Cognition 154 (2016) 102–117

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

The historical context in conversation: Lexical differentiation and memory for the discourse history

Si On Yoon *, Aaron S. Benjamin, Sarah Brown-Schmidt

Department of Psychology, University of Illinois, Champaign, IL 61820, USA

A R T I C L E I N F O

Article history:
Received 14 October 2015
Revised 9 May 2016
Accepted 16 May 2016
Available online 31 May 2016

Keywords:
Dialogue
Discourse
Lexical differentiation
Memory

A B S T R A C T

When designing a definite referring expression, speakers take into account both the local context and certain aspects of the historical context, including whether similar referents have been mentioned in the past. When a similar item has been mentioned previously, speakers tend to elaborate their referring expression in order to differentiate the two items, a phenomenon called lexical differentiation. The present research examines the locus of the lexical differentiation effect and its relationship with memory for the discourse. In three experiments, we demonstrate that speakers differentiate to distinguish current from past referents; there was no evidence that speakers differentiate in order to avoid giving two items the same label. Post-task memory tests also revealed a high level of memory for the discourse history, a finding that is inconsistent with the view that failures of memory underlie low differentiation rates. Instead, memory for the discourse history, while necessary, is not sufficient for speakers to design language with respect to the historical context. Speakers must additionally view the discourse history as relevant to design language with respect to this broader context. Finally, measures of memory for past referents point to asymmetries between speakers and listeners in their memory for the discourse, with speakers typically remembering the discourse history better.

1. Introduction

For communication to be successful, conversational partners must take into account each other's general knowledge and memory for the ongoing discussion. Consider the process of designing a definite referring expression. For a speaker to successfully communicate her meaning, she must take into account properties of the intended referent, as well as contextual information, in order to uniquely identify the intended referent (Roberts, 2003). The contextual information that shapes the design of referring expressions includes information in the *immediate context*, such as the characteristics of other candidate referents (Olson, 1970; Osgood, 1971). The way in which a given object will be described, then, depends on the properties of the other items in the local context (Beun & Cremers, 1998; Brennan & Clark, 1996; Brown-Schmidt & Tanenhaus, 2006; Horton & Keysar, 1996; Nadig & Sedivy, 2002; Wardlow Lane & Ferreira, 2008). For example, imagine a situation where some friends are at a shoe shop, and one friend wants to point out the shoes she would like to buy. In such a situation, she would have to distinguish her intended referent from the many other items in the local context, likely through the use of a modified referring expression, as in "*The leopard-print heels are super cute!*", rather than "*The shoes are super cute*!"

Identifying the features that distinguish the intended referent from those in the local context is a cognitive process that unfolds over time and must be coordinated with utterance planning. As a result of this interplay between contextual encoding and language planning, speakers sometimes produce over-informative or under-informative expressions (Deutsch & Pechmann, 1982; Engelhardt, Bailey, & Ferreira, 2006; Ferreira, Slevc, & Rogers, 2005; Maes, Arts, & Noordman, 2004; Olson, 1970; Sedivy, 2003). The likelihood of producing a locally overinformative expression varies with adjective class (Sedivy, 2005; Brown-Schmidt & Konopka, 2011), and can occur when analysis of the local context lags behind production planning (Pechmann, 1989).

Another source of contextual constraint in conversation is the *historical discourse context*, which includes information that was discussed in the past (Brennan & Clark, 1996). A speaker who takes into account both the immediate discourse context as well as the historical discourse context may produce an expression that is overspecified with respect to the immediate context, but appropriately informative if the discourse history is taken into

* Corresponding author at: Department of Psychology, 603 E. Daniel St., Champaign, IL 61820, USA.

E-mail addresses: syoon@illinois.edu (S.O. Yoon), asbenjam@illinois.edu (A.S. Benjamin), sarahbrownschmidt@gmail.com (S. Brown-Schmidt).

http://dx.doi.org/10.1016/j.cognition.2016.05.011
0010-0277/

account. Consider the case of our shoe-shopping friends. If they were to continue their conversation over a cup of coffee, reference to the newly-purchased shoes in the new context would no longer require the modifier "leopard-print". Yet studies of language use in dialogue show once a term has been established, that speakers persist in the use of these terms even if the context changes and the modifier is no longer necessary, as in "*I just love my new* <u>*leopard-print*</u> *heels*" produced in a context with only a single pair of shoes (Brennan & Clark, 1996; Van der Wege, 2009). This tendency to persist in the use of previous terms, termed *lexical entrainment*, is one example of the influence of the historical discourse context. In the present research we examine a related effect of historical influence on referring—the process of *lexical differentiation*, in which speakers take into account past reference when designing new referring expressions.

1.1. Lexical differentiation and its source

Lexical differentiation refers to a discourse phenomenon in which speakers differentiate two sequentially presented objects from the same category (Van der Wege, 2009). For example, imagine a situation in which a speaker describes one shirt in the context of several unrelated objects. In a context like this one, the speaker is likely to refer to the shirt with a bare noun phrase, as in "*the shirt*". However, if she were to later refer to a second, distinct shirt in the context of several unrelated objects, she might differentiate the second shirt from the first by using a modifier, as in "*the striped shirt*," even though the modifier is not necessary in the local context (i.e., "*the shirt*" would suffice to uniquely identify the intended referent). Speakers also sometimes use a different noun to differentiate the second object from the first, as in "the blouse" rather than "the shirt" (see Yoon & Brown-Schmidt, 2013). These findings show that the discourse history influences how speakers design referring expressions.

According to Van der Wege (2009), speakers lexically differentiate in order to avoid giving the same label to two different entities, a process termed "pre-emption by similar form" (also see Clark & Clark, 1979). When speakers refer to the second object, they prefer to use a distinct label that contrasts with the previously established label; in other words, the speaker avoids giving the label, "*the shirt*", to two different entities. By calling the second shirt "*the striped shirt*" or "*the blouse*", the speaker is able to differentiate the two labels. The idea behind pre-emption by similar form is that the previously used label "*the shirt*" pre-empts the subsequent use of the same label to refer to a different item, thus creating the need for lexical differentiation. If this view of the lexical differentiation effect is correct, differentiation should not be observed if the speaker had not labeled the first entity with the basic object label. For example, if the speaker were to refer to the first shirt with a locative expression, such as "*the top left one*", there would be no pre-emption of the label "*the shirt*", and thus no need to describe the second shirt with a modifier (instead "*the shirt*" would be an appropriate label for the second shirt). Alternatively, the locus of the differentiation effect might be an attempt to distinguish current from past *referents*, regardless of how they had been named. If so, any previous reference to a shirt—with a locative or a descriptive expression—should increase the likelihood that the speaker would differentiate the second shirt from the first.

1.2. The relationship between lexical differentiation and memory for past referents

A necessary precondition to designing referential expressions with respect to the historical discourse context is successfully accessing a memorial representation of the previous referent when planning a description of the current referent. If a speaker fails to remember describing a shirt in the past, they would be unlikely to differentiate the current shirt from the previous shirt, instead producing the same expression, "*the shirt*", to refer to both shirts. Failures to remember the past discourse context may explain why differentiation is relatively infrequent: Speakers differentiated only 7.5% of the time in Yoon and Brown-Schmidt (2013)'s study and 19–33% in Van der Wege (2009)'s study in the "real audience" condition (19% for atypical objects; 33% for typical objects; see Van der Wege, 2009, Fig. 5). The larger effect in Van der Wege's study is likely due to methodological differences. In particular, the use of atypical category exemplars in the stimulus set, and a design in which each of 12 items in a display were referenced may have encouraged speakers to distinguish current from past referents.

Yoon and Brown-Schmidt (2013) extended the lexical differentiation effect in a paradigm that examined how listeners interpreted these lexically differentiated expressions. In that experiment, the first of two key referents (e.g., two different shirts) was labeled, e.g., "*the shirt*", or located, e.g., "*the top left one*", and then in the critical condition, the speaker referred to the second (distinct) shirt with a modified expression, as in "*the striped shirt*". Somewhat surprisingly, analysis of listener eye-gaze as they interpreted these expressions found no evidence that listeners expected speakers to differentiate the two objects in the naming condition more than the locating condition. One interpretation of these findings is that both labeling and locating the first shirt prompted listeners to expect differentiation when listening to a description of the second shirt. Whether speakers differentiate in both of these circumstances is an open question that the present work is designed to address.

Another possibility is that listeners may have not remembered the previous discourse referent when interpreting these utterances. When considering the role of memory in producing or expecting lexical differentiation, a relevant phenomenon is the *generation effect*, which refers to the finding that the act of generating information promotes better memory for that information when compared to reading (Slamecka & Graf, 1978). In an analysis of what information tends to get repeated over the course of a conversation, Knutsen and Le Bigot (2014) report that referring expressions like "*the market*" are more likely to be repeated in a conversation by the person who first introduced that topic into the conversation, consistent with a generation effect. McKinley, Brown-Schmidt, and Benjamin (in preparation) similarly report a generation benefit for item recognition in a natural conversation paradigm where the "items" were pictures that participants discussed with one of two conversational partners. Based on these findings, speakers may have better memory for what has been said in conversation, compared to listeners. Listeners' comparatively worse memory for past referents, then, may explain the apparent lack of differentiation in comprehension (Yoon & Brown-Schmidt, 2013).

1.3. The present research

The goal of the present research is to examine the locus of the lexical differentiation effect and its relationship with memory for past discourse referents. In Experiments 1 and 2, we elicit a differentiation effect in language production, and examine the situations in which it does and does not occur in order to understand the influence of the historical discourse context on referring. In Experiment 3, we examine the same question in situations that include an unmentioned, but target-related context item. Measures of memory for the discourse history in Experiments 1–3 are used to evaluate whether poor memory for past referents explains the low incidence of differentiation, and listeners' consequent lack of expectation for it (Yoon & Brown-Schmidt, 2013).

2. Experiment 1

In the current experiment, we manipulate the way in which the first referent was referenced—with a locative, "*the top left one*", or a descriptive phrase, *"the shirt"*—in order to test the locus of the lexical differentiation. In addition, we measure speakers' and listeners' memory for the discourse history.

2.1. Method

2.1.1. Participants

Seventy-two undergraduates (thirty-six pairs) at the University of Illinois at Urbana-Champaign participated in the experiment in return for partial course credit or cash payment ($8). Participants were native speakers of North American English with normal hearing and normal or corrected-to-normal vision.

2.1.2. Procedure

The experiment consisted of two phases: a referential communication task followed by a memory test. The entire experiment lasted approximately 50 min.

