Ordering adjectives in referential communication

Kumiko Fukumura

Psychology, Faculty of Natural Sciences

University of Stirling, UK

Address correspondence to:

Kumiko Fukumura

Psychology, Faculty of Natural Sciences

University of Stirling

Stirling FK9 4LA

UK

kumiko.fukumura@stir.ac.uk

Abstract

We contrasted two hypotheses concerning how speakers determine adjective order during referential communication. The *discriminatory efficiency* hypotheses claims that speakers place the most discriminating adjective early to facilitate referent identification. By contrast, the *availability-based ordering* hypothesis assumes that speakers produce most available adjectives early to ease production. Experiment 1 showed that speakers use more pattern-before-color modifier orders (than the reversed) when pattern, not color, distinguished the referent from alternatives, providing support for the discriminatory efficiency hypothesis. Participants also overspecified color more often than pattern, and they generally favored color-before-pattern orders, in support of the availability-based ordering hypothesis. Experiments 2 and 3 replicated both effects in a dialogue setting, where speakers' adjective ordering was also primed by their partner's ordering, using conjoined and non-conjoined constructions. We propose a novel model (*PASS*) that explains how discriminability and availability simultaneously influence adjective selection and ordering via competition in the speaker's message representation.

Key words: adjective, word order, referential communication, language production, dialogue, priming

Introduction

When producing multiple modifiers, speakers do not order them randomly. In English and other similar languages, *the large red car* is much more preferred than *the red large car.* Previous research on adjective order has focussed on identifying underlying *semantic constraints* (e.g., Dixon, 1982; Hetzron 1978; Quirk, Greenbaum, Leech, & Svartik, 1985; Scott, 2002; Whorf, 1945), whereby size typically occurs further from the noun than shape (*large round table* rather than *round large table*) or color (*large red car* rather than *red large car*). The main conclusion from these studies has been that adjectives that denote more *absolute* (Martin, 1969b), *intrinsic* (Danks & Glucksberg, 1971), *inherent* (Quirk et al., 1985; Whorf, 1945) or *subjective* (Hetzron, 1978; Scontras, Degen, & Goodman, 2017) semantic properties occur closer to the noun. Yet speakers do not always follow such semantic rules (Ney, 1986; Mitchell, 2009). An important question for psychological theories of language production concerns what cognitive and communicative constraints might underlie adjective ordering. Whereas research has concentrated on *what* property speakers tend to choose in referential descriptions (e.g., Deutsch & Pechmann, 1982; Dale, 1989; Dale & Reiter, 1995; Frank & Goodman, 2012), much less attention has been paid to *how* speakers express those properties (i.e., *formulation* as opposed to *message generation*, Levelt, 1989). The purpose of the present study was to address this, by uncovering the processes and representations that underlie adjective ordering.

Property selection

When the context contains similar referential candidates, speakers normally refer to the referent's distinguishing property to avoid referential ambiguity. Research has shown that the likelihood of speakers referring to a particular property of the referent (e.g., size as in *large circle*) is higher when the context contains similar category exemplars (e.g., multiple circles) than when the referent is the only category exemplar (e.g., V.Ferreira, Slevc, & Rogers, 2005; Horton & Keysar,

1996; Pechmann, 1989). Some computational models of object specification therefore focus on contextual discriminability to predict which properties speakers are likely to refer to (Frank & Goodman, 2012; Dale, 1989). Frank and Goodman (2012), for instance, proposed that the likelihood in which speakers choose a particular property of the referent, amongst alternatives, is inversely related to the number of referential candidates to which the property can apply: The fewer referential candidates the property can be applied to, the more informative it is for identifying the referent, leading to a higher likelihood of reference. Suppose color discriminates your intended referent from other referential candidates, whereas size does not (e.g., the referent is the only red circle in the context, but there are several small circles). The chance that you will then refer to color rather than size is higher compared to when size but not color is discriminating. The preference for choosing the most discriminating property over less discriminating ones is in accord with Dale (1989), who regarded it as a heuristic for producing the shortest-possible distinguishing referring expressions, in keeping with the Gricean assumption that cooperative speakers should be *optimally* informative (Grice, 1975).

However, speakers do not always produce *minimally-specified* descriptions; instead, they often *overspecify* (Pechman, 1989). Speakers often add a modifier when other properties of the description alone can identify the referent uniquely (e.g., Deutsch & Pechmann, 1982; Engelhardt, Beiley, & F.Ferreira, 2006; Pechmann, 1989) or when the additional modifier does not rule out any referential alternative (Belke & Meyer, 2002; Gatt, Krahmer, Van Deemter, & Van Gompel, 2016; Koolen, Krahmer, & Swerts, 2015). A possible reason for *overspecification* is that speakers try to be informative by providing additional cues that might help referent identification. For instance, Arts, Maes, Noordman and Jansen (2011) showed that participants were more likely to overspecify when the task required more precise instructions than not. An alternative possibility is that when starting to speak, speakers do not necessarily know exactly which property is fully distinguishing. Using eye-tracking, Pechmann (1989) showed that nearly all participants began articulating referring

expressions before fixating all objects in the referential scene. Comparing the referent against every single referential alternative in the context requires time. Hence, Pechmann suggested that selection might begin with properties that are most easily recognisable rather than those that can rule out as many alternatives as possible. Dale and Reiter (1995) formalized such model, termed the *incremental algorithm*. In this model, the selection of properties is primarily led by *preference* rather than discriminability, as the algorithm simply iterates through a list of properties ranked according to the degree of preference. Preferred properties will be incrementally added to the description, ruling out at least one referential competitor that has not yet been ruled out, until the referring expression is fully distinguishing. An added property cannot be *backtracked* even if it results in redundancy. Hence, assuming that speakers select properties according to their preference, the model explains why speakers sometimes overspecify.

Adjective ordering

The aforementioned models are primarily concerned about *what* properties speakers tend to select in a given referential context, avoiding the question of *how* speakers sequence the chosen properties. Nevertheless, we can derive predictions pertaining to adjective ordering, by assuming that constraints that affect adjective selection also influence adjective ordering. Specifically, we may hypothesize that discriminability affects not only property selection but also word order. The communicative benefits of early discrimination are evident from research showing that comprehenders interpret speech incrementally. Eberhard, Spivey-Knowlton, Sedivy and Tanenhaus (1995) reported that the position of the discriminating word in a referential description determines latencies of eye fixations onto the target referent: When listening to instructions like *Touch the starred yellow square*, addressees were quicker at fixating the target square when the first adjective (*starred*) rather than the second adjective (*yellow*) provided the discriminating cue. Hence, if speakers formulate referring expressions to facilitate early referent identification, they should be

more likely to place the most discriminating modifier (i.e., one that rules out most referential alternatives) in an earlier position. We call this the *discriminatory efficiency hypothesis*.

However, language production research generally suggests that speakers choose a particular word order to ease production rather than comprehension. Specifically, availability-based production models assume that speakers preferentially choose word orders that allow the earlier placing of words or phrases that are more available to them to facilitate production (Bock, 1982; 1986a; Bock & Irwin, 1980; Bock & Warren, 1985; V.Ferreira & Yoshita, 2003; McDonald, Bock, & Kelly, 1993; Prat-Sala & Branigan, 2000; Tanaka, Branigan, McLean, & Pickering, 2011). Studies have found that availability due to conceptual salience (Bock & Warren, 1985; McDonald et al., 1993; Tanaka et al., 2011), semantic priming (Bock, 1986a), or discourse *givenness* (Bock & Irwin, 1980; V.Ferreira & Yoshita, 2003) influences the choice between active and passive voice (e.g., *The doctor administered the shock* vs. *The shock was administered by the doctor*) and dative alternations (e.g., *The rancher sold the cowboy the horse* vs. *The rancher sold the horse to the cowboy*). The assumption is that the incremental production of highly available words or phrases facilitates production processes, as it minimizes the need of buffering available information in memory. Hence, if similar production constraints underlie adjective ordering, speakers should preferentially select and order more salient or available properties early to facilitate reference production. We call this the *availability-based ordering hypothesis*.

Previous research

Experimental work on adjective ordering has mostly focused on understanding the nature of the preference for the size-before-color order observed in many languages. Although the findings from these studies were inconclusive, as we discuss below, the researchers proposed important predictions concerning how discriminability and availability might influence adjective ordering.

First, Danks and colleagues examined the link between discriminability and the ordering preference of size-before-color adjectives in the 1970s. Specifically, Danks and Glucksberg (1971) proposed that properties that are more intrinsic to the referent identity, such as color, tend to occur closer to the noun, because they are less informative for discrimination compared to those that are less intrinsic, such as size (e.g., *a large yellow banana;* bananas are usually yellow, whereas their size may vary). That is, speakers place less intrinsic adjectives before more intrinsic ones to facilitate discrimination (see Oller & Sales, 1969, for a similar proposal). In languages with pre-nominal modification, this means that less intrinsic and hence more discriminating adjectives should occur earlier, consistent with the discriminatory efficiency hypothesis. Danks and Schwenk (1972) tested this proposal by having English-speaking participants describe an object in a visual array, which contained a referential competitor that contrasted with the target in either color, size or both. Regardless of which property discriminated the target from the competitor, participants almost invariantly produced size-before-color orders (see also Belke, 2001, for similar results in German). When participants had to choose, amongst alternatives, a spoken instruction that would correctly identify the referent, they sometimes shifted their preference to the color-before-size order if color was distinguishing in the context. Yet this happened only when the first-mentioned adjective was stressed. When color was distinguishing, the color-before-size order had the distinguishing property stressed, whereas the size-before-color order had the non-distinguishing property stressed. The reverse was the case when size was distinguishing; the distinguishing property was stressed in the size-before-color order, but the non-distinguishing property was stressed in the color-before-size order. Engelhardt and F.Ferreira (2014) recently found that distinguishing modifiers tend to have a longer articulatory duration than non-distinguishing modifiers. Hence, because Danks and Schwenk had no counterbalancing condition, where the second-mentioned adjective was stressed, it remained unclear whether the preferences were due to the word order or stress (Richards, 1975). Moreover, we do not know whether and to what extent *comprehension* preferences might inform production

preferences (Pechmann, 1989; Schriefers & Pechmann, 1988). Hence, it remains open whether discriminability affects adjective ordering.

