Article

Experimental Approaches to Referential Domains and the On-Line Processing of Referring Expressions in Unscripted Conversation

Sarah Brown-Schmidt [1],* and Agnieszka E. Konopka [2]

[1] University of Illinois at Urbana-Champaign, 603 E Daniel St. Champaign, IL 61820, USA

[2] Max Plank Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands;
   E-Mail: Agnieszka.Konopka@mpi.nl

* Author to whom correspondence should be addressed; E-Mail: brownsch@illinois.edu.

Abstract: This article describes research investigating the on-line processing of language in unscripted conversational settings. In particular, we focus on the process of formulating and interpreting definite referring expressions. Within this domain we present results of two eye-tracking experiments addressing the problem of how speakers interrogate the referential domain in preparation to speak, how they select an appropriate expression for a given referent, and how addressees interpret these expressions. We aim to demonstrate that it is possible, and indeed fruitful, to examine unscripted, conversational language using modified experimental designs and standard hypothesis testing procedures.

Keywords: conversation; modification; scalar; eye-tracking; production; comprehension

1. Introduction

Historically, a divide has existed in the linguistic and psycholinguistic literatures between studies of *conversation* and investigations of language *processing*. Language processing research is typically concerned with how language is understood as it unfolds in time (e.g., [1–3]), while experimental research on conversation often focuses on how interlocutors coordinate dialog and jointly create meaning in rich contexts. Accordingly, both theoretical and methodological differences between these lines of inquiry have separated the research in the two traditions [4,5]. For example, early psycholinguistic processing techniques, such as analyses of lexical decision times or reading times,

often required metalinguistic judgments or used repetitious, pre-scripted materials—approaches that do not readily afford the study of language use in conversation. The selection of these techniques was justified by processing models that proposed that the efficiency of language processing is due, in part, to the encapsulation of syntactic processes from other sources of information, such as the discourse context ([6], see discussion in [3]). In contrast, research on conversation traditionally assumed that the context in which language occurred was central to the language itself [4,7]. Thus techniques such as the referential communication task [8] and the analysis of unscripted conversations (or Conversation Analysis; e.g., [9]) required the study of language in conversational contexts and could not be easily combined with the on-line measurement techniques available at the time.

Over the last 10–15 years, the divide between the two research traditions has been quickly weakening as researchers develop techniques for studying real-time language processing in rich visual contexts. One of the most obvious properties of language, and of communication in general, is that processes involved in production and comprehension unfold on a very fine time scale, exemplified by the tight coordination between the speech of conversational partners [10,11] and the speed of cognitive processes like grammatical encoding (e.g., [12]). Not surprisingly then, one long-term focus of research in this domain has been on the relationship between these processes during normal language use, and a wide range of questions of interest to linguists and psycholinguists alike hinge on an understanding of the temporal dynamics of linguistic processes. In this respect, the crucial turning point in the study of on-line language use came largely from the adaptation of eye-tracking technology to the study of language processing using the visual-world paradigm ([13]; also see [14–16]), where participants produce or listen to linguistic input about the items (usually pictures) presented in a visual display while their gaze is being recorded. The advantage of the visual world paradigm eye-tracking technique is that it affords investigation of on-line language processing in rich contexts, without requiring participants to make an explicit judgment, which might interfere with the phenomena of interest.

Since that time, this technique has been extended to increasingly naturalistic and unscripted conversational settings. For example, several researchers have used tasks in which pairs of naive participants, or participants paired with an experimenter or confederate (someone pretending to be a naive participant), give instructions to each other over a series of trials. This approach has been successfully applied to questions concerning the time-course of producing and interpreting referring expressions [17–19] as well as producing and interpreting syntactically ambiguous sentences [20–22]. Other researchers have used the link between gaze and speech to study unscripted conversations that are not constrained by an experimental trial structure. For example, Richardson and Dale ([23]; also see [24,25]) examined the correlation between the gaze of a naive speaker-and-listener pair as they conversed about a TV show to test hypotheses about the link between gaze coordination and conversational success. Others [26–28] examined how giving one dialog partner information about the other partner's gaze (real or simulated) influences language use in conversation. Another technique uses lengthy, unscripted, task-based conversations that are treated as a rich corpus of linguistic and eye-tracking data to test hypotheses about the on-line coordination of producing and interpreting referring expressions in dialog [29–31].

In this article, we discuss how several of these lines of research have contributed to our understanding of real-time language processing in conversation. In particular, the focus of our

discussion is on the modification of noun phrases in unscripted conversation (e.g., *the truck* vs. *the yellow truck*): How do speakers plan referential expressions in normal, every-day exchanges to ensure they are talking about the same person, object, or idea as their interlocutor? From the perspective of research emphasizing the importance of conversational context in language use, noun modification is a rich test bed test for theories about inter-speaker coordination in dialog. From the perspective of experimental work on message formulation and sentence understanding, noun modification involves processes critical to understanding information flow in the language system during production and comprehension. We aim to illustrate how both lines of inquiry can benefit from examining speakers' use of modification in relatively complex but unconstrained dialog.
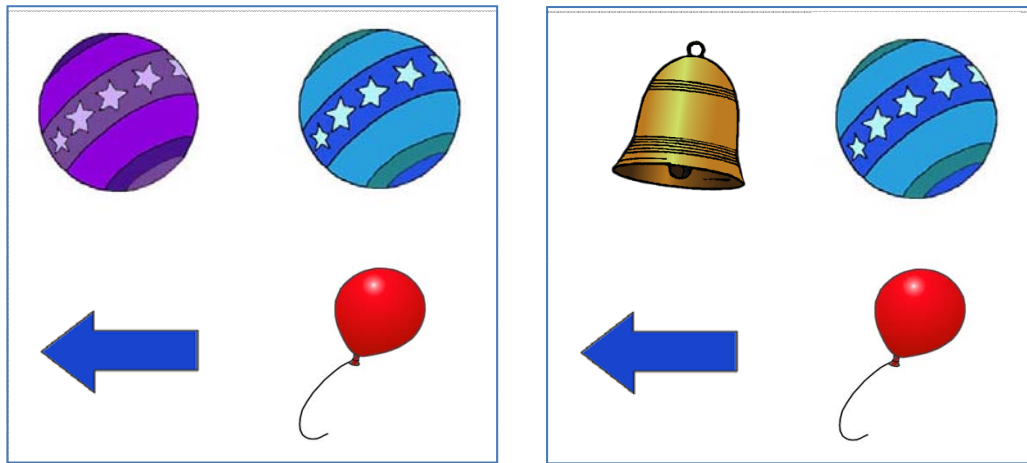
Traditional approaches to modification suggest that noun phrases uttered by cooperative speakers [32] should be uniquely identifiable by the addressee [33]; thus in many contexts, producing an informative expression requires producing a modifier. For example, if James and Otto were playing with a set of trucks that included two front loaders, one black and one yellow, James would need to use the modifier *yellow* to pick out the one he wanted, e.g., *I want the yellow front loader*. Further, according to the Gricean maxim of quantity, which states that speakers should make their contributions only as informative as necessary (*i.e.*, that they should avoid being over-informative), modifiers should be used if they are needed to uniquely identify the referent and not otherwise: the speaker should avoid producing a sentence like *I want the yellow front loader* if there is only one front loader in the referential domain.

Research on conversation suggests that the construction of referring expressions is decidedly more complex in interactive language use. If conversation is an interactive process, then the construction of referring expressions is also situated in an interactive exchange, with expressions created *jointly* by speaker and addressee, often in rich *contexts* [34]. This adds a layer of complexity to language use that more traditional tasks where speakers produce unrelated sentences in isolation do not tap into. In fact, studies of language use in conversational settings have uncovered at least two noteworthy departures from assumptions regarding unique identifiability and reference construction. First, the link between the referential context and modification is not uniform across different classes of modifiers, with some types of modifiers showing stronger contextual dependency than others. Second, identification of the relevant referential domain turns out to be a non-trivial issue, yet it is one that interlocutors seem to solve effortlessly. We discuss each of these issues in turn, and then illustrate our experimental approach to these questions.

1.1. Modification

One key departure from Gricean norms revealed in experimental studies of modification is that there is a non-equivalence across different types of modifiers in terms of their sensitivity to referential context. Consider the case of *over*modification. Overmodifications include the use of an adjective in a noun phrase when it is not needed to uniquely identify the referent [35] or the use of a proper name when a pronoun would do [36,37].

Figure 1. Example displays used to study modification rates in production. The target referent is the blue ball (top right). Left panel: A contrasting item (purple ball, top left) is present, requiring speakers to say *the blue ball* to refer to the target. Right panel: No contrasting item is present, so speakers can identify the target by saying *the ball*.

