LEXICAL DIFFERENTIATION IN LANGUAGE PRODUCTION AND COMPREHENSION

BY

SI ON YOON

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Master's Committee:

Assistant Professor Sarah Brown-Schmidt, Chair
Professor Gary Dell

Abstract

This paper presents the results of three experiments that explore the breadth of the relevant discourse context in language production and comprehension. Previous evidence from language production suggests the relevant context is quite broad, based on findings that speakers differentiate new discourse referents from similar referents discussed in past contexts (Van Der Wege, 2009). Experiment 1 replicated and extended this "lexical differentiation" effect by demonstrating that speakers used two different mechanisms, modification, and the use of subordinate level nouns, to differentiate current from past referents. In Experiments 2 and 3, we examine whether *addressees* expect speakers to differentiate. The results of these experiments showed no evidence of an expectation for differentiation, for either lexically differentiated modified expressions (Experiment 2), nor for subordinate level nouns (Experiment 3). Taken together, the present findings suggest that the breadth of relevant discourse context differs across language production and comprehension. Speakers show more sensitivity to things they have said before, possibly due to better knowledge of the relevant context. In contrast, listeners have the task of inferring what the speaker believes is the relevant context; this inferential process may be more error-prone.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Reference is a central component of language use and is often accomplished through the use of definite referring expressions such as *the shirt*. Reference occurs with respect to a context and as such, when using a definite referring expression, the speaker must identify the intended referent with respect to the context. Choosing how much information to provide in order to identify the referent with respect to the context thus becomes a critical task that the speaker must accomplish. According to Grice (1975, p.45), these referential processes are facilitated by a mutual expectation on the part of conversational participants to obey the following cooperative principle:

> *Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.*

The present research is concerned with a particular aspect of the cooperative principle called the maxim of *Quantity*, which states that speakers should make their contribution as informative as is necessary without being overinformative. By following these principles, the speaker can ensure successful communication by including just enough information in order to allow an addressee to understand the speaker's intended meaning. Providing too much information would be inefficient, and potentially confusing. Providing too little information would not allow the addressee to grasp the speaker's meaning. Importantly, since the amount of necessary information is constrained by the characteristics of the context, the key question, then, is what constitutes the context?

Consider for example, if a speaker wants to refer to a dress shirt in a context containing a single shirt, a ball and a pair of pants. In such a context, she would likely use the basic level noun "*shirt*" because all other things equal, this basic level noun is the easiest to retrieve and produce, despite the fact that a superordinate noun such as *clothing* would suffice (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Brennan & Clark, 1996). By contrast, in a context that contains a dress shirt and a sleeveless shirt, a more specific term such as *dress shirt* would be required. Thus, the amount of information required to uniquely identify the referent depends on what we will refer to as the immediate referential context (Olson, 1970; Osgood, 1971). Indeed, speakers show striking sensitivity to the informativity of referential expressions given the immediate context (Brennan & Clark, 1996; Beun & Cremers, 1998; Brown-Schmidt & Tanenhaus, 2006), even in cases where evaluating the context requires attention to multiple dimensions of the scene and task at hand (Brown-Schmidt & Tanenhaus, 2008). This sensitivity surfaces both in the production of modifiers (e.g., *the striped shirt*), and subordinate nouns (e.g., *the blouse*), to distinguish the intended referent from the alternatives. While these findings are robust, it bears note that referential form is not always rigidly determined by context, as speakers readily produce unnecessary modifiers in situations where they have not fully analyzed the context (Deutsch & Pechmann, 1982; Pechmann, 1989), and in cases where they have additional communicative goals (Ford & Olson, 1975; Mangold & Pobel, 1988; Brown-Schmidt & Konopka, 2011; Brennan & Clark, 1996).

What is the context?

Given that utterance form depends on the referential alternatives in the context, a key question is what constitutes the context that interlocutors consider while they communicate. A case in point is face-to-face conversation, where despite being in the same environment, each person has different information in their field of view, and comes to the conversation with different background knowledge. According to one view, conversational partners assume that the relevant discourse context is restricted to jointly known information, also known as *common ground* (Clark, 1992, 1996; Clark & Wilkes-Gibbs, 1986; Stalnaker, 1978). Common ground is thought to be composed of information that is, or was, physically co-present, linguistically co-present and culturally co-present for the two interlocutors (Clark & Marshall, 1978; 1981). While extensive evidence suggests that interlocutors represent information about common ground, and use this information to guide language processing (Hanna, Tanenhaus, & Trueswell, 2003; Wilkes-Gibbs & Clark, 1992), the scope of the influence of common ground in on-line language processing is debated (e.g., Keysar, et al., 1998). Further, the ways in which different types of contextual information differentially guide language processes are poorly understood.

Here we focus on contextual changes throughout a conversation, and examine whether previous contexts shape the perception of the immediate context. In particular, we examine the distinction between information in the immediate or *ahistorical* context, as compared to information jointly experienced in previous contexts, or the *historical* context (Brennan & Clark, 1996). Following Brennan and Clark, we suggest that the relevant context for a discourse includes both the immediate context as well as the

relevant historical context. Evidence that interlocutors are sensitive to information in the immediate context is extensive. For example, in a situation in which both a speaker and a listener are able to see a single shirt, the speaker can readily refer to that shirt (e.g., "*Pass me the shirt*.", Clark & Marshall, 1981), without having to contrast her expression against shirts that only the speaker knows about (e.g., the shirts in her closet at home). Indeed, in the absence of external pressures (e.g., time pressure, task demands), in situations such as this one, speakers easily design expressions with respect to jointly visible objects (Horton & Keysar, 1996; Wardlow Lane & Ferreira, 2008; Nadig & Sedivy, 2002), and listeners primarily restrict their domain of interpretation to those objects that are jointly visible (Hanna, et al., 2003; Heller, Grodner, & Tanenhaus, 2008; also see Brown-Schmidt, Gunlogson & Tanenhaus, 2008; cf. Keysar, et al., 1998).

Less well understood is how the joint *historical* context guides language processing decisions. The historical context is thought to include information that was previously, and jointly experienced by the interlocutors, and thus in the interlocutors' common ground (Brennan & Clark, 1996). The historical context includes previously experienced linguistic and visual events, such as jointly created conceptual pacts for how to refer to specific entities (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986), as well as lexically entrained terms (Garrod & Anderson, 1987). Indeed, some recent evidence suggests that addressees are sensitive to common ground for previously-mentioned linguistic information (Hanna, et al., 2003; Brown-Schmidt, 2012; Metzing & Brennan, 2003; Brown-Schmidt, 2009). In what follows, introduce two specific aspects of historical context that are the focus of the present research namely, the process of lexical entrainment and lexical differentiation.

Lexical Entrainment and Lexical Differentiation

In a classic experiment, Brennan and Clark (1996) asked speakers to describe one of multiple objects from a given category (e.g., one of three fish) to an addressee. Through conversational grounding processes (Clark & Schaefer, 1989), partners developed specific, entrained terms (e.g., *the curved round fish*) to distinguish among the category exemplars. Interestingly, when the local context changed to include only a single category exemplar, speakers tended to continue to use the previous entrained expressions even though they overspecified the referent given the immediate context (e.g., speakers continued to say *the curved round fish,* despite the fact that *the fish* would have uniquely identified the referent). This overmodification pattern was large (it occurred at a rate of 25-81%, depending on the number of times and the contexts in which the speaker had previously referred to the object), suggesting that the historical context was powerful enough to override local informativity expectations.

There are at least two mechanisms which could support the sensitivity to historical context. One mechanism, suggested by Brennan and Clark (1996), is that the speakers agreed on a name or *conceptual pact* for how to refer to the entity (Brennan & Clark, 1996). On this account, conversational partners learn shared names for entities they repeatedly refer to, and historical context effects emerge because speakers re-use these learned names when speaking with their same partner (Wilkes-Gibbs & Clark, 1992; Schober & Clark, 1989). Thus, the aspect of the historical discourse record that guides reference is knowledge of how a current referent was previously referred to with the current partner. While the evidence for these shared names or conceptual pacts is

clear, an open question is whether speakers design expressions with respect to other aspects of the historical context, such how *different* entities were previously referred to.

If speakers do take into consideration how different entities were referred to when designing subsequent expressions, this would suggest that referential processes consider an even broader view of the historical record. Thus an alternative, but non-mutually exclusive explanation for Brennan & Clark's finding is that conversational partners encode information about each of the entities in the historical context, and subsequently design referring expressions with respect to all of the entities in both the local and historical context. On this alternative, *context-memory* view, a speaker might persist in calling a given fish *the curved round fish* in a context with a single fish, because the relevant context included both that single fish, as well as other previously-referred to fish. Importantly, such a view is not incompatible with Brennan and Clark's (1996) proposal that partners form conceptual pacts. Instead, the context-memory view would claim that *in addition to* using entrained terms, speakers design referential expressions to contrast with descriptions of previous referents, even when those referents are not present in the immediate context.

Is there any evidence that referents from previous contexts guide subsequent referring? According to the principle of pre-emption by homonymy, a well-established term preempts the use of the same term to signify a new meaning (Clark & Clark, 1979; while these principles were originally applied to denominal verbs, they clearly extend to nouns; Brennan & Clark, 1996; Kronmüller & Barr, 2007). According to this principle, once interlocutors arrive at an entrained term for a given entity, that term should not be used to refer to anything but the original referent. Van Der Wege (2009) identified a

related phenomenon in the production of referring expressions, and termed it *lexical differentiation*. In her experiment, participants described a series of objects in contexts with one typical exemplar and one atypical exemplar of a given category (e.g., a tree with leaves and a tree without leaves). To test for differentiation, speakers first established an entrained term to refer to a unique referent (e.g., a tree in a context that did not contain other trees was referred to as *the tree*) and then subsequently described a different unique referent from the same basic level object category (e.g., a tree without leaves in a context that did not contain other trees). Speakers in live conversation were significantly more likely to describe the second object with a more specific referential expression that only applied to the current target (e.g., *the tree without leaves*), even though a basic level noun would have been sufficient for the immediate context (note that the results were slightly different in non-conversational settings).

This lexical differentiation finding is consistent with the *context-memory* view because it shows that speakers take into consideration how referents from the discourse history were previously described when designing new expressions. However, a complication of these findings is that the lexical differentiation phenomenon was strongly influenced by stimulus type and order: When participants were exposed to the typical item first and then the atypical item later, the percentage of precise reference phrases increased from 7% for typical items to 62% for atypical items, which is consistent with a large lexical differentiation effect. However, when speakers were exposed to atypical items first, the percentage of precise reference phrases *decreased* from 43% for the atypical item to 40% when later referring to the typical item from the same category. While it is the case that within a category speakers were grew more precise from the first

to the second phase (typical: 7% to 40%; atypical: 43% to 62%), it is unclear to what extent the differentiation effect was driven by the contrast in typicality between the two pairs. Thus, while these findings do show that speakers consider the broader historical context (and not just the entrained name for the current referent), an open question is whether this effect is limited to contrasts with atypical category exemplars, or whether it would generalize to contrasting pairs of typical referents.