Second, Schriefers and Pechmann (1988) discussed the role of availability in adjective ordering when they conjectured that the preference for size-before-color orders reflects the scope of initial conceptual planning. In their study, speakers almost always referred to an object's semantic category (denoted by the head noun), regardless of whether it was required for unique identification. Likewise, speakers frequently mentioned an object's color when it did not contribute to its unique identification, whilst they rarely overspecified the object's size. Subsequent research also reported higher rates of overspecification with color relative to size (Belke, 2006; Belke & Meyer, 2002; Engelhardt & Ferreira, 2014; Engelhardt et al., 2006; Gatt et al., 2016; Pechmann, 1989), though the rate of color overspecification varies depending on color (a)typicality for given objects (Sedivy, 2003; Westerbeek, Koolen, & Maes, 2015). Importantly, Schriefers and Pechmann suggested that speakers overspecify the referent's category as well as its color, because these properties are conceptually more available, constituting an entity's *basic conceptual unit*, based on which speakers plan other descriptive features[1]. Size is less conceptually available than color, and it tends to be included when this initial conceptual representation fails to identify the referent uniquely. Hence, size typically occurs further away from the noun than color in order to modify the combination of the color and noun. Belke (2006) elaborated on Schriefers and Pechmann's account further, arguing that adjective ordering reflects *perceptual grouping*: Speakers of pre-nominal languages preferentially produce size-before-color orders, as they first classify an object with *contextually-independent* properties, such as color and shape ([colour[object]]), before classifying them with *context-dependent* properties, such as size ([size[colour[object]]]), for discrimination.

[1] In keeping with this, some computational algorithms explicitly set the cost for producing the noun (category) and sometimes color modifiers to zero (e.g., Krahmer, Theune, Viethen, & Hendrickx, 2008).

Indeed, an object's color tends to be determined more quickly than its size (Belke & Meyer, 2002). A possible reason for this is that the properties of color and size differ in *absoluteness.* Martin (1969a; 1969b) pointed out that adjectives that denote *absolute* properties, such as color, are more available than relative properties, such as size, because they require fewer comparisons to select a particular property; for instance, whether a car is red or not is not dependent on the color of other cars, whereas the size of an object is dependent on its comparison set. Similarly, Bierwisch (1987) argued that size is a *scalar* or *dimensional* adjective that denotes a relative property, whereas color is only partially scalable; although its intensity may vary, its property can be determined without comparison with other objects. Indeed, in comprehension, Sedivy, Tanenhaus, Chambers, and Carlson (1999) found that the presence of a size-contrasted competitor led to faster referent identification following size-modified descriptions, whereas Sedivy (2003) reported no facilitation following the presence of a color contrast for color-modified descriptions in her unpublished experiment. More recently, Fukumura and Van Gompel (2017) found that the absence of a referential competitor immediately delays reading times for size modifiers, but not for color modifiers, which led them to argue that the redundancy of size modifiers is more costly because the semantic processing of size modifiers requires a contrast set.

A critical question for us is whether adjectives that are more frequently overspecified and hence more available *generally* occur closer to the noun. Martin (1969a; 1969b) made such a claim, proposing that the order of adjectives reflects the *inverse* of the order in which it is selected. He argued that absoluteness affects adjective ordering, because adjectives that denote more absolute properties, such as color, are more available, and the more available the adjective is, the closer to the noun it will be. However, as discussed earlier, availability normally affects word order such that more available words or phrases occur earlier, *not later*. Hence, his account assumes that availability affects adjective ordering differently from other word order choice, contra the availability-based ordering hypothesis we discussed earlier. However, the preference for size-before-color orders may

result from the difference in meaning (e.g., absoluteness) rather than the difference in availability. Moreover, research suggests that color-before-size orders in English are processed as a grammatical-semantic violation (e.g., Kemmerer, Weber-Fox, Price, Zdanczyk, & Way, 2007). Hence, investigating the ordering preference of size and color adjectives can be problematic, because the preference may be determined semantically or it is too conventionalized to be easily overridden by communicative and cognitive constraints.

Finally, Haywood, Pickering and Branigan (2003) examined the role of speakers' *helpfulness* to their addressee, another communicative constraint on adjective ordering. They did this, using color and pattern modifiers. In their study, participants took turns with their confederate in describing and identifying simple geometric figures. The figures varied in color and pattern, and the interlocutors had to find the correct figure from a box, as described by their partner. Describing the figures using a pattern-before-color order (e.g., *stripy orange square*) was more helpful to the addressee when the figures in the box were grouped by color first before being subdivided by pattern than when the figures were grouped by pattern before color. When the confederate matched adjective ordering with the figure arrangement order of the box, participants showed a marginal tendency for a similar behaviour, but the effect did not generalize across items. When the confederate did not match adjective order with the sorting order of the figures, participants showed no sensitivity to the context, indicating that participants were merely responding to their partner's "prompt". Indeed, overall, participants tended to repeat the adjective order produced by the confederate; they were more likely to choose color-before-pattern orders rather than pattern-before-color orders when the confederate produced color-before-pattern orders (as opposed to the reversed orders) in the preceding trial. In this study, participants described the target figure in isolation from the referential alternatives. In such situations, the discriminating function of adjectives might not have been sufficiently clear to speakers, leaving it open that helpfulness, specifically discriminability, affects adjective ordering.

Current study

The primary goal of the current study was to examine whether speakers vary adjective order depending on discriminability for early referent identification or whether adjective ordering is primarily determined by availability. To this end, the three experiments we report below examined the ordering of color and pattern modifiers. Both modifiers denote context-independent, absolute properties, which neither depend on the meaning of the noun nor require a comparison set (e.g., whether a bow is green or striped is not dependent on the object's identity or the properties of other objects). Hence, the ordering of these properties may be more flexible compared to the ordering of size and color. As a preview, Experiment 1 showed that both discriminability and availability affected adjective ordering, and we replicated the findings in a dialogue setting in Experiment 2. Experiment 3 further examined the effects of discriminability and availability in conjoined adjectives. In the General Discussion, we propose a model that explains how discriminability and availability might influence the selection as well as the ordering of adjectives.

Experiment 1

Experiment 1 examined referential communication between a pair of naïve participants; one took part as *the speaker* and the other as *the listener*. The referential context was manipulated as in Fig.1, in which the objects marked in the box are the target referents and the other three objects are the referential competitors. In the *color-relevant* context (Fig.1A), the target's color (green) rules out all three competitors, whereas its pattern (spotted) rules out only one competitor. In the *pattern-relevant* context (Fig.1B), pattern rules out all three competitors, whereas color rules out only one alternative. In these conditions, the use of a single distinguishing adjective minimally specifies the target (e.g., *green bow* in Fig.1A; *spotted bow* in Fig.1B) and producing both properties (e.g., *green spotted bow*, *spotted green bow*) overspecifies the referent. In the *both-relevant* context (Fig.1C), each property rules out only two competitors, so the inclusion of both properties is necessary for full discrimination.

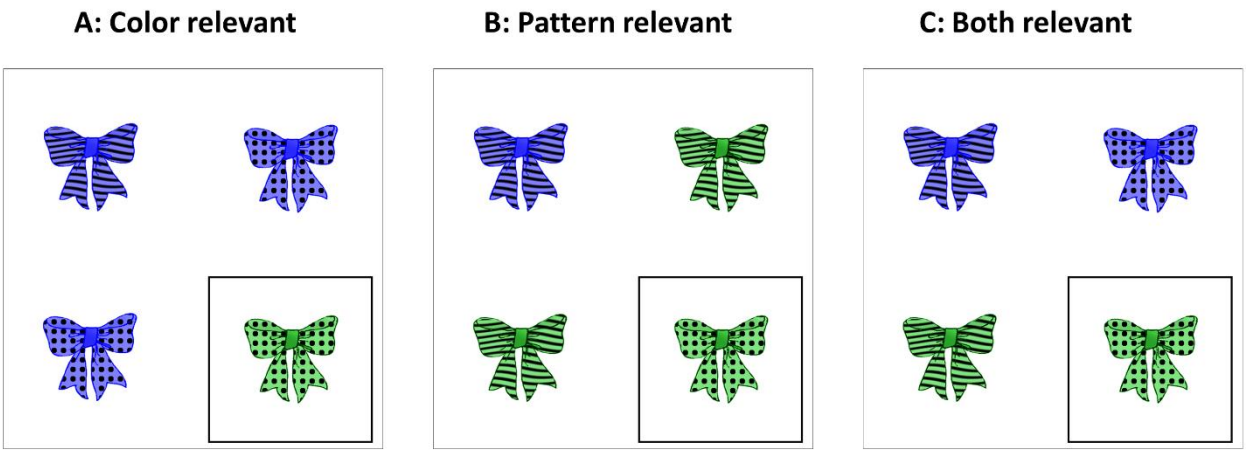

Fig.1 Example stimuli by condition

The *discriminatory efficiency hypothesis* predicts that discriminability affects adjective selection as well as ordering. Hence, the probability of choosing pattern over color will be higher in the pattern-relevant context than in the color-relevant context. Conversely, the probability of choosing color over pattern should be higher in the color-relevant context than in the pattern-relevant context. Thus, in these conditions, participants may use the fully discriminating property as the *first and only* adjective, by not including the least discriminating adjective. Crucially, when producing both modifiers, speakers should produce the most discriminating adjective before the less discriminating one; the chance of choosing *pattern-before-color* orders (rather than *color-before-pattern* orders) should be higher in the pattern-relevant context than in the color-relevant context. Overall, the discriminatory efficiency hypothesis predicts that the chance of choosing *pattern-initial* descriptions, comprising *pattern-only* responses (*spotted bow*) as well as *pattern-before-color* responses (*spotted green bow*) (relative to *all color-initial* descriptions comprising *color-only* responses and *color-before-pattern* responses) should be higher in the pattern-relevant context than in the both-relevant context, followed by the color-relevant context.

The *availability-based ordering hypothesis*, on the other hands, predicts that availability rather than discriminability affects adjective selection and ordering; adjectives that are more

available are more likely to be selected and ordered earlier. Tarenskeen, Broersma, and Geurts (2015) recently reported that participants overspecified color more often than pattern when the noun alone sufficiently identified the referent, which led them to argue that color was conceptually more salient than pattern. Thus, in the current study, speakers may overspecify color in the pattern-relevant context more frequently than pattern in the color-relevant context. That is, there should be more multiple modifier responses (i.e., fewer single modifier responses) in the pattern-relevant context than in the color-relevant context. Crucially, the availability-based ordering hypothesis predicts that properties that are overspecified more frequently are more available, and they should be produced earlier than those that are overspecified less often. Hence, across conditions, color-before-pattern orders should be favored more than pattern-before-color orders.