In the case of modified noun phrases, Sedivy [35] found that speakers frequently used color adjectives in the absence of contextual support when directing a partner to select one of four visually co-present objects: on approximately 40% of all trials, speakers said *the blue ball* to refer to a blue-colored ball in a context containing a single ball (Figure 1, right panel). In this case, the target referent (the ball) could be identified by the noun alone, making the color modifier redundant. The high modification rate in the absence of a contrasting item in the display (such as a purple ball; Figure 1, left panel) suggests that the use of color adjectives is not always motivated by unique identifiability. In contrast, other types of modifiers, such as scalar (e.g., *small*, *long*) and material adjectives (e.g., *wooden*, *plastic*), were used far less often in the absence of contextual support—on about 7% of all trials each. Interestingly, Sedivy [38] reports that the contextual independence of color modifiers varies with the characteristics of the intended referent: when the referent was of a predictable color, such as a yellow banana or green peas, color overmodification rates dropped to <10%.

Listeners appear to be exquisitely sensitive to these production tendencies. Using a visual-world eye-tracking task, Sedivy [39] found that the interpretation of an expression such as *the tall glass* in the context of another tall object (e.g., a tall pitcher) is facilitated if the context contains a size-contrasting object, such as a *short* glass: In contexts including both a short glass and a tall glass, participants preferred to look at the tall glass over the tall pitcher shortly after the onset of the word *tall*, suggesting that listeners are well aware that scalar adjectives are typically reserved for situations in which the context requires them (e.g., when a scalar contrast is present). Further, this knowledge influences comprehension immediately. This interpretation benefit is also observed for material adjectives, which similarly show a high degree of contextual sensitivity. In contrast, interpretation of color adjectives is not facilitated by the presence of a color-contrast item (e.g., the purple ball in Figure 1, left panel; [35]), unless the speaker is describing the color of an object with a predictable color (e.g., *the yellow banana* in the context of a yellow and a brown banana; [38]). In addition, the benefit for scalars is eliminated if the speaker routinely uses a scalar adjective in non-contrastive settings, most likely because repeated overmodification ceases to be informative [40]. According to Sedivy [38], this

pattern of results is evidence that the interpretation of adjectives depends not on adjective type, but rather on their information status in the discourse: in situations where the adjective provides contextually-relevant information, it can be exploited by the listener as an early cue to the speaker's referential intent.

1.2. Referential Domains

In the studies described above, speakers described and interpreted referring expressions in relatively circumscribed situations—the display included a set of four objects, and the participant was either asked to describe or to interpret a reference to one of these objects. In such settings, the referential domain, or the domain of interpretation for the referring expression, can be easily identified as the set of objects shown in the display. Identification of the referential domain in unscripted conversation is likely to be significantly more complex [41], particularly when the set of potential referents is large, when the interlocutors have different perspectives on the referential domain, and when the potential discourse referents have different affordances. For example, Chambers and colleagues [42] tested whether the affordances of a potential referent and their consistency with a spatial preposition guided listeners' interpretation of instructions. Given an instruction like *Put the cube <u>inside</u> the can*, evidence from the gaze of listeners in a visual-world paradigm task suggested that, upon hearing the preposition *inside*, they quickly constrained the domain of interpretation to goals that were compatible with the preposition (e.g., containers, but not other objects) and that were big enough to contain the particular cube in question. Similarly Heller, Grodner, and Tanenhaus [43] found that in situations where the speaker and addressee had different perspectives on a display (parts of the display were visible to both participants and other parts were not), the listener constrained the referential domain for interpretation of an imperative such as *Pick up the small duck* to objects that were in jointly visible scalar-contrast sets (also see [44]). While these studies were all conducted using interactive but non-conversational implementations of the visual-world paradigm, they suggest that both contextual and pragmatic information are critical for evaluating the referential domain.

Additional factors may come into play during unscripted conversation. For example, in an analysis of referring patterns in a task-based conversation, Beun and Cremers [45] found that the number of referents that participants considered as possible targets was limited by the focus of attention. In their task, one participant instructed another to create a building out of a set of blocks that differed in shape and size. The focus of attention in this study was modulated by factors such as pointing gestures to these blocks, linguistic mention, and functional relevance of the blocks to the task at hand (*i.e.*, the usefulness of a particular block at the moment of speaking). Despite the complexity of the task, participants produced and were able to successfully interpret potentially ambiguous expressions like *the yellow one* by considering only referents within a reduced referential domain. As we shall see, factors such as these guide not only the production of referring expressions in dialog, but also their comprehension in real-time.

In what follows, we provide examples of two experimental approaches within which we can address testable hypotheses about language processing in conversation. The experiments we report examined how the referential context guides production and interpretation of referential expressions. In doing so, our goals are two-fold. First, we aim to illustrate how on-line techniques give critical insight into the

time-course of these processes. Our second goal is to examine how the referential context guides production and interpretation of referential expressions during interactive conversation.

In the first experiment, we examine the hypothesis that the degree to which speakers consult the referential domain when planning a referring expression differs across different types of modifiers (see Sedivy and colleagues). The specific prediction was that speakers should be less likely to attend to the referential domain when producing a color-modified or number-modified expression than a scalar-modified expression. We hypothesized that color and number adjectives are planned in response to factors other than the referential domain, so we also consider the question of how speakers decide to use color and number modification in the first place. In the second experiment, we examine the use of modification in a different task, and then examine how dynamic changes in what interlocutors consider to be the referential domain at the moment of speaking change the way addressees interpret otherwise ambiguous expressions.

2. Experiment 1: Real-Time Querying of Referential Context in Conversation

In order to examine the mechanisms involved in the *unscripted* production of modified noun phrases, we created situations in which naive speakers described pictured referents to a partner. We evaluated the link between message formulation and utterance planning for three different types of noun phrase modification: size modification (*the small butterfly*), color modification (*the yellow book*), and number modification (*the three cherries*). By systematically including or not including contrasting sets of pictures, we were able to elicit modified noun phrases from participants without explicitly instructing speakers on what they were allowed to say. Our goal was to understand how the process by which speakers interrogate the referential domain and design a referring expression varies as a function of the expression itself.

In previous work [17,18], we observed an intriguing link between speakers' production of scalar-modified expressions and their interrogation of the referential domain. Consistent with previous findings by Sedivy and colleagues, we found that speakers were more likely to use a scalar modifier if the referential domain contained a scalar contrast item (e.g., a picture of a large butterfly when speakers were describing a small butterfly). Further, we made a novel observation: there was a strong link between the time when speakers fixated the scalar-contrast item (e.g., large butterfly) and the form of the ensuing expression. When speakers fixated the contrast well before speech onset, they tended to produce fluent referring expressions, such as *the small butterfly*. However, when they fixated the scalar-contrast just before or after the onset of their expression, they tended to produce disfluent repairs, such as *the butterfly...small one*. Delayed contrast fixations were also observed for fluent expressions with post-nominal modifiers, such as *the square with small triangles* [18], as well as fluent, post-nominal modifiers in Spanish, such as *la mariposa pequeña* [17]. These findings are clear evidence that scalar modification is prompted by the speaker's attention to a scalar-contrasting item in the display. Delayed attention to the contrast, relative to speech onset, is associated with delayed modification, whether it be a disfluent repair, a postnominal modifying phrase, or a postnominal adjective (as in Spanish). This gaze-speech link is important for two reasons: First, it demonstrates that the first fixation to a scalar contrast can be used as an indicator for when the speaker first formulated the intention to mention size. Second, this tight link between noticing the contrast and producing the
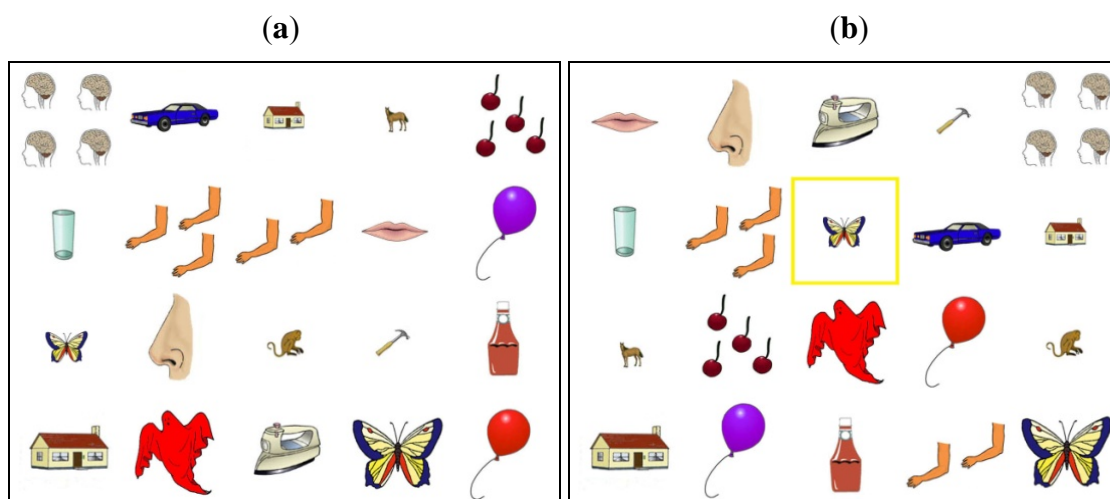
adjective suggests that speakers can plan components of a noun phrase (e.g., size information), as they are articulating the expression itself [17]. In the present research, we asked if the same process was responsible for the production of color and number modifiers.

