Temporary ambiguity and memory for the context of spoken language

Kaitlin Lord and Sarah Brown-Schmidt

Vanderbilt University, Department of Psychology and Human Development

Corresponding Author: Sarah Brown-Schmidt

Vanderbilt University Department of Psychology and Human Development

230 Appleton Place · Nashville, TN 37203, USA

615-875-8081

sarah.brown-schmidt@vanderbilt.edu

Abstract

Spoken language is interpreted incrementally, with listeners considering multiple candidate meanings as words unfold over time. Due to incremental interpretation, when a speaker refers to something in the world, there is often temporary ambiguity regarding which of several candidate items in the referential context the speaker is referring to. Subsequent tests of recognition memory show that listeners have good memory for referenced items, but that listeners also sometimes recognize non-referenced items from the referential context that share features with items that were mentioned. Predicted or inferred (but not experienced) interpretations of what was said are also sometimes retained in memory. While these findings indicate that multiple items from the referential context may be encoded in memory, the mechanisms supporting memory for the context of language use remain poorly understood. This paper tests the hypothesis that a consequence of temporary ambiguity in spoken language is enhanced memory for the items in the referential context. Two experiments demonstrate that periods of temporary referential ambiguity boost memory for non-referenced items in the referential context. Items that temporarily matched the unfolding referring expression were better remembered than those that did not. The longer the period of ambiguity, the stronger the memory boost, particularly for items activated early in the expression. In sum, the fact that spoken language unfolds over time creates momentary ambiguity about the speaker's intention; this ambiguity, in turn, allows the listener to later remember not only what the speaker did say, but also what they could have, but did not.

Keywords: Temporary ambiguity, eye-tracking, incremental processing, recognition memory

It is well-known that the use of language to refer to things in the world is guided by the context of language use, also known as the referential context (Altmann & Steedman, 1998; Tanenhaus, et al. 2000). Against this referential context, the interpretation of signed and spoken languages is incremental (Allopenna, Magnuson, & Tanenhaus, 1998; Lieberman, Borovsky, & Mayberry, 2018), with the unfolding linguistic signal creating temporary ambiguity between candidate meanings (Eberhard, et al. 1995). Over longer time-scales, memory for what was said allows conversational partners to form common ground (Clark & Marshall, 1978; Horton & Gerrig, 2005), supports the use of pronouns (Foraker & McElree, 2011; Karimi, Swaab, & Ferreira, 2018), and facilitates ongoing communication (Yoon, et al., 2016; McKinley, et al., 2017; Brennan & Clark, 1996). Given that language is interpreted incrementally, with listeners considering multiple candidate meanings over time, what becomes of these temporarily-considered meanings in memory?

Consider the process of interpreting a referring expression like "*the dotted bag*" or "the tall glass", given a referential context defined by pictured items in a visual display (Tanenhaus, et al., 2000). If that referential context contains multiple items that partly or fully match the expression, such as a picture of a dotted bag, striped bag, and dotted shirt, the fact spoken language unfolds over time creates temporary ambiguity regarding which item the speaker is referring to. Measures of on-line language processing reveal that when the listener processes the initial part of the expression, e.g., *the dotted*, listeners gaze at objects matching the initial words (dotted bag/shirt). As the expression unfolds, listeners also sometimes fixate objects matching subsequent words before identifying the referent (Sedivy, 2003; Eberhard, et al., 1995; Sedivy, et al., 1999; Fukumura & Carminati, 2021), though to the best of our knowledge, the literature lacks a direct comparison of fixations to early and later-matching competitors. *Semantically* related items also become activated. For example, when interpreting "*the key*" in a referential context with a lock and key, listeners fixate the lock significantly more than semantically unrelated items (e.g., an apple, Yee & Sedivy, 2006; Yee, et al., 2011). Likewise,

when hearing "the salt, uh I mean...", listeners fixate the pepper in the scene (Lowder &

Ferreira, 2016), suggesting listeners activate multiple types of related items when processing

referring expressions.