Historical and Immediate contexts in language comprehension

The studies by Brennan and Clark (1996) and Van Der Wege (2009) focused on the process of language production, and found that speakers tend to maintain precedents they established with their conversational partners. We now turn to the question of the addressee's expectations of the speaker. Do addressees expect speakers to use established precedents even when the local context changes?

The existing research, while limited, suggests that addressees do expect speakers to maintain referential precedents even in situations in which the precedent would overspecify the referent for the current display. Barr and Keysar (2002) presented listeners with a series of pictures and auditory recordings that instructed them to click on one of two pictures in a display. In the baseline phase of the experiment, listeners interpreted basic level nouns such as *"car"* in contexts including, e.g., a sportscar and a carnation. Following this baseline, the same objects were referred to using subordinate level nouns, such as "*carnation*" in the context of a daisy and a carnation, or "*sportscar*" in the context of a station wagon and a sports car. A final set of post-test trials were identical to the pre-test, and included two objects from different categories with entrained

subordinate labels (e.g., a sportscar and a carnation). Critically, the stimuli were designed such that the onset of the basic level term "*car*" phonologically overlapped the subordinate level noun "*carnation*" (i.e., they were cohort competitors, Allopenna, et al., 1998). When interpreting *car,* listeners fixated the subordinate term (e.g., *carnation*) more at the post-test compared to the pre-test, even though the immediate context did not necessitate the subordinate term in either case. However, this entrainment effect did not override the expectation for a basic level term: Even after training, listeners still fixated the basic-level target "car" more than the subordinate-level competitor "carnation". Further, because this entrainment effect relied on a pre-test vs. post-test comparison, it is unclear if the increase in "carnation" fixations at post-test was due to learning of the entrained name, or because participants had clicked on the carnation multiple times during entrainment.

While the experiment by Barr and Keysar (2002) provides initial evidence that listeners maintain expectations for entrained terms across contextual changes, what is not known is whether interpretation processes are sensitive to other aspects of the historical context, such as how *other*, non-present objects were referred to. This is the focus of the present research. Specifically, we ask whether historical conversational precedents for how previous, non-present discourse referents can override quantity expectations (Grice, 1975), and thus render a locally overspecified, but *unentrained* referential expression felicitous.

In three experiments, we examined the question of whether historical expectations can override quantity violations by examining the process of lexical differentiation (Van Der Wege, 2009) and the scope of contexts that afford overmodification. Experiment 1

was designed to test for lexical differentiation in language production for reference to typical category exemplars. In Experiments 2-3, we examined whether addressees generate lexical differentiation expectations, focusing on the interpretation of differentiation in modified expressions (Experiment 2), and in subordinate nouns (Experiment 3).

CHAPTER 2: EXPERIMENT 1

The lexical differentiation results reported by Van Der Wege (2009) suggest that the scope of relevant historical context might be quite broad. Recall that speakers tended to use a more precise term when describing a new exemplar from a previously-described category (e.g., *the tree…. the tree without leaves*), but that the differentiation effect was largely driven by trials on which speakers first described a typical item (e.g., leafy tree) and then later described an atypical item (e.g., tree without leaves). Here we test the generalizability of these findings by testing for the differentiation effect in situations where speakers described two *typical* exemplars from the same category.

Method

Participants

Sixteen undergraduates at the University of Illinois at Urbana-Champaign participated in the experiment as speakers in return for partial course credit or either cash payment ($8). Participants were native speakers of North American English.

Procedure

A participant and an experimenter played a version of a referential communication task (Krauss & Weinheimer, 1966) while sitting at separate computers in the same room. On each trial, the participant and the experimenter viewed a computer screen with four pictures, and the participant gave the experimenter an instruction aloud to click on one of four objects. The experimenter followed the participant's instructions on her own screen. Recordings of the participant's voice were saved directly to disk. The

target object was indicated to the participant with a black box (see Figure 1). Participants were allowed to describe the target object in any way they saw fit except for using locative phrases (locatives, e.g., *the top left one*, instead of *the shirt*, were disallowed because they would not have set up the critical preconditions for differentiation, whereby the basic level object name for the target object was already used for another object). On the experimenter's screen, the target was not indicated. The task was interactive and the experimenter asked questions for clarification as needed (this was uncommon). The experiment lasted approximately 50 minutes.

Figures 1. Experiment 1: Example stimuli from entrainment trials (left) and test trials (right). The target is indicated to the speaker by the black rectangle. This example shows the "contrast" condition; in the "no-contrast" condition, the target during entrainment would be an unrelated item such as a towel.

Materials

Each participant completed a total of 448 trials including 32 target trials, 192 entrainment trials and 224 filler trials (see Figure 2). Experiment 1 consisted of two phases: entrainment and test. In entrainment trials, participants described either the contrast object (e.g., checkered shirt) or an unrelated object (e.g., towel) six times to

establish entrained terms. Following entrainment, test displays had four new objects, including a target object and three unrelated objects. The target was a second exemplar of the same category (e.g., striped shirt) as the contrast. Filler trials contained two contrasting objects from the same category (e.g., two fish) and two unrelated objects. On half of the filler trials, the target was one of two contrast objects and on the other half, the target was one of two unrelated objects. Filler trials with contrasting objects were included so that speakers would sometimes need to produce modified noun phrases. Entrainment and test trials for the different item sets were interspersed with filler trials in a pseudo-random order such that from the participant's 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, 1 test) occurred within a span of at most 100 trials.

The primary manipulation was whether speakers had previously described the contrast object or not, and was manipulated within-subjects. In the contrast condition, participants described the contrast object (e.g., checkered shirt) in entrainment trials and then described the target object (e.g., striped shirt) at test. In the non-contrast condition, participants described an unrelated object (e.g., towel) in entrainment, and then described the target object (e.g., striped shirt) at test. Note that for items in the non-contrast condition, speakers never saw the contrast object.

There were a total of 288 visual stimuli, all of which were pictures of every-day objects. The stimuli included the 32 triplets of target (e.g., checkered shirt), contrast (e.g., striped shirt), and unrelated items (e.g., towel) that were used as critical items in the contrast and no-contrast conditions. The remaining visual stimuli were used on filler trials and included 32 triplets contrasting objects (e.g., opened box, wrapped box, stacked

box) and 64 pairs of contrasting objects (e.g., sitting dog, jumping dog) from the same categories. For filler object triplets, two of the pictures were presented together on filler entrainment trials and the third was subsequently presented alone at test. For filler object pairs, the two objects were presented on separate trials during entrainment and test. The filler trials were designed to create the necessary conditions such that on target trials all four objects were novel—that is, the speaker had not seen or described them before—whereas each of the four objects came from a category that was familiar to participants (see illustration in Figure 2). All items were counterbalanced across conditions across two lists. Each participant completed the items in one list.

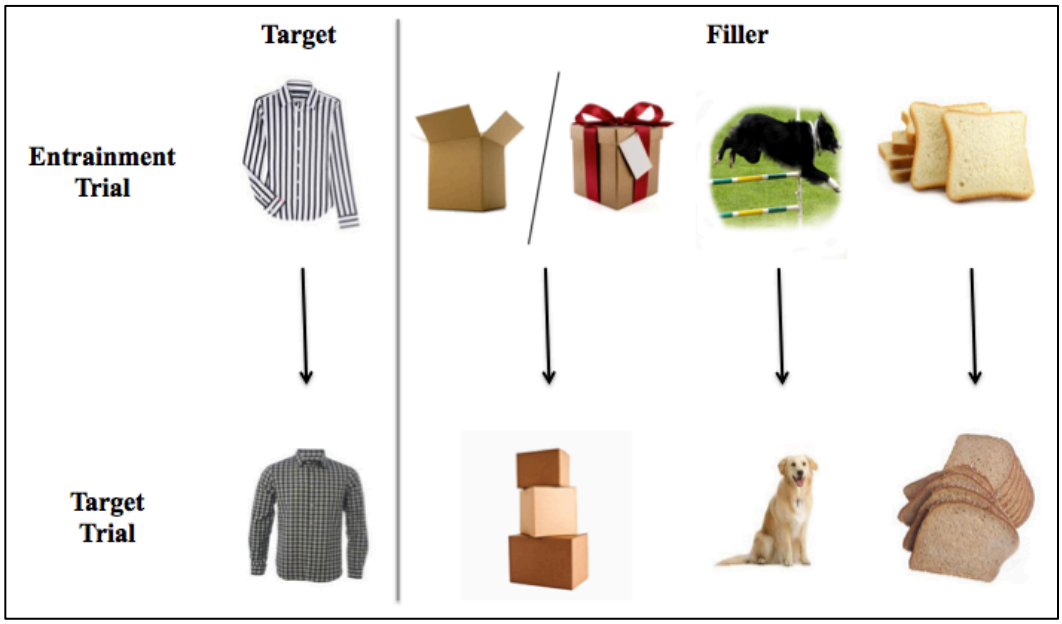

Figure 2. Illustration of how entrainment trials established the necessary pre-conditions for lexical differentiation in Experiments 1 and 2. Upper row: example targets on (four different) entrainment trials. During entrainment, the striped shirt, jumping dog, and white bread would be presented on separate trials, and the two boxes would be presented on the same trial. Bottom row: target trials contained four new exemplars of previously-experienced categories. In this example, the four (previously unseen) images in the bottom row would be arranged on a single target trial as shown in Figure 1(right).

Predictions

If previous reports of lexical differentiation were restricted to atypical category exemplars, speakers should not be sensitive to historical context and instead describe the target with respect to the current display. Thus we would expect speakers to use bare noun phrases both in entrainment and target trials. However, if speakers do lexically differentiate typical exemplars, we would expect speakers to use bare nouns to refer to target objects in entrainment trials, and more specific (differentiated) expressions on target trials.

Analysis and Results

The recordings of the speakers' productions were transcribed and coded for whether they used a bare noun phrase or a modified noun phrase. In listening to the recordings, we noticed that in some cases speakers alternated from one noun to another noun when describing the contrast object and then the target. For example, if they first referred to the contrast object in entrainment trials as *the knife*, in the test trials, speakers sometimes used a different noun, such as *the sword*, to refer to the target. This finding suggests that speakers may differentiate using multiple mechanisms and in some cases prefer an alternate, unmodified term.