2.1. Method

The materials and procedure of this study were identical to those reported in Brown-Schmidt and Konopka [17], thus only key details are described here. Monolingual English-speaking participants ($n = 16$) engaged in the task with an experimenter. The participant wore a lightweight head-mounted EyeLink II eye-tracker (SR Research), and both the experimenter and participant wore headset microphones. The participant's computer controlled the recording of both eye-tracking data and audio files, allowing gaze and speech to be synchronized.

On each trial ($n = 480$ over two sessions), the participant and experimenter saw 20 pictures on their respective computer screens, and took turns describing a highlighted target picture (see Figure 2a,b) so that their partner could find it on her screen and click on it. Critical trials included targets that encouraged speakers to use color, number, and size modifiers. On contrast-present trials ($n = 120$), the participant's screen contained the target referent along with a contrast-picture that was identical to the target, differing only in color ($n = 40$), number ($n = 40$), or size ($n = 40$). On contrast-absent trials (or control trials; $n = 120$), a similar set of target pictures was presented without a contrast-picture. Finally, a separate but similar set of target pictures was designed for the 240 trials on which the experimenter was speaking.

Figure 2. Example scene for a contrast-present trial from (**a**) the speaker's perspective and (**b**) the listener's perspective.

In order to create variability in the point in time when speakers noticed the contrast picture, the 20 pictures were arranged on the screen randomly with the exception that target and contrast pictures were never immediately adjacent. To prevent participants from using locative descriptions like *the top right picture* (without explicitly telling them what to say), the arrangement of pictures on the participant's and experimenter's computer screens was different (Figures 2a,b).
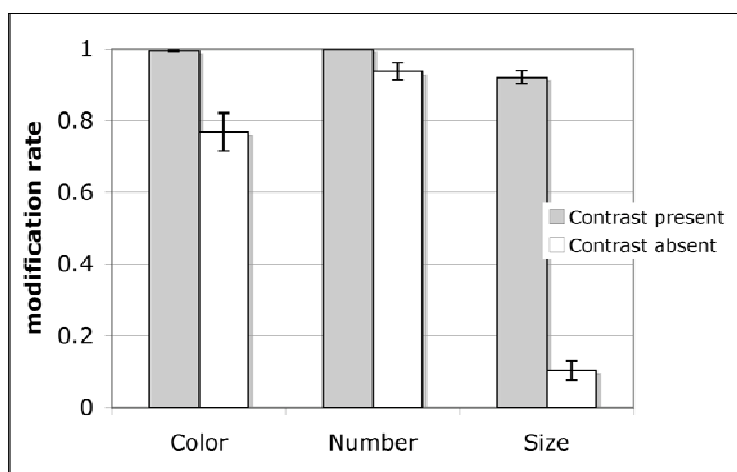
2.2. Analysis

Participants' descriptions of the 240 target objects were transcribed. We eliminated trials in which the speaker was unable to name the picture (e.g., *I have no idea what that is*) and trials in which the experimenter interrupted the participant, leaving 3551 good trials for our analysis (or 92% of all trials). For each utterance, we recorded the time from the beginning of the trial to the onset of each word in the noun phrase.

For contrast-present trials, we coded whether the speaker used a contrast-appropriate adjective to describe the target. For example, if the target picture (e.g., a yellow book) was presented with a color-contrast picture (a red book), we coded whether the speaker mentioned the color of the target. For no-contrast trials, we coded whether the speaker used a color adjective (and not any other modifier) to describe the target as well. Similarly, descriptions of number- and size-contrast targets in no-contrast trials were coded for the use of number and size modifiers. These no-contrast trials provide the necessary baseline to evaluate the role of the contrast picture in eliciting modification in contrast-present trials.

2.3. Modification Rates, Referential Context, and Gaze

In general, speakers were sensitive to the referential domain, using more color, number, and size modifiers when a contrast picture was present than when it was absent. However, this effect was much stronger for size-contrast pairs than color- or number-contrast pairs (see Figure 3), suggesting that speakers' use of adjectives on contrast-present trials is driven both by the presence of a contrast and by the type of modifier required.

Figure 3. Modification rates for color, number, and size trials in Experiment 1.

The effects of picture type and presence of a contrast on modification rates were analyzed in a 2 (presence of contrast) × 3 (modifier type) analysis of variance (ANOVA). All reported results are significant at the $\alpha = 0.05$ level (two-tailed), unless indicated otherwise. The main effects of modifier type and presence of a contrast were qualified by a significant interaction, $F_1(2,30) = 191.43$, $F_2(2,234) = 300.29$. Planned comparisons demonstrated a significant effect of referential context for

each type of modifier, but the effect size (Cohen's *d*; [46]) was strongest for size-contrast pictures: Color, $t_1(15) = 4.30$, $t_2(81) = 8.37$, $d = 1.50$; Number, $t_1(15) = 2.52$, $t_2(75) = 6.25$, $d = 0.89$; Size, $t_1(15) = 26.99$, $t_2(78) = 30.56$, $d = 9.03$. Most of these modifiers were pre-nominal (e.g., *the yellow book*), although speakers sometimes used disfluent post-noun repairs (e.g., *the book... uh yellow one*). While the rate of post-noun repairs was low for color (39 trials, or 3% of color trials) and number-contrast pictures (4 trials, or .4% of number trials), this was a common construction for scalar-contrast pictures (216 trials, or 18% of scalar trials).

The fact that speakers used number and color modifiers even when the display did not contain a contrast shows that these modifiers were planned independently of consideration of the referential domain. As a further test of this hypothesis, we compared modification rates for contrast-present trials when speakers did and did not fixate the contrast ($n = 1809$, 86% with a contrast fixation). Modification rates on contrast-present trials in the absence of a contrast fixation are important to consider because they provide a validation of using the first contrast fixation as an indicator of when speakers first noticed the contrast. If the planning of a modifier is contextually dependent, and reliably begins following the first contrast fixation, there should be significantly more modification when the contrast is fixated (of course, this is not a perfect measurement, as speakers sometimes notice the contrast through peripheral viewing, and some contrast fixations are not recorded due to track loss). Indeed, for both color and number contrast trials, speakers were no more likely to use a modifier if they fixated the contrast (color: 99%; number: 100%) than if they did not (color: 100%; number: 100%). Scalar contrast trials showed a different pattern: speakers were significantly more likely to use a modifier when they fixated the contrast (95.9% of the time) than when they did not (26.4% of the time), with a 95% CI of ±19%. This result is consistent with the claim that scalar, but not color or number modifiers, are planned with respect to the speaker's consideration of the referential domain.
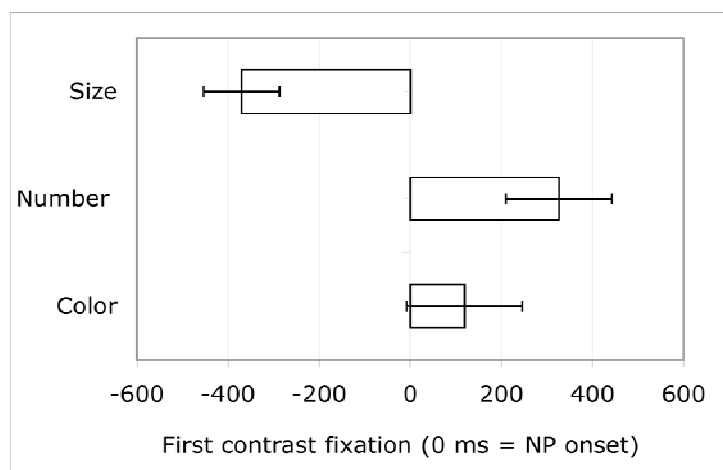
2.4. Referential Form and Timing of Fixations

Previous research [17,18] established that for scalar adjectives, the timing of the speaker's first fixation to the contrast object is highly predictive of the form of the ensuing referring expression. For delayed or postnominal modifiers (*the flower...big one; the square with big triangles;* or *la mariposa pequeña* in Spanish), first contrast fixations occurred significantly *later* than for constructions with early or prenominal modifiers (*the big flower*, *the big square with triangles*, *the small butterfly*). These findings suggest that, at least for scalars, planning of the adjective begins with attention to the contrasting item in the referential domain. Following from our analysis of modification rates, we expected that the relationship between context and the planning of a referential expression would take an entirely different form for number and color adjectives, such that the presence and fixation of a contrast would not determine the use of color or number modification.