2.1.3. Referential communication task

During the referential communication task (cf. Krauss & Weinheimer, 1966), two naïve participants were randomly assigned to the roles of speaking and listening and sat at separate computers in the same room. The computers were arranged so that participants could see each other's faces but could not see each other's computer screens. On each trial, the speaker and listener viewed a computer screen with four pictures. The pictures appeared in one of four different random positions on each screen. The target object was indicated to the speaker with a black box (see Fig. 1). On the listener's screen, the target was not indicated. The speaker's task was to give the listener an instruction aloud to click on one of the four pictures (e.g., *Click on the sock*). The listener followed the speaker's instructions on her own screen. The task was interactive and participants were allowed to ask questions for clarification (clarification requests were rare). After the listener clicked the target, the speaker clicked the mouse to advance to the next trial. Participants maintained their assigned role (i.e., speaker or listener) throughout the task. Recordings of the speaker's voice were saved directly to the computer.

2.1.4. Materials

During the referential communication task, each participant completed a total of 462 trials, including 198 entrainment trials, 33 test trials, and 231 filler trials. During the test trials (the focus of our analyses), the speaker described a "target" object from a particular category (e.g., a dotted sock). Each test trial was associated with 6 previous entrainment trials (Fig. 1). The type of entrainment trial served as the key experimental manipulation, and was manipulated within-subjects. Entrainment trials always preceded test trials, and allowed us to manipulate whether the basic object label for a given target object at test, e.g., "sock" had already been used to describe a different item during entrainment. During both entrainment and test trials, the target object was presented with three unrelated objects such that given the *local* context, a bare noun was sufficient to identify the target. These unrelated objects were rotated pseudo-randomly such that the three unrelated items varied from trial to trial. The critical question was whether speakers would design referring expressions at test based on the *historical* discourse context.

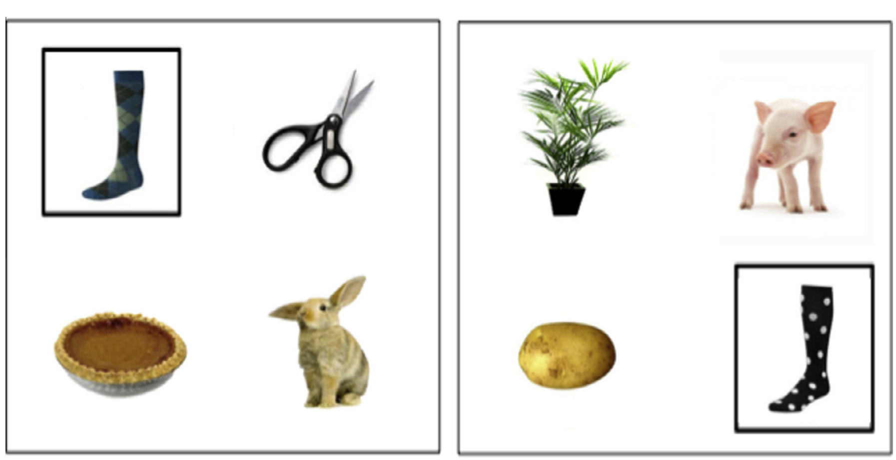
In the Contrast-naming condition, the target on entrainment trials was an exemplar from the same category as the target on the test trial. For example, the speaker would describe an argyle sock 6 times during entrainment trials prior to describing the target dotted sock at test.

In the Non-contrast condition, participants described an unrelated object (e.g., apple) 6 times in entrainment, and then described the target (e.g., dotted sock) at test. Note that for items in the Non-contrast condition, speakers never saw the contrast object.

Finally, in the Contrast-location condition, speakers described the location of the contrast object (e.g., *top left one*) 6 times, and then described the target object at test (e.g., *dotted sock*). The speaker was cued to use a locative phrase through an on-screen text prompt (e.g., "LOCATION"). This condition was included in order to test whether it is naming *per se* that is critical to elicit the differentiation effect.

Following entrainment, test displays showed four new (previously unseen) objects, including the target and three unrelated objects. While the target and the unrelated objects were all new tokens, the categories that they came from (e.g., socks, pigs, etc.) had all been previously experienced during the entrainment trials.

Filler trials contained two contrasting objects from the same category (e.g., two fish) and two unrelated objects. Half of the time



(a) (b)

Fig. 1. (a and b) Experiments 1 and 2: Example stimuli from the referential communication task, including entrainment trials (a) and test trials (b). The target is indicated to the speaker by the black rectangle. The addressee's screen would show the same four pictures, but in different locations and without the black box. This example shows the "Contrast-naming" condition; in the "Non-contrast" condition, the target during entrainment would be an unrelated item such as an apple.

the target was one of the two contrasting objects and on the other half of filler trials, the target was one of the two unrelated objects. Filler trials with contrasting objects were included so that speakers would sometimes need to produce modified noun phrases based on the immediate context.

Entrainment and test trials for the different item sets were interspersed with filler trials in a pseudo-random order such that from the participants' point of view, there was no distinction between the trial types, and so that all 7 trials associated with a single item set (6 entrainment trials followed by 1 test trial) occurred within a span of at most 120 trials (the span ranged from 73 to 116, M = 100.37, SD = 12.4). Note that for a given object set, the maximum time between the final entrainment trial and the test trial was 15 trials (range of 1–15, M = 7.39, SD = 4.46).

All visual stimuli were pictures of common objects. The stimuli included the 33 triplets of target (e.g., dotted sock), contrast (e.g., argyle sock), and unrelated items (e.g., apple) that were used as critical items across the three conditions. The remaining visual stimuli were used on filler trials and included 33 triplets of contrasting objects (e.g., opened box, wrapped box, stacked box) and 66 pairs of contrasting objects (e.g., sitting dog, jumping dog) from the same categories. All target items were counterbalanced across conditions across three lists. Each participant completed the items on one list.

2.1.5. Memory test

Following the communication task, participants performed a surprise recognition memory test on their own computer (Fig. 2). Each trial presented two pictures from the same category: one was an "old" picture that was presented during the communication task, and the other was a "new" picture from the same basic object category. Participants were asked to click on the old picture on each trial. The old and new pictures were rotated across two lists, so that the old picture on one list was the new picture on the other list. A total of 99 trials tested memory for the 33 contrast items (e.g., argyle sock or apple), 33 target items (e.g., dotted sock), and 33 unrelated filler items (e.g., drum). The correct answer for each trial was always the object that had either been named or referenced with a locative phrase during the referential communication task. The speaker and listener were not allowed to talk to each other during this phase of the experiment.

2.1.6. Predictions

The communication task was designed to replicate the lexical differentiation effect in language production (Van der Wege, 2009; Yoon & Brown-Schmidt, 2013). If speakers lexically differentiate, the modification rate (e.g., production of a modified expression like *dotted sock* vs. an unmodified expression like *sock*) should be higher in the Contrast-naming condition than in the Non-contrast condition. If the locus of the differentiation effect is to avoid giving two different items the same *label*—for example to avoid labeling both the argyle sock and the dotted sock, "*the sock*" (pre-emption by similar form; Van der Wege, 2009)—speakers should be more likely to modify test expressions in the Contrast-naming condition than the Contrast-location condition. This is because it is only in the Contrast-naming condition that the basic object label for the target had previously been used (e.g., "sock" had been used on entrainment trials for a different referent). However, if speakers differentiate to distinguish current from past *referents*, whether or not the target label had been used at entrainment should not matter, and the rate of modifiers at test should be higher in both the Contrast-naming and Contrast-location conditions compared to the Non-contrast condition.

Analysis of the relationship between memory for contrast objects, and speakers' referential design choices in the referential communication task will allow us to address why the differentiation rate in previous studies (Van der Wege, 2009; Yoon & Brown-Schmidt, 2013) was so low. If successful differentiation requires memory for the contrast object, recognition of contrast objects should be related to the differentiation rate. Such a finding would point to memory as a key factor in determining whether or not speakers will design expressions with respect to the historical discourse record. Comparisons between speakers' and listeners' memory for the discourse history will speak to the apparent lack of a differentiation effect in language comprehension (Yoon & Brown-Schmidt, 2013). If the generation effect in memory research (Marsh, Edelman, & Bower, 2001) extends to more natural conversational settings, speakers may consistently outperform listeners in their memory of the discourse history, offering a memory-based explanation why listeners would exhibit less consideration of the discourse history.

2.2. Results

2.2.1. Referential communication task

Recordings of speakers' object descriptions at entrainment and test were transcribed and coded for whether their referring expressions were modified or not (Fig. 3). In the (rare) cases where listeners asked a clarification question, we coded only the speaker's initiating description of the referent (see Duff, Gupta, Hengst, Tranel, & Cohen, 2011). On entrainment trials, speakers produced modified noun phrases 0.03% of the time in the Non-contrast condition and 14.0% of the time in the Contrast-naming condition. Note that the referents during entrainment in the Non-contrast condition (e.g., apple) were not rotated across conditions; thus this difference in modification rate is likely due to item differences. Because Non-contrast objects were unrelated to test objects, and the focus of our analyses is on modification rate at test, we do not discuss these differences further.

At test, there were two potential ways speakers could differentiate current from past referents: Speakers could produce a modified noun phrase (e.g., "*shirt*" → "*striped shirt*"), or they could produce an alternative bare noun phrase (e.g., "*shirt*" → "*blouse*", Yoon & Brown-Schmidt, 2013). However, it was difficult

Fig. 2. Experiments 1 and 2: Example stimuli from the memory test.

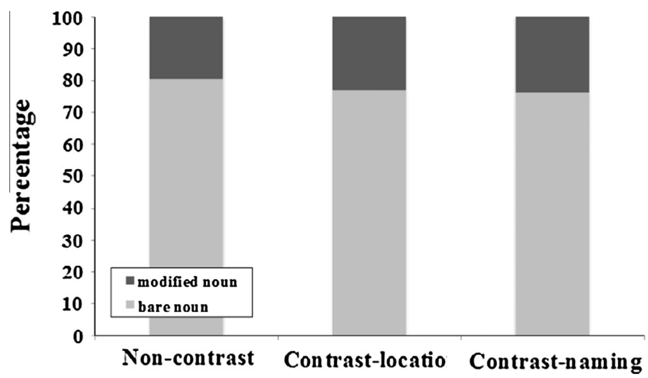


Fig. 3. Experiment 1: Percentage of each noun phrase type on target trials during the referential communication task.

to measure the proportion of alternative, bare noun phrases in the Non-contrast and Contrast-location conditions, because objects in the target category were referenced only once, making it impossible to know if the speaker was differentiating or not. As a result, our planned analyses follow the procedure of Yoon and Brown-Schmidt (2013) and focus on the rates of modified noun phrases.