In the non-contrast condition, on test trials speakers typically referred to the target with a bare noun phrase (89.9%), and used a modified noun phrase on only 10.1% of trials. In the contrast condition, on target trials speakers used modified noun phrases on 17.6% of trials, and noun phrases that were different than the ones used to refer to the contrast during entrainment on 16.4% of trials (e.g., *knife* → *sword*). While bare noun phrases were common in the contrast condition (66%), the percentage was substantially

lower compared to the non-contrast condition (see Table 1). Thus, we found that speakers lexically differentiate using two mechanisms; modified noun phrases and different bare noun phrases. Note that it was not possible to measure the rate at which different noun phrases were used in the non-contrast condition because target object categories were only named once.

Table 1. Percentage (SE) of each noun phrase type on test trials in Experiment 1.

	Bare noun phrase (e.g., knife)	Modified noun phrase (e.g., fancy knife)	Different bare noun phrase (e.g., knife/sword)
Non-contrast	89.9% (11.05)	10.1% (11.15)	NA
Contrast	66% (18.81)	17.6% (7.26)	16.4% (6.83)

We analyzed the data in a logistic mixed effects model with contrast condition as a fixed effect, and subjects and items as crossed random intercepts (see Table 2). The dependent variable was whether the speaker used a modifier or not (note that this metric does not capture differentiation using a different bare noun). Evaluation of AIC (Akaike, 1974) and the deviance test statistic indicated that adding random slopes for the condition effect did not improve model fit. This model revealed a main effect of contrast (z=2.81, p<.01): Speakers were more likely to use modifiers when they had previously referred to contrast objects (contrast condition) compared to situations where they had not (non-contrast condition).

Table 2. Mixed effect model with contrast (contrast vs. no contrast) as a fixed effect in Exp1.

		Estimate	SE	*z*-value	Pr(>|z|)
(a) overall					
Fixed effects					
(intercept)		-2.7370	0.4005	-6.834	8.28e-12
Contrast		0.8626	0.3067	2.813	**0.00491**
				Variance	Std.Dev.
Random effects					
Groups					
Subject	(Intercept)			0.97742	0.98865
Item	(Intercept)			2.06136	1.43574

In summary, the present findings replicate and extend Van Der Wege (2009), demonstrating lexical differentiation in situations where speakers are sequentially describe two typical exemplars of the same category. Our findings bear resemblance to other findings that speakers are more likely to avoid homophonic ambiguity when they have previously referred to another item (e.g., referring to a baseball bat as "*bat"* reduces subsequent likelihood of using "*bat"* to refer to a flying mammal bat; Ferreira, Slevc, & Rogers, 2005). More generally, our results suggest that when speakers design their referential expressions, they take into account the previous discourse record, even in situations where the local context has changed.

The fact that speakers themselves had previously referred to the contrast object may be a key contributor to their resistance to re-use the critical term for a different referent. Upon viewing the target referent on test trials, the similarity of the production experience to the contrast trials (i.e., describing a different shirt) may evoke the appropriate conditions to elicit differentiation. An open question, then, is whether lexical differentiation would extend to language comprehension. In Experiments 2-3, we address this question by examining whether *addressees* expect speakers to lexically differentiate. We examine the two mechanisms speakers used to differentiate separately, focusing on

interpretation of modified noun phrases in Experiment 2 and different noun phrases in Experiment 3.

CHAPTER 3: EXPERIMENT 2

The lexical differentiation effect in Experiment 1 sets the stage for the investigation of lexical differentiation during language comprehension in Experiment 2. Specifically, we ask whether addressees expect speakers to lexically differentiate their expressions, even when it would result in local overspecification.

According to the historical view of conversation, the relevant discourse record is broad. It leads speakers to maintain entrained precedents (Brennan & Clark, 1996), and lexically differentiate current expressions from previously used expressions (Van Der Wege, 2009). In both cases, entrainment and differentiation result in expressions that are appropriately modified with respect to the historical discourse record, but overmodified or too specific (in the case of subordinate terms) for the immediate context. Given that overmodification can be confusing for the listener (Engelhardt, Bailey, & Ferreira, 2006; Engelhardt, Demiral, & Ferreira, 2011), one might expect that comprehension could be facilitated if listeners took a similarly broad view of the discourse record, within which, such locally overmodified expressions would be appropriately informative.

Further, if, as some have claimed, language production and comprehension draw on the same processes (Pickering & Garrod, 2004, in press; Pickering & Branigan, 1998), the scope of relevant discourse history should be the same in comprehension and production, particularly in interactive conversation where interlocutors are thought to prime one another at multiple levels of representation, including common ground (Pickering & Garrod, 2004). Indeed, some evidence suggests that listeners do take a historical view of the discourse record. For example, listeners show facilitated interpretation of previously entrained referential terms (Metzing & Brennan, 2003;

Brown-Schmidt, 2009; Brennan & Hanna, 2009; Kronmuller & Barr, 2007), and show competition from previously entrained terms even when the local referential context does not require them (Barr & Keysar, 2002). However, these experiments examined situations in which the speaker and listener had established a term for a particular referent and then tested whether the listener expected the speaker to continue using that expression for the very same referent. While these findings offer insight into the learned mappings between specific referents and names, they say less about the scope of the relevant discourse context per se. Indeed, according to some accounts, learned associations between expressions, referents and discourse partners reflect a type of associative priming (Horton, 2007; Horton & Gerrig, 2005). On such a view, these entrainment effects are prompted by the referent, and offer little insight into whether listeners bring to mind contextual entities from previous joint experience when interpreting referential expressions.

If the lexical differentiation phenomenon did extend to language comprehension, it would likely require the listener to recall past discourse contexts, and generate an expectation that the speaker would contrast current referents with previous referents. Unlike entrainment, such a process would require memory for multiple items in the discourse context, as after all, there is no way of predicting which of the previous context objects would be contrasted in future trials.

In order to test the hypothesis that listeners generate lexical differentiation predictions during comprehension, we compared language production and comprehension processes in two situations: one in which the conversational history supports lexical differentiation (i.e., the basic level object term which would ordinarily be

used to refer to the current referent was previously used to refer to different exemplar from same category), and one in which it does not (i.e., the basic level object term which would ordinarily be used to refer to the current referent was not previously used in the discourse). If a speaker's previous use of a basic level object term to refer to one category exemplar results in an expectation for differentiation when referring to other members of the same category, we should observe differentiation-based expectations when the conversational history includes a previous use of that basic level term.

Method

Participants

Thirty-two undergraduates at the University of Illinois at Urbana-Champaign participated in the experiment as listeners in return for either cash ($8) or partial course credit. Participants were native speakers of North American English and had normal or corrected-to-normal hearing and vision. None of the participants in Experiment 2 had previously participated in Experiment 1.

Procedure

A participant and an experimenter sat at separate computers facing each other. Across participants, a total of 5 native English-speaking lab assistants (two female) played the role of experimenter. Lab assistants were unaware of the specific experimental predictions. Each trial began with a fixation cross for 1000ms, then the cross disappeared and 4 different objects appeared on the screen. On each trial, the experimenter's screen showed the same four objects as the participant's screen, as well as a text prompt

indicating what the critical instruction was for that trial (the computer screens were positioned such that the participant was unaware of the text prompts). Once the pictures appeared on their respective computer screens, participants followed the experimenter's instructions to manipulate one of the four objects (e.g., "*Click on the shirt*."). While this instruction was scripted, the participant and experimenter were free to otherwise converse. Thus, if the instructions were not clear enough, participants were encouraged to ask for clarification (this was uncommon).

The participants' eye movements were monitored during the experiment with an Eyelink 1000 desktop mounted eye tracker. It sampled eye movements monocularly at 1000hz and recorded them to disk. The experimenter wore a headset microphone and their voice was recorded directly to disk. The experiment lasted approximately 50 minutes.

Materials

Each participant completed a total of 448 trials including 32 target trials, 192 entrainment trials and 224 filler trials (See Figure 3). As in Experiment 1, this experiment consisted of two phases: entrainment trials and test trials. Filler trials were randomly intermixed between these two phases. In entrainment trials, participants were exposed to the contrast object (e.g., checkered shirt) in the context of three unrelated objects six times. The contrast and the three unrelated objects were equally likely to be referred to during entrainment. Following entrainment, test trials presented the target object and three unrelated objects. As in Experiment 1, all four objects on the test trial were new (i.e., they had not been seen before in the experiment), but the categories of all four

objects were familiar. Test instructions were equally likely to refer to the target (critical trials) as one of the three unrelated objects (filler trials).

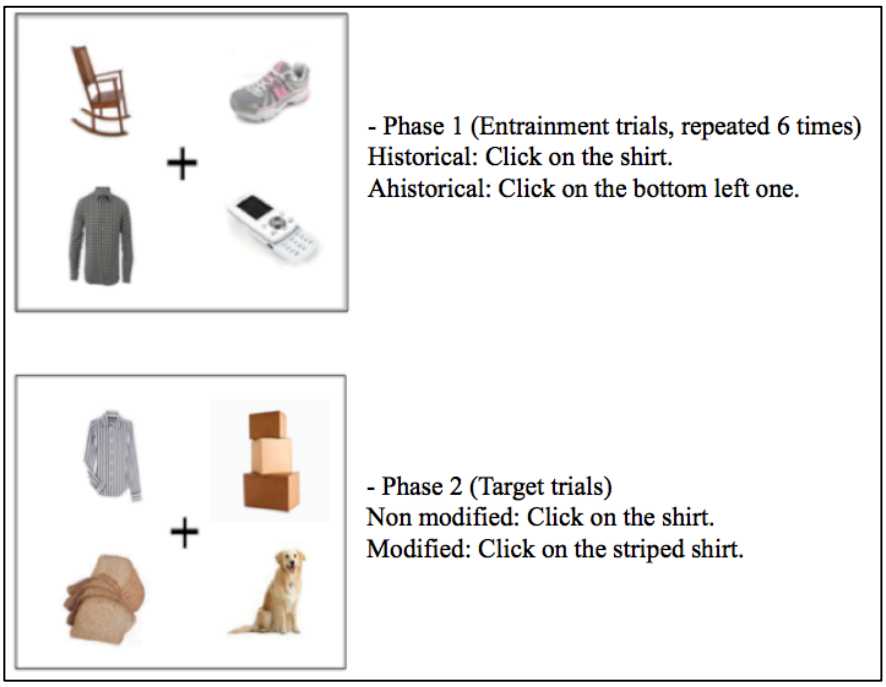

Figure 3. Example stimuli for Experiment 2.