If noun modification is prompted by consideration of a contrast present in the display, first contrast fixations should occur well before the onset of the noun phrase; if not, first contrast fixation times should be relatively delayed relative to noun onset. First contrast fixations were thus measured relative to noun phrase onset (see Figure 4). Trials with disfluencies (e.g., *thee, uh, um*) were excluded because disfluency is likely to be related to delayed planning of modifiers [18], along with trials where speech

onset times were more than 2 SDs outside the grand mean for all contrast-present trials (M = 1640 ms, SD = 661 ms). This left 1247 trials for the analysis (80% of all trials with a fixation to the contrast).

A one-way ANOVA revealed a significant effect of trial type (color, number, scalar), $F_1(2,30) = 23.23$, $F_2(2,116) = 18.16$. When producing a size adjective, first contrast fixations occurred significantly earlier than on trials where speakers used a color adjective, $t_1(15) = 4.74$, $t_2(77) = 3.97$, or a number modifier, $t_1(15) = 5.88$, $t_2(77) = 5.83$. Contrast fixations for number-modified phrases were also delayed compared to color-modified phrases, $t_1(15) = 2.21$, $t_2(78) = 2.25$. Because speakers did not fixate the contrast until well after the onset of the noun phrase when producing color- and number-modified expressions, noun modification in these cases cannot be due to their noticing that modification is required by the referential context. Instead, it is more likely that features of the target object alone prompted speakers to use these modifiers.

Figure 4. First contrast fixation times (ms relative to noun phrase onset) for fluent size-, number-, and color-modified expressions.

Further, replicating previous findings [18], we observed that for disfluent size repairs, as in *the butterfly ... uh small one*, first contrast fixations were significantly delayed (595 ms post-onset), compared to fluent, pre-nominally modified expressions (396 ms *before* speech onset), $t_1(15) = 7.78$, $t_2(36) = 4.53$. For disfluent, pre-nominally modified expressions (*thee uh small butterfly*), the average first fixations times were −234 ms (a 162 ms delay compared to fluent expressions), $t_1(13) = 1.79$, $p < 0.05$, one-tailed, $t_2(33) = 0.80$ (this effect was weaker, likely due to the small number of disfluent trials, $n = 85$). Because color and number repairs were infrequent in this dataset, it was not possible to evaluate the role of disfluency in the production of color- and number-modified phrases.

2.5. Time to Begin Speaking

Additional evidence that the process of constructing a modified noun phrase differs across modifier types comes from an analysis of speech onset times. Speakers were fastest to begin speaking on color-contrast trials, at 1471 ms (SE = 40 ms) after display onset. Number-contrast trials had intermediate speech onsets, at 1575 ms (SE = 45 ms), and size-contrast trials were the slowest at 1647 ms (SE = 49 ms). Noun phrase onset times for fluent, pre-nominally modified trials were

submitted to a one-way ANOVA, which revealed the expected main effect of trial type, $F_1(2,30) = 11.10$, $F_2(2,116) = 5.88$: Onset times were shorter for pictures in color contrasts than in size contrasts, $t_1(15)=4.63$, $t_2(77) = 3.96$, and marginally shorter (in the by-participant analysis only) in number contrasts than size contrasts, $t_1(15) = 1.94$, $p = 0.07$, $t_2(77) = 1.69$, $p = 0.10$. Onsets for color-contrast expressions were also shorter than those for number contrasts, $t_1(15) = 2.77$, $t_2(78) = 2.71$. The delay for size-contrast trials is consistent with our analysis of gaze, and suggests that it takes speakers more time prior to speech onset to inspect the referential domain and design an appropriately modified expression, compared to cases where the adjective is not designed with respect to the referential domain. This relationship may be specific to size modifiers, consistent with other findings that subjects spend more time planning size information than color information even when instructed to avoid overmodification [47].

2.6. Conclusions

In this unscripted communication task, we observed that the relationship between the referential domain and referential form depended on the type of modification the speaker used. Speakers planned scalar modifiers more often when the referential domain contained a contrast and when they had fixated it than when they had not. Further, the timing of fixations to this contrast was clearly linked to the form of the modified expression, with early contrast fixations resulting in fluent, pre-nominally modified noun phrases, and delayed contrast fixations resulting in disfluent expressions. Color and number modifiers showed a different pattern: speakers planned these modifiers independently of the referential domain, even when the domain contained a relevant contrast object, showing the lack of a gaze-modification link at least for the purposes of modification itself. The results of our time-course analysis add to previous findings that speakers use color adjectives in the absence of contextual support [35] and show a similar (if not stronger) contextual independence for number modifiers.

Why did speakers choose to use a color or number modifier without considering the referential domain? Clearly, speakers in everyday conversation do not modify all noun phrases with color and number modifiers. We suspect that intrinsic properties of the intended referents are key to understanding this process. Features like color are easier to detect than a relative dimension like size [47]. Also, our stimuli were designed as parts of scalar, color, or number contrast pairs. As a result, the color targets were highly colorized on both contrast-present and contrast-absent trials, and likewise the stimuli for number targets contained multiples of the item on contrast-present and contrast-absent trials. These relatively salient properties of the target pictures may have prompted modification irrespective of properties of other pictures in the display. This hypothesis can be evaluated by considering overmodification rates on no-contrast trials for the three classes of stimuli. Indeed, the overmodification rate for color adjectives was 78% on no-contrast color trials, but only 7% on number and size trials where the stimuli were not highly colorized. Similarly, the overmodification rate for number modifiers was 94% on no-contrast number trials, but only 1% on color and size trials (*i.e.*, speakers rarely produced expressions like *one butterfly*). Scalars again appeared sensitive to context, with an 11% overmodification rate on no-contrast scalar trials, but only a 1% modification rate on color and number trials.

In short, the findings of this experiment suggest that in unscripted speech, number and color adjectives are largely planned in response to stimulus properties, whereas the use of scalars is directly tied to interrogation of the referential domain. The fact that speakers were quite willing to include extra color and number modifiers is inconsistent with the Gricean maxim of quantity, as these modifiers were clearly not required for unique identification. This result also represents an apparent departure from the principle of least collaborative effort [34], as it would seem more effortful to produce an additional modifier than not, even if listeners did not interpret those modifiers contrastively. Why, then, do speakers produce these modifiers?

Perhaps the answer is that color and size modifiers do, in fact, help listeners identify the target more quickly, possibly by highlighting a feature that both interlocutors are likely to notice quickly [16], which would be consistent with the principle of least collaborative effort. Alternatively, perhaps the goal of referring is not always to uniquely identify, but instead to *characterize*, or express a viewpoint on some object. For example, Brennan and Clark [48] argued that when speakers refer to something, they propose a conceptualization of the referent—a conceptualization that can then be accepted or modified by the addressee. In the present case, when the speaker chose to describe a balloon as *blue*, the inclusion of an adjective that was not necessary for identification *per se* might have been important for the communicative act the speaker had intended (e.g., to discuss *blueness*).

Finally, as the present research and previous findings illustrate [17,18], overmodification may also occur when information about size, color, or number is added to the preverbal message at different points in time. In some circumstances, speakers can begin talking having encoded the minimum amount of information necessary for production (e.g., a bare noun) and at the same time continue scanning the display to plan what to say next ([17,18], also see [49]). Radically incremental message planning of this sort means that speakers may occasionally add a redundant modifier if they do not notice the contrast picture early enough to produce the more relevant or context-motivated modifier instead. For example, if they do not notice a size contrast early in the planning process, they may choose to produce a different kind of adjective first, perhaps one that refers to a salient property of the referent or that can be planned before completing interrogation of the referential domain, in an effort to be as specific as possible.

This last hypothesis is relevant to debates about planning scope in production. While some aspects of production are undoubtedly the product of processes specific to the production architecture, like information flow between the different levels of the production system during lexical access (e.g., see [12,50,51]), others are subject to external pressures such as conversational constraints or meeting communicative goals [49]. Thus the process of planning any expression may reflect a compromise between finding the minimal phrase that uniquely identifies the target and planning that utterance as quickly as possible. So, in situations like the one in Experiment 1, where unique identification requires scanning of the referential domain, concerns about speed of communication may motivate the use of context-independent modifiers that can be planned more quickly (such as color or number). An important question for future research is to understand when the balance between planning larger *vs.* smaller chunks of a message before speaking might shift towards a "quick and easy" strategy of mentioning salient but potentially redundant features of an object in the interest of time.
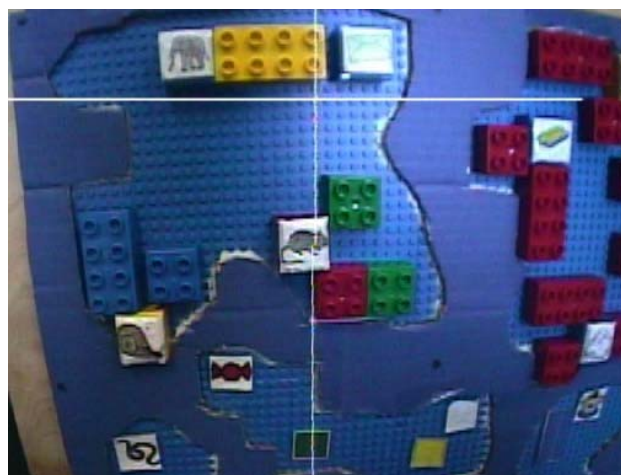
3. Experiment 2: Referential Domain Circumscription in Conversation

Experiment 1 used a trial-based unscripted conversational task to examine the relationship between modification and consideration of the referential domain. In that study, the referential domain was a scene made up of 20 pictures shown on a computer screen, and it changed from trial to trial. In this section, we describe a different technique for examining how referential domains are shaped over the course of a conversation when the display used for the task is very complex but consistent across time. This creates a situation where interlocutors can use discourse history as an additional constraint on the production and interpretation of referential expressions (also see [34]). Within this context, modification rates should reflect sensitivity to properties of the referential domain in question as well as to speakers' awareness of their shared knowledge about the task.