We analyzed the data in a logistic mixed effects model with entrainment type (Non-contrast, Contrast-location, and Contrast-naming) as a fixed effect, and subjects and items as random intercepts. The model was fit using the lmer package in R, with the maximal random-effects structure for subjects and items. In cases where the maximal model did not converge, a backwards-fitting procedure was used to identify the model with the largest random effects structure that would converge (see Barr, Levy, Scheepers, & Tily, 2013). The dependent variable was whether the speaker used a modifier or not. Full model details are presented in the Appendix (Table 1A).

While bare noun phrases were common in all three conditions, speakers produced modified noun phrases significantly more often when the contrast object had previously been discussed (21.9% in the Contrast-naming vs. 20.8% in the Contrast-location condition) than when it had not (14.1%, Non-contrast condition), ($z = 2.194$, $p < 0.01$). The rate of modified noun phrases did not significantly differ between the two contrast conditions ($z = 0.067$, $p > 0.05$). The size of the differentiation effect was 7.8%,[1] similar to our previous findings.

In summary, these findings demonstrate a small lexical differentiation effect (Van der Wege, 2009; Yoon & Brown-Schmidt, 2013), indicating that speakers consider the historical discourse record when designing referring expressions (Brennan & Clark, 1996). Critically, however, the fact that modification rates in the Contrast-location and Contrast-naming conditions did not significantly differ suggests that speakers differentiate not to avoid lexical conflict (cf., Van der Wege, 2009), but rather to distinguish current referents from past *referents*, regardless of how those referents had been labeled.

2.2.2. Memory test

We evaluated memory for the contrast items to address the question of whether the low rate of lexical differentiation is due to memory failures. We also compared memory performance by speakers and listeners to evaluate whether they developed distinct memory representations of the discourse. Overall, both speakers and listeners successfully recognized contrast and target items over 80% of the time (Fig. 4), suggesting that failures to retrieve past referents cannot be the primary cause of the low rate of differentiation.

The memory accuracy data were analyzed in a logistic mixed effects model with role (listener vs. speaker), referent (contrast vs. target), and entrainment type (Non-contrast vs. Contrast-location vs. Contrast-naming) as fixed effects (see Appendix for model details, Table 2A). Overall, speakers had better memory than listeners ($z = 2.935$, $p < 0.05$). Memory was also better for contrast than target items, which is unsurprising, as contrast items were referenced 6 times during the communication task, whereas target items were referenced only once ($z = -6.658$, $p < 0.05$).

A significant interaction between role, referent, and entrainment type ($z = -1.993$, $p < 0.05$) was explored by examining the role and entrainment effect for contrast and target items separately. Memory for contrast items was high for both speakers and listeners (>95%) and did not differ from one another ($z = 1.062$, $p > 0.05$), providing evidence against the hypothesis that listeners' failure to differentiate is due to poor memory for the contrast. In addition, contrast items in the Contrast-naming condition were better remembered than items in the Contrast-location condition ($z = 2.994$, $p < 0.05$, see Table 3A). Memory for target items was better for speakers than listeners (94.9% vs. 86.5%; $z = 5.090$, $p < 0.05$, see Table 4A), consistent with a generation benefit.

2.2.3. Relationship between referential form and memory

The fact that recognition of contrast items was uniformly high is inconsistent with the idea that speakers fail to differentiate when memory for the contrast item is weak. We tested whether speaker memory for the contrast predicted differentiation. A mixed-effects model included whether the speaker correctly recognized the contrast object during the memory test as a fixed effect; the dependent measure was whether the associated target item was modified. This analysis did not support the hypothesis that the differentiation rate is influenced by the speaker's memory for the contrast ($z = -0.426$, $p > 0.05$).

While there was no evidence that contrast memory modulated use of a modifier on test trials, an open question is whether use of a modifier on test trials improves memory for the target. One reason to think that modification might improve memory for the referent is that the process of selecting an appropriate modifier necessarily requires an elaborate encoding of that object (e.g., saying *striped shirt* requires conceptualizing the shirt as striped; Bradshaw & Anderson, 1982). A mixed-effects model included role, entrainment type, and whether the target expression was modified as fixed effects; the dependent measure was whether the target object was correctly recognized during the memory test. The results of this analysis, however, showed that when speakers described target objects with modifiers, memory for targets did not improve for speakers ($z = 0.010$, $p > 0.05$) or listeners ($z = -0.200$, $p > 0.05$).

2.3. Summary

In summary, the results of Experiment 1 showed that speakers were more likely to use modifiers when they had previously referenced an item from the same basic object category, replicating previous findings (Van der Wege, 2009; Yoon & Brown-Schmidt, 2013, Experiment 1). However, naming was not necessary to elicit lexical differentiation. This finding is inconsistent with the argument that differentiation is motivated by pre-emption by a similar referential form (Van der Wege, 2009). Instead, speakers differentiate in order

[1] One question about this finding is whether the lag between the last entrainment trial and the test trial modulates the differentiation effect (the lag varied between 1 and 15 trials). If speakers fail to differentiate after long lags, this would be consistent with the memory-based explanation for why speakers differentiate so infrequently. The results of this post hoc analysis, however, showed that neither the main effect of lag ($z = -0.826$, $p > 0.05$), nor the interactions with condition (Non-contrast vs. Contrast-location & Contrast-naming: $z = -0.301$, $p > 0.05$; Contrast-location vs. Contrast-naming: $z = 0.994$, $p > 0.05$) were significant.

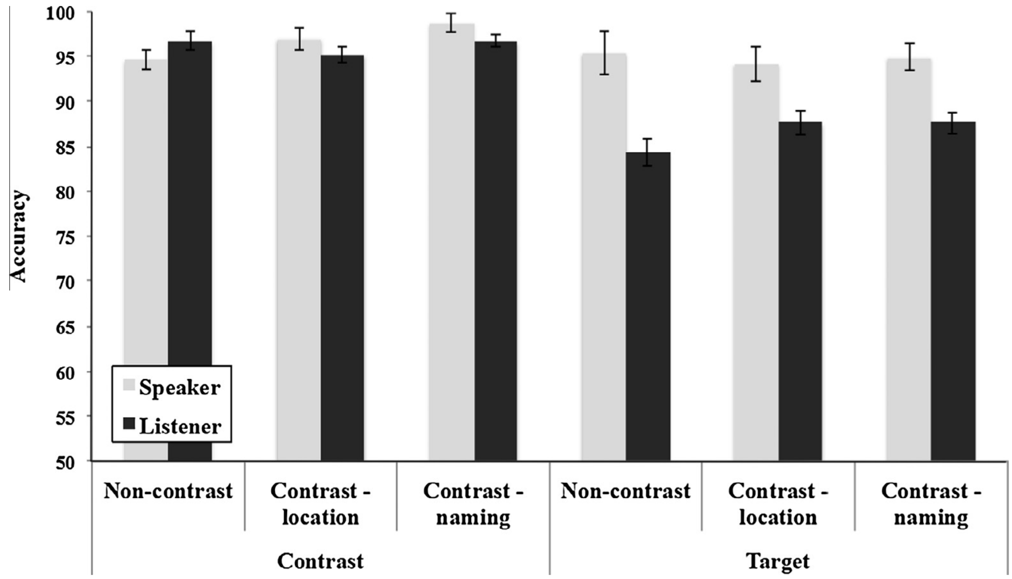


Fig. 4. Experiment 1: Accuracy on two-alternative forced choice recognition memory task for contrast items (on the left) and target items (on the right) by speakers and listeners. Error bars indicate by-participant standard error.

to distinguish the current referent from past referents—regardless of how those past referents had been labeled. In addition, this finding offers insight into a previous study of listeners' expectations for differentiation (Yoon & Brown-Schmidt, 2013, Experiments 2–3). In that study, listeners were equally likely to expect differentiated (e.g., modified) expressions when the contrast had previously been located vs. labeled. Rather than a failure to expect differentiation, the present work suggests that listeners may instead be sensitive to the fact that any reference to a contrast object can prompt subsequent differentiation. The fact that speakers and listeners had equivalent memory for contrast objects also supports this argument.

Recognition memory for past referents was high, consistent with findings in the memory literature indicating that recognition memory for pictures tends to be quite good (Shepard, 1967). We also observed that speakers had better memory for target objects; whether memory for contrast objects follows a similar pattern was obfuscated by near-ceiling performance. The lack of a relationship between memory for the contrast item and the differentiation rate suggests that failures to link the current context with the past context are not due to failures to remember that past context.

3. Experiment 2

The aim of Experiment 2 was to replicate Experiment 1 with a lower number of entrainment trials in order to equate exposure to target and contrast items. In addition, participants in Experiment 2 alternated roles of speaking and listening to address the possibility that lower recognition rates on the part of listeners was due to poor task engagement. While the speaker-memory advantage in Experiment 1 is consistent with a generation effect, speakers also had a more active role in the conversation. The required commitment to both roles in a conversation in Experiment 2 should decrease the possibility of differences in engagement.

3.1. Method

3.1.1. Participants

Ninety-six undergraduates (forty-eight pairs) at the University of Illinois at Urbana-Champaign participated in return for partial course credit or cash ($8). Participants were native speakers of North American English and had normal hearing and normal or corrected-to-normal vision. None had participated in Experiment 1.

3.1.2. Materials and procedure

The procedure of Experiment 2 was identical to Experiment 1 with the following exceptions: First, during the communication task, the two participants first completed 20 practice trials, followed by two blocks of trials in the communication task. Participants alternated between the roles of speaking and listening in each block. The aim of the practice trials was to introduce the participants to both the speaking and listening task roles. Second, we included a 20-min break between the communication task and the memory test in an additional attempt to avoid a ceiling effect on the memory test.

The design of the communication task was also slightly different. Each participant completed a total of 200 trials, including 20 practice, 36 entrainment, 36 test, and 108 filler trials. Filler trials were of the same format as in Experiment 1; on half of filler trials the speaker referred to one of two contrasting objects, and on the other half the speaker referred to one of two unrelated objects. Note that, unlike Experiment 1, each target item in Experiment 2 was associated with a single entrainment trial in order to equate exposure to contrast and target objects prior to the memory test. The experimental conditions were the same as in Experiment 1 (Non-contrast, Contrast-location, Contrast-naming conditions), and were manipulated within-subjects.