Method

Participants. Thirty-pairs of native speakers of British English, who reported to be monolingual and aged under 35, took part. They were recruited from the University of Strathclyde and the University of Dundee student community in exchange for course credits or cash. One pair of participants, who mostly used nouns rather than adjectives to describe patterns ("has stripes", "has polka dots", "has cheques") were replaced.

Materials and design. We created 48 experimental items. Each item consisted of three versions of a visual display (1000 × 1000 pixels) of four common objects such as socks, ties, umbrellas and tables. Fig.1 shows example stimuli, where the objects marked in the box were the target object that participants had to describe. The remaining objects served as referential competitors. The objects were taken from shaded images in Rossion and Pourtois (2004), and we applied one of three different colors (red, blue, green) and patterns (striped, spotted, chequered). In the *color-relevant* condition (Fig.1A), color ruled out all three competitors (only the target is green in Fig.1A), whilst pattern ruled out only one competitor, as the remaining two competitors had the same pattern as the target (the target is one of the three spotted objects in Fig.1A). Conversely, in the

pattern-relevant condition (Fig.1B), pattern ruled out all three competitors (no competitor is spotted in Fig.1B) and color ruled out only one (the target is one of the three green objects in Fig.1B). In the *both-relevant* condition (Fig.1C), each property ruled out only two competitors and the combination of both properties identified the referent uniquely (in Fig.1C, neither *green* nor *spotted* alone identifies the target uniquely). We counterbalanced, across conditions, the color and pattern of the target and competitor objects, as well as the target's positions in the display. To elicit different nominal constructions, we constructed 66 filler displays with objects that were distinguished in terms of size, orientation and category.

Thus, there were three conditions in total: (1) color-relevant condition, (2) pattern-relevant condition, and (3) both-relevant condition. Forty-eight experimental items and 66 filler items were randomly distributed across three lists, with each list containing 16 experimental items from each condition, with one version of each item, subject to the constraint that the same color or pattern of the target object should not occur in succession and there should be at least one filler between the experimental items.

Procedure. Participants drew lots to determine who would be the speaker or the addressee. The speaker and the addressee sat on opposite sides of a table, each facing a 1280 × 1024 pixels computer monitor. Before the experiment, the experimenter gave oral instructions. Participants who received the speaker's role were asked to describe objects, such that their addressee was able to identify them, and those who received the addressee's role were asked to select the objects using a computer mouse. The experimenter then presented example objects with varied colors and patterns, and named each color and pattern to establish naming agreement. After the instructions, five practice trials followed, where speakers (hereafter *participants*) referred to the objects that differed from the rest in color, pattern, size or direction. In the first three trials, the target objects were uniquely identifiable using either color or pattern. When participants produced multiple adjectives, the experimenter gave oral feedback that an alternative order was also possible. This was done to

discourage participants from strategically labelling properties in a fixed order. In the fourth practice trial, the target object was uniquely identifiable with the combination of size, color or pattern, and in the final practice trial, the combination of color and direction identified the object. To discourage spatial descriptions such as *the one on the left*, participants were told that their partner might not always see the same objects in the same position. Participants were free to use any word order, including post-nominal modification.

At the beginning of each trial, a large cross in the middle of the screen indicated the beginning of the trial. Speakers pressed a spacebar on a keyboard to proceed, triggering the presentation of a visual display, together with a short beep prompting a response. Speakers then described a target object indicated by a box in the display. The box did not appear in the addressee's display and the addressee had to select the described object using a computer mouse. The selected object was then highlighted by a grey box on both participants' displays. Speakers then determined whether the addressee identified the referent correctly by pressing *yes* or *no* key on their keyboard. This was done to ensure that speakers paid attention to their partner's comprehension. The speaker and addressee always kept the same roles throughout the experiment, and they were given as much time as needed to carry out their tasks. The experiment lasted about 30 mins and there was a brief break halfway through.

Scoring. We scored if participants produced a *color-before-pattern* (*CP*) order (n = 614) or a *pattern-before-color* (*PC*) order (n = 359). When they produced only one adjective, the response was coded as a *Color-only* (n = 271) or a *Pattern-only* (n = 94) description. These included cases where participants substituted the head noun with "one" (e.g., *the blue one*) (n = 24 with Color-only responses; n = 27 with CP order responses). Descriptions were coded as *Other* (7% of total responses) when participants did not produce nouns (n = 44); they revised the initial response that would alter our scoring classification (e.g., *the stripy book...blue*) (n = 38); they produced relative clauses (e.g., *the chequered pipe… that's red*) (n = 11), post-nominal modifiers (*cap green*

chequered) (n = 1) or used prepositional modifiers (e.g., *the red socks with black stripes*) (n = 8). These *Other* responses were excluded from analyses.

Results

Fig.2 presents the percentage of the different referring expressions produced by participants. As Fig.2 shows, in the color-relevant and the pattern-relevant context, participants produced *single modifier responses* (comprising Color-only responses and Pattern-only responses), by referring to the fully-discriminating property only, as well as *multiple modifier responses* (comprising both CP and PC order responses), referring not only to the fully-discriminating property but also to the least-discriminating one.

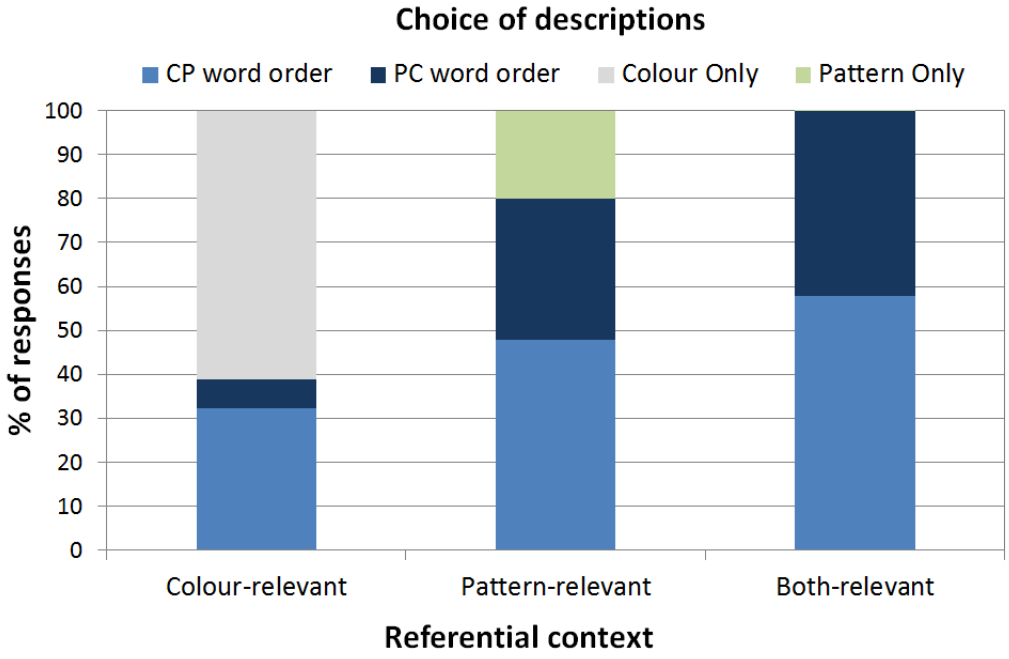

Fig.2. Percentages of descriptions by referential context (Experiment 1). CP = color-before-pattern, PC = pattern-before-color.

Throughout the study, we analysed choice of different referring expressions using logit mixed effects modelling (e.g., Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013) in R (R Core Team, 2016) using the lme4 package (Bates, Machler, Bolker, & Walker, 2015). Unless

stated otherwise, we adopted the maximal random effect structure justified by the design, by including by-participants and by-items random intercepts as well as random slopes for all the relevant fixed factors (Barr et al., 2013). When the maximal model did not converge, we included random slopes that improved the model fit (relative to intercepts-only models) (Baayen et al., 2008). For significance testing, we ran log-likelihood ratio χ^2 tests using model comparison, and we report the resulting *p* values, together with β values associated with the significant fixed effects.

In Experiment 1, we contrasted three conditions; color-relevant context, pattern-relevant context, and both-relevant context. We carried out planned comparisons by comparing a model that distinguished all three conditions, which we call the *full model*, with models that collapsed the relevant comparison conditions, which we call *collapsed models*; for instance, to contrast the color-relevant context with the pattern-relevant context, we compared a model that collapsed across these two contexts with a model that distinguished all three contexts. For the random slopes for condition, the relevant context pair was coded as 1 and -1 and the irrelevant context as 0 in both models. If the full model significantly improved the fit of the collapsed model, we conclude that the distinction between these two contexts had an effect on predicting the choice of referring expressions.

Discriminability. We examined if discriminability affected the choice between PC and CP orders when participants produced multiple modifiers (by excluding single modifier responses). First, we compared the likelihood of choosing PC and CP orders in the pattern-relevant condition and in the color-relevant condition, by comparing a model that collapsed these two conditions to a full model. PC and CP order responses were coded as 1 and 0 respectively. Model comparison revealed that a full model showed a significant better fit than a collapsed model, $\chi^2(1) = 22.50$, $p < .001$ ($\beta = 1.85$, $SE = 0.40$, $z = 4.61$, $p < .001$). This confirmed that there were more PC orders (fewer CP orders) in the pattern-relevant condition ($M = 40\%$, $SE = 3\%$) than in the color-relevant condition ($M = 17\%$, $SE = 3\%$). A full model showed no improvement relative to a model that collapsed across the pattern-relevant and both-relevant conditions, $\chi^2(1) = 0.01$, $p = .930$. Hence, the likelihood of

producing PC orders (relative to CP orders) did not significantly differ between the pattern-relevant context (*M* = 40%, *SE* = 3%) and the both-relevant context (*M* = 42%, *SE* = 2%). A full model also improved the model fit relative to a model that collapsed across the color-relevant and both-relevant conditions, $\chi^2(1) = 24.17$, $p < .001$ ($\beta = 1.86$, *SE* = 0.40, $z = 4.70$, $p < .001$), indicating that there were more PC orders (fewer CP orders) in the both-relevant context (*M* = 42%, *SE* = 2%) than in the color-relevant context (*M* = 17%, *SE* = 3%).