There were total of 7 trials per target item (e.g., "shirt"): 6 entrainment trials and 1 target trial. All trials associated with one target item, from the first entrainment trials to the target trial, occurred within at most 120 trials. Filler trials always presented two contrast objects from the same category (e.g., two birds) and two unrelated objects. On half of the filler trials, the instructions referred to one of the contrasting objects using a modified noun phrase and on the other half of filler trials, instructions with bare noun phrases were given. There were a total of 288 visual stimuli. All stimuli were color photographs of every-day objects. The 32 pairs of target and competitor objects were identical to the critical pairs used in Experiment 1 (e.g., striped shirt and checkered shirt).

The remaining stimuli were fillers, and like the target and competitor objects, were organized into sets. The filler sets were of two types: triplets of contrasting objects (e.g., opened box, wrapped box, stacked box) and pairs of contrasting objects (e.g., sitting dog, jumping dog). These filler stimuli were identical to those presented in Experiment 1 (see Figure 2). Contrast sets (e.g., two boxes or two dogs) were never displayed on the same screen except for on filler trials.

In order to test for lexical differentiation in language comprehension, two factors were manipulated within-subjects: The first factor was whether the basic level object term for the target object (e.g., *shirt*) was previously used to refer to a different token from the same category during entrainment (historical condition), or not (ahistorical condition). In the historical condition, during entrainment, the experimenter consistently described the contrast object with the critical basic level term (e.g., *Click on the shirt*). In the ahistorical condition, during entrainment the experimenter referred to the contrast using a locative phrase (e.g., *Click on the bottom right object*). Importantly, the use of a locative in the ahistorical condition allows us to control for the fact that the contrast object was referred to and was clicked on by the subject, without producing the necessary preconditions for lexical differentiation. The fact that addressees interpreted an instruction to click on the contrast object in this control condition is essential because this would allow us to attribute any interpretation benefit in the historical condition at test, to the fact that the listener expected the speaker to differentiation the target expression, and not because the listener has more experience clicking on objects from that category.

The second factor was whether the test instruction differentiated the target object from the previously-experienced contrast object using a modified noun phrase (e.g., *Click*

on the striped shirt), or not (e.g., *Click on the shirt*). Note that the modified noun phrase was over informative for the current target display, but appropriately informative given the historical context. The critical expressions on test trials were designed following the production patterns in Experiment 1 (recall that the same test images were used), such that the most popular noun used to describe each target images in Experiment 1 was used as the critical noun on test trials in Experiment 2.

The 32 target trials were rotated through the four conditions across four experimental lists. Each participant saw the materials in one list.

Predictions

If language production and comprehension use similar, historical representations of the discourse (Pickering & Garrod, 2004; Pickering & Branigan, 1998; Brennan & Clark, 1996), then listeners in Experiment 2 should show evidence of lexical differentiation-based expectations. If so, interpretation of the modified noun phrase (e.g., *the striped shirt*) should be significantly facilitated in the historical, compared to the ahistorical condition, because the listener would be more likely to expect a differentiated, modified expression in the historical condition compared to the ahistorical condition. In the absence of an expectation for differentiation (ahistorical condition), the modifier would be an unnecessary, and likely to result in comprehension difficulty (e.g., Engelhardt, et al., 2006; Engelhardt, et al., 2011).

A different pattern of results is predicted for the non-modified expressions. When interpreting non-modified expressions on test trials (e.g., *the shirt*), interpretation of the bare noun phrase should be facilitated in the ahistorical condition because listeners would

be more likely to expect an unmodified (non-differentiated) expression. In fact, if listeners in the historical condition strongly expect a modifier, they might make fewer target fixations when interpreting non-modified expressions than in the ahistorical condition, due to the mismatch between their expectation and the unmodified expression.

In contrast, if listeners do *not* generate a lexical differentiation expectation, we would expect an early benefit for the historical condition that is the same for modified and non-modified expressions, due to the previous exposure to the target noun in the historical condition (i.e. a priming effect).

Analysis and Results

We first analyzed the recorded instructions to determine whether there were differences in prosodic form across conditions. Because all instructions were naturally produced live in the context of an experiment, potential differences in prosodic form across conditions could make it difficult to tease apart effects of historical experience from effects of prosody. All instructions produced by the experimenter were transcribed and the critical adjective and noun were identified in Praat (Boersma & Weenink, 2003). The prosodic properties of pitch (minimum, maximum and average), intensity and duration of the critical adjective and noun were analyzed. There were no significant differences in any prosodic properties across conditions (t's<1.00).

We now turn to our primary analysis, which focuses on the eye movements that listeners made as they interpreted the experimenter's descriptions on target trials. The pattern of eye movements on target trials was considerably different depending on whether the participant heard a modified or a non-modified instruction as, after all, the

modifier provided unique information about the target prior to the noun (e.g., Sedivy, et al., 1999). As a result, our planned comparisons analyzed modified and non-modified noun phrases separately (see Figures 4-5).

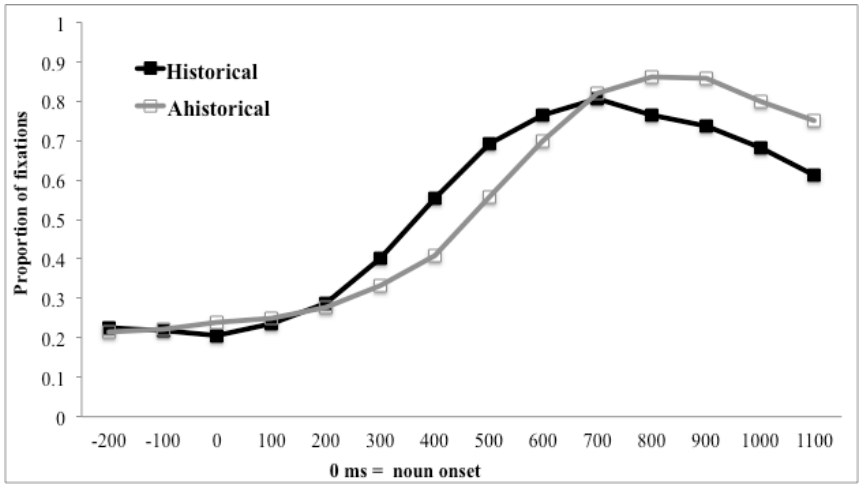

Figure 4. The proportion of target fixations following noun onset for non-modified instructions in Experiment 2.

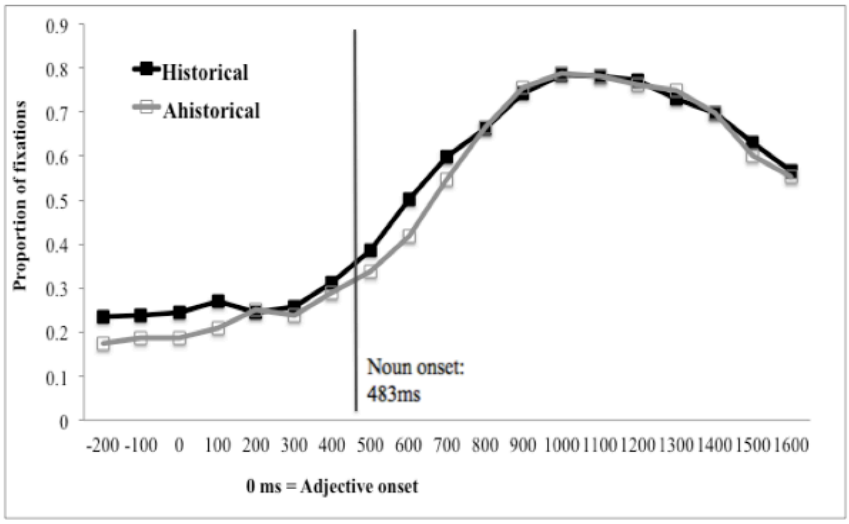

Figure 5. The proportion of target fixations following noun onset for modified instructions in Experiment 2.

Non-modified expressions

Fixations to the target started to increase approximately 200-300ms following noun onset. Contrary to predictions, when listeners had previously interpreted the non-modified expression as referring to the contrast object (historical condition), there was no penalty for subsequently interpreting a non-modified reference to the target using the same term. Eye movements for non-modified expressions were analyzed in two 500ms analysis regions (see Table 3); regions were planned in advance to capture early and late comprehension effects. The first 500ms time-window was from 100ms to 600ms after noun onset and the second was from 600ms to 1100ms after noun onset. The first window was offset by 100ms given that it takes some time to program and launch an eye movement (see Altmann, 2011; Hallet, 1986).

We analyzed the proportion of fixations to the target in a series of mixed effects models with time-window and historical condition as fixed effects, coded with orthogonal contrast codes (see Table 3).

Table 3. Mixed effect model with time-window and historical condition as fixed effects for non-modified expression in Exp2. The dependent measure is the proportion of fixations to the target object.

	Estimate	SE	*t*-value	pMCMC			Variance	Std.Dev.
(a) overall								
Fixed					*Random*			
(intercept)	11.1525	0.5017	22.228	0.0001				
Window	8.2602	0.5103	16.187	**0.0001**	*Subject*	(Intercept)	1.2661	1.1252
Entrainment	1.9062	0.7237	2.634	**0.0080**	*Trial*	(Intercept)	2.5393	1.5935
Window*Entrainment	-3.8555	1.0207	-3.777	**0.0001**	*Residual*		60.5477	7.7812
(b) 100-600ms								
Fixed					*Random*			
(intercept)	11.0753	0.6366	17.399	0.001				
Entrainment	1.8954	0.7371	2.571	**0.008**	*Subject*	(Intercept)	5.7116	2.3899
					Trial	(Intercept)	2.8304	1.6824
					Residual		62.4290	7.9012
(c) 600-1100ms								
Fixed					*Random*			
(intercept)	19.4624	0.5051	38.53	0.001				
Entrainment	-1.9654	0.6810	-2.89	**0.006**	*Subject*	(Intercept)	0.60332	0.77674
					Trial	(Intercept)	3.75454	1.93766
					Residual		53.52448	7.31604

Table 4. Mixed effect model with time-window and historical factor as fixed effects for modified expressions in Exp2. The dependent measure is the proportion of fixations to the target object.