We also manipulated the lag between the entrainment and the test trial (1 vs. 10 trials) to examine whether the differentiation effect might be more pronounced at shorter lags. In the 1-lag condition, the entrainment trial for a given item (e.g., the argyle sock) was immediately followed by the test trial for that item (e.g., the dotted sock). By contrast, in the 10-lag condition, the test trial occurred 10 trials after the entrainment trial.

After the communication task, there was a 20-min break during which participants completed an unrelated distractor test (math test). Following the break, participants completed the unexpected memory test. As in Experiment 1, each test trial on the memory test showed two pictures from the same category, one of which was old and one of which was new. The task was to click on the old picture. There were a total of 108 trials that were presented in a randomized order that tested memory for the 36 contrast items, 36 target items, and 36 of the filler items. The experiment lasted about 60 min.

3.1.3. Predictions

On test trials in the communication task, we predicted that speakers would produce modified noun phrases more often in the Contrast-naming and Contrast-location conditions than the Non-contrast condition, replicating the results of Experiment 1. However, if one entrainment trial is not sufficient to elicit differentiation, the rate of modified noun phrases should not differ among the three conditions.

In the present experiment, participants alternated roles between speaking and listening to better equate task engagement. If speakers continue to outperform listeners during the memory test, this would provide convincing evidence that speaking in a conversation yields memory benefits for discourse referents. In contrast, if the superior memory performance of speakers in Experiment 1 was due to a higher level of engagement, then the new task should eliminate this benefit. By equating the number of exposures to contrast and target items, we anticipated that any effects of participant role (speaker vs. listener) should obtain for both targets and contrasts. Finally, by relating measures of contrast memory to referential form, this experiment provides another opportunity to test the hypothesis that the low differentiation rates observed in our previous work are related to failures to remember the past context.

3.2. Results

3.2.1. Referential communication task

The speakers' productions on entrainment and test trials were coded in the same way as in Experiment 1 (Fig. 5). On entrainment trials, speakers produced modified noun phrases for 0.02% of the time in the Non-contrast condition and 17.0% of the time in the Contrast-naming condition. As in Experiment 1, the higher modification rate in the Non-contrast condition is likely due to uncontrolled item differences. Because the focus of our analyses is on modification rate at test, we do not discuss these differences further.

The focus of our analyses is on the rate of modified noun phrases at test, which we use as a measure of differentiation. In the Non-contrast condition, speakers modified their expressions 19.4% of the time on test trials, whereas modification rates were higher in both the Contrast-location (23.0%) and Contrast-naming conditions (23.7%).

We analyzed the modification rates in a logistic mixed effects model with a binomial link function, using entrainment type as fixed effects (see details in Appendix Table 5A). An initial analysis included lag between entrainment and test as a factor, however, neither the main effect of lag ($z = 0.814$, $p > 0.05$), nor the interaction with entrainment type were significant ($ps > 0.05$); further, adding lag to the models did not improve model fit. As a result, we do not discuss the manipulation of lag further. The primary analysis revealed a significant main effect of entrainment type (Non-contrast vs. Contrast-location & Contrast-naming; $z = 2.270$, $p < 0.05$), consistent with the results of Experiment 1. Speakers tended to differentiate the current referent from the past referent more when they had previously referenced the contrast item (regardless of whether it was named or spatially located), compared to when they had not been exposed to the contrast. The modification rates in the Contrast-location and Contrast-naming conditions did not differ significantly ($z = 0.203$, $p > 0.05$).

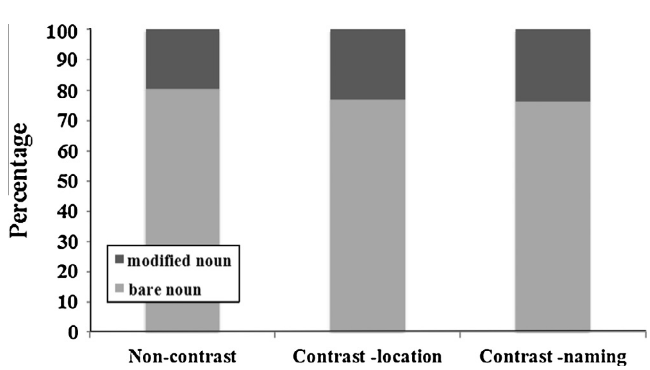


Fig. 5. Experiment 2: Percentage of each noun phrase type on test trials during entrainment.

3.2.2. Memory test

Accuracy on the memory test was overall high for both speakers and listeners, though not at ceiling, unlike Experiment 1 (see Fig. 6). We analyzed accuracy in a maximal mixed-effects model with participants' role (listener vs. speaker), referent (contrast vs. target) and entrainment type (Non-contrast vs. Contrast-location vs. Contrast-naming) as fixed effects (see details in Appendix, Table 6A). Overall, speakers had better memory than listeners ($z = 4.181$, $p < 0.05$). Main effects of entrainment type were qualified by a significant interaction between referent type (contrast vs. target) and entrainment type (Contrast-location vs. Contrast-naming, $z = -2.413$, $p < 0.05$).

Separate planned analyses for contrast and target items were used to explore the interactions. For <u>target</u> items, there was only a main effect of role, due to better memory performance by speakers ($z = 3.842$, $p < 0.05$, see Table 7A). For <u>contrast</u> items, there was a significant interaction between role and entrainment type ($z = 3.292$, $p < 0.05$, see Table 8A): Whereas speakers had better memory in the Contrast-naming than the Contrast-location condition ($z = 2.363$, $p < 0.05$), listeners' memory did not differ across these conditions ($z = 0.340$, $p > 0.05$). These findings show that naming boosted speakers' memory more for the contrast item than it did for listeners.

3.2.3. Relationship between referential form and memory

In Experiment 1, we did not observe a significant relationship between memory for the discourse context and referential form, possibly due to very high levels of memory performance for the contrast item. However, despite off-ceiling memory performance for the contrast item in Experiment 2, memory for the contrast item still did not significantly predict the differentiation rate ($z = 1.382$, $p > 0.05$).

As in Experiment 1, we also examined whether the way a target was described affected how it was remembered. The results of this analysis revealed that when speakers described the target with a modifier, *both* speakers and listeners remembered the referent better, compared to situations when the referent was identified by a bare noun phrase ($z = 3.916$, $p < 0.05$). This result suggests that modification is helpful for future memory. Of course, it is worth remembering that the relationship between memory and the occurrence of a modifier is subject to an uncontrolled item-selection effect—it is possible that there are item characteristics that both attract the use of modifiers and make it easier to remember. There was no interaction with role or entrainment type.

3.3. Summary

In summary, the results of Experiment 2 demonstrated a lexical differentiation effect that was not contingent on naming, replicating Experiment 1. The small magnitude of the effect (a ~4% increase in modification when an object from the same basic object category had previously been referenced) may be due to the

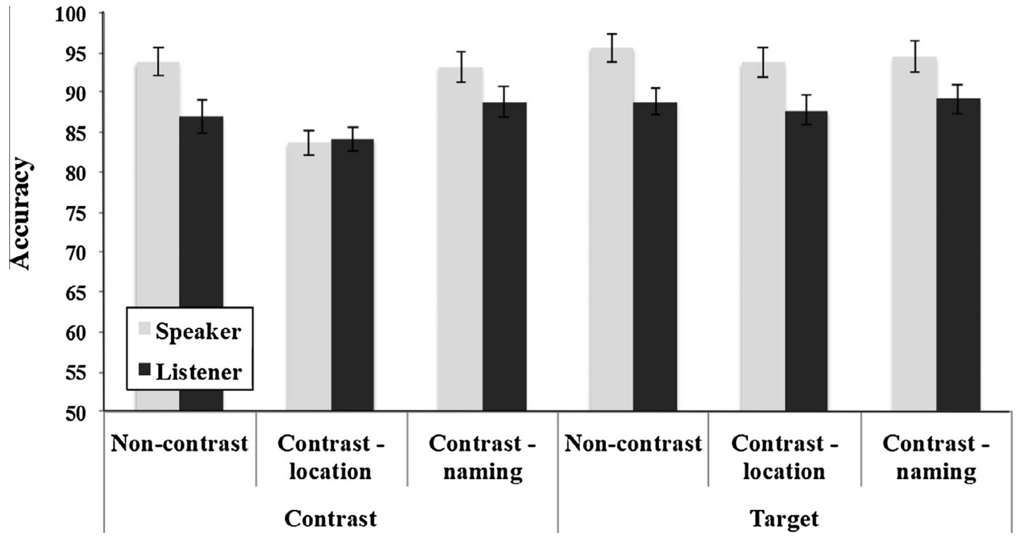


Fig. 6. Experiment 2: Accuracy on two-alternative forced choice recognition memory task for contrast items (on the left) and target items (on the right) by speakers and listeners. Error bars indicate by-participant standard error.

fact that the contrast item had only been seen once during entrainment (vs. 6 times in Experiment 1). As in Experiment 1, performance on the surprise memory test demonstrated that speakers had better memory for past referents than did listeners. In addition, naming (rather than locating) past referents boosted speakers' but not listeners' memory for contrast items. Lastly, memory was improved for items that were described at test with modified noun phrases, demonstrating a link between referential language and memory for the discourse.

4. Experiment 3

Experiments 1 and 2 showed that naming is not necessary to elicit differentiation. Speakers lexically differentiate two entities not to avoid giving the same label to two different entities, but to distinguish current from past referents. In the current experiment, we seek lexical differentiation in a novel situation where, during entrainment, the local context contains an object from the same basic level object category as the intended referent. Speakers are highly sensitive to the contents of the local discourse context (Olson, 1970; Osgood, 1971), and should produce a modified noun phrase to distinguish the intended referent from the contrasting context object, e.g., "*the argyle sock*", in a context with both an argyle sock and a striped sock. On test trials, speakers must describe a third exemplar from the same category (dotted sock) that is unique within the local context (as in Experiments 1–2). If speakers differentiate current referents from previous referents that are in the same basic object category, we would expect speakers to produce significantly more modified noun phrases in the Contrast-naming condition compared to the Non-contrast condition, despite the fact that the bare label "*the sock*" had not previously been used (because the contrast object would have been described with a modified expression like "*the argyle sock*" during entrainment). Such a finding would add to the evidence that speakers differentiate current from past referents, not past labels.