The discriminatory efficiency hypothesis predicts that discriminability affects not only adjective ordering but also adjective selection. In the pattern-relevant context, participants frequently produced single modifier responses, specifying pattern as the first and only modifying property. Likewise, the color-relevant context saw frequent usage of color adjectives as the first and only modifier. We thus analysed if the choice between all *pattern-initial* responses, comprising both PC orders and Pattern-only responses (coded as 1), and all *color-initial* responses, comprising both CP and Color-only responses (coded as 0), was affected by the relative discriminability of the modifiers, as predicted by the discriminatory efficiency hypothesis. A full model provided a better fit than a model that collapsed the pattern-relevant context and the both-relevant context, $\chi^2(1) = 9.01$, $p = .003$ ($\beta = -0.63$, *SE* = 0.20, $z = -3.18$, $p = .001$). This indicated more pattern-initial responses (i.e., fewer color-initial responses) in the pattern-relevant context (*M* = 52%, *SE* = 2%) than in the both-relevant context (*M* = 42%, *SE* = 2%). Furthermore, the distinction between the both-relevant context and the color-relevant context improved the model fit, $\chi^2(1) = 41.44$, $p < .001$ ($\beta = 3.78$, *SE* = 0.69, $z = 5.44$, $p < .001$), reflecting more pattern-initial responses (i.e., fewer color-initial responses) in the both-relevant context (*M* = 42%, *SE* = 2%) than in the color-relevant context (*M* = 7%, *SE* = 1%).

Availability. We then analysed the likelihood of overspecification in the color-relevant and pattern-relevant contexts as an index of availability. The likelihood of producing multiple modifier responses (coded as 1) relative to single modifier responses (coded as 0) in these conditions indicated the degree of overspecification. Random effects included by-participants and by-items intercepts.

Analyses on the choice between multiple modifier and single modifier responses showed that a full model showed a significant improvement relative to a model that collapsed across the color-relevant context and the pattern-relevant context, $\chi^2(1) = 233.80$, $p < .001$ ($\beta = 2.96$, $SE = 0.24$, $z = 12.23$, $p < .001$). This confirmed more multiple modifier responses (i.e., fewer single modifier responses) in the pattern-relevant context than in the color-relevant context. In other words, participants overspecified color in the pattern-relevant context more ($M = 80\%$, $SE = 2\%$) than pattern ($M = 39\%$, $SE = 2\%$) in the color-relevant context. We then analysed which property speakers generally tended to mention in an earlier position, by examining the choice between PC and CP order responses *across conditions* (by excluding single modifier responses). A significant negative intercept in a random-intercepts-only model with no fixed effects indicated that *overall*, participants produced CP orders ($M = 63\%$, $SE = 2\%$) more frequently than PC orders ($M = 37\%$, $SE = 2\%$), $\beta = -0.70$, $SE = 0.33$, $z = -2.14$, $p = .032$. Hence, participants generally placed in an earlier position the property that they overspecified more frequently.

In sum, the results provided clear support for the discriminatory efficiency account. Participants were more likely to choose pattern-before-color orders (rather than color-before-pattern orders) when pattern was fully discriminating than when color was, indicating that speakers vary adjective ordering to deliver more discriminating information early. Moreover, the relative discriminability of the adjectives affected the choice of descriptions, with the discriminating adjective occurring as the first and often only modifier. Finally, overall, participants overspecified color more often than pattern, indicating that color was generally more available than pattern. Crucially, availability also affected adjective ordering: Across conditions, participants tended to produce color before pattern. That is, the more frequently an adjective was overspecified, the more likely it was to be produced earlier, providing support for the availability-based ordering account that availability promotes the earlier placing of an adjective.

Experiment 2

Haywood et al. (2003) reported that contextual helpfulness only marginally affected the order of color and pattern adjectives, and the tendency was only observed when the confederate produced the helpful adjective orders. Overall, participants persisted in the adjective order produced by their partner, demonstrating an effect of *priming* on adjective ordering. A similar priming effect was more recently reported by Goudbeek and Krahmer (2012), who reported that speakers tended to repeat the order of attributes produced by their partner. According to Pickering and Garrod (2004), such priming results from enhanced *alignment* of representations between the interlocutors, facilitating not only production but also comprehension. Adversely, priming might render referential choice more *formulaic*, affecting the range of descriptions speakers consider for a given partner (e.g., Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987). For instance, Brennan and Clark showed that if an object had been introduced with a subordinate-level expression rather than a basic-level expression (e.g., *pennyloafer* rather than *shoe*) as the context contained another category exemplar (e.g., another shoe), the interlocutors tended to repeat the subordinate expression in subsequent trials, even though a basic-level term (*shoe*) was optimally informative. Hence, whilst referential choice becomes easier in a dialogue due to representational alignment (Garrod & Pickering, 2004), priming might alter the speaker's communicative strategy or underlying representations, such that discriminability plays little role.

The goal of Experiment 2 was to test the robustness of the discriminability effect in a dialogue setting. In Experiment 1, only one interlocutor produced descriptions; in a real dialogue, both interlocutors speak as well as comprehend (Pickering & Garrod, 2004). Hence, in Experiment 2, participants took turns with a confederate in describing and identifying objects, as in Haywood et al. (2003). In the *prime* trial, the confederate produced a prime description, using either a color-before-pattern (*CP prime*) or pattern-before-color (*PC prime*) order in the both-relevant context. In the *target* trial, participants described a target object in the color-relevant or pattern-relevant context. To

maximize the chance of obtaining a priming effect, prime and target always used the same nouns (cf. Cleland & Pickering, 2003; Pickering & Branigan, 1998). Hence, for example, the confederate described the green chequered sock in Fig.3C as *green chequered sock* or *chequered green sock*, and participants described the red striped sock presented either in Fig.3A or in Fig.3B. If priming affects adjective ordering, speakers should produce PC orders more often after a PC prime than a CP prime. Importantly, if the effect of discriminability generalizes to a dialogue setting, speakers should produce PC orders (e.g., *stripy red sock*) more often in the pattern-relevant context than in the color-relevant context.

Method

Participants. Thirty-two native speakers of British English were recruited from the same population as in Experiment 1.

Procedures and materials. The participant and the confederate took turns in describing and identifying pictures. This contrasts with Experiment 1, where the participants kept the same roles throughout the experiment. In the confederate's turn, the confederate produced, in a neutral tone, a scripted description (*prime*), which appeared in the middle of the confederate's monitor. Participants then identified the corresponding referent using a computer mouse. Once participants selected a picture, the confederate pressed the spacebar on a keyboard, triggering the presentation of a fixation cross, which indicated the beginning of the next trial. As in Experiment 1, participants referred to a target object, marked by a box in their display. The confederate then selected the referent on their display using their own computer mouse.

We used the same 48 sets of experimental items used in Experiment 1. Fig.3 shows example stimuli. The prime and target object pair always had the same nominal category but differed in color and pattern. The prime trials occurred in the *both-relevant* context (Fig.3C), where both properties were necessary for discrimination, and the confederate produced a color-before-pattern (*CP*) or

pattern-before-color order (*PC*) prime description in a neutral tone. The target object was presented either in the *color-relevant* context (Fig.3A), where color distinguished the target, or in the *pattern-relevant* context (Fig.3B), where pattern distinguished the target. For instance, the confederate described the green chequered sock in Fig.3C as *green chequered sock* or *chequered green sock*, and participants described the red striped sock, which was presented in either Fig.3A or Fig.3B.

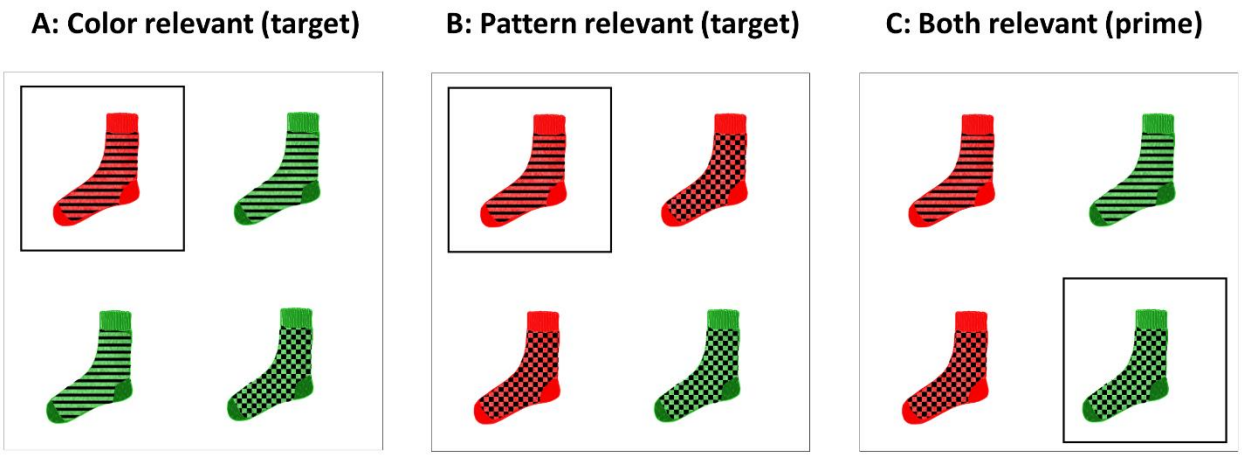

Fig.3. Example stimuli by condition.

Design. We used a 2 (prime: CP vs. PC prime) × 2 (referential context: color-relevant vs. pattern-relevant) within-subjects and within-items repeated measures design. The four experimental conditions were randomly distributed across four lists, each having 12 experimental items from each condition, one version from each item, and 96 filler items.

Scoring. As in Experiment 1, we scored whether participants produced *CP* (*n* = 828) or *PC* orders (*n* = 272). Single modifier responses were coded as *Color-only* (*n* = 368) or *Pattern-only* (*n* = 66) responses. We included cases where participants replaced the noun with "one" (30 in Color-only responses, 30 in CP orders, 9 in PC orders). Responses were coded as *Other* when participants changed the response type (e.g., *the spotty trousers… the blue spotty trousers*) (*n* = 2), and these were excluded before analyses.

Results

Fig.4 reports the means by prime (CP vs. PC prime) and referential context (color-relevant vs. pattern-relevant). As in Experiment 1, participants frequently produced single modifier responses, though the rates were relatively lower (by approximately 10%); color-only responses occurred 48% of the time in the color-relevant context and pattern-only responses 9% of the time in the pattern-relevant context. This might be because in the current experiment, the confederate produced multiple modifier descriptions as prime and participants produced more multiple modifiers in these conditions, influenced by the confederate's prime description.