	Estimate	SE	*t*-value	pMCMC			Variance	Std.Dev.
Fixed					*Random*			
(intercept)	10.7931	0.4809	22.442	0.0001				
Window	3.6801	0.2804	13.123	**0.0001**	*Subject*	(Intercept)	1.3598	1.1661
Entrainment	0.9861	0.7241	1.362	0.1772	*Trial*	(Intercept)	1.8481	1.3595
Window*Entrainment	-0.5664	0.5609	-1.010	0.3136	*Residual*		79.9009	8.9387

The dependent measure was the proportion of target fixations within the time-window. Backwards evaluation of AIC (Akaike, 1974) and the deviance test statistic indicated that models with random slopes for the fixed effects did not improve model fit. In this and all subsequent analyses, we report the critical t-statistic along with a Monte Carol Markov Chain (MCMC) based *p*-value estimate, which provides a more conservative estimate of the significance level than would be obtained from using the degrees of freedom associated with the test statistic (see Baayen, Davidson, & Bates, 2008; Barr, 2008). The first model included time-window and entrainment type as factors and revealed two main effects and a significant interaction. To explore the interaction, two planned analyses focused on the effects at the first and second time-window separately.

During the first analysis window (100-600ms after the onset of the unmodified noun), participants fixated the target more often in the historical than the ahistorical condition ($p<.01$). This result suggests that if listeners did generate a lexical differentiation expectation, it did not affect interpretation of an expression that was appropriately modified for the immediate context. Instead, there appeared to be a small benefit for previous experience with the target noun. In the second time-window (600-1100ms following noun onset), the pattern of fixations flipped, with a larger proportion of fixations to the target in the ahistorical condition ($p<.01$). The flip in the direction of the effect is likely due to delays in target identification (e.g., see Brown-Schmidt, 2009), as the average target fixation latency was (non-significantly) delayed in the ahistorical (m=412ms, SD=282) compared to the historical condition (m=389ms, SD=311ms). The locus of the effect may be priming of the critical noun (Ferreira, Kleinman, Kraljic, &

Siu, 2012), as critical nouns in the historical condition were repeatedly interpreted during entrainment, whereas for items in the ahistorical condition, participants were never exposed to the target noun prior to the critical trial.

Modified expressions

Eye movements in response to modified expressions were analyzed in three planned time-windows. The first time-window was from 100ms after adjective onset to the average noun onset time (483ms following adjective onset), and captured fixations in response to the critical adjective. The second window was from noun onset to 500ms after noun onset, and captured fixations in response to the noun. The final window was from 500ms to 1000ms after noun onset, and was selected to capture late effects. We analyzed the proportion of fixations to the target in a series of mixed effects models with time-window and historical condition as fixed effects (see Table 4). As in the model for non-modified expressions, fixed effects were coded using orthogonal contrast codes and random slopes did not improve model fit. The first model included time-window and entrainment type as factors and revealed a main effect of time-window only, due to an increase in target fixations over the trial.

If listeners expected the speaker to lexically differentiate, we would expect more target fixations in the historical vs. the ahistorical condition. However, neither the effect of entrainment, nor the interaction with time was significant. Further, separate planned analyses at each time-window showed no evidence of a historical context effect (*ps*>.05). This result suggests that listeners did not generate an expectation for lexical differentiation when sequentially interpreting reference to two different exemplars of

same category. The fact that the results in comprehension are in contrast with the production findings of Experiment 1 and Van Der Wege (2009) suggests that the scope of relevant discourse history may be different in comprehension and production.

Relationship between production and comprehension

The lack of a differentiation effect in Experiment 2 suggests that listeners strongly expected to hear basic level names to refer to the target referent. Indeed, while speakers in Experiment 1 showed significant differentiation effects, they repeated basic level terms on over half of the differentiation trials. Perhaps, then, speakers and listeners alike are reluctant to deviate from using a basic name, particularly for referents with well-established names. To examine this possibility, a post-hoc analysis was conducted to examine if the consistency with which targets were named in Experiment 1 predicted the strength of the differentiation prediction in Experiment 2. We calculated the number of nouns that were used by speakers in Experiment 1 to refer to each target object across all participants, for both the contrast (where differentiation is expected) and no-contrast (where differentiation is not expected) conditions. If participants were consistent in how they named targets, name agreement would be 100%; if participants were inconsistent, the percent agreeing on a single noun might be lower. The name agreement rates in Experiment 1 ranged from 56.25% to 100% (total: $M = 87.89$, $SD = 13.65$; contrast condition: $M = 87.89$, $SD = 15.71$, no-contrast condition: $M = 88.28$, $SD = 15.53$), suggesting substantial variability in the participant's impression about basic level nouns. We analyzed the proportion of fixations to the target in a mixed effects model with name agreement (from Experiment 1), time-window, and entrainment as fixed effects when

participants heard modified noun phrases in Experiment 2. Subject and item were included as random intercepts; random slopes did not improve model fit. The analysis revealed significant effects of both name agreement ($p<0.05$) and time-window ($p<0.01$); neither the effect of entrainment, nor the interaction with name agreement was significant. The naming agreement effect was due to a *decrease* in target fixations with *increasing* naming agreement. This result suggests that locally overinformative modifiers may be particularly dispreferred when an object's name is highly predictable. However, the lack of an interaction with the entrainment effect suggests that high name agreement was not to blame for a failure to generate differentiation expectations.

In summary, the results of Experiment 2 offered no evidence that listeners generate lexical differentiation expectations. Instead, both modified and non-modified nouns were interpreted equally rapidly, regardless of the historical context. However, an outstanding possibility is that listeners in Experiment 2 *did* expect speakers to lexically differentiate, but that they were unable to make a specific enough prediction about the form of the differentiated, modified expression in order to guide comprehension. Consider that if listeners *did* expect speakers to use a more specific expression, several different modifiers, or alternative subordinate nouns could have been used and thus perhaps listeners could not make a useful prediction (for a related discussion, see Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Federmeier, 2007; Hoeks, Stowe, & Doedens, 2004). For example, when listeners in the historical condition saw the striped shirt at test, they might have successfully contrasted this item with their historical representation of the checkered shirt, and the fact that the contrast was referred to as *the shirt*. However, to gain a processing benefit, the listener would have had to anticipate the

critical expression *the striped shirt,* and not *dress shirt, pinstripe shirt,* or *tailored shirt.* If the addressee predicted the wrong differentiated term, they would not show any benefit when interpreting *the striped shirt,* due to the mismatch between their expectation and the current input.

CHAPTER 4: EXPERIMENT 3

In Experiment 3, we test for differentiation effects in comprehension again, but in a paradigm where the differentiation prediction should be easier to make. Recall that the results of Experiment 1 showed that speakers used two mechanisms to differentiate their expressions: modification and the use of a different noun. In Experiment 3 we asked whether listeners expected speakers to differentiate using a different, subordinate-level nouns, which might be easier to make predictions about compared to the modified expressions that we tested in Experiment 2.

The critical stimuli in this experiment were unmodified noun phrases such as *the poodle,* or *the dock.* We used a cohort-competitor paradigm (Allopenna, Magnuson, & Tanenhaus, 1998; Barr & Keysar, 2002), which allowed us to test for competition between temporary cohort competitors such as *poodle* and *pool,* and *dock* and *dog*. As in Experiment 2, participants completed both entrainment and test trials. During entrainment, listeners interpreted basic level nouns such as *the dog* (as referring to, e.g., a German Shepard). Test scenes included a second exemplar from the same object category (e.g., a poodle), a subordinate-level cohort competitor (pool), a basic-level cohort competitor (dock, where *dock* and the basic-level term *dog* are cohort competitors), and an unrelated item (racquet). When interpreting the subordinate term *poodle,* which is temporarily ambiguous with the basic-level term *pool,* listeners should be more likely to interpret the temporarily ambiguous phoneme sequence *poo-* as referring to the subordinate-level name (poodle) if they expected the speaker to lexically differentiate. When interpreting *dock*, which is temporarily ambiguous with the basic-level term *dog*, listeners should be more likely to interpret the initial ambiguous phonemes *do-* as

referring to the dock if they expected the speaker would have referred to the dog using a differentiated term (such as *poodle*).

Importantly, by including a competitor during test trials, Experiment 3 offers a stronger test of lexical differentiation effects, because test scenes contain referents temporarily consistent with both the differentiation, and non-differentiation interpretation (see Barr & Keysar, 2002 for a similar approach). Thus, if listeners in Experiment 2 did in fact generate lexical differentiation expectations, but failed to predict the specific modifier that the speaker used, in Experiment 3 the pattern of cohort competition during the interpretation of subordinate (e.g., *poodle*), and basic level competitors (e.g., *pool*) should reveal the differentiation-based expectations.

Method

Participants

Thirty-two undergraduates at the University of Illinois at Urbana-Champaign participated in the experiment as speakers in return for partial course credit or either cash ($8). Participants were native speakers of North American English. No participant took part in Experiments 1 or 2.

Materials and Procedure

As in Experiment 2, a participant and an experimenter played a version of a referential communication task. The procedure of Experiment 3 was identical to Experiment 2. Across a series of trials, the experimenter gave instructions to the participant to click one of the four objects on the screen. The participant's eye

movements were monitored and the experimenter's voice was recorded during the experiment.

Each participant completed a total of 256 trials including 16 target trials, 96 entrainment trials and 144 filler trials. Like Experiment 2, there were two phases: entrainment trials and test trials. During the entrainment trials, participants were shown to the contrast object (e.g., dog) with three unrelated objects six times. The experimenter gave instructions referring the contrast object or one of the unrelated objects using a bare noun phrase (e.g., *Click on the dog*).

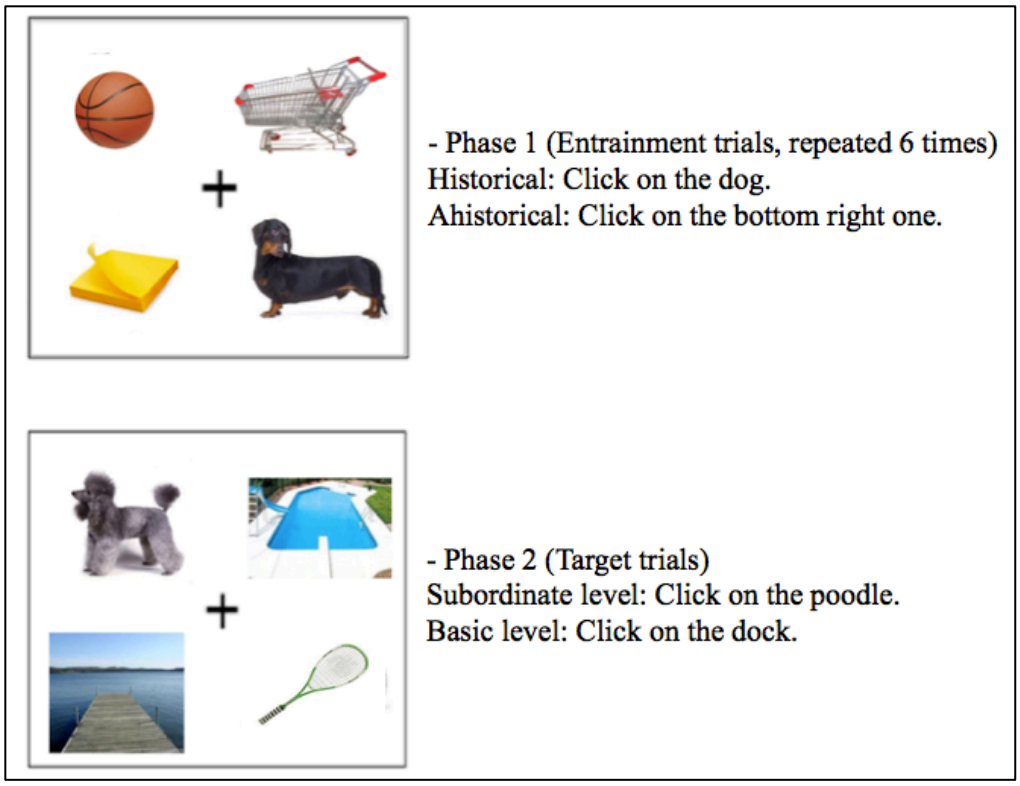

Figure 6. Example stimuli for Experiment 3.