In addition, the use of context objects during the entrainment trials allows us to further explore the speaker advantage in memory for the discourse history. A question not addressed by Experiments 1 and 2 is whether speakers have superior memory for undiscussed aspects of the referential context as well. We address this question by comparing speakers' and listeners' memory for context items that were unmentioned in the conversation, but that were relevant to the way the intended referent was described.

4.1. Method

4.1.1. Participants

Ninety-six undergraduates (forty-eight pairs) at the University of Illinois at Urbana-Champaign participated in return for partial course credit or cash payment ($8). Participants were native speakers of North American English and had normal hearing and normal or corrected-to-normal vision. No participant had previously taken part in Experiments 1 or 2.

4.1.2. Materials and procedure

The general procedure of Experiment 3 was identical to Experiment 2. Pairs of participants completed a referential communication task during which they alternated between the roles of speaker and listener in each block, followed by a 20-min break and then a surprise memory test.

As before, the three types of entrainment trials (Non-contrast, Contrast-location, and Contrast-naming) formed our conditions of interest. One small change to the format of the entrainment trials was made in order to test speakers' and listeners' memory for previously unmentioned items from the discourse context. Unlike Experiments 1 and 2, we added a contrasting context item (e.g., striped sock) to the entrainment trials (see Fig. 7a). In the entrainment trials of the Contrast-location and Contrast-naming conditions, participants viewed two items from the same category (e.g., argyle sock and striped sock), as well as two unrelated items. Speakers referred to one of these two items (e.g., argyle sock) during the entrainment trial; the other object (e.g., striped sock) was never mentioned. Because two items from the same basic object category were in the immediate context, we expected speakers to use modifiers during the entrainment phase in the Contrast-naming condition (e.g., "argyle sock", rather than "sock"). For consistency in terminology across the experiments, we will refer to the mentioned object as the "contrast item", and the unmentioned object from the same basic object category as the "context item". As in Experiments 1–2, in the Non-contrast condition, neither the contrast nor the context object was shown to participants, and speakers named an unrelated object during entrainment (e.g., apple).

Following the communication task, participants performed the same filler task as in the previous experiments for 20 min and then completed an unexpected memory test. Unlike Experiments 1 and 2, we used a yes/no recognition memory test, in which participants

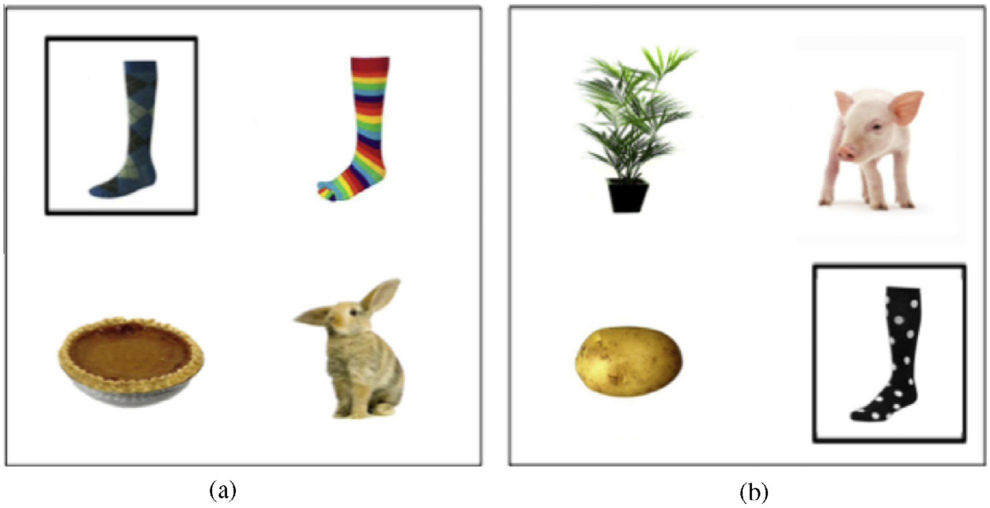


Fig. 7. (a and b) Experiment 3: Example stimuli from entrainment trials (a) and test trials (b). Note that the context item (e.g., striped sock) in entrainment trials was never mentioned.

made a recognition judgment about objects one at a time. In the two-alternative forced choice task used in Experiments 1 and 2, each memory test trial contained an old object; thus participants were forced to choose the more familiar picture even in when they were highly uncertain. In contrast, the yes/no recognition task used in Experiment 3 allows participants to reject pictures (and say "new") in cases where they fail to recognize an item. On each trial, there was a single picture on the screen and participants were instructed to press the "Y" key if the picture was an old one that they had seen during the communication task and to press "N" if the picture was new. Participants completed 264 recognition test trials. Half of the pictures were old items and the other half were new. The old items included 36 contrast (e.g., argyle sock or apple), 36 target (e.g., dotted sock), 24 context (e.g., striped sock), and 36 filler items. The 132 new objects were drawn from the same category as each old item (e.g., three new socks). Thus, during the memory test, participants were exposed to 6 different items from the same category; three were old and the three new. The order of test trials was random.

4.1.3. Predictions

If speakers lexically differentiate in order to distinguish the current referent from past referents, the modification rate should be higher in the two contrast conditions compared to the Non-contrast condition, despite the fact that the unmodified basic noun was not used in entrainment.

With respect to memory for the discourse, we expected that speakers would continue to exhibit better performance than listeners for both target and contrast items, and that contrast memory would not predict differentiation. The new question is whether this speaking benefit to memory for past referents extends to unmentioned items in the visual context. In the list-learning literature, retrieval practice of word pairs enhances memory for original spatial location of the word pairs, suggesting that the benefits of retrieval extend beyond the retrieved information itself (Stanley & Benjamin, in preparation). If successful reference involves consideration of the referential alternatives in the discourse context (Olson, 1970; Pechmann, 1989; see also Fraundorf, Benjamin, & Watson, 2013; Fraundorf, Watson, & Benjamin, 2010), then speakers should have better memory for relevant items in the local context. However, some findings in the memory literature suggest that *listeners* should have better context memory than speakers (Gopie & MacLeod, 2009; Jurica & Shimamura, 1999; cf. McKinley et al., in preparation). According to this view, the act of speaking puts the speaker's attentional focus on the referent, at the expense of attention to the context (Koriat, Ben-Zur, & Druch, 1991). By contrast, listeners distribute attention more broadly, supporting better context memory.

4.2. Results

4.2.1. Communication task

Speakers' descriptions of the referents were coded in the same way as in Experiments 1 and 2. Speakers almost always used modifiers (99.5%) to identify the target on entrainment trials in the Contrast-naming condition, but not in the Non-contrast condition (0.03%). On the remaining 0.5% of entrainment trials in the Contrast-naming condition, speakers uniquely identified the intended referent using a subordinate bare label, such as "*parrot*".

The measure of differentiation was the same as that in Experiments 1–2, and was defined as the use of a modified noun phrase on test trials. On test trials in the Non-contrast condition, speakers modified their expressions 16.3% of the time, while modification rates were higher in both the Contrast-location (20.2%) and Contrast-naming conditions (24.7%), consistent with lexical differentiation (Fig. 8). Modification rates were analyzed in a mixed effects model with entrainment type as a fixed effect (see Appendix, Table 9A). A main effect of entrainment type (Non-contrast vs. Contrast-location & Contrast-naming: $z = 3.170$, $p < 0.05$), was due to significantly higher rates of modified noun phrases in both contrast conditions compared to the Non-contrast condition. Consistent with the results of Experiments 1–2, the modification rate did not differ significantly between the two contrast conditions ($z = 1.694$, $p > 0.05$).

4.2.2. Memory task

Performance on the memory task is plotted in Fig. 9 in terms of the participants' ability to discriminate old from new items, or d′. Discriminability (d′) was calculated by subtracting the standardized false alarm rate from the standardized hit rate. The use of the d′ measure is preferred over other accuracy measures because it allows us to partial out the effects of response bias. The data were analyzed in three mixed effects models, which examined memory separately for target, contrast, and context items (note that an omnibus model with all three object types was not possible because the Non-contrast condition did not contain a context item). In all models, role (speaker vs. listener), entrainment (Non-contrast, Contrast-naming, Contrast-location) and item type

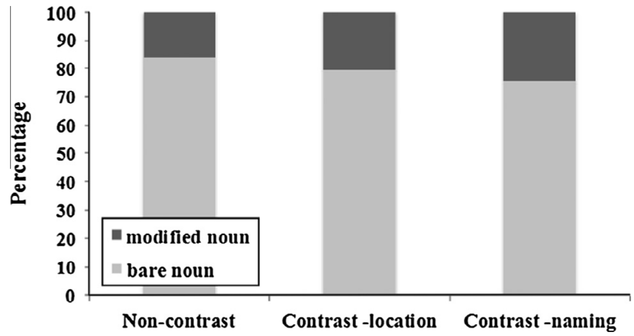


Fig. 8. Percentage of each noun phrase type on test trials in Experiment 3.

(old vs. new) were included as fixed effects. Item type (old vs. new) was included as a fixed effect in order to separate correct acceptance of old items and correct rejections of new items from response bias (see Fraundorf et al., 2010; Wright, Horry, & Skagerberg, 2008). Significant effects of item type show that participants were highly successful at discriminating old from new items; condition by item type interactions test whether condition influenced sensitivity to this distinction between old and new items. The dependent measure was binary; it coded whether the response on the memory test was 'yes' (old) or 'no' (new) (See Appendix for full model details).

For target items (Table 10A), speakers had better memory than listeners ($z = -2.961$, $p < 0.01$), consistent with Experiment 2. Target memory was also better in the Non-contrast condition than the other two conditions ($z = 7.113$, $p < 0.01$), possibly due to confusion created by exposure to multiple items from the same category in the contrast conditions (i.e., Benjamin, 2001; Roediger & McDermott, 1995).

For contrast items (Table 11A), speakers had better memory than listeners ($z = -3.503$, $p < 0.05$). In addition, memory performance was substantially worse in the Contrast-location condition compared to the other two conditions ($z = -9.618$, $p < 0.05$). These effects were qualified by a three-way interaction between item type, role (speaker vs. listener), and entrainment type ($z = 2.046$, $p < 0.05$): Speakers had better memory than listeners when the contrast had been labeled (Non-contrast: $z = -3.500$, $p < 0.05$; Contrast-naming: $z = -3.773$, $p < 0.05$), but not when it had been located (Contrast-location: $z = -0.812$, $p > 0.05$).