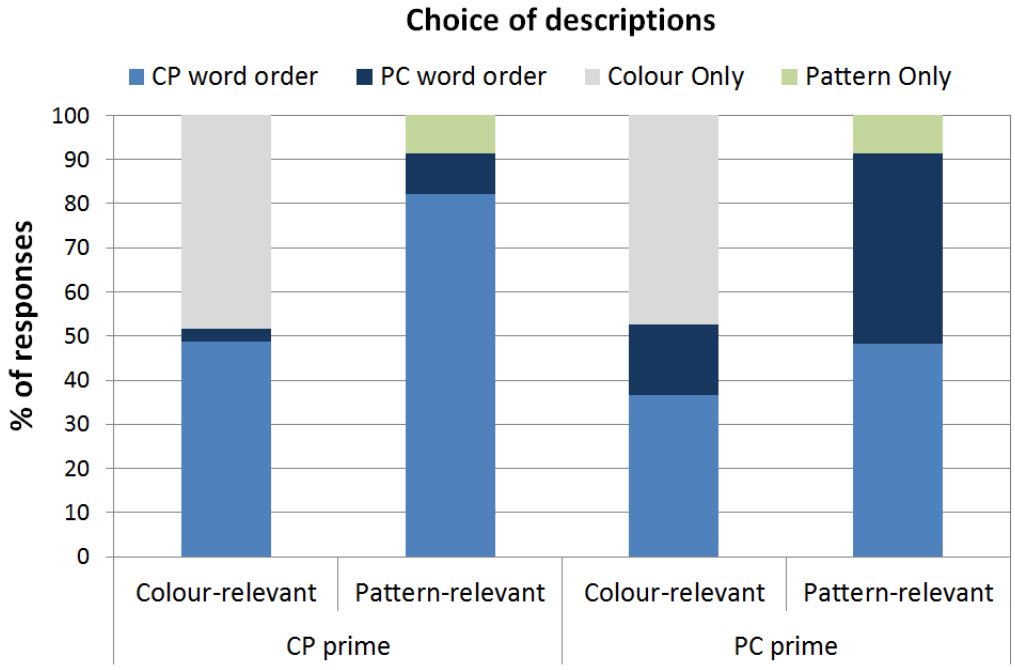

Fig.4. Percentages of descriptions by prime and referential context (Experiment 2). CP = color-before-pattern, PC = pattern-before-color.

Discriminability and priming. We first examined the choice between PC and CP word orders (excluding single modifier responses) as a function of prime (PC prime vs. CP prime) and referential context (color-relevant vs. pattern-relevant). These fixed effects were mean-centred to interpret the results in terms of main effects and interactions. As before, we used log-likelihood ratio χ^2 tests to

compute p-values, and PC and CP responses were coded as 1 and 0 respectively. Random effects included by-participants and by-items random intercepts and a by-participants random slope for prime. Inclusion of prime to a model significantly improved the fit relative to a model containing only referential context, confirming a main effect of prime, $\chi^2(1) = 36.38$, $p < .001$; participants produced more PC orders (i.e., fewer CP orders) following a PC prime (M = 41%, SE = 2%) rather than a CP prime (M = 8%, SE = 1%). Crucially, inclusion of referential context (discriminability) to a model containing only prime also improved the model fit, indicating a significant main effect of referential context; participants chose more PC orders (i.e., fewer CP orders) in the pattern-relevant context (M = 29%, SE = 2%) than in the color-relevant context (M = 18%, SE = 2%), $\chi^2(1) = 23.78$, $p < .001$. The inclusion of the prime × referential context interaction did not improve the model fit relative to a model containing the two main effects only, $\chi^2(1) = 0.05$, $p = .827$. Table 1 summarizes the final model.

Table 1. Coefficients for priming and discriminability in the choice between color-before-pattern and pattern-before-color orders (Experiment 2).

Predictor	B	*SE*	Wald *z*	*p*
(Intercept)	-2.20	0.37	-5.99	<.001
Priming	1.50	0.21	7.10	<.001
Discriminability	0.55	0.12	4.66	<.001

Availability. As in Experiment 1, we examined the likelihood of overspecification as an index of availability, by examining the choice between all multiple modifier responses (coded as 1) and single modifier responses (coded as 0). Random effects included by-participants and by-items random intercepts and a by-participants random slope for referential context. Inclusion of referential context significantly improved the model fit (relative to a model with prime only), with more

multiple modifier responses (relative to single modifier responses) in the pattern-relevant context (*M* = 91%, *SE* = 1%) than in the color-relevant context (*M* = 52%, *SE* = 2%), $\chi^2(1) = 25.92$, $p < .001$ (β = 2.59, *SE* = 0.61, *z* = 4.24, $p < .001$). In other words, participants overspecified color more frequently in the pattern-relevant context than pattern in the color-relevant context. Neither priming, $\chi^2(1) = 0.04$, $p = .834$, nor the context × prime interaction, $\chi^2(1) = 0.13$, $p = .722$, had an effect on the choice of multiple modifier responses. As in Experiment 1, we then examined the overall preference of CP and PC orders. A significant negative intercept in Table 1 indicates that overall, participants chose CP word orders (*M* = 75%, *SE* = 1%) more frequently than PC word orders (*M* = 25%, *SE* = 1%).

In sum, as in Experiment 1, color was overspecified more frequently and speakers generally tended to produce the more available property earlier, in keeping with the availability-based ordering account. In addition, participants persisted in the adjective order produced by their addressee, demonstrating an effect of priming. Crucially, these did not eliminate the effect of discriminability observed in Experiment 1. Although the coefficients in Table 1 indicate that the preference for the color-first order and priming made larger contributions, speakers reliably placed the more discriminating adjective earlier, providing support for the discriminatory efficiency hypothesis.

Experiment 3

The first two experiments demonstrated that adjective ordering is affected by discriminability, availability and priming. Here we examined the generalizability of these effects using coordinated modifiers. A common assumption is that adjective order reflects hierarchical relations (e.g., Everaert, Huybregts, Chomsky, Berwick, & Bolhuis, 2015). Consider Fig.5A as an illustration of the hierarchical representation that may underlie *green spotted bow*. In this diagram, *green* occupies the highest position in the nominal structure and modifies *spotted* and *bow* as a phrase, which is located lower in the structure. Researchers have assumed that the preference for size-before-color orders over color-before-size orders in languages with pre-nominal modification

arises because hierarchical structuring interfaces with conceptual or perceptual grouping (Belke, 2006; Shriefers & Pechmann, 1988). Likewise, when researchers linked the semantic restrictions on adjective ordering to discriminability (Dankes & Gluckesberg, 1971; Oller & Sales, 1969) and availability (Martin, 1969b), they typically assumed the involvement of hierarchical modification. For instance, Oller and Sales argued: "Relative to a particular context of situation, modifiers in the English NP are ordered from the least limiting to the most limiting proceeding away from the head noun, such that each additional modifier denotes a sub-class included by the denotation of the unit it modifies" (p.222).

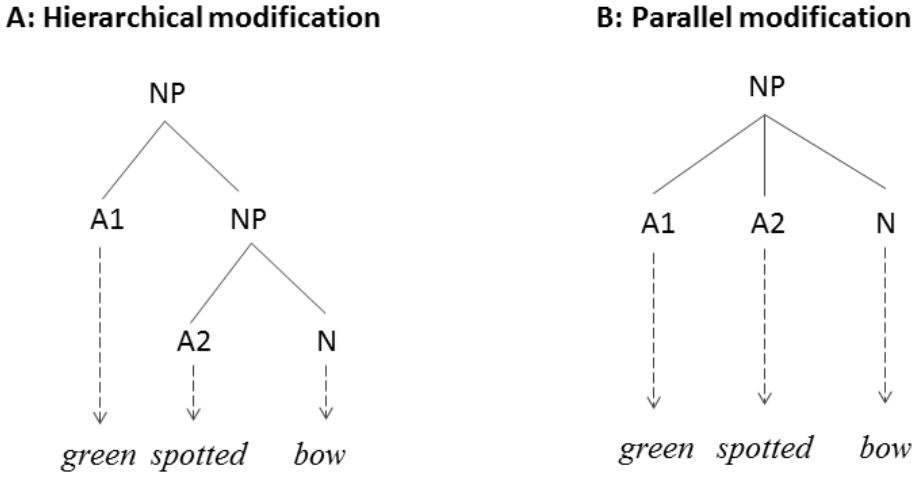

Fig.5. Underlying representations for green spotted bow. NP = noun phrase, N = noun, A1 = first-mentioned adjective, A2 = second-mentioned adjective.

Yet adjective order might not necessarily be hierarchically structured; each adjective can separately and independently modify the noun, as shown in Fig.5B, adapted from what Sproat and Shih (1991) called *parallel modification*. Sproat and Shih argued that under this nominal structure, the semantic restrictions on adjective ordering, such as the preference for the size-before-color order as opposed to the reversed order, should not apply. For instance, the ordering of size and colour adjectives should be interchangeable in Mandarin when the particle *de* following the adjectives

blocks hierarchical modification. Sproat and Shih also predicted no ordering restrictions for French, where, they assumed, post-nominal modifiers take the form of parallel modification, often coordinated by the conjunct *et* ('and'). Similar observations were reported by Hezron (1978), who found no preference in Somali, where the conjunct *oo* ('and') obligatorily occurs between the adjectives, and in spoken Arabic and modern Hebrew, in which a conjunct precedes the final adjective (though Hezron noted that the presence of a definite article before each adjective may also be responsible). We must note that syntactic accounts often assume *asymmetry* in conjunction, postulating a hierarchy similar to the one in Fig.1A (see Cormack & Smith, 2005, for a review). Yet the reported insensitivity of conjoined adjectives to semantic restrictions raises the possibility that neither discriminability nor availability determine ordering preferences for coordinated adjectives, *if* the communicative and cognitive constraints affect adjective ordering similar to semantic constraints.

Moreover, previous experimental work suggests that availability does not always influence the ordering of conjoined noun phrases. Bock and Warren (1985) showed that conceptual availability due to higher imageability has an effect on the choice between active and passive voice as well as dative alternations, such that the structures that allowed the earlier occurrence of the nouns with higher imageability were more favored. Interestingly, imageability did not influence the order of nouns within conjuncts (e.g., *time and winter* vs. *winter and time*), consistent with the view that conceptual availability affects a "hierarchy of grammatical relations", but not word order directly. However, using freer word order languages like Greek and Japanese, Branigan, Pickering, and Tanaka (2008) found that participants tended to produce animate patients in sentence-initial position by using either a passive or shifted active voice, indicating that animacy affects function assignment as well as linear positioning (see also, Kempen & Harbusch, 2004, for similar results in German). Although Tanaka et al. (2011) subsequently reported that animacy does not reliably influence the order of nouns within conjuncts in Japanese, McDonald et al. (1993) found animate-first preference

for noun phrase conjuncts in English that occurred in isolation, outside of sentence contexts, suggesting that conceptual availability can influence the ordering of conjoined nouns.