3.1. Method

Here we present some novel analyses of an experiment that was originally reported by Brown-Schmidt and Tanenhaus ([30], Experiment 2), so only critical details are described here. The experiment used an unscripted conversational task that, unlike Experiment 1, did not have a trial structure. Instead, pairs of naive participants (one of whom was eye-tracked) were seated on opposite sides of a display (see Figure 5) and worked together to arrange blocks on each side of the display. At the beginning of the task, each partner had stickers on their display representing the correct location for only half of the blocks in the display, so throughout the study, the participants had to tell each other where to place blocks in order to eventually arrive at the same arrangement of blocks on their respective boards. To elicit modified expressions, the blocks varied in color, size, location, and orientation. The task lasted about 2 h, so the referential domain or the speakers' focus of attention shifted repeatedly as their task focus shifted during the conversation [45,52].

> **Figure 5.** Partial view of the display for the eye-tracked participant, approximately half-way through the task. The cross-hair represents the eye-tracked participant's gaze.

Here we describe some of the authors' original findings in this experiment, and then present some new analyses regarding the characteristics of definite referring expressions produced by eight pairs of participants, and their interpretation throughout the discourse. (Note that Brown-Schmidt and

Tanenhaus [30] report data from 12 pairs of participants. In the new analyses presented here, data from only 8 pairs is reported, as the coding format for the 4 excluded pairs made the new analysis difficult). Our goals are two-fold: First, we intend to highlight features of referring patterns in a lengthy, unscripted conversation that differ from expressions produced in the very short, trial-based discourse segments of Experiment 1. Our second goal is to illustrate that it is feasible to collect experimental data on language processing in completely unscripted conversational settings.

3.2. Analysis and Results

All conversations were transcribed. Our analyses focus on expressions produced by the non-eye-tracked participant, as the goal was to understand how the eye-tracked participant interpreted these expressions. This participant's board was divided into five sub-areas (e.g., in Figure 5, the top-left sub-area is the part of the board that contains 10 blocks, including one yellow, one red, two blue and two green blocks). Each expression was coded for the type of modification used, if any, and for whether it uniquely identified the target referent given other blocks in that sub-area. An expression was considered linguistically "ambiguous" if it did not uniquely specify a referent with respect to the sub-area it was located in (e.g., *the blue*, given the top left sub-area in Figure 5). In contrast, an expression was considered "disambiguated" if it uniquely identified the target referent with respect to the other potential referents in the sub-area of the display (e.g., with a modifier such as *the yellow block* or *the blue block that's vertical* given the top-left sub-area in Figure 5). Thus ambiguity was defined with respect to the expression <u>and</u> the context.

Table 1. Distribution of different modifier types for ambiguous and disambiguated expressions (*n* = 989). Percentages sum to more than 100% because many expressions had multiple modifiers. The relevant modifier is underlined in the examples.

Modifier type	Ambig	Disambig	Difference (95% CI)	Example
none	11.2%	4.0%	7.2% (5%)	*thuuh the square*
color	84.3%	92.0%	−7.7% (8%)	*the <u>yellow</u> block*
size	6.6%	5.1%	1.5% (8%)	*the <u>small</u> block*
location	8.4%	5.6%	2.8% (4%)	*therr red block <u>that sits on top of the lamp</u>*
orientation	6.2%	8.5%	−2.3% (6%)	*the red <u>horiz- vertical</u> one*
past action	2.3%	6.0%	−3.7% (5%)	*the red rectangle <u>that we just p-placed</u>*
"other"	2.7%	0.4%	2.3% (3%)	*the <u>other</u> red block underneath it*

In contrast to Experiment 1, where participants routinely overmodified their expressions, speakers in this study routinely *under*specified their expressions. Of the 989 definite references to color blocks produced by the non-eye tracked participant, a total of 439 (44%) were underspecified with respect to the sub-area of the board under discussion. For example, when talking about the upper-left area in Figure 5 with the yellow, blue and green blocks, speakers frequently used ambiguous expressions like *the green block* to describe a referent when another green block was present. This 44% estimate is comparable to the 43% proportion of references that were ambiguous with respect to the whole domain reported by Beun and Cremers [45], and only slightly lower than the 47% estimate reported by Brown-Schmidt and Tanenhaus [30]. (The difference between the current estimate and that reported by

Brown-Schmidt and Tanenhaus [30] is likely due to the inclusion of trials in the present analysis which were previously excluded due to track loss, or because no competitors were present on the board, e.g., producing an expression like *the white block* in an area with only a single white block).

A closer look at the definite references used in this experiment (see Table 1) indicates that 84% of the ambiguous and 92% of the disambiguated expressions contained color modifiers, a difference that was not significant. The fact that speakers were no more likely to include color modification when producing a disambiguated expression is consistent with the results of Experiment 1, despite the fact that the stimuli and task structure were quite different in that study. Completely unmodified expressions (e.g., *the square)* were also common in this corpus. Not surprisingly, this occurred significantly more often in ambiguous expressions (11%) than disambiguated ones (4%). Unmodified expressions like *the square* (see Table 1) were disambiguated when the noun itself identified the target (e.g., a context with one square and one rectangle), and were ambiguous when the context contained multiple squares.

Whereas stimulus features seemed to drive the use of adjectives in Experiment 1, the conversational goals and task demands of this experiment played a key role in shaping the form and interpretation of referring expressions. A direct example of this is the frequency with which partners referred to what they had done with particular blocks in the past (e.g., *the one we just placed*), suggesting that they were tracking and exploiting task history as a key source of information to facilitate identification of intended referents. Brown-Schmidt and Tanenhaus [30] quantified these observations in an analysis of referential patterns and their relationship to properties of the display. The results indicated that speakers tended to use undermodified expressions—e.g., saying *the green one* in contexts with multiple green blocks—if the competitor blocks (*i.e.*, blocks that were consistent or temporarily consistent with the referring expression) were <u>not</u> (a) proximal to the last-mentioned block, (b) relevant to the current task, or (c) recently mentioned. In other words, speakers only disambiguated their expressions with respect to competitors when the competitors were salient and relevant. Similar constraints have been identified as critical to narrowing the referential domain in other task-based conversations [45]. Thus it seems that, having limited their referential domain to small areas of the workspace circumscribed by these three factors, modification use was then based on these smaller referential domains. For example, an expression like *the rectangle* might in fact uniquely identify the yellow rectangle in Figure 5 if the competitor (the blue rectangle) had not been discussed in some time and was irrelevant to the current conversational goals. In this sense, these "ambiguous" expressions may not have been ambiguous at all when the referential domain at the moment of speaking was much smaller than the physical space of a sub-area itself (also see [45]).
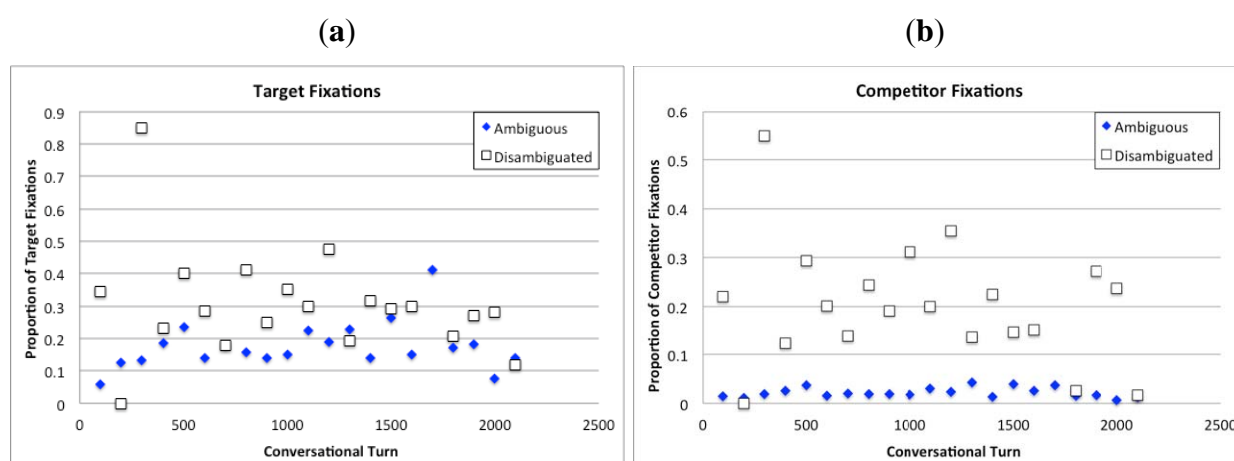
Supporting evidence for this claim comes from the analysis of eye movements of the eye-tracked addressees. Brown-Schmidt and Tanenhaus [30] reported that when hearing linguistically ambiguous noun phrases, addressees nevertheless successfully interpreted these expressions and fixated the intended referent. In fact, the average proportion of fixations to competitor blocks that were consistent with the expression was equivalent to the proportion of fixations to unrelated blocks that did not match the expression. In contrast, for disambiguated expressions, addressees fixated competitors more than unrelated blocks that did not temporarily match the referring expression before the point-of-disambiguation. The point-of-disambiguation is the point in the expression that uniquely identifies the target referent given the local domain. For an expression like *the blue*