For context items (Table 12A), which were a newly included item category in Experiment 3, the analysis only included the two entrainment conditions for which a context item was present during entrainment (Contrast-naming and Contrast-location). Unlike target and contrast memory, speakers did not remember context items better ($z = -0.173$, $p > 0.05$). While context item memory was better in the Contrast-naming condition than in the Contrast-location condition ($z = -9.162$, $p < 0.05$), this effect did not differ as a function of participant role ($z = -0.82$, $p = 0.42$). Thus, naming the referent improved context memory for both speakers and listeners. While speakers exhibit superior memory for past referents, this benefit does not extend to memory for unmentioned context items.

4.2.3. Relationship between referential form and memory

As in Experiments 1 and 2, the relationship between referential form during the communication task and memory for past referents was examined. Consistent with the results of Experiments 1–2, the lexical differentiation rate was not related to contrast memory ($z = -0.579$, $p > 0.05$). Consistent with Experiment 2, both speakers and listeners remembered target items better when the target had been described with a modifier than when it had been described with a bare noun phrase ($z = -2.026$, $p < 0.05$).

4.3. Summary

Experiment 3 demonstrated a clear, but small differentiation effect in the communication task that was not linked to naming. In addition, memory for previous discourse referents was better for speakers than for listeners. Though this benefit for past referents did not extend to unmentioned aspects of the discourse context (context items) selectively for speakers, *naming* (rather than locating) an intended referent did boost memory for unmentioned aspects of the discourse context for both speakers and listeners.

5. General discussion

In the present research, we used the lexical differentiation effect as a test case to examine the way in which memory for the discourse history influences the design of referring expressions. In what follows we discuss the implications of our findings for language, memory, and the relationship between the two.

5.1. Lexical differentiation

The way in which we use language is strongly shaped by the contexts of its use, where the context is broadly defined and

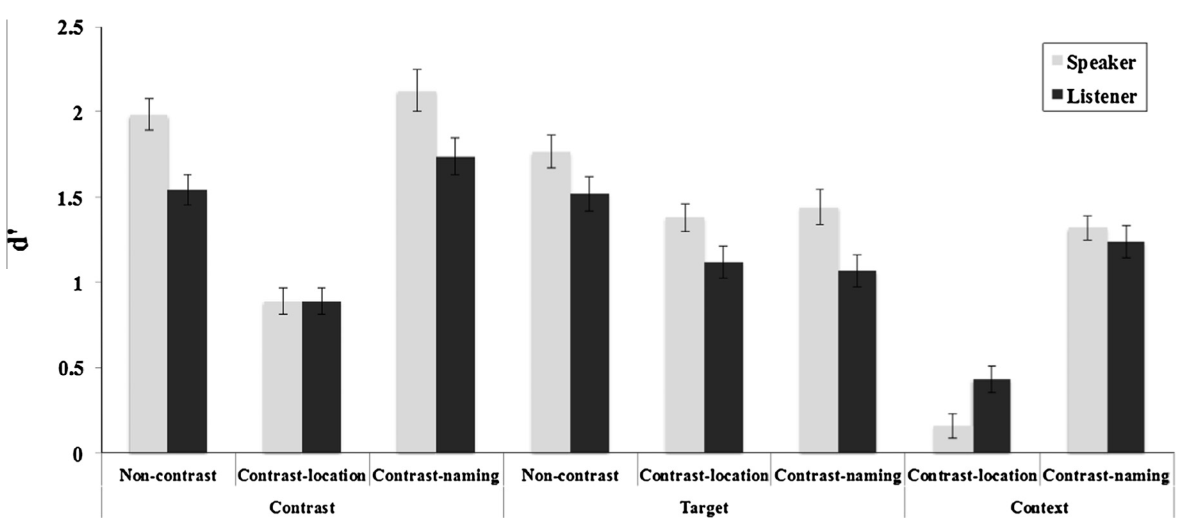


Fig. 9. Discriminability (d′) on the memory test in Experiment 3. Error bars indicate by-participant standard error of the mean.

includes both immediately available information as well as previously experienced information (Brennan & Clark, 1996; Olson, 1970). The starting point for our research was to use the phenomenon of lexical differentiation as a tool to understand how the past and the present are brought to bear on the process of language use in conversation. In particular, we aimed to understand why speakers are so reluctant to differentiate current from past referents. The results of three experiments replicate the lexical differentiation effect and reveal several new phenomena that explain the origins and magnitude of this effect.

One hypothesis about the origin of lexical differentiation is that speakers differentiate in order to avoid using the same label (e.g., *sock*) for two different items (Van der Wege, 2009). To test this hypothesis, we examined a condition in which the speaker used a locative phrase to identify the contrast object (e.g., *the top left one*), thus avoiding the potential for lexical conflict at test. Across the three experiments, however, the differentiation rate did not differ as a function of whether the contrast item had been located or labeled, suggesting that lexical conflict is not the source of differentiation (averaged across the three experiments, the differentiation rate was 21.3% in the Contrast-location vs. 23.4% in the Contrast-naming condition). Instead, our findings show that speakers differentiate in order to distinguish current from past *referents*.

The fact that speakers differentiated equally often in the locative and labeling conditions also clarifies an earlier puzzling finding. Yoon and Brown-Schmidt (2013) showed that listeners were equally likely to expect speakers to produce a modified noun phrase to describe a target referent when a contrasting referent had previously been located vs. labeled. Although they suggested that listeners did not expect speakers to differentiate, the present results offer a different interpretation of those findings. In that study, listeners may have been sensitive to the fact that differentiation can be prompted by previous reference to a similar object, regardless of the form of that reference.

We also demonstrate that lexical differentiation is observed under conditions in which the previous referent (the contrast object) was described with a modified noun phrase. In the Contrast-naming condition of Experiment 3, describing the target with a bare noun phrase would be sufficient to distinguish the current label from previous labels as the contrast item had been described during entrainment with a modifier 99.5% of the time (e.g., *argyle sock*). As a result, an unmodified description of the target (e.g., *sock*, rather than *dotted sock*) would contrast with this earlier label. The fact that speakers produced modifiers on test trials at similar rates in Experiments 2 and 3 (23.7% and 24.7%, respectively) adds to the evidence that it is the distinction between *referents* and not *labels* that gives rise to the lexical differentiation effect.

Why then, is the differentiation rate so low? Memory for the contrast item was excellent for speakers and listeners alike. This finding suggests that the low differentiation rate is not caused by the speaker's failure to remember the contrast. While remembering the historical discourse record is a necessary precondition for the accommodation of current language with respect to it, merely remembering past referents was insufficient to motivate speakers to lexical differentiate. Instead, the low rates of lexical differentiation (about 5% in these experiments) suggests that speakers were more influenced by the local than the historical discourse context, possibly because speakers viewed the historical discourse context as irrelevant. The use of distinct visual scenes on each trial may have resulted in the perception of event boundaries between trials (see Kurby & Zacks, 2008) that discouraged contextual integration across trials. In a conversation in which the past and present were linked through a coherent unfolding event, differentiation might be more pronounced. When the historical discourse context is seen as relevant to the present conversation speakers may also be inclined to design language with respect to the discourse history. More generally, our findings show that considerations of memory alone cannot fully explain utterance design. While access to the relevant memory representations is necessary to design language with respect to mutual knowledge (Horton & Gerrig, 2005), limitations of memory for the discourse history do not explain when speakers will and will not design with respect to the shared discourse history.

A further consideration is that memory was tested using a recognition paradigm in which participants were presented with objects and made judgments about whether they were old or new. An open question is whether the relationship between memory and lexical differentiation would be more pronounced if it was memory for the labels that was probed instead, or if memory were probed in a free-recall task. While the recall rate for conversational memory is generally low (Stafford & Daly, 1984), the act of recalling information from memory may be more similar to what speakers do when they bring to mind past discourse contexts and relate them to the current topic of conversation.

5.2. Distinct memory representations for speakers and listeners

Interlocutors are thought to maintain representations of the discourse context in the form of rich representations of joint experiences (Brown-Schmidt, 2012; Clark & Marshall, 1978), and through the automatic association of partners and referents (Horton & Gerrig, 2005). Here we examined whether speakers and listeners develop distinct memory representations while communicating.

Memory performance was quite good for both speakers and listeners, even in Experiments 2 and 3 where a 20-min break was introduced between the communication task and the unexpected memory test. This finding is generally consistent with the idea that discourse representations are maintained over time, even in a task-based dialogue where there is little pressure to remember previous topics (Brennan & Clark, 1996; Yoon & Brown-Schmidt, in press). However, consistent differences in memory performance between speakers and listeners imply that the representation of the discourse record varies as a function of one's role in a discourse. The fact that the speaker advantage obtained even in Experiments 2 and 3, where participants alternated roles of speaking and listening, suggests that this speaker-benefit operates at the level of individual utterances within the discourse (see also McKinley et al., in preparation). Notably, the speaker benefit did not extend to unmentioned aspects of the discourse context (Experiment 3, context memory), a topic which we return to below.

Why does speaking improve referent memory? One explanation is that speakers invest more effort into the planning of utterances compared to the amount of effort needed to interpret the same utterance. This asymmetric effort explanation is consistent with the idea that the act of generating material increases the mental effort or depth of processing during encoding (Graf, 1980; McFarland, Frey, & Rhodes, 1980; Slamecka & Graf, 1978). Generation is also known to enhance the memory for cued items (e.g., a given word *rapid* when generating a synonym, *fast* in "rapid-f___") as well as generated items (Marsh, 2006; McDaniel & Waddill, 1990). The current findings of speakers' superior memory for past referents are consistent with the memory enhancement for cued items in generation. This view predicts that speakers who invest more effort during encoding should perform better on subsequent memory tests for both cued and generated items. While we have no direct evidence that would speak to this hypothesis, the fact that the use of a modified noun phrase on target trials (e.g., *spotted sock* vs. *sock*) improved subsequent target memory is generally consistent with this idea.

An alternative possibility is that interlocutors develop different strategies to encode information depending on their role as speaker or listener. The act of speaking requires utterance planning and may place more focus on the referent at the expense of attention to the immediate context (see Gopie & MacLeod, 2009; Jurica & Shimamura, 1999; Koriat et al., 1991; cf. Brown-Schmidt & Tanenhaus, 2006; McKinley et al., in preparation). By contrast, listening requires evaluating the unfolding speech signal with respect to the candidate referents in the discourse context (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995), and as a result may improve encoding of both the intended referent as well as the alternatives in the local context. While our findings do show the predicted speaker benefit on this account, there was no evidence that this was at the expense of memory for the context. Similar to recent findings by McKinley et al. (in preparation), speakers and listeners showed equivalent context memory that was improved—regardless of role—when the speaker named the object rather than located it.