The third and final experiment thus aimed to determine whether discriminability and availability influence the ordering of conjoined adjectives, such as *red and stripy sock*, resulting in the earlier occurrence of the more discriminating or more available adjective. To this end, the confederate varied the ordering of color and pattern adjectives using the coordinating conjunction *and* (e.g., *green and chequered sock* vs. *chequered and green sock* in Fig.3C). The referential context was manipulated as before, such that participants referred to the target object in the either color-relevant (Fig.3A) or pattern-relevant context (Fig.3B).

Method

Participants. Thirty-two participants from the same population as in the previous experiments took part. Data from two participants, who reported to have problem with visual processing, were replaced.

Materials, design, procedure. These were the same as in Experiment 2, except that the confederate's primes always contained coordinating conjunct *and*.

Scoring. Conjoined multiple modifier responses were scored as *C&P* responses (n = 250) if color preceded pattern and *P&C* responses (n = 31) if pattern preceded color. As before, non-conjoined responses were coded as *CP* (n = 630) or *PC* (n = 169) orders and single modifier responses were classified as either *Color-only* (n = 388) or *Pattern-only* (n = 66) responses. These included cases where the head noun was substituted with "one" (44 in Color-only, 28 in CP, 3 in PC and 1 in P&C responses). Other responses were coded as *Other* (n = 2) and excluded from analyses.

Results & discussion

Fig.6 reports the means. The rates of single modifier responses were comparable to those in Experiment 1; they occurred approximately 51% of the time in the color-relevant context, and 9% in

the pattern-relevant context. Multiple modifier responses comprised two types; conjoined modifier responses (C&P and P&C) and non-conjoined modifier responses (CP and PC).

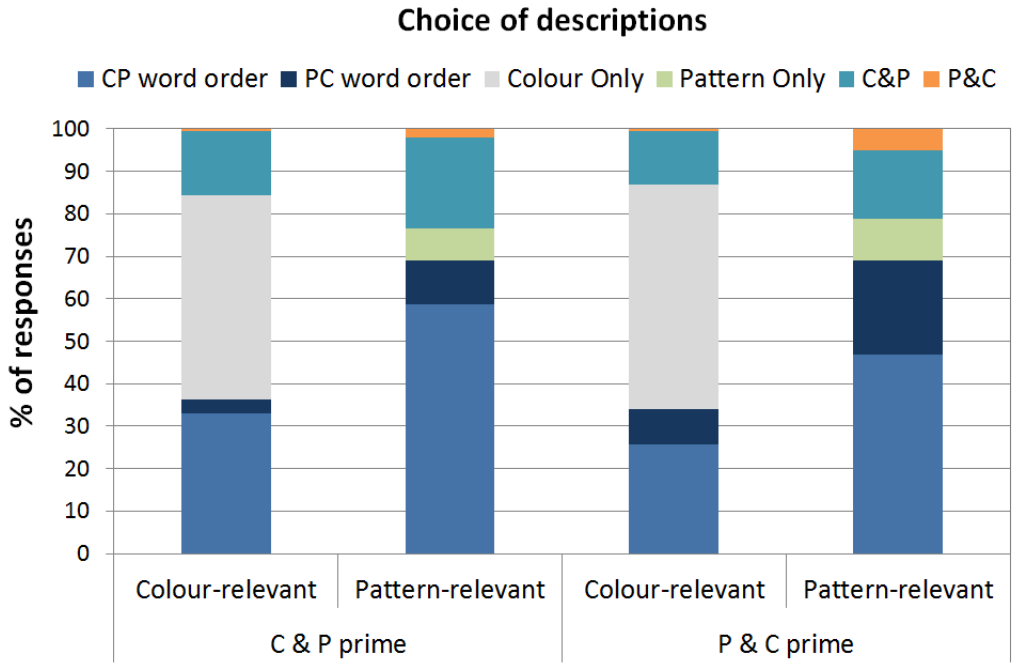

Fig.6. Percentages of descriptions by prime and referential context (Experiment 3). CP = color-before-pattern, PC = pattern-before-color, C&P = color-and-pattern, P&C = pattern-and-color.

P&C vs. C&P conjoined modifiers: We first analysed the ordering choice of the conjoined modifiers (by excluding other target responses) as a function of referential context (color-relevant vs. pattern-relevant) and prime (C&P prime vs. P&C prime). P&C responses were coded as 1 and C&P as 0. Random effects included by-participants and by-items random intercepts. Inclusion of referential context significantly improved the model's fit (relative to a model containing prime only), confirming a main effect of referential context; participants produced more P&C (fewer C&P) responses in the pattern-relevant context (M = 16%, SE = 3%) than in the color-relevant context (M = 4%, SE = 2%), $\chi^2(1) = 9.37$, $p = .002$. Inclusion of prime also contributed to the fit (relative to a model containing only referential context); participants produced more P&C (fewer C&P) responses following P&C primes (M = 16%, SE = 3%) than following C&P primes (M = 7%, SE = 2%), $\chi^2(1)$

= 6.56, p = .010. These main effects did not interact, $\chi^2(1) = 0.33$, p = .564. Table 2 provides the final model.

Table 2. Coefficients for priming and discriminability in the ordering of conjoined adjectives (Experiment 3).

Predictor	β	*SE*	Wald *z*	*p*
(Intercept)	-4.06	1.28	-3.16	.002
Priming	0.71	0.31	2.33	.020
Discriminability	1.02	0.38	2.66	.008

PC vs. CP word orders: We then analysed the choice between PC orders and CP orders as before. Random effects included by-participants and by-items random intercepts and a by-participants random slope for context. Model comparison revealed a significant main effect of prime, with more PC orders (fewer CP orders) after a P&C prime (M = 30%, SE = 2%) than a C&P prime (M = 13%, SE = 2%), $\chi^2(1) = 45.84$, $p < .001$. A significant main effect of referential context indicated that there were more PC orders (fewer CP orders) in the pattern-relevant context (M = 24%, SE = 2%) than in the color-relevant context (M = 16%, SE = 2%), $\chi^2(1) = 8.97$, p = .003. There was a marginally significant prime × display interaction, $\chi^2(1) = 3.54$, p = .060, with only a slightly stronger (2%) effect of referential context after the P&C prime than the C&P prime. Table 3 summarizes the fixed effects.

Table 3. Coefficients for priming and discriminability in the ordering of non-conjoined adjectives (Experiment 3).

Predictor	β	*SE*	Wald *z*	*p*
(Intercept)	-2.35	0.46	-5.12	<.001
Priming	0.76	0.12	6.49	<.001
Discriminability	0.65	0.25	2.56	.011

Availability: We examined the choice between single and multiple modifier responses as before. Random effects included by-participants and by-items random intercepts and by-participants slopes for the main effects and interaction. Inclusion of referential context contributed to the model fit, with more multiple modifier responses (fewer single modifier responses) in the pattern-relevant context (M = 91%, SE = 1%) than in the color-relevant context (M = 49%, SE = 2%), $\chi^2(1) = 15.25$, $p < .001$ ($\beta = 1.47$, $SE = 0.26$, $z = 5.56$, $p < .001$). Hence, as in the previous experiments, participants overspecified color in the pattern-relevant context more often than pattern in the color-relevant context. Neither priming, $\chi^2(1) = 0.97$, $p = .325$, nor the interaction, $\chi^2(1) = 0.85$, $p = .358$, significantly influenced the likelihood of overspecification. Importantly, the significant negative intercepts reported in Table 2 and 3 suggest that when producing multiple modifiers, participants tended to produce color before pattern in both conjoined modifier responses (M = 89%, SE = 2%) and in non-conjoined modifier responses (M = 79%, SE = 1%). Hence, availability affected the ordering of adjectives in both constructions.

In sum, Experiment 3 showed that both discriminability and availability affect the ordering of conjoined and non-conjoined modifiers. In both nominal constructions, the coefficients for the intercepts indicate a strong preference for color-before-pattern orders. In conjoined adjectives, discriminability also had a larger coefficient (relative to priming), demonstrating strong sensitivity to

discriminability. Additionally, the ordering of conjoined modifiers primed that of non-conjoined adjectives. Although syntactic theories may differ on the exact syntactic representations for these constructions, our results point to the shared underlying representations for the ordering of conjoined and non-conjoined adjectives. We discuss the implications of these findings in the General Discussion.

General Discussion

All three experiments showed that the discriminability of adjectives affects adjective order: Speakers produced more discriminating adjectives before less discriminating adjectives. Adjective ordering was also affected by availability: The property of color, which was overspecified more often and hence more available than pattern, was more likely to be placed earlier. Moreover, in contrast with research that suggests that conjoined adjectives are free from semantic restrictions on adjective ordering, both discriminability and availability affected the ordering of conjoined adjectives, indicating that these communicative and cognitive constraints influence adjective ordering differently from semantic constraints. In the following section, we first discuss the mechanisms that may underlie the effects of discriminability and availability.

According to some models of object description (Frank & Goodman, 2012; Dale, 1989), speakers preferentially choose more discriminating properties for inclusion. Indeed, in the current study, participants almost always included the most discriminating property, whereas they often omitted the least discriminating property by producing minimally-specified referring expressions. Yet crucially, the rates of inclusion varied across different properties when they were least discriminating; speakers consistently overspecified color more often than pattern. Moreover, although the discriminatory efficiency hypothesis we set out earlier explains why participants placed the more discriminating adjective early, it does not explain why participants also included a less discriminating adjective at all, especially after having produced a fully discriminating one. These findings are hard to account for if we only focus on discriminability. Alternative models assume that

property selection is led by preference or availability (cf. Dale & Reiter, 1995; Pechmann, 1989), so speakers overspecified color more often than pattern, because color was more preferred or available. Yet the incremental algorithm proposed by Dale and Reiter (1995) specifically assumes serial selection, starting with the most preferred property: If the preferred property rules out any referential competitor, it should always be included, and property selection completes once the description contains fully distinguishing information. In other words, no other properties should be selected if the most preferred property rules out all referential competitors. In the current study, all properties ruled out at least one competitor. Hence, if property selection was led by a fixed order of preference, as predicted by the incremental algorithm, then color should always have been included. However, when pattern alone discriminated the referent, participants sometimes avoided color. Moreover, after having produced color, participants often went on to overspecify with pattern even when color was fully discriminating.