During test trials each display contained four objects: the subordinate-level object (e.g., poodle/dog), two potential cohort competitor objects (e.g., pool and dock), and an unrelated object (e.g., racquet). Of the two potential cohort competitor objects, one (e.g., dock) was a cohort competitor with the basic-level noun (e.g., dog) and the other was a cohort (e.g., pool) with the subordinate-level noun (e.g., poodle). Among the 16 test trials, the target was a subordinate-level object for the half of trials (e.g., poodle) and for the other half trials, the target was a cohort object (e.g., dock). Target trials never used the basic-level term for the subordinate-level object (e.g., the term *dog* was never used to refer to the poodle).

Two factors were manipulated within-subjects: Entrainment type (historical or ahistorical) and target type (subordinate level object or cohort object with the basic-level noun). During entrainment trials, in the historical condition, participants heard instructions that referred to the target with a bare noun phrase, while in the ahistorical condition, a locative phrase was used. Again, note that the use of a locative in the ahistorical condition critically controls for the fact that participants have experience clicking on the contrast object, thus allowing us to focus on the lexical differentiation process per se. During target trials, target type was manipulated based on the level of specificity of the target noun. Unlike Experiment 2, all target trials used bare, unmodified NPs. Target trials either referred to the subordinate-level object (e.g., poodle) using the subordinate level-term *poodle*, or the basic-level object (e.g., dock), using the basic level term *dock*. The 16 target trials were rotated through the four conditions across four experimental lists. Each participant completed the items on a single list. The experiment lasted about 40 minutes.

Predictions

If listeners generate a lexical differentiation expectation, during interpretation of the ambiguous portion of the subordinate level noun (e.g., *poo-*), the listener should initially interpret this as referring to the subordinate target (e.g., poodle) in the historical condition, whereas in the ahistorical condition, the listener should initially interpret this sequence as referring to the basic-level target (e.g., pool). Such a result would obtain if listeners expected the subordinate term (e.g., *poodle*) in the historical condition, and the basic-level-term (e.g., *dog*) in the ahistorical condition.

In contrast, in the basic-level noun condition, where the participant hears a basic noun like *dock*, if the listener makes a differentiation-based prediction, he or she should be more likely to interpret the ambiguous portion (e.g., *do-*) as referring to "dock" in the historical condition, because the speaker would not re-use the term *dog* to refer to the poodle. If so, this makes the prediction that there should be more fixations on the basic-level target (e.g., dock) than the critical object (e.g., dog/poodle) in historical condition compared to the ahistorical condition.

By contrast, if listeners are not sensitive to historical context and do not generate lexical differentiation expectations, when they hear a basic-level noun, we would expect equivalent fixations between cohort object and subordinate level object immediately following word onset. Similarly, there would be no difference in fixations to the subordinate level object between the historical and ahistorical conditions when listeners hear a subordinate level noun.

Analysis and Results

As in Experiment 2, an initial analysis of the acoustic properties of the target stimuli demonstrated that there were no significant prosodic differences across conditions (*t's*<1.00).

Eye movement analyses focused on the eye-movements made in response to the interpretation of the initial phonemes of the target word on test trials. Because the difference in the frequency of basic- and subordinate-level nouns is likely to affect the time-course of interpretation (Magnuson, Dixon, Tanenhaus, & Aslin, 2007), we analyzed test trials separately for subordinate and basic-level terms (see Figures 7-8).

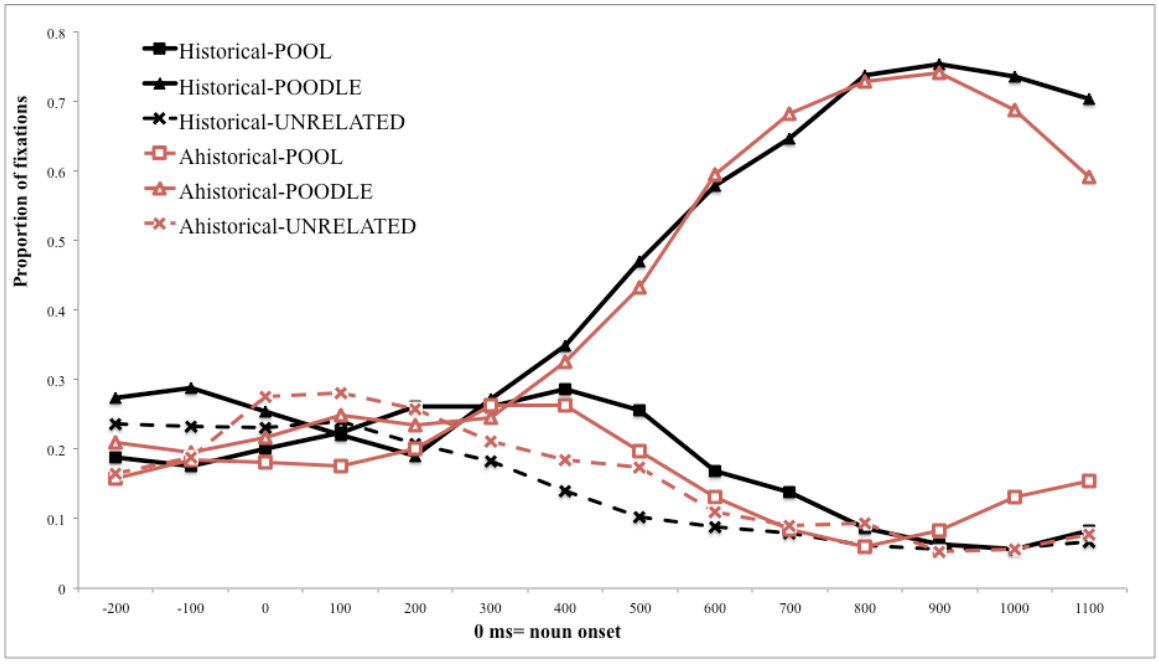

Figure 7. The proportion of target (poodle), competitor (pool), and unrelated (racquet) fixations for subordinate level nouns in Experiment 3. The target word is *poodle.*

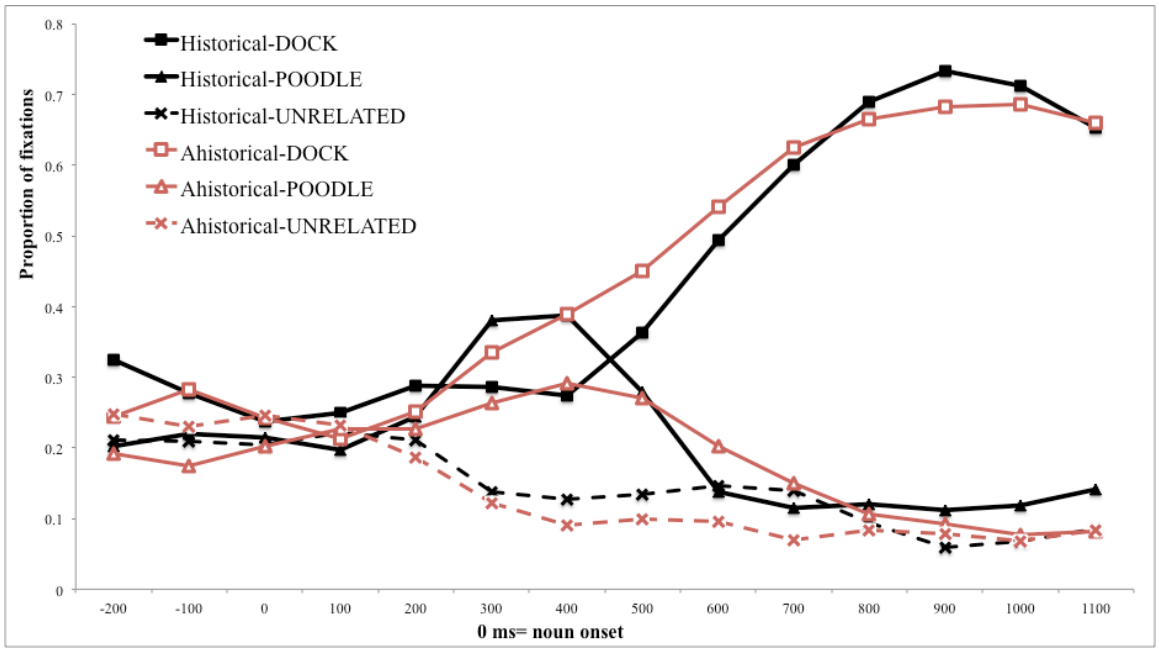

Figure 8. The proportion of target (dock), competitor (dog/poodle), and unrelated (racquet) fixations for basic level nouns in Experiment 3. The target word is *dock*.

Subordinate level targets (e.g., Click on the poodle)

When participants interpreted subordinate level nouns, fixations to the target began to rise approximately 400ms following noun onset. If listeners expected differentiation, we would expect a larger preference for the subordinate target (poodle) in the historical compared to the ahistorical condition. Contrary to predictions, when listeners had previously experienced the contrast object (historical condition), there was no advantage for interpreting a subordinate-level reference to the target. Eye movements for subordinate level expressions were analyzed in two planned 500ms analysis regions, 100-600ms and 600-1100ms after noun onset (Table 5). To capture both target and competitor fixations in a single analysis, we calculated the empirical logit for the ratio of target fixations to competitor fixations (.5 was added both the numerator and the denominator as the calculation is problematic in the absence of a competitor fixation; see

Barr, 2008 for a similar approach). The empirical logit was calculated for the two time windows for each trial separately. The resulting dependent measure, hereafter "target advantage" is symmetrical about zero with positive values indicating a target preference and negative values indicating a competitor preference.