5.3. Context memory

It is known that the historical record of a discourse (i.e., memory for the discourse history) includes representations of events and their participants, as well as how past referents have been described (Brennan & Clark, 1996; Greene, Gerrig, McKoon, & Ratcliff, 1994; Nieuwland, Otten, & Van Berkum, 2007). Less clear is whether unmentioned objects in the discourse context are encoded in memory as well, and if so, how they are bound or linked to referenced items. In Experiment 3, when speakers had to use a modified noun phrase to distinguish the target from the context item (Contrast-naming condition), memory for the context items was comparable to memory for targets, and similar for speakers (target $d' = 1.45$; context $d' = 1.32$) and listeners (target $d' = 1.12$; context $d' = 1.23$). By contrast, memory for context items decreased dramatically when the speaker produced a locative (Context-location condition; speaker: context $d' = 0.16$, listener: $d' = 0.43$). These findings show that there is flexibility in how a discourse context is (or is not) encoded. Speakers and listeners do not automatically encode everything in the context, but instead selectively encode information that is conversationally-relevant.

Context in the present studies was defined as the *referential context*. Designing an appropriate referring expression in Experiment 3 required taking into consideration the properties of both the target and the context item in order to select words that would uniquely identify the target. For example, given the scene at entrainment in Fig. 7a, modified expressions such as *the tall sock* or *the multicolored sock* would not suffice to identify the target; instead the speaker would have to take both the target and the context item into consideration to select an expression that uniquely picks out the target such as *the argyle sock*. This joint consideration of an intended referent and its local context may support context memory in natural conversation. Whether other types of contexts (such as memory for one's discourse partner) may be similarly boosted in natural conversation remains an open question. The present research also does not address how unmentioned and irrelevant aspects of the discourse context are encoded in memory. For example, in Fig. 7a, the bunny and the pie were never mentioned and were largely irrelevant to describing the target (sock); whether naming (rather than locating) the target would boost memory for these irrelevant context objects is unknown. If the memory boost for the context object in the naming condition was due to the relevance of the context object to the target, this naming benefit would likely not extend to irrelevant aspects of the context.

6. Conclusion

The present research replicates and extends previous findings of lexical differentiation in conversation, and provide new insights into the origin of this effect. We demonstrate that speakers differentiate in order to distinguish current from past referents; there was no evidence that speakers differentiate current from past labels (cf., Van der Wege, 2009). Yet the differentiation rate was low. By investigating the relationship between language and memory in dialogue, we show that failures to differentiate cannot be explained by a failure to remember the relevant past context. Although remembering the past is necessary to design language with respect to the discourse history, it is not sufficient to elicit differentiation. Further, by measuring memory in conversation, we demonstrate a generation effect for item memory, pointing to an asymmetry in memory between speakers and listeners. The fact that memory for past referents varied on an item-by-item basis depending on a person's role at the time in the conversation, and how that item was described, points to a high degree of flexibility (and variability) in how discourse referents are encoded in conversation. This finding implies that successfully designing utterances based on the knowledge state of one's addressee—the process of *audience design* (Clark & Murphy, 1982)—likely requires conversational partners to appreciate the fact that listener memory may be fleeting. In some cases then, a failure to assume common ground may accurately reflect the listener's comparatively poor memory.

In conclusion, the present research exemplifies how understanding the relevant memory processes involved in dialogue is crucial for fleshing out a theory of how discourse history guides language use. This integrative approach to the study of language use in dialogue with measures of memory for the discourse history represents a key step forward in developing a unified theoretical framework of the cognitive processes underlying language use in dialogue.

Acknowledgments

This material is based on work supported by National Science Foundation Grant NSF BCS 12-57029 to Sarah Brown-Schmidt.

Appendix A

See Tables A1–A12.

Table 1A
Experiment 1 modification rates: Mixed effect model with entrainment type during the communication task as a fixed effect. The dependent measure is binary – whether the expression on the test trial was modified or not. (NC: Non-contrast, CN: Contrast-naming, CL: Contrast-location). Values in bold indicate significant results.

	Estimate	SE	*z*-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	−2.047	0.278	−7.362	<0.001	*Subject*	(Intercept)	1.371	1.171
Entrainment 1 (NC vs. CN & CL)	0.439	0.120	**2.194**	**0.028**		Entrainment 1	0.103	0.320
Entrainment 2 (CN vs. CL)	0.014	0.211	0.067	0.946		Entrainment 2	0.011	0.107
					Trial	(Intercept)	0.959	0.979
						Entrainment 1	0.009	0.094
						Entrainment 2	0.059	0.243

Table 2A
Experiment 1 memory performance: Mixed effect model with role, referent, and entrainment type as fixed effects. The dependent measure is binary - whether the response on the memory test was correct or not. Values in bold indicate significant results.

	Estimate	SE	*z*-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	3.560	0.157	22.708	<0.0001	*Subject*	(Intercept)	0.662	0.814
Role (Speaker vs. Listener)	0.775	0.264	**2.935**	**0.003**		Referent	1.075	1.037
Referent (Target vs. Contrast)	−1.827	0.274	**−6.658**	**<0.0001**		Entrainment 1	0.119	0.345
Entrainment 1 (NC vs. CN & CL)	0.172	0.181	0.950	0.342		Entrainment 2	0.005	0.072
Entrainment 2 (CN vs. CL)	0.290	0.207	1.401	0.161	*Trial*	(Intercept)	0.183	0.427
Role * Referent	0.691	0.432	1.600	0.110		Referent	0.904	0.951
Role * Entrainment 1	0.100	0.343	0.291	0.771		Entrainment 1	0.147	0.383
Role * Entrainment 2	0.328	0.410	0.799	0.425		Entrainment 2	0.017	0.130
Referent * Entrainment 1	−0.564	0.351	−1.605	0.108				
Referent * Entrainment 2	−0.537	0.413	−1.299	0.194				
Role * Referent * Entrainment 1	−1.328	0.666	**−1.993**	**0.046**				
Role * Referent * Entrainment 2	−0.380	0.820	−0.463	0.643				

Table 3A
Experiment 1 contrast memory: Mixed effect model with role and entrainment type as fixed effects. The dependent measure is binary – whether the contrast item was correctly recognized during the memory test. Values in bold indicate significant results.

	Estimate	SE	*z*-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	4.816	0.286	16.827	<0.0001	*Subject*	(Intercept)	1.782	1.335
Role (Speaker vs. Listener)	0.503	0.474	1.062	0.288		Entrainment 1	0.180	0.424
Entrainment 1 (NC vs. CN & CL)	0.318	0.416	0.765	0.444		Entrainment 2	0.664	0.815
Entrainment 2 (CN vs. CL)	1.197	0.400	**2.994**	**0.003**	*Trial*	(Intercept)	0.654	0.808
Role * Entrainment 1	0.768	0.610	1.260	0.208		Entrainment 1	2.183	1.477
Role * Entrainment 2	0.532	0.795	0.670	0.503		Entrainment 2	0.020	0.141

Table 4A
Experiment 1 target memory: Mixed effect model with role and entrainment type as fixed effects. The dependent measure is binary – whether the target item was correctly recognized during the memory test. Values in bold indicate significant results.

	Estimate	SE	*z*-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	2.654	0.139	19.038	<0.0001	*Subject*	(Intercept)	0.410	0.641
Role (Speaker vs. Listener)	1.157	0.227	**5.090**	**<0.0001**		Entrainment 1	0.200	0.447
Entrainment 1 (NC vs. CN & CL)	−0.207	0.190	−1.092	0.275		Entrainment 2	0.027	0.165
Entrainment 2 (CN vs. CL)	−0.004	0.195	−0.021	0.983	*Trial*	(Intercept)	0.209	0.456
Role * Entrainment 1	−0.607	0.379	−1.602	0.109		Entrainment 1	0.004	0.063
Role * Entrainment 2	0.114	0.391	0.292	0.770		Entrainment 2	0.001	0.034

Table 5A
Experiment 2 modification rates: Mixed effect model with entrainment type during the communication task as fixed effects. The dependent measure is binary – whether the expression on the test trial was modified or not. (NC: Non-contrast, CN: Contrast-naming, CL: Contrast-location). Values in bold indicate significant results.

	Estimate	SE	*z*-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	−1.821	0.258	−7.052	<0.0001	*Subject*	(Intercept)	1.373	1.172
Entrainment 1 (NC vs. CN & CL)	0.345	0.166	**2.081**	**0.037**		Entrainment 1	0.146	0.383
Entrainment 2 (CN vs. CL)	0.056	0.184	0.303	0.762		Entrainment 2	0.236	0.486
					Trial	(Intercept)	1.137	1.066
						Entrainment 1	0.057	0.238
						Entrainment 2	0.035	0.187

Table 6A
Experiment 2 memory performance: Mixed effect model with role, referent, and entrainment type as fixed effects. The dependent measure is binary – whether the response on the memory test was correct or not. Values in bold indicate significant results.

	Estimate	SE	*z*-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	2.746	0.101	27.169	<0.0001	*Subject*	(Intercept)	0.376	0.613
Role (Speaker vs. Listener)	0.577	0.138	**4.181**	**<0.0001**		Role	0.730	0.854
Referent (Target vs. Contrast)	0.248	0.162	1.527	0.127		Referent	0.196	0.443
Entrainment 1 (NC vs. CN & CL)	−0.403	0.158	**−2.552**	**0.011**		Entrainment 1	0.147	0.384
Entrainment 2 (CN vs. CL)	0.473	0.129	**3.677**	**0.000**		Entrainment 2	0.314	0.561
Role * Referent	0.341	0.203	1.677	0.093	*Trial*	(Intercept)	0.186	0.431
Role * Entrainment 1	−0.366	0.202	−1.815	0.070		Role	0.139	0.373
Role * Entrainment 2	0.399	0.213	1.873	0.061		Referent	0.304	0.551
Referent * Entrainment 1	−0.103	0.304	−0.339	0.735		Entrainment 1	0.862	0.928
Referent * Entrainment 2	−0.543	0.225	**−2.413**	**0.016**		Entrainment 2	0.085	0.291
Role * Referent * Entrainment 1	0.439	0.399	1.100	0.271				
Role * Referent * Entrainment 2	−0.693	0.421	−1.646	0.099				

Table 7A
Experiment 2 target memory: Mixed effect model with role and entrainment type as fixed effects. The dependent measure is binary – whether the target item was correctly recognized during the memory test. Values in bold indicate significant results.