Thus, the existing models do not explain how speakers select and sequence adjectives. In the following, we therefore propose an alternative model, which we call *PASS* (Parallel-Activation-for-Selection-and-Sequencing). Unlike previous models, PASS assumes that both discriminability and availability *simultaneously* influence property selection. Specifically, both constraints influence adjective selection by influencing the activation of to-be-selected conceptual properties in the speaker's message representation. Unlike in the incremental algorithm, multiple properties are *activated in parallel* rather than in sequence for selection in PASS. If the activation of a conceptual property exceeds threshold, then its information will be passed onto the production processes, resulting in the production of a modifier. If two properties exceed the threshold activation, then both properties should be verbalized, even if it results in overspecification. Hence, speakers may select a less discriminating property if its availability boosts its activation. Likewise, a less available modifier can also be produced if discriminability enhances its activation.

Crucially, PASS proposes that the level of activation that determines property selection also influences word order. When two properties are chosen, the difference in the activation level determines their relative positioning: The more strongly the property is activated, the sooner the information is passed onto the production process and thus the more likely it is to become linguistically available earlier. In keeping with availability-based production models, PASS assumes that the more available an adjective is, the earlier it tends to occur in speech. If speakers focus their attention on discriminability, the most discriminating property should receive higher activation than a less discriminating property, leading to its earlier position in speech. If a less discriminating property receives higher activation than a more discriminating property, due to high saliency or availability, then it can also *win the race* and take an earlier position. Thus, although discriminability and conceptual salience are different conceptual constructs, they both simultaneously influence adjective selection and ordering via the same competitive mechanism, wherein their relative contribution varies depending on the speaker's attention allocation to discriminability and the degree of availability of specific properties. Such a mechanism not only enables the incremental production of the most available adjective, but also facilitates referent identification, at least in some situations, as it allows the earlier production of the most discriminating cue. Simultaneous contribution from discriminability and availability to property selection as well as ordering explains why speakers include the least discriminating adjective, sometimes even after having produced a fully discriminating one.

Importantly, adjective ordering is guided by an *abstract conceptual representation* that specifies property order. Most research on word order priming has focused on *syntactic* or *structural priming*, or the repetition of constituent structure, as a way of exploring the representation of syntax (e.g., Bock, 1986b; Bock & Loebell, 1990; Pickering & Branigan, 1998; see Branigan & Pickering, 2017; Pickering & V.Ferreira, 2008, for reviews). Cleland and Pickering (2003), for instance, showed that speakers are more likely to repeat a noun phrase with a relative clause (*the square that's*

red) when the prime description also contained a relative clause (*the square that's green*) than not (*the green square*). Cleland and Pickering interpreted the priming effect as resulting from shared abstract syntactic representations. Yet evidence also points to the involvement of conceptual representations in structural priming (Bock, Loebell, & Morey, 1992; Chang, Bock, & Goldberg, 2003; Heydel & Murray, 2000; Pappert & Pechmann, 2014). For instance, Bock et al. (1992) showed that speakers tend to persist in the animacy order of the prime sentence; they are more likely to choose *An alarm clock awakened a boy* (rather than *A boy was awakened by an alarm*) after *A boat carried five people* rather than *Five people carried a boat*. Moreover, Pappert and Pechmann (2014) showed in German that the order of thematic roles can be primed independent of grammatical function assignment (but see also Pickering, Branigan, & McLean, 2002). Researchers have also found that the assignment of *emphasis* to particular thematic roles persists across different constituent structures, raising the possibility that word order priming can operate completely at the conceptual level (Bernolet, Hartsuiker, & Pickering, 2009; Vernice, Pickering, & Hartsuiker, 2012; cf. Heydel & Murray, 2000). In the current study, it was the order of *conceptual* properties that was primed, and priming occurred across different adjectives and between different constituent structures, demonstrating the presence of abstract conceptual representations for adjective ordering. Presumably, participants persisted in particular adjective ordering as the prime reinforced these conceptual representations. The presence of such abstract representations might mean that discriminability and availability affect adjective order because competition between the conceptual properties influences the selection of the order that enables the earlier placing of the most activated and hence "winning" property. Hence, whereas priming directly influences the selection of such order, discriminability and availability affects ordering via competition between the properties.

Previous research on size and color adjectives in English and German has failed to observe an effect of discriminability on adjective ordering, as participants hardly reversed the preferred size-before-color orders (Belke, 2001; Danks & Schwenk, 1972). A possible reason, suggested by

researchers, is that color is more salient or available than size, as demonstrated by higher rates of overspecification for color (e.g., Schriefers & Pechmann, 1988), and the more available an adjective is, the closer it occurs to the modified noun (Martin, 1969b). However, this would mean that availability affects adjective ordering differently from other word order choice. Our experiments showed that availability affects adjective ordering in a manner consistent with availability-based language production models, with more available adjectives occurring earlier and hence further from the noun in English. It is therefore unlikely that the size-before-color order preference in languages with pre-nominal modifiers results from availability promoting closer proximity to the noun, thereby leading to the late placement of color adjectives. Instead, the size-before-color preference is better construed in terms of conceptual or perceptual grouping (Belke, 2006; Schriefers & Pechmann, 1988) or the difference in absoluteness interacting with hierarchical modification (e.g., Sproat & Shih, 1991).

Danks and colleagues attributed the semantic restrictions on adjective ordering to the pragmatic and communicative effect of discriminability; adjectives that denote properties that are more absolute or intrinsic to the modified noun occur closer to the noun, because less intrinsic or more relative properties are more informative for discrimination, and speakers preferentially place less intrinsic adjectives in an earlier position. That is, the effect of semantic absoluteness on adjective ordering reflects the general tendency to place more discriminating adjectives earlier. Crucially, the semantic restrictions on adjective ordering predict distance from the noun (e.g., Dixon, 1982; Hetzron, 1978; Martin, 1969b), and they are assumed to interact with hierarchical aspects of language (Scott, 2002; Sproat & Shih, 1991). Some linguistic observations indeed suggest that languages with post-nominal modifiers display the reverse ordering preferences from languages with pre-nominal modifiers (e.g., Greenberg, 1963; Hezron, 1978; Martin, 1969a); less intrinsic and more relative properties occur further from the modified noun but appear *later* in linear ordering due to the position of the modified noun. Danks (1976) therefore speculated that in languages with post-

nominal modifiers, more discriminating adjectives might occur *later* than less discriminating ones. By contrast, the discriminatory efficiency hypothesis predicts that the effect of discriminability is driven by a general, non-linguistic communicative constraint for facilitating early referent identification, so higher discriminability *cross-linguistically* promotes the earlier placing of the adjective. Indeed, other data from our lab suggests that adjectives that are more discriminating occur earlier in linear order even in Basque, a language with post-nominal modifiers (Fukumura & Santesteban, 2017). Hence, the semantic restrictions on adjective ordering, which interacts with the hierarchical structure of language, cannot be equated with communicative and cognitive constraints, which affect adjective ordering independent of the structure of language.

Furthermore, the current study sheds light on our understanding of the role of word order choice in effective communication. There has been very limited evidence that speakers vary word order to help comprehension. For instance, previous research has shown that speakers often fail to avoid syntactic ambiguity. Arnold, Wasow, Asudeh, and Alrenga (2004), for instance, found that speakers do not vary constituent order to avoid a temporal ambiguity such as *The foundation sent Grant's letters to Lincoln to a museum*, where comprehenders may initially wrongly analyze *to Lincoln* as the destination of *Grant's letters,* unlike in the unambiguous alternative, *The foundation sent a museum Grant's letters to Lincoln*. Moreover, V.Ferreira and Dell (2000) showed that the insertion of the optional complementizer *that* is not dependent on whether the following noun creates a temporary ambiguity, as in *The coach knew you missed practice,* or not, as in *The coach knew I missed practice.* Instead, speakers tended to include the complementizer *that* to maintain fluency when the upcoming material was less available. Haywood, Pickering, and Branigan (2005) showed that participants avoided temporary ambiguous prepositional phrases (e.g., *Put the penguin in the cup on the star*, where *in the cup* can be wrongly analyzed as the destination of *the penguin*) by using the relativizer *that's* (*Put the penguin that's in the cup on the star*) in the context of referential ambiguity (two penguins). Yet research has shown that listeners do not misinterpret these prepositional phrases

in the presence of referential ambiguity (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The referential context might have affected the use of the relative clause in Haywood et al., because speakers favored the restrictive relative clause over prepositional phrase modifiers to avoid referential ambiguity. Other studies have shown that speakers use prosody to avoid syntactic ambiguity (Kraljic & Brennan, 2005; Snedeker & Trueswell, 2003), though the use of prosody may merely reflect the speaker's representation of the phrase structure rather than syntactic ambiguity avoidance per se. To avoid syntactic ambiguity, speakers first need to construct and parse the to-be-produced syntactic structure, whereas speakers may not routinely engage in such linguistic monitoring during production processes (cf., V.Ferreira et al., 2005). By contrast, discriminability is a non-linguistic construct, presumably available from the early stage of the production processes (hence influencing the speaker's message generation). We have argued that discriminability affects adjective ordering via competition between different conceptual properties in the speaker's message representation, without the need of consulting alternative orders.

Finally, although we have focused on the ordering of pre-nominal modifiers, modifiers can occur post-nominally, as in *the bow that's green and striped* or *the green bow that is striped.* One might wonder how discriminability and accessibility affect adjective ordering in such constructions. The rate of post-nominal constructions using color and pattern adjectives was extremely low in the current study (0.6%); when participants failed to produce the most discriminating adjective first, it was either produced in the second-mentioned position pre-nominally, in keeping with the observation that modifiers typically precede the noun that they modify in English (e.g., Greenberg, 1963), or in a post-production repair (e.g., *the green bow… striped)*. Was there anything about the task that might have constrained participants' performance? We consider this very unlikely. First, speakers in the current experiments were free to use any word order, including post-nominal modification. Second, other studies that examined languages with pre-nominal modification also did not report that participants frequently produced adjectives post-nominally (Haywood et al., 2003;

Van Gompel, Van Deemter, Gatt, Snoeren, & Krahmer, under revision). In Van Gompel et al., for instance, both English and Dutch speakers produced post-nominal modification, such as *big light bulb that's green*, less than 1% of the time. Hence, the infrequent use of post-nominal adjectives does not appear to be specific to the current experiments. Any study that aims to examine post-nominal constructions in English would therefore need to use a task that specifically elicits these constructions.