The presence of non-referenced items in the context is functional, facilitating

interpretation of modified referring expressions like "*the dotted bag*". This is because it is the

presence of contrasting items from the same category as the intended referent (e.g., the striped

bag when the speaker says "dotted bag") that supports use a modifier like "dotted" in the first

place (Fernald, Thorpe, & Marchman, 2010; Grodner & Sedivy, 2011). After all, if there was only

one bag in the display, the speaker could simply say "*bag*" (Olson, 1970). Studies of language

production show that when describing objects, *speakers* primarily gaze at the referent (e.g.,

dotted bag), but also fixate contrasting items from the same category (e.g., striped bag). When

speakers do not fixate the contrasting item, they are less likely to use an adjective (Brown-

Schmidt & Tanenhaus, 2006; Pechmann, 1989), reflecting the influence of competing items on

language production.

Studies of spoken word recognition offer potential insights, by analogy, to the

mechanisms of processing modified referential phrases. Models of spoken word recognition

posit a continuous mapping process where the unfolding speech stream is continuously mapped

to lexical candidates as language unfolds over time. Much of this modeling work focuses on

interpretation of individual words. When interpreting a word like *beaker*, analyses of fixations to

objects in a corresponding scene show early fixations to items that match the initial sounds of

the word (e.g., *beetle, beeper*). At the end of the word *-ker*, listeners temporarily consider

candidates that match the latter half of the word (e.g., *speaker*). Of note, this early competition

is generally stronger than later competition (Allopenna, et al., 1998; Creel, Aslin, & Tanenhaus,

2008; also see Burt, et al., 2017). Allopenna, et al. (1998) model these activation dynamics

using the TRACE model of spoken word recognition, which assumes that multiple alternative

candidate interpretations of words are activated as speech unfolds over time. Indeed, empirical

findings show that even when initial portions of a word or phrase are inconsistent with a candidate referent, if subsequent linguistic material is consistent with that item, it is retained (or recovered) as an interpretation of what was said (Connine, et al., 1991; McMurray, Tanenhaus & Aslin, 2009).

Memory for temporarily considered meanings?

The present research probes the implication of temporary referential ambiguity on subsequent memory for what *could have been* said, but wasn't. Some prior work indicates that listeners form memories of temporarily-considered candidate meanings. In a series of experiments examining task-based conversation, participants took turns describing objects to each other in visual displays with multiple items from the same category, such as a striped shirt and a polka-dot shirt (e.g., "*Click on the striped shirt*"). Subsequent tests of memory showed that speakers and listeners correctly recognized images of both the referenced item (striped shirt) and the contrasting non-referenced item (polka-dot shirt) at above-chance levels (Yoon, Benjamin, & Brown-Schmidt, 2016; 2021). Further, Yoon, et al. (2016) demonstrated that memory for the non-referenced item was significantly better when the speaker used an expression that partially matched the non-referenced item (e.g., "*the striped shirt*"), compared to a locative construction that did not (e.g., "*the top left one*"). This raises the possibility that the process of interpreting the noun "*shirt*" in "*the striped shirt*" was responsible for the memory boost to the non-referenced item (the polka-dot shirt).

Other work shows that *alternative* meanings may be remembered as well. Readers sometimes falsely recognize words that were predicted in a sentence (Hubbard et al., 2019). Similarly, following disfluent repairs (e.g., *the bowl, I mean the ladle*), the repaired item (bowl) is maintained in memory (Karimi, Diaz, & Ferreira, 2020; Ferreira, et al., 2004). Further, when a sentence is temporarily ambiguous in meaning, candidate meanings that turn out to be incorrect

are retained in memory, particularly when the period of ambiguity is long (Christianson, et al., 2001; Ferreira, Lau, & Bailey, 2004; Lau & Ferreira, 2005).

