”: Semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience* 32(4). 514–531.

Spalding, Thomas L. & Christina L. Gagné. 2014. Relational diversity affects ease of processing even for opaque English compounds. *The Mental Lexicon* 9(1). 48–66.

Spalding, Thomas L. & Christina L. Gagné. 2015. Property attribution in combined concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(3). 693–707.

Springer, Ken & Gregory L. Murphy. 1992. Feature availability in conceptual combination. *Psychological Science* 3(2). 111–117.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632–1634.

Tarenskeen, Sammie, Mirjam Broersma & Bart Geurts. 2015. Overspecification of color, pattern, and size: Salience, absoluteness, and consistency. *Frontiers in Psychology* 6. 1703.

Westerbeek, Hans, Ruud Koolen & Alfons Maes. 2015. Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology* 6. 935.

Appendix

Table 2: Full list of target items showing the unusual properties, normal properties, and the head noun.

Properties		Head noun
blue	brown	dog
fresh	wilted	flowers
rotten	red	apples
soggy	crisp	crackers
polluted	blue	lake
curdled	white	milk
rubbery	savory	chicken
soap	shoe	shop
melted	frozen	ice cream
blurry	glossy	photographs
explicit	meaningful	lyrics
green	orange	fire
stale	soft	buns
yellow	green	grass
clown	public	school
carrot	sweet	candy
snake	soft	slippers
cold	hot	shower
sour	sweet	honey
monster	school	friends
crashing	flying	planes
candy	boreal	forests
gravy	train	station
plastic	coloured	chalk
purple	morning	sunshine
plaid	school	pants
smokeless	smoky	cigarettes
pickle	juice	pitcher