rectangle, uttered when a sub-domain includes a blue square and a blue rectangle (e.g., see Figure 5), the point-of-disambiguation would be the onset of the word *rectangle*, as this is the point in the utterance that uniquely identifies the target.

These results illustrate how listeners used task knowledge to circumscribe the referential domain and identify the target referent. How does this knowledge emerge during the course of a conversation? If speakers and addressees explicitly work to coordinate domains during the course of the conversation, one might expect that coordination would improve over the course of the conversation. Alternatively, speakers might immediately rely on task knowledge and very recent discourse history to constrain referential domains. We evaluated these two possibilities in a new analysis by comparing the fixations that addressees made to potential referents in the scene as they conversed over the entire experimental session. If coordination takes time and helps speakers develop ways to perform the task more efficiently, the preference to fixate the target should increase over the course of the conversation.

A conversational turn was defined as a contribution to the discourse that was not interrupted by the partner (with the exception of backchannels, such as *uh-huh*). For each conversational turn, eye movements were analyzed in two 800 ms analysis regions during which we observed significant competition effects for disambiguated expressions (see [30] for details). The time region for disambiguated expressions encompassed the 800 ms immediately before the point-of-disambiguation, plus 200 ms to account for the time it takes to program and launch a saccade [53]. Ambiguous expressions by definition contained no disambiguating point, so an analogous 800 ms time-window, relative to expression onset, was used for this analysis. Fixation proportions were analyzed in a mixed effects regression model with expression type (ambiguous, disambiguated) and conversational turn as fixed effects (see Table 2).

Figure 6. Proportion of fixations to (**a**) target and (**b**) competitor blocks by conversational turn for linguistically ambiguous and disambiguated expressions.

(**a**)   (**b**)

Consistent with previous findings, the overall proportion of fixations to the target was higher than the average proportion of fixations to the competitors, for both ambiguous (0.18 *vs.* 0.02, 95% $CI_{difference} = 0.05$) and disambiguated expressions (0.3 *vs.* 0.2, 95% $CI_{difference} = 0.06$). The target preference, even for ambiguous expressions, was likely because targets were highly salient—that is,

they were usually recently mentioned, proximal to what had been mentioned recently, and task relevant [30].

Table 2. Parameter estimates for model of target fixations and competitor fixations, with expression type (ambiguous, disambiguated) and conversational turn as fixed effects. Subject pair was included as a random intercept in both models and as a random slope in the model of competitor fixations (the model of target fixations was not improved by adding a random slope). The *t*-values are reported with MCMC estimated *p*-values (see [54], for discussion).

Fixation type	Estimate	SE	*t*-value	pMCMC
(a) Target Fixations				
Fixed effects				
(intercept)	$1.51E^{-01}$	$3.28E^{-02}$	4.592	
Expression Type	$1.44E^{-01}$	$2.83E^{-02}$	5.086	0.001*
Conversational Turn	$3.08E^{-05}$	$2.90E^{-05}$	1.062	0.288
Random effects				
Pair	0.001	0.035		
Residual	0.107	0.326		
(b) Competitor Fixations				
Fixed effects				
(intercept)	$1.51E^{-02}$	$1.48E^{-02}$	1.016	
Expression type	$1.89E^{-01}$	$2.79E^{-02}$	6.752	0.001*
Conversational turn	$9.16E^{-06}$	$1.37E^{-05}$	0.671	0.744
Random effects				
Pair	0	0		
Pair * Type	0.007	0.081		
Residual	0.024	0.155		

The present analysis revealed that addressees were more likely to fixate the target referent when they heard a *disambiguated* expression (see Figure 6a and Table 2a). There was no effect of conversational turn, suggesting that identification of the target did not improve over the course of the conversation. The analysis of constraints mentioned above found that competitors were also more salient and task-relevant when speakers produced disambiguated expressions. Why then, did addressees fixate the target more when interpreting disambiguated expressions? After all, the analysis region for the eye movement data only captured fixations made *before* disambiguated expressions were linguistically disambiguated. The answer likely relates to the fact that there were many more opportunities for addressees to look away from the target to a competitor for ambiguous expressions than for disambiguated expressions, thereby reducing the number of fixations to the target. Not surprisingly, the number of competitor blocks consistent or temporarily consistent with the target expression was twice as high for ambiguous expressions than for disambiguated ones (4.23 *vs*. 2.41).

The results for the analysis of competitor fixations were similar, with a higher average proportion of fixations to competitors when they heard disambiguated expressions, and no effect of turn (see Figure 6b and Table 2b). This result is consistent with the analysis of conversational constraints in Brown-Schmidt and Tanenhaus [30], which found that competitors for disambiguated expressions had

been more recently mentioned, were more proximal to the target, and were more task-relevant. The fact that addressees were no faster at fixating the target or eliminating competitors as the conversation progressed, suggests that referential domains were well established even from the earliest definite references, rather than being slowly narrowed across the course of the conversation. This may have occurred because partners spent a good deal of time at the beginning of the task establishing joint representations of the referential domain. Thus, by the time speakers referred to blocks placed on the board, the referential domain had been thoroughly negotiated and constrained by the previous discourse. For instance, in Example 1, which was taken from the very beginning of one of the conversations, Partners 1 and 2 very carefully establish joint knowledge about their referential domain before finally producing a definite reference to a block (underlined).

Example 1. Example excerpt from one conversation in Experiment 2.

1: Alright, uhhh do you wanna start inn…*thee*

2: *a quadrant*

1: like … upper left?

2: *your upper left or my upper left*?

1: uhh

2: *let's do YOUR *upper left**

1: *my upper left*

2: *ok so my upper right*

1: yep

2: *and you've got ay ... like ... I don't know it's a shape ... pond type thing*

1: yeah

2: *alright. Umm th-which corner ... of that*?

1: ch-ch-choo ... so ... let's make sure we're looking at the actual same thing

2: *ok*

1: I haave...in like the uppermost...and...furthest...to the edge part of that?

2: *yeah*

1: umm i have an indentation

2: *yeah*

1: which is about...uhh siix...little dots high? and about six dots in...do you have one of those?

2: *sure*

1: d- does it NOT look like that?

2: *yeah it does*

1: ok...um...so if you go over...I haave one column of dots between THAT aand...a long yellow block?

2: *really*?

1: yes

2: *k*

1: uhh and the...long yellow block iiis...going up and down. like aa skyscraper?

By clearly establishing their domain at the beginning of the conversation, partners likely improved interpretation of each other's definite expressions, and thus as a result, interpretation of subsequent definite expressions did not improve significantly across the conversation.

Finally, if speakers routinely used ambiguous expressions and addressees were able to understand them easily, why did nearly 90% of these "ambiguous" expressions contain other types of modifiers (*i.e.*, modifiers that did not serve the purpose of disambiguation)? After all, if modification is used to identify the intended referent [32,33], then why did speakers bother planning uninformative modifiers at all? We can imagine two possibilities. Perhaps, as in Experiment 1, speakers planned these modifiers on the basis of target properties alone. If so, the modifiers could have been used by speakers to provide a characterization of the referent. Consistent with this possibility is the fact that color modification rates on ambiguous trials in Experiment 2 and on no-contrast color trials in Experiment 1 were similar (~80%)—in both cases, speakers may have been planning modification on the basis of target properties alone. A second possibility is that the modifiers on ambiguous expressions *did* uniquely identify the target, but did so within a much smaller referential domain.