The target advantage scores were analyzed in a mixed effects model with time-window and entrainment type as fixed effects, and subject and item as random intercepts (random slopes did not improve model fit). The model revealed only a main effect of time window, with no evidence that participants were more likely to fixate the basic-level cohort object in the ahistorical condition, and no interaction with time-window (ps>.5). If anything, listeners were *more* likely to fixate the cohort competitor in the historical condition. This result suggests that listeners did not generate lexical differentiation expectations, despite the fact that speakers consistently differentiated their expressions.

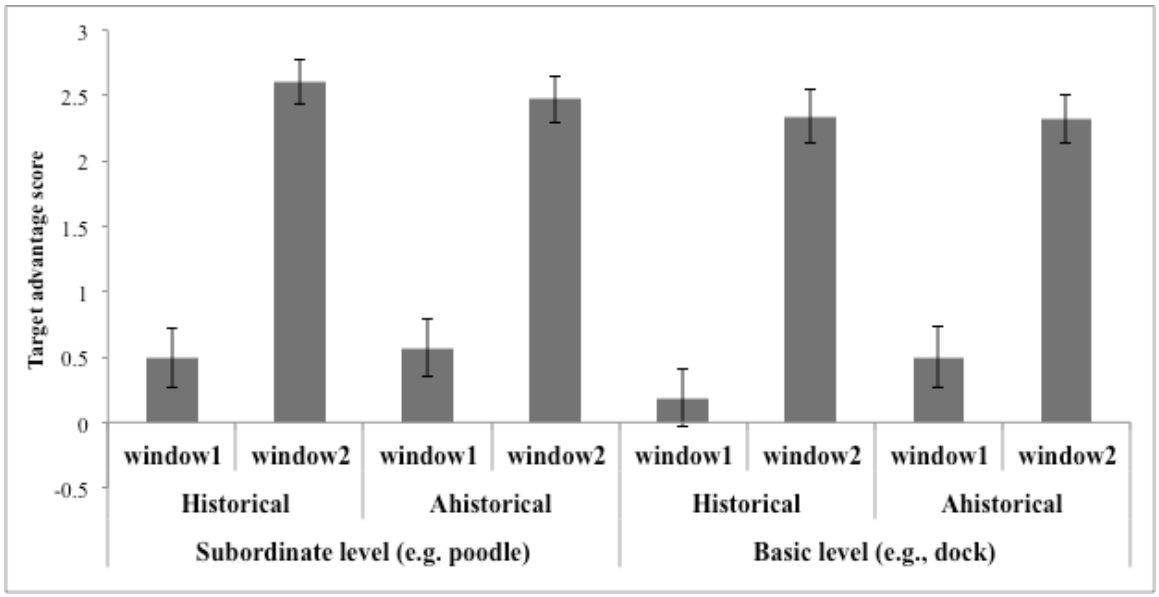

Figure 9. Target advantage scores in each condition.

Table 5. Mixed effect model with time-window and historical condition as fixed effects for target advantage score when participants heard a subordinate level noun (poodle) in Exp3.

	Estimate	SE	*t*-value	pMCMC			Variance	Std.Dev.
(a) overall								
Fixed effects					*Random effects*			
(intercept)	0.52553	0.21545	2.439	0.0164	Groups	Name		
Window	2.00701	0.19119	10.497	**0.0001**	Subject	(Intercept)	0.20665	0.45459
Entrainment	0.06805	0.27047	0.252	0.8064	Trial	(Intercept)	0.34652	0.58866
Window*Entrainment	-0.20998	0.38239	-0.549	0.5896	Residual		4.58753	2.14185

Table 6. Mixed effect model with time-window and historical condition as fixed effects for target advantage score when participants heard a basic level noun (dock) in Exp3.

	Estimate	SE	*t*-value	pMCMC			Variance	Std.Dev.
(a) overall								
Fixed effects					*Random effects*			
(intercept)	0.3110	0.1943	1.600	0.1106	Groups	Name		
Window	1.9808	0.2205	8.983	**0.0001**	Subject	(Intercept)	0.072625	0.26949
Entrainment	0.3181	0.3126	1.017	0.3126	Trial	(Intercept)	0.172710	0.41558
Window*Entrainment	-0.3304	0.4411	-0.749	0.4492	Residual		5.421875	2.32849

Basic-level targets (e.g., Click on the dock)

When participants interpreted basic level nouns (e.g., *dock*) that were cohort-competitors with the basic-level critical item (e.g., *dog*), we observed a similar lack of differentiation. For basic-level targets, if listeners expected differentiation, we would expect a larger preference for the basic-level target (dock) in the historical compared to the ahistorical condition (i.e., because of an expectation to refer to the dog as a *poodle*). The eye movement data were analyzed in the same two time-windows as for the analysis of interpretation of subordinate-level expressions; 100-600ms and 600-1100ms following noun onset. Target advantage scores (calculated as before) were analyzed in a mixed effects model with time-window and target type as fixed effects (see Table 6), and subjects and items as random intercepts (random slopes did not improve model fit). The model included time-window and entrainment as factors and revealed a main effect of time-window only, due to increasing target fixations throughout the trial. Contrary to predictions, target advantage scores were non-significantly larger in the ahistorical condition than historical condition, possibly due to priming of the cohort-competitor term *dog* during entrainment trials.

CHAPTER 5: GENERAL DISCUSSION

Our findings show that speakers design their referential expressions with respect to a broadly-defined historical context that includes not only previously-established names for a given item, but in addition, how other items not present in the immediate context had been named in the past. By contrast, two different experiments found no evidence that listeners generated complementary expectations about how the speaker would refer to a given entity. Instead, comprehension processes were primarily driven by the accessibility of the interpreted term, with no apparent penalty for applying the same term to multiple referents. These findings suggest that the scope of the relevant discourse context for speakers and listeners differs, resulting in subtle asymmetries in the relationship between reference and context.

Mechanisms of lexical differentiation

The results of Experiment 1 showed that speakers differentiated 24% of all critical expressions through the use of two different methods (34% modified and different expressions in the contrast condition vs. 10% in the no-contrast condition). This differentiation rate was comparable to the effective rates reported by Van Der Wege for either typical (33%) or atypical items (19%) in interactive conversation. While both studies show clear evidence of differentiation, it is worth noting that speakers were content to re-use bare noun phrases to refer to different referents much of the time (66% in our experiment), suggesting that, at least in the case of reference to basic objects, *not* differentiating may not be the norm.

Indeed, a combined analysis of Experiments 1 and 2 found that listeners in Experiment 2 showed *smaller* differentiation expectations for target objects that exhibited higher name agreement in Experiment 1. This finding suggests that highly established names may be resistant to contextual influences. If so, speakers may be reluctant to deviate from that name, even if it were to create the potential for confusion by homonymy (cf. Clark & Clark, 1979). Thus, differentiation might be more common for referents with ill-established names, or for atypical referents. And, when speakers do differentiate referents with well established, or highly predictable names, the salience of the established name may lead speakers to produce the established name first, followed by a description (Gorman, Gegg-Harrison, Marsh, & Tanenhaus, in press; Heller, Gorman, & Tanenhaus, 2012), instead of using prenominal modification or subordinate terms. If so, this might contribute to listeners' initial expectation for the unadorned name, regardless of the need to differentiate.

The coherence of the context may matter as well. Both Experiment 1 and Van Der Wege (2009) tested for differentiation in situations where the context (pictures on a computer screen) changed from entrainment to test. Speakers and listeners alike might be more likely to draw comparisons across contexts and generate stronger differentiation predictions in contextually coherent situations, because contextual overlap between entrainment and test facilitates retrieval of the previous discourse context (see Horton, 2007).

What is the context of conversation?

According to the classic view of conversational processes, interlocutors store rich, diary-like representations of jointly held knowledge and beliefs that include representations of previously and jointly experienced contexts (Clark & Marshall, 1978; Clark, 1992; 1996). In subsequent linguistic exchanges these representations of joint experience, or common ground, support communication by providing a shared context within which reference can be understood. This view has recently been challenged by an alternative view that suggests that sensitivity to joint experience may not require access to these rich representations. Instead, according to the association view, associations between names, referents and partners serve as simple representations of common ground (Pickering & Garrod, 2004; Horton, 2007). Support for this view comes from findings that speakers are faster to name an object when the object name (or a semantic associate) had been jointly experienced by the dialog partners in the past (Horton, 2007).

While simple associations might capture some aspects of sensitivity to joint experience, our findings suggest that simple associations are not enough to capture the breadth of the relevant discourse context in language production. The fact that speakers used a *different* referring expression when they had previously described a referent from the same category suggests that the historical representation of the discourse contained more than an association between referents and partners. After all, if speakers had only formed an association between the contrast and their partner, the semantic association between the contrast and target would likely have *increased* the use of the same basic level term. The fact that speakers tended to use a different term instead suggests the

historical record included enough information to distinguish the target from the way in which the contrast was previously named.

Indeed, the idea that interlocutors maintain rich representations of the discourse context is consistent with a variety of related findings in a growing literature on the processing of dialog. For example, speakers and listeners draw on their knowledge of what information has and has not been discussed in the conversation when interpreting utterances (Brown-Schmidt, et al., 2008; Brown-Schmidt, 2009; 2012). Further, detailed analyses of interactive conversation show that speakers' and listeners' representations of common ground are highly coordinated, facilitating successful use and interpretation of potentially ambiguous expressions (Brown-Schmidt & Tanenhaus, 2008; Beun & Cremers, 1998; Brennan, 2005). Further, when interlocutors do share more common ground, they tend to distribute visual attention in a shared screen in a more similar manner (Richardson, Dale, & Kirkham, 2007; Richardson, Dale, & Tomlinson, 2009).

While these findings point to a high degree of coordination in conversation, the results of the present research suggest that representations of the historical discourse record may be slightly different for speakers and addressees. Speakers, as the authors of their utterances, might encode more detailed representations of the historical discourse context, resulting in more easily-observable lexical differentiation effects. In contrast, listeners have the task of inferring what the speaker believes is the relevant context, and then making a prediction about how the speaker would distinguish the intended referent against this context. This inferential process may be more error-prone.