We have argued that competition between different modifying properties affects adjective ordering. One might ask whether the position of the noun can also be affected by conceptual competition; color and pattern occurred before the noun in the current study, because these conceptual properties were more available than the object's category. As mentioned earlier, speakers of post-nominal language typically produce those properties post-nominally (Fukumura & Santesteban, 2017), in line with other experiments that contrasted pre-nominal and post-nominal modification (Brown-Schmidt & Konopka, 2008). Hence, the position of the noun in the current study seems to have resulted from the syntactic preferences in English. The question is how such syntactic representation comes into play. One possibility is that the syntactic preferences interact with conceptual ordering, such that the grammar can guide which conceptual properties compete; if the position of the noun is relatively fixed in the language, the object's category is very unlikely to compete with the properties for the modifiers. Alternatively, conceptual ordering may be at least partly independent from syntactic preferences, so the object's category could compete with other properties such as color and pattern. For instance, speakers sometimes replace a full noun phrase with an indefinite pronoun (e.g., *the green one*) when the modifier but not the noun identifies the referent uniquely, as in our study, and inclusion of the modifiers is often dependent on the discriminability of the noun. Moreover, using a priming task that elicited post-nominal modification, Haywood et al. (2003) found that participants were more likely to produce *noun-first* (e.g., *square that's red*) rather than *color-first* (*red square*) descriptions, when the referents were grouped by

shape rather than color. Although the post-nominal use of adjectives in this study was very rare despite priming, it suggests that our account (PASS) might be generalizable to the ordering of the properties of different grammatical categories (e.g., nouns vs. adjectives); when a post-nominal constructions is primed, the object category and the property of the adjectives can compete not only for selection but also sequencing.

Conclusion

The current study showed that speakers do not always produce the shortest possible referring expressions to achieve efficient referential communication; when producing seemingly redundant referring expressions, speakers can also vary adjective order to deliver the more discriminating information early to facilitate referent identification. Yet adjective ordering is also susceptible to a production-centered constraint, sometimes leading speakers to utter and place the most available adjective first, even if it is least informative. What speakers are likely to refer to and how is thus determined by these two conflicting forces, arising from the need to deliver helpful as well as fluent communication.

Acknowledgements

We thank Bill Woodside for implementing the experiments and Maria Nella Carminati and Roger van Gompel for their comments on an earlier version of this manuscript. We also thank Alice Wilson-McFarlane, Bianca Nardini, Bethan Gray, Sangeeta Bhopal, Kirsty Monaghan, Zahra Ahmed, Grace McPherson, Andrea Goh for their assistance with data collection and transcription and Vladislava Segen for her assistance with material construction. This work is supported by the Leverhulme Trust [RPG-2016-253].

References

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language, 51*, 55-70.

Arts, A., Maes, A., Noordman, L.G.M., & Jansen, C. (2011). Overspecification in written instruction. *Linguistics*, *49*, 555-574.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, B., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Belke, E. (2001). *On the time course of naming multidimensional objects in a referential communication task: Analyses of eye movements and processing times in the production of complex object specifications*. Doctoral dissertation, University of Bielefeld, Germany.

Belke, E. (2006). Visual determinants of preferred adjective order. *Visual Cognition*, *14*, 261-294.

Belke, E., & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during "same"-"different" decisions. *European Journal of Cognitive Psychology*, *14*, 237-266.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2009). Persistence of emphasis in language production: A cross-linguistic approach. *Cognition*, *112*, 300–317.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.

Bock, J. K. (1986a). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575–586.

Bock, J. K. (1986b). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*, 467–484.

Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*, 1–39.

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review, 99*, 150–171.

Bock, J.K., & Warren, R.K. (1985). Conceptual availability and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.

Branigan, H., & Pickering, M.J. (2017). Structural priming and the representation of language. *Behavioral and Brain Sciences.*

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*, 172-189.

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

Brown-Schmidt, S. & Konopka, A. E. (2008). Little houses and casas pequeñas: message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition, 109*, 274-280.

Chang, F., Bock, J. K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition, 90*, 29–49.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Cleland, A. A., & Pickering, M. J. (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,* 214–230.

Cormack, A. & Smith, N. (2005). What is coordination? *Lingua*, *115,* 395-418.

Dale, R. (1989). Cooking up referring expressions. In *Proceedings of the 27th annual meeting on Association for Computational Linguistics* (ACL'89) (pp. 68-75). Vancouver, BC: Association for Computational Linguistics.

Dale, R., & Reiter, E. (1995). Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, *19,* 233-263.

Danks, J. (1976). Language-processing mechanisms are not unique: a reply to Richards' critique. *American Journal of Psychology, 89,* 137-145.

Danks, J. H., & Glucksberg, S. (1971). Psychological scaling of adjective orders. *Journal of Verbal Learning and Verbal Behavior, 10,* 63-67.

Danks, J. H., & Schwenk, M. A. (1972). Pre-nominal adjective order and communication context. *Journal of Verbal Learning and Verbal Behavior, 11*, 183-187.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11,* 159-184.

Dixon, R. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax*. Amsterdam: Mouton de Gruyter.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409-436.

Engelhardt, P.E., & F.Ferreira, F. (2014). Do speakers articulate over-described modifiers differently from modifiers that are required by context? Implications for models of reference production, *Language and Cognitive Processes, 29,* 975-985.

Engelhardt, P.E., Bailey, K., & F.Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language, 54*, 554-573.

Everaert, M.B.H., Huybregts, M.A.C., Chomsky, N., Berwick, R.C., Bolhuis, J.J. (2015). Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences, 19*, 729-743.

V.Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340.

V.Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, *32*, 669–692.

V.Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*, 263-284.

Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science, 336*, 998.

Fukumura, F., & Santesteban, M. (2017). *Ordering adjectives for communicative efficiency in English and Basque.* In Proceedings of the 2017 conference on Architectures and Mechanisms in Language Processing (AMLaP 17).

Garrod, S., & Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*, 8-11.

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

Gatt, A., Krahmer, E., Van Deemter, K., & Van Gompel, R.P.G. (2016). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science, 41*, 1457–1492.

Goudbeek, M. and E. Krahmer (2012). Alignment in interactive reference production: Content planning, modifier ordering and referential overspecification. *Topics in Cognitive Science*, *4*, 269-289.

Greenberg, J.H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In: J.H. Greenberg (Ed.), *Universals of Language* (pp. 73-113). Cambridge, MA: MIT Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics: Vol. III. Speech acts* (pp. 41–58). New York: Academic Press.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2003). Co-operation and co-ordination in the production of noun phrases. In R. Alterman, & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 533-538). Mahwah: Lawrence Erlbaum Associates.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science, 16,* 362–366.

Hetzron, R. (1978). On the relative order of adjectives. In: H. Seiler (Ed.), *Language Universals* (pp. 165-184). Tübingen: Gunter Narr Verlag.

Heydel, M., & Murray, W. S. (2000). Conceptual effects in sentence priming: A cross-linguistic perspective. In M. De Vincenzi & V. Lombardo (Eds.), *Cross-linguistic perspectives on language processing* (pp. 227–254). Dordrecht: Kluwer.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.

Kemmerer, D., Weber-Fox, C., Price, K., Zdanczyk, C., & Way, H. (2007). Big brown dog or brown big dog? An electrophysiological study of semantic constraints on pre-nominal adjective order. *Brain and Language, 100,* 238–256.

Kempen, G., & Harbusch, K. (2004). A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 173–181). Berlin: Mouton de Gruyter.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, *43*, 3231-3250.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The Effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, *37*, 395-411.

Krahmer, E., Theune, M., Viethen, J., & Hendrickx, I. (2008). GRAPH: The costs of redundancy in referring expressions. In *Proceedings of the Fifth International Natural Language Generation Conference* (pp. 227-229). Association for Computational Linguistics.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50,* 194–231.

Martin, J. E. (1969a). Some competence-process relationships in noun phrases with pre-nominal and post-nominal adjectives. *Journal of Verbal Learning and Verbal Behavior*, *8*, 471-480.

Martin, J. E. (1969b). Semantic determinants of preferred adjective ordering. *Journal of Verbal Learning and Verbal Behaviour*, *8*, 697-704.

McDonald, J., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, *25*, 188–230.

Mitchell, M. (2009, March). Class-based ordering of pre-nominal modifiers. In Proceedings of the 12th European Workshop on Natural Language Generation (pp. 50-57). Association for Computational Linguistics.

Ney, J.W. (1986). Optionality and choice in the selection of adjectives in English. *General Linguistics*, *23*, 135–148.

Oller, J. W., & Sales, B. D. (1969). Conceptual restrictions on English: A psycholinguistic study. *Lingua, 23,* 209-232.

Pappert S, & Pechmann T. (2014). Priming word order by thematic roles: No evidence for an additional involvement of phrase structure. *The Quarterly Journal of Experimental Psychology, 67,* 2260-2278.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27,* 89-110.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39,* 633–651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–225.

Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one stage. *Journal of Memory and Language, 46,* 586–605.

Pickering, M.J. & V.Ferreira, V.S. (2008). Structural priming: A critical review. *Psychological Review, 134,* 427-459.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language, 42,* 168–182.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

R Core Team (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Richards, M. M. (1975). The pragmatic rule of adjective ordering: a critique. *American Journal of Psychology, 88,* 201-215.

Schriefers, H., & Pechmann, T. (1988). Incremental production of noun-phrases by human speakers. In M. Zock & G. Sabah (Eds.), *Advances in natural language generation* (pp. 172-179). London: Pinter.

Scontras, G., Degen, J., and Goodman, N.D. (2017). Subjectivity predicts adjective ordering preferences. *Open mind: discoveries in cognitive science*. *1*, 53-66.

Scott (2002). Stacked adjectival modification and the structure of nominal phrases. In G. Cinque (Ed.), *Functional structure in DP and IP: The cartography of syntactic structures* (Vol.1, pp.91-120). New York: Oxford University Press.

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*, 109–147.

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

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48,* 103–130.

Sproat R., & Shih C. (1991). The cross-linguistic distribution of adjective ordering restrictions. In C. Georgopoulos & R. Ishihara (Eds.), *Interdisciplinary Approaches to Language* (pp.565-593). Dordrecht: Springer.

Tanaka, M. N., Branigan, H. P., Mclean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*, 318-330.

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

Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: Salience, absoluteness, and consistency. *Frontiers in Psychology*, *6*, 1-16.

Van Gompel, R.P.G., Van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. (under revision). Conceptualisation in reference production: Probabilistic modelling and experimental testing.

Vernice, M., Pickering, M. J., & Hartsuiker, R. J. (2012). Thematic emphasis in language production. *Language and Cognitive Processes, 27*, 631–664.

Westerbeek, H., Koolen, R., Maes, A. (2015). Stored object knowledge and the production of referring expressions: the case of color typicality. *Frontiers Psychology, 6*, 1–12.

Whorf, B.L. (1945). Grammatical categories. *Language, 21*, 1-11.