3.3. Conclusion

Previous work demonstrated that speakers interpret referring expressions within referential domains circumscribed by factors such as affordances of the referent [42] and the perspective of the speaker and addressee [43,44]. Consistent with Beun and Cremers' [45] off-line analyses of production patterns, the results of this experiment demonstrate that interlocutors engaging in task-based conversation arrive at similarly constrained referential domains that are circumscribed by recency of mention in the discourse, task relevance, and spatial location. Constraining the referential domain was likely the cause of one of the most striking observations of this experiment—the high undermodification rate (44%), in conjunction with the fact that listeners apparently had no problem understanding these expressions.

Consistent with this finding, Brown-Schmidt and Tanenhaus [30] also observed that phonological (cohort) competitors of a pictured referent were eliminated from consideration during the course of similar dialogs: During interpretation of a word like *penguin*, the majority of fixations in their experiment were on the target itself (penguin), with no more fixations to the cohort competitor (pencil) than to non-competitors (e.g., comb). These findings demonstrate that the context of a single conversation can effectively eliminate phonological and referential competition, thereby reducing the complexity of decoding a speaker's communicative intent. Thus, the context of the conversation itself allowed speakers and addressees to communicate effectively by dramatically reducing the potential for ambiguity. This result is surprising in light of the standard view that phonological competition is a central problem of speech perception (e.g., [55]), but is consistent with research examining how background knowledge or discourse history help readers process otherwise ambiguous words or sentence constructions (e.g., [56,57]).

Interestingly, however, speakers were no more likely to use color modifiers when they disambiguated their expressions than when they produced these ambiguous expressions. Might the color adjectives have disambiguated the linguistically ambiguous expressions with respect to smaller, constrained domains described above? We suspect that, in many cases, they would not: for example, two blocks of the same color were often immediately adjacent in the display, so even a very small

domain would not afford disambiguation with a color adjective (e.g., the top area in Figure 5 contains mostly red blocks—in that area, producing the modifier "red" would not uniquely identify the target even within that small domain). Further, the rate of color modification was comparable to that observed in Experiment 1, when use of color adjectives led to *overmodification*. These results lend further support to the conclusion of Experiment 1, that use of modification in cases when it is not contextually motivated is not for the purposes of unique identifiability, but perhaps to provide a characterization of the referent.

The results of the current experiment also demonstrate how quickly listeners can make use of discourse-level information in a conversation. The reduction of competition from phonological and referential competitors in such tasks suggests that word recognition in conversation may be simpler than once thought if the domain is sufficiently constrained—which is arguably often the case in conversations where partners jointly build common ground [7]. Thus it is likely that phenomena considered central to language comprehension in many psycholinguistic models may be fundamentally altered or weakened in the most basic context of language use—in normal, interactive conversation— and that a comprehensive understanding of language processing cannot be achieved with techniques that probe language processing solely in constrained, non-interactive settings.

4. General Discussion

Language use during everyday conversation requires the fine-grained coordination of a range of processes. The types of exchanges we discuss here involve the construction and interpretation of referential expressions. A large number of processes implicated in producing simple expressions have been investigated in non-interactive contexts, and thus outside the realm of conversational dynamics and constraints. However, taking the example of reference-building in conversation, designing a modified expression to guarantee communicative success turns out to be a more complex problem: it is essential to first understand what the referential domain is and then to search this domain to construct an appropriately informative expression. For many researchers, the challenge has been to find ways of investigating such questions empirically in rich contexts, and recent advances in experimental studies of conversation have provided key evidence about how these processes unfold in real time. We focused on eye-tracking studies, where the gaze-speech link shows how and when the referential domain motivates the use of modification and how referential domains are constrained during conversation, facilitating understanding.

Modification is common in conversation but does not always obey Gricean maxims: speakers do not always produce contextually unambiguous expressions and do not always omit contextually redundant information. For example, not all modifiers are planned based on inspection or sensitivity to the referential domain. Experiment 1 showed that the strong gaze-speech relationship for scalar adjectives does not extend to color and number modifiers, which were consistently added to noun phrases in the absence of contextual support. We suggest that this may be because size and number modifiers have meanings that are not dependent on context, and are thus planned on the basis of target features alone. More generally, the problem of overmodification shows that processes other than gaze-speech coordination are at play in such tasks, and that unique identification might not be the only communicative goal in the production of modified noun phrases (*cf.* [33,58]).

Alternative goals could be, in order of complexity, proposing a conceptualization of the target that illustrates one's viewpoint or interpretation of the referent (*i.e.*, [48]), describing the target in detail, or simply highlighting a feature of the target. Persistence of naming conventions as a result of priming [59] or jointly established reference [34] may also result in overmodification in cases where the domain changes [49,60]. Finally, in some circumstances, speakers might adopt a "quick and easy" approach of mentioning salient features before the referential domain is fully analyzed in order to describe the target rapidly. Such a strategy could serve the dual role of allowing the speaker to begin speaking quickly, as well as guiding the addressee through a complex referential domain with multiple identifying pieces of information (e.g., [61])—even if some of those pieces of information are less informative than others [16]. This strategy may be less evident in languages with postnominal adjective placement, where speakers might have more time to consult the referential domain and thus weigh the informativeness of different types of modifiers before encoding them linguistically (see [47] for a discussion).

Experiment 2 showed how some aspects of modification—such as the rate of color overmodification—are unchanged by large differences in task and conversational structure. Again, color modification appeared to be largely driven by properties of the intended referent, and less by the need to uniquely identify this referent. Further, speakers exploited conversational history and task constraints to their advantage. Noun phrases that were entirely consistent with multiple referents within a ~6 inch radius were easily and uniquely interpreted by addressees. The dialog partners achieved successful communication in this manner by constraining referential domains to even smaller areas of the board, based on representations of the discourse history and task demands. Whether speakers and addressees are mutually aware of each other's reliance on the same constraints is an important question for future work.

More generally, the present research offers new perspectives on the study of the mental representations and processes traditionally emphasized in psycholinguistic models. Language processing is assumed to be rapid, efficient, and relatively error-free, and it has now been repeatedly shown that contextual aspects of language use in conversation make essential contributions to the efficiency of producing and interpreting simple expressions. This raises a number of questions about language processing in connected discourse. For example, if language processing is incremental, it is important to assess how incremental planning draws on the rich sources of information and constraints available in the conversational context. If language production requires monitoring, it is important to assess how discourse context changes, or most likely simplifies, this process. Clearly, speakers design expressions *for* addresses, since after all a central goal in discourse is for the speaker and addressee to exchange information. To a large extent, then, communicative success may depend as much on speakers' ability to exploit information accumulated during interactive exchanges as it does on the efficiency of processes like lexical retrieval and structure building.

5. Conclusions

In this article, we have presented experimental findings from two different approaches to the study of reference in conversation. Our goals were two-fold. First, we aimed to demonstrate that it was possible, and in fact, fruitful, to examine questions concerning language use in unscripted conversation

using experimental paradigms and standard hypothesis testing procedures. Second, we aimed to examine how conversational partners formulate and interpret referential expressions in rich dialog contexts. The first approach used a task-based conversation structured by a series of trials to examine how and when speakers planned to include information about features of a referent in their expressions. The second approach used a completely unscripted task-based conversation to examine how interlocutors jointly and implicitly shaped referential domains during conversation, and how these domains influenced on-line language use. We view the two approaches as complementary: we can gain insight into problems of language use by pairing techniques that require an experimenter-controlled trial structure, with techniques that allow interlocutors to navigate through the referential domain on their own. The problem of noun modification in unscripted conversation is thus one example of a domain where interactive aspects of communication can now be quantified with dependent measures traditionally employed in psycholinguistic research. Results obtained by testing, for example, the sensitivity of different modifiers to context, and speakers' cooperative construction of referential domains in unscripted conversation, highlight possible new areas of inquiry into the interface of context and language use in normal, everyday exchanges.

Acknowledgements

References

1. Frazier, L. Sentence processing: A tutorial review. In *Attention and Performance XII: The Psychology of Reading*; Coltheart, M., Ed.; Erlbaum: Hillsdale, NJ, USA, 1987; pp. 559–586.
2. Garnsey, S.M.; Pearlmutter, N.J.; Myers, E.; Lotocky, M.A. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *J. Mem. Lang*. **1997**, *37*, 58–93.
3. Trueswell, J.C.; Tanenhaus, M.K. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang.* **1994**, *33*, 285–318.
4. Clark, H.H. *Arenas of Language Use*; University of Chicago Press: Chicago, IL, USA, 1992.
5. Trueswell, J.C.; Tanenhaus, M.K. *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*; MIT Press: Cambridge, UK, 2005.
6. Ferreira, F.; Clifton, C. The independence of syntactic processing. *J. Mem. Lang*. **1986**, *25*, 348–368.
7. Clark, H.H. *Using Language*; Cambridge University Press: Cambridge, MA, USA, 1996.
8. Krauss, R.M.; Weinheimer, S. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *J. Pers. Soc. Psychol*. **1966**, *4*, 343–346.
9. Sachs, H., Schegloff, E.A.; Jefferson, G. A simplest systematics for organization of turn-taking for conversation. *Language* **1974**, *50*, 696–735.
10. Poesio, M.; Rieser, H. Completions, coordination, and alignment in dialogue. *Dialogue Discourse* **2010**, *1*, 1–89.