Given that speakers did show sensitivity to the historical discourse context, a critical question is what aspects of the historical discourse record are encoded in

conversation. According to the classic view, the relevant discourse context includes rich representations of the context of past experience, including information about unmentioned, but jointly perceived people and entities (Clark & Marshall, 1978). At a minimum, lexical differentiation effects show that speakers encode and subsequently retrieve information about how previous discourse referents had been referred to. However, an open question is whether speakers consider other aspects of the previous discourse context, such as unmentioned properties of previous referents, as well as unmentioned contrast objects.

For example, consider a past context that included a striped cotton shirt and a dotted polyester shirt, and the speaker uniquely identified the intended referent by saying *the striped shirt*. Now imagine a future context containing a dotted silk shirt. If the speaker only contrasts her expressions with respect to previously-mentioned entities, and not unnamed contrasted entities, then the speaker might simply say *the dotted shirt* to differentiate the new referent from the previous referent. However, if the contrast object from previous contexts is encoded into the relevant historical context as well, speakers might instead say *the silk shirt*, to distinguish the intended referent from both the previous referent and the previous contrast.

Some recent evidence from a recognition memory paradigm suggests that contrasted items are remembered in some cases: Fraundorf, Watson, and Benjamin (2010) played stories that introduced contrasting pairs of entities such as *British scientists and French scientists,* and then continued with a sentence that re-mentioned one of the entities using a contrastive pitch accent (or not, e.g., *The FRENCH scientists…*). Results from a subsequent memory test showed that invoking contrast with a pitch accent

increased memory for the contrasted item (British scientists). This result suggests that at least when a contrast has already been established in the discourse, a referring expression that invokes contrast can alter the representation of the contrasted item itself.

Overspecification

Speakers in the Experiment 1 frequently overspecified for the immediate context, providing more modifiers or more specific terms than necessary in order to contrast with the previous discourse record. The fact that speakers were willing to modify with respect to a broader discourse context may explain some apparent overmodification effects in the literature, including findings that speakers produce modifiers in single-referent contexts about 30% of the time for sentences like *"Put the apple that's on the towel in the box*" (Engelhardt, et al., 2006). While overspecified with respect to the current trial, speakers in these situations might be modifying to contrast the current referent with previous referents (e.g., the single apple is currently on the towel but was previously on a sock in a different trial). Thus, such apparent violations of the Maxim of Quantity (Grice, 1975) may in fact reflect modification with respect to a broader discourse context.

This evidence of local overmodification is in apparent contrast to evidence that in other domains, speakers seem insensitive to ambiguity, even when ambiguity avoidance is necessary from the listeners' perspective (Ferreira & Dell, 2000; Ferreira, 2008). For example, while Ferreira, Slevc, and Rogers (2005) found that speakers in their second and third experiments used locally ambiguous homophones less often after using those homophones to refer to a contrast object, the speakers were still using ambiguous homophones about 1/3 of the time. Further, even in cases where speakers do provide

disambiguating information, for example, as in the use of helpful prosody, speakers may not be doing so to reduce ambiguity for the listener per se (Kraljic & Brennan, 2005). One explanation for why speakers in our experiment disambiguated their expressions with respect to the historical context may be the type of ambiguity we examined. Speakers might have more awareness or control over the potential for ambiguity when describing entities from the same basic level object category (e.g., two shirts) compared to objects from different categories that have homophonic names, or compared to syntactic ambiguity.

Speakers may also produce additional information when they are not certain whether their expression provides enough information for the listener to identify the referent (Van Der Sluis & Krahmer, 2007), or when the speaker does not receive feedback from the listener (Arts, 2004). Overspecified expressions may also be preferred in cases where they reduce the speaker's effort in assessing the context (Engelhardt et al., 2006), or when task success is important (Arts et al., 2011a).

With respect to over-modification in comprehension, the theoretical landscape is mixed. Whereas some argue that extra information facilitates comprehension (Arts, Maes, Noordman, & Jansen, 2011b; Levelt, 1989), others argue that extra information harms comprehension (Engelhardt, Bailey, & Ferreira, 2006; Engelhardt, Demiral, & Ferreira, 2011). For example, Arts et al. (2011b) compared expressions that included information only about the shape of a target item (e.g., *the square button*), with expressions that mentioned three attributes (size, color, and shape, e.g., *the large square white button*). They found that the additional attribute information facilitated comprehension and led to faster identification of the target, relative to visual scene onset.

Similarly, Paraboni et al. (2007) showed that speakers prefer to use redundant expressions when the referential context is potentially confusing and thus the listener might select the wrong entity. One reason that redundant information might facilitate comprehension in such cases is that the redundant cues can help rule out distractors, and thus reducing the effort required for the listener to identify the intended referent.

However, other studies suggest that unnecessary information can harm comprehension. Engelhardt and his colleagues (2006) reported that an unnecessary prepositional phrase modifier, "on the towel" in an instruction such as "*Put the apple on the towel in the box."* impaired comprehension compared to the appropriately informative instruction such as "*Put the apple in the box*". The locus of these effects might be the fact that not only was the prepositional phrase modifier supported by the context (i.e., only one apple was in the scene), but in addition, the verb *"Put"* is often immediately followed by a prepositional phrase argument. Thus, in cases where the context supports an interpretation of the modifier that is inconsistent with noun-overmodification—for example, as an argument of the verb (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), or as a modifier for a different object that is in a locally-available contrast set (Sedivy, Tanenhaus, Chambers, & Carlson, 1999), comprehension of overmodification is likely to be impaired.

CHAPTER 6: CONCLUSIONS

Here we explored the breadth of the relevant discourse context, examining the process of lexical differentiation (Van Der Wege, 2009) in both production and comprehension. We identified two different mechanisms through which speakers lexically differentiate their referents from other referents in the historical discourse context. This result demonstrates that the historical context can license violations of local informativity (Olson, 1970; Grice, 1975), and offers support for a *context memory* view of historical context effects, where speakers design expressions with respect to both information about how the current referent had previously been referred to (Brennan & Clark, 1996), but also information about how contrasting referents from the same category had previously been referred to (Van Der Wege, 2009). More generally, our findings support a view of common ground in which speakers maintain rich representations of the joint discourse context (Clark & Marshall, 1981). Interestingly, while speakers showed exquisite sensitivity to the historical discourse record, we found no evidence that listeners generated complementary expectations for lexical differentiation. Thus, while listeners do show historical sensitivity when a previous referent is referred to again (Barr & Keysar, 2002; Metzing & Brennan, 2003; Brown-Schmidt, 2009), the present results suggest they are relatively insensitive to how other referents were previously referred to.

Our findings point to an asymmetry in the role of context of language production and comprehension, and suggest that while the historical context can shape production, this context might not be sufficient to generate specific enough predictions to guide comprehension processes. Speakers, who are the authors of their ideas may be more

sensitive to things they have experienced and said before, as a result of active engagement in assessing the relevant context to design utterances. By contrast, an addressee's representation of the context may be less detailed, and the inferential process required to make predictions about how the speaker will design expressions with respect to the historical context may be more error-prone.

REFERENCES

Akaike, Hirotugu (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419-439.

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica.*

Arts, A. (2004). *Overpsecification in instructive texts*. Tilburg: Tilburg University dissertation.

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

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011b). Overspecification facilitates object identification. *Journal of Pragmatics, 43*, 361–374.

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

Barr, D., & Keysar, B. (2002). Anchoring and comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391–418.

Barr, D. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457-474.

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

Boersma, Paul & Weenink, David (2003). Praat: doing phonetics by computer (Version 4.1.3) [Computer program]. Retrieved Jul. 2nd, 2003, from http://www.praat.org/

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions (pp. 95-129). Cambridge, MA: MIT Press.

Brennan, S. E., & Clark, H, H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *22*, 482-493.

Brennan, S. E. & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science, 1*, 274-291.

Brown-Schmidt, S. (2009). The role of executive function in perspective-taking during on-line language comprehension. *Psychonomic Bulletin and Review*, *16*, 893-900.

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., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition, 107*, 1122-1134.

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.

Brown-Schmidt, S. & Tanenhaus, M.K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language gam approach. *Cognitive Science, 32*, 643-684.

Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

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 (pp. 57-63). In D. L. Waltz (Ed.), *Theoretical issues in natural language processing* (Vol. 2). New York: Association for Computing Machinery.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science* , *13,* 259-294.

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

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition, I1,* 159-184.

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.

Engelhardt, P. E., Baris Demiral, S. B., Ferreira, F. (2011). Over-specified referring expression impair comprehension: An ERP study. *Brain and Cognition, 77*, 304-314.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491-505.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146,* 75-84.

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation: Advances in Research and Theory, 49*, 209-246.

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

Ferreira, V. S., Kleinman, D., Kraljic, T., Siu, Y. (2012). Do priming effects in dialogue reflect partner- or task-based expectations? *Psychonomic Bulletin & Review, 19*, 309-316.

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

Ford, W., & Olson, D. (1975). The elaboration of the noun phrase in children's descriptions of objects. *Journal of Experimental Child Psychology*, *19*, 371-382.

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.

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

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (in press). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Grice, H.P. (1975). Logic and conversation. In Core P., & Morgan, J.L. (Ed.), Syntax and semantics 3: Speech arts (pp.41-58). New York: Academic Press.

Hallett, P. E. (1986). Eye movements. *Handbook of perception and human performance (Vol. 1),* eds Kenneth R. Boff, Lloyd Kaufman, and James P. Thomas, pages 10.1-10.112. Wiley, New York, New York.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory & Language*, *49*, 43-61.

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science, 4*, 290–305.

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

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research, 19,* 59-73.

Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes, 22,* 1114-1139.

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.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language, 39*, 1-20.

Kraljic, T., & Brennan, S. E. (2005). Using prosody and optional words to disambiguate utterances: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194-231.

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.

Kronmüller E. & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56,* 436-455.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*, 133-156.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology*, *7*, 181-191.

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

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*, 329-336.

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* (pp. 497–529). Cambridge, MA: Cambridge University Press.

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

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

Pickering, M. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27,* 169-226.

Pickering, M. & Garrod, S. (in press). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*.

Richardson, D. C., Dale, R. & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science, 18,* 407-413.

Richardson, D.C., Dale, R. & Tomlinson Jr, J.M. (2009). Conversation, gaze coordination and beliefs about visual context. *Cognitive Science*, *33,* 1468-1482.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21,* 211-232.

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

Stalnaker, R. (1978). Assertion. In P. Cole (ed.), Syntax and Semantics 9: Pragmatics. Academic Press.

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

Van Der Sluis, I. & Krahmer, E. (2007). Generating multimodal references. *Discourse Processes, 44*, 145-174.

Van Der Wege, M. M. (2009). Lexical entrainemtn 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.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Cognition* , *31,* 183-194.