	Estimate	SE	z-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	2.955	0.121	24.338	<0.0001	*Subject*	(Intercept)	0.342	0.584
Role	0.789	0.205	**3.842**	**<0.001**		Role	1.328	1.152
Entrainment 1 (NC vs. CN & CL)	−0.328	0.185	−1.780	0.075		Entrainment 1	0.250	0.450
Entrainment 2 (CN vs. CL)	0.139	0.207	0.671	0.502		Entrainment 2	0.888	0.942
Role ∗ Entrainment 1	−0.453	0.303	−1.495	0.135	*Trial*	(Intercept)	0.201	0.448
Role ∗ Entrainment 2	0.053	0.321	0.167	0.867		Referent	0.244	0.493
						Entrainment 1	0.283	0.532
						Entrainment 2	0.249	0.499

Table 8A
Experiment 2 contrast memory: Mixed effect model with role and entrainment type as fixed effects. The dependent measure is binary – whether the contrast item was correctly recognized during the memory test. Values in bold indicate significant results.

	Estimate	SE	z-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	2.756	0.141	19.587	<0.0001	*Subject*	(Intercept)	0.551	0.742
Role	0.444	0.181	**2.456**	**0.014**		Role	1.084	1.041
Entrainment 1 (NC vs. CN & CL)	−0.222	0.261	−0.850	0.395		Entrainment 1	0.008	0.090
Entrainment 2 (CN vs. CL)	0.803	0.223	**3.605**	**<0.001**		Entrainment 2	1.586	1.260
Role ∗ Entrainment 1	−0.470	0.276	−1.707	0.088	*Trial*	(Intercept)	0.294	0.542
Role ∗ Entrainment 2	1.017	0.309	**3.292**	**0.001**		Role	0.122	0.349
						Entrainment 1	1.619	1.272
						Entrainment 2	0.183	0.428

Table 9A
Experiment 3 modification rates: Mixed effect model with entrainment type as a fixed effect. The dependent measure is binary – whether the expression on the test trial was modified or not. (NC: Non-contrast, CN: Contrast-naming, CL: Contrast-location). Values in bold indicate significant results.

	Estimate	SE	z-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept)	−1.792	0.211	−8.496	<0.0001	*Subject*	(Intercept)	0.888	0.942
Entrainment 1 (NC vs. CN & CL)	0.643	0.203	**3.170**	**0.002**		Entrainment 1	0.038	0.195
Entrainment 2 (CN vs. CL)	0.329	0.194	1.694	0.090		Entrainment 2	0.013	0.991
					Trial	(Intercept)	0.982	0.991
						Entrainment 1	0.017	0.131
						Entrainment 2	0.008	0.091

Table 10A
Experiment 3 target memory: Mixed effect model with role, entrainment, and item type as fixed effects. The dependent measure is binary – whether the response on the memory test was 'yes' (old) or 'no' (new). Values in bold indicate significant results.

	Estimate	SE	z-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept) (response bias)	0.647	0.088	7.334	<0.0001	*Subject*	(Intercept)	0.385	0.621
(effect on response bias)						Role	0.001	0.033
Role	−0.185	0.064	−2.888	0.004		Item	1.077	1.038
Entrainment 1 (NC vs. CN & CL)	0.116	0.086	1.346	0.178	*Trial*	(Intercept)	0.113	0.337
Entrainment 2 (CN vs. CL)	−0.042	0.085	−0.490	0.624		Role	0.003	0.052
Role ∗ Entrainment 1	−0.048	0.142	−0.337	0.736		Item	0.726	0.852
Role ∗ Entrainment 2	0.030	0.150	0.197	0.843				
Item (old vs. new) (sensitivitiy)	−2.802	0.162	**−17.267**	**<0.0001**				
(effect on sensitivity)								
Item ∗ Role	−0.381	0.129	**−2.961**	**0.003**				
Item ∗ Entrainment 1	1.222	0.172	**7.113**	**<0.0001**				
Item ∗ Entrainment 2	−0.004	0.170	−0.021	0.983				
Item ∗ Role ∗ Entrainment 1	−0.167	0.284	−0.587	0.557				
Item ∗ Role ∗ Entrainment 2	−0.389	0.300	−1.297	0.195				

Table 11A
Experiment 3 contrast memory: Mixed effect model with role, entrainment, and item type as fixed effects. The dependent measure is binary – whether the response on the memory test was 'yes' (old) or 'no' (new). Values in bold indicate significant results.

	Estimate	SE	z-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept) (response bias)	0.491	0.088	5.564	<0.0001	*Subject*	(Intercept)	0.349	0.590
(effect on response bias)						Role	0.011	0.103
Role	−0.099	0.069	−1.441	0.149		Item	1.056	1.028
Entrainment 1 (NC vs. CN & CL)	0.256	0.087	2.932	0.003	*Trial*	(Intercept)	0.194	0.440
Entrainment 2 (CN vs. CL)	−0.908	0.091	−9.948	<0.0001		Role	0.017	0.130
Role * Entrainment 1	0.158	0.142	1.112	0.266		Item	0.528	0.726
Role * Entrainment 2	−0.446	0.161	−2.771	0.006				
Item (old vs. new) (sensitivity)	−2.994	0.166	**−18.018**	**<0.0001**				
(effect on sensitivity)								
Item * Role	−0.477	0.136	**−3.503**	**<0.001**				
Item * Entrainment 1	0.690	0.173	**3.964**	**<0.0001**				
Item * Entrainment 2	−1.754	0.182	**−9.618**	**<0.0001**				
Item * Role * Entrainment 1	0.580	0.283	**2.046**	**0.041**				
Item * Role * Entrainment 2	−0.550	0.322	−1.706	0.088				

Table 12A
Experiment 3 context memory: Mixed effect model with role, entrainment, and item type as fixed effects. The dependent measure is binary – whether the response on the memory test was 'yes' (old) or 'no' (new). Values in bold indicate significant results.

	Estimate	SE	z-value	Pr(>|z|)			Variance	Std. Dev.
Fixed					*Random*			
(intercept) (response bias)	1.032	0.088	11.789	<0.0001	*Subject*	(Intercept)	0.360	0.599
(effect on response bias)						Role	0.020	0.143
Role	0.287	0.082	3.518	<0.001		Item	0.147	0.383
Entrainment (CN vs. CL)	−0.726	0.085	−8.508	<0.0001	*Trial*	(Intercept)	0.098	0.313
Role * Entrainment	−0.380	0.153	−2.486	0.013		Role	0.062	0.249
						Item	0.488	0.699
Item (old vs. new) (sensitivity)	−1.726	0.131	**−13.189**	**<0.0001**				
(effect on sensitivity)								
Item * Role	−0.028	0.161	−0.173	0.862				
Item * Entrainment	−1.540	0.168	**−9.162**	**<0.0001**				
Item * Role * Entrainment	−0.249	0.306	−0.815	0.415				

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.05.011.

References

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.

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 941–947.

Beun, R. J., & Cremers, A. H. M. (1998). Object reference in a shared domain of conversation. *Pragmatics & Cognition, 6*, 121–152.

Bradshaw, G. L., & Anderson, J. R. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal learning and Verbal Behavior, 21*, 165–174.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22*, 482–493.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes, 27*, 62–89.

Brown-Schmidt, S., & Konopka, A. E. (2011). Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information, 2*, 302–326.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language, 54*, 592–609.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55*, 767–811. Reprinted in: F. Katamba (Ed.), Critical concepts in linguistics, Vol. 5: Morphology: Its relation to semantics and the lexicon. London: Routledge, 2003. pp. 128–183.

Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.). *Theoretical issues in natural language processing* (Vol. 2, pp. 57–63). New York: Association for Computing Machinery.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. L. Ny & W. Kintsch (Eds.), *Language and comprehension*. New York: North Holland.

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

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The use of definite reference signals declarative memory: Evidence from patients with hippocampal amnesia. *Psychological Science, 22*, 666–673.

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 Psycholinguistics Research, 24*, 409–436.

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

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

Fraundorf, S. H., Benjamin, A. S., & Watson, D. G. (2013). What happened (and what didn't): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language, 69*, 196–227.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language, 63*, 367–386.

Gopie, N., & MacLeod, C. M. (2009). Destination memory: Stop me if I've told you this before. *Psychological Science, 20*, 1492–1499.

Graf, P. (1980). Two consequences of generating: Increased inter and intra word organization of sentences. *Journal of Verbal Learning and Verbal Behavior, 19*, 316–327.

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language, 33*, 511–526.

Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1–35.

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

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition, 27*, 648–656.

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin and Review, 21*, 1590–1599.

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and output events in memory. *Psychological Research, 53*, 260–270.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343–346.
Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*, 72–79.
Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes, 37*, 117–144.
Marsh, E. J. (2006). When does generation enhance memory for location? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1216–1220.
Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition, 29*, 798–805.
McDaniel, M. A., & Waddill, P. J. (1990). Generation effects for context words: Implications for item-specific and multifactor theories. *Journal of Memory and Language, 29*, 201–211.
McFarland, C., Frey, T., & Rhodes, D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior, 19*, 210–225.
McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2016). Common ground in conversation: Evidence from item and context memory (in preparation).
Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329–336.
Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, 228–236.
Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77*, 257–273.
Osgood, C. E. (1971). Where do sentences come from? In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*. New York: Cambridge University Press.
Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27*, 89–110.
Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguistics and Philosophy, 26*, 287–350.
Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 803–814.
Sedivy, J. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32*, 3–23.
Sedivy, J. (2005). Evaluating explanations for referential context effects: Evidence for Grecian mechanisms in online language interpretation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language as product and language as action traditions* (pp. 153–171). Cambridge, MA: MIT Press.
Shepard, R. N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4*, 592–604.
Stafford, L., & Daly, J. A. (1984). Conversational memory: The effects of recall mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 10*, 379–402.
Stanley, S. & Benjamin, A. S. (2016). Testing improves memory for context (in preparation).
Van der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language, 60*, 448–463.
Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 1466–1481.
Wright, D. B., Horry, R., & Skagerberg, E. M. (2008). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods, 41*, 257–267.
Yoon, S. & Brown-Schmidt, S. (in press). The influence of the historical discourse record on language processing in dialogue. *Discourse Processes*. http://dx.doi.org/10.1080/0163853X.2016.1193429.
Yoon, S., & Brown-Schmidt, S. (2013). Lexical differentiation effects in language production and comprehension. *Journal of Memory and Language, 69*, 397–416.