11. Purver, M.; Kempson, R. Incrementality, alignment and shared utterances. In *Catalog '04: Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue*; Ginzburg, J., Vallduví, E., Eds.; Barcelona, Spain, 2004; pp. 85–92.

12. Bock, K. Exploring levels of processing in language production. In *Natural Language Generation*; Kempen, G., Ed.; Martinus Nijhoff: Dordrecht, The Netherlands, 1987; pp. 351–363.

13. Tanenhaus, M.K.; Spivey-Knowlton, M.J.; Eberhard, K.M.; Sedivy, J.C. Integration of visual and linguistic information in spoken language comprehension. *Science* **1995**, *268*, 1632–1634.

14. Cooper, R.M. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn. Psychol.* **1974**, *6*, 84–107.

15. Meyer, A.S.; Sleiderink, A.M.; Levelt, W.J.M. Viewing and naming objects: Eye movements during noun phrase production. *Cognition* **1998**, *66*, B25–B33.

16. Pechmann, T. Incremental speech production and referential overspecification. *Linguistics* **1989**, *27*, 89–110.

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

18. Brown-Schmidt, S.; Tanenhaus, M.K. Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *J. Mem. Lang.* **2006**, *54*, 592–609.

19. Hanna, J.E.; Brennan, S.E. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *J. Mem. Lang.* **2007**, *57*, 596–615.

20. Cutler, A., Dahan, D.; van Donselaar, W. Prosody in the comprehension of spoken language: A literature review. *Lang. Speech* **1997**, *40*, 141–201.

21. Kraljic, T.; Brennan, S.E. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cogn. Psychol.* **2005**, *50*, 194–231.

22. Spivey, M.J.; Tanenhaus, M.K.; Eberhard, K.M.; Sedivy, J.C. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cogn. Psychol.* **2002**, *45*, 447–481.

23. Richardson, D.C.; Dale, R. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cogn. Sci.* **2005**, *29*, 1045–1060.

24. Richardson, D.C.; Dale, R.; Kirkham, N.Z. The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychol. Sci.* **2007**, *18*, 407–413.

25. Richardson, D.C.; Dale, R.; Tomlinson, J.M. Conversation, gaze coordination, and beliefs about visual context. *Cogn. Sci.* **2009**, *33*, 1468–1482.

26. Bard, E.G.; Anderson, A.H.; Chen, Y.; Nicholson, H.B.M.; Havard, C.; Dalzel-Job, S. Let's you do that: Sharing the cognitive burdens of dialog. *J. Mem. Lang.* **2007**, *57*, 616–641.

27. Brennan, S.E.; Chen, X.; Dickinson, C.; Neider, M.; Zelinsky, G. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* **2008**, *106*, 1465–1477.

28. Carletta, J.; Hill, R.L.; Nicol, C.; Taylor, T.; de Ruiter, J.P.; Bard, E.G. Eyetracking for two-person tasks with manipulation of a virtual world. *Behav. Res. Methods* **2010**, *42*, 254–265.

29. Brown-Schmidt, S.; Gunlogson, C.; Tanenhaus, M.K. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition* **2008**, *107*, 1122–1134.

30. Brown-Schmidt, S.; Tanenhaus, M.K. Real-time investigation of referential domains in unscripted conversation: a targeted language game approach. *Cogn. Sci.* **2008**, *32*, 643–684.

31. Watson, D.; Arnold, J.A.; Tanenhaus, M.K. Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition* **2008**, *106*, 1548–1557.

32. Grice, H.P. Logic and conversation. In *Syntax and Semantics 3: Speech Arts;* Core, P., Morgan, J.L., Eds.; Academic Press: New York, NY, USA, 1975; pp. 41–58.

33. Olson, D.R. Language and thought: Aspects of a cognitive theory of semantics. *Psychol. Rev.* **1970**, *77*, 257–273.

34. Clark, H.H.; Wilkes-Gibbs, D. Referring as a collaborative process. *Cognition* **1986**, *22*, 1–39.

35. Sedivy, J.C. Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In *Approaches to Studying World-Situated Language Use: Bridging the Language as Product and Language as Action Traditions*; Trueswell , J.C., Tanenhaus, M.K., Eds.; MIT press: Cambridge, MA, USA, 2005; pp. 153–171.

36. Arnold, J.E.; Griffin, Z. The effect of additional characters on choice of referring expression: Everyone competes. *J. Mem. Lang.* **2007**, *56*, 521–536.

37. Fukumura, K.; van Gompel, R.P.G.; Pickering, M.J. The use of visual context during the production of referring expressions. *Q. J. Exp. Psychol.* **2010**, *63*, 1700–1715.

38. Sedivy, J.C. Pragmatic *versus* form-based accounts of referential contrast: Evidence for effects of informativity expectations. *J. Psycholinguist. Res*. **2003**, *32*, 3–23.

39. Sedivy, J.C.; Tanenhaus, M.K.;Chambers, C.G.; Carlson, G.N. Achieving incremental semantic interpretation through contextual representation. *Cognition* **1999**, *71*, 109–147.

40. Grodner, D.; Sedivy, J.C. The effects of speaker-specific information on pragmatic inferences. In *The Processing and Acquisition of Reference*; Pearlmutter, N., Gibson, E., Eds.; MIT Press: Cambridge, MA, USA, in press.

41. Landragin, F. Visual perception, language and gesture: A model for their understanding in multimodal dialogue systems. *Signal Process.* **2006**, *86*, 3578–3595.

42. Chambers, C.G.; Tanenhaus, M.K.; Eberhard, K.M.; Filip, H.; Carlson, G.N. Circumscribing referential domains during real-time language comprehension. *J. Mem. Lang.* **2002**, *47*, 30–49.

43. Heller, D.; Grodner, D.; Tanenhaus, M.K. The role of perspective in identifying domains of reference. *Cognition* **2008**, *108*, 831–836.

44. Hanna, J.E.; Tanenhaus, M.K.; Trueswell, J.C. The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* **2003**, *49*, 43–61.

45. Beun, R.J.; Cremers, A. Object reference in a shared domain of conversation. *Pragmatics Cognition* **1998**, *6*, 121–152.

46. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed; Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ, USA, 1988.

47. Belke, E. Visual determinants of preferred adjective order. *Vis. Cogn*. **2006**, *14*, 261–294.

48. Brennan, S.E.; Clark, H.H. Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol. Learn. Mem. Cogn.* **1996**, *22*, 1482–1493.

49. Ferreira, F.; Swets, B. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* **2002**, *46*, 57–84.

50. Dell, G.S.; O'Seaghdha, P.G. Mediated and convergent lexical priming in language production: A comment on Levelt, *et al.* (1991). *Psychol. Rev.* **1991**, *98*, 604–614.

51. Levelt, W.J. M.; Roelofs, A.; Meyer, A.S. A theory of lexical access in speech production. *Behav. Brain. Sci.* **1991**, *22*, 1–75.

52. Grosz, B.J.; Sidner, C.L. Attention, intentions, and the structure of discourse. *Comput. Linguist.* **1986**, *12*, 175–204.

53. Hallett, P.E. Eye movements. In *Handbook of Perception and Human Performance*; Boff, K.R., Kaufman, L.; Thomas, J.P., Eds.; Wiley: New York, NY, USA, 1986; pp. 10.1–10.112.

54. Baayen, R.H.; Davidson, D.J.; Bates, D.M. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **2008**, *59*, 390–412.

55. Allopenna, P.D.; Magnuson, J.S.; Tanenhaus, M.K. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *J. Mem. Lang.* **1998**, *38*, 419–439.

56. Greene, S.B.; Gerrig, R.J.; McKoon, G.; Ratcliff, R. Unheralded pronouns and management by common ground. *J. Mem. Lang.* **1994**, *33*, 511–526.

57. Kintsch, W. The role of knowledge in discourse comprehension: A construction-integration model. *Psychol. Rev.* **1988**, *95*, 163–182.

58. Roberts, C. Uniqueness in definite noun phrases. *Ling. Philos.* **2003**, *26*, 287–350.

59. Cleland, A.A.; Pickering, M.J. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *J. Mem. Lang.* **2003**, *49*, 214–230.

60. van der Wege, M.M. Lexical entrainment and lexical differentiation in reference phrase choice. *J. Mem. Lang.* **2009**, *60*, 448–463.

61. Spivey, M.J.; Tyler, M.J.; Eberhard, K.M.; Tanenhaus, M.K. Linguistically mediated visual search. *Psychol. Sci.* **2001**, *12*, 282–286.