Present Research

This paper tests the hypothesis that a consequence of temporary ambiguity in spoken language is enhanced memory for items in the referential context. While work by Yoon and colleagues shows that items that are temporarily consistent with a referring expression are encoded to memory (Yoon, et al., 2016; 2021), the mechanisms supporting memory for the referential context remain poorly understood, and the role of referential form is unexplored. The work by Yoon and colleagues (2016), for example, did not test memory for different types of items in the referential context, such as items that matched the early part of the expression (e.g., striped pants), or items that were from a related category but did not match the expression (e.g., dotted pants). In particular it is unknown if it is the temporary ambiguity in speaker meaning that boosts memory for temporarily-considered items in the context. Alternatively, the memory boost for these items may simply be due to the fact that temporarily-considered items are semantically related to the referent.

Experiment 1

Methods

This experiment was preregistered on the Open Science Framework (https://osf.io/wthfu). Participants in this study first listened to instructions that contained referential descriptions that were either pre-nominally modified (e.g., *Click on the dotted bag*), or post nominally-modified (e.g., *Click on the bag with dots*). Each scene contained 6 candidate referents, one of which fully matched the referential expression (the target), 2 that partially matched the expression (competitors), 1 that matched each competitor on one dimension but did not match the target (non-competitor), and 2 that did not match (fillers). The task was to click

on the image referenced in the instruction. A subsequent surprise recognition memory test probed memory for the critical target and non-target referents.

Participants

Participants were recruited through the online platform, Amazon Mechanical Turk (MTurk). Participation was restricted to participants who were native speakers of English, located in the United States, at least 18 years old, and who had over 500 approved HITs and at least a 95% acceptance rate on MTurk. Based on the pre-registered analysis plan, data from participants who reported that they were non-native English speakers were excluded prior to analysis. Eighty-seven participants reported their gender identity as male, fifty-seven reported their gender identity as female, one as Genderqueer, and two did not report their gender. The average age was 38 years (SD = 12). Though our planned sample size was 128 participants, due to oversampling the final sample size was 147 participants.

Materials and Procedure

To create the manipulations of interest, we created 40 sets of 4 critical images (totaling 160 critical images). Each set of 4 images contained two items from each of two basic level object categories that exhibited one of two distinct and contrasting features. The features and objects were carefully selected to afford either pre-nominal or post-nominal modification (cf., Edwards & Chambers, 2011). Item sets were designed such that one feature (e.g. dotted) matched two of the four referents, and a second feature (e.g. striped) matched the other two referents, as illustrated in Table 1:

	Target	Early competitor	Late competitor	Non-competitor
pre-nominal: *Click on the dotted bag*	dotted bag	dotted bowtie	striped bag	striped bowtie
post-nominal: *Click on the bag with dots*	bag with dots	bag with stripes	bowtie with dots	bowtie with stripes

Table 1. <u>Experiment 1</u>: Illustration of experimental manipulation with example item set. The 4 critical images in the pre- and post-nominal conditions are the same; for illustration purposes, the images are described using the corresponding modification frame. Across experimental lists, we counterbalanced which of the 4 critical images was the target, whether it was referenced using pre- or post-nominal modification, and which image versions were viewed at study. The filler items for this set were ice cream cones.

To depict these items, we collected a large set of photographic and clipart images from the internet. These images were reviewed in-lab multiple times in order to select two clear depictions of each of the 160 critical items[1]. One version of each of the 160 critical items was shown to participants during the study trials; participants later viewed both versions at the test phase and were tasked with distinguishing the image they had seen from the one they had not seen. Each image set was paired with 2 unrelated filler images for a total of 6 images per set (Figure 1). These filler images were included to make the experimental manipulation less noticeable. In sum, the final critical image set consisted of 320 critical images, with two exemplars of each of the 160 critical items, plus 80 filler images.

The first phase of the task was the study phase, during which participants were shown 40 image sets, one at a time. On each trial, the 6 images in the set were arranged in a 3 x 2 array, with the position of the individual images randomly arranged within the array. Including the critical and filler images, across the 40 study trials, participants viewed a total of 240 images. The 40 study trials appeared in a different random order for each participant. During each trial, participants clicked a button that played auditory instructions telling them to select one of the six images on the screen. The critical noun phrase was either pre-nominally modified, e.g. *Click on the dotted bag*, or post-nominally modified, e.g. *Click on the bag with dots*. Given the target referent and the form of the noun phrase, within each image set one item was designated as the target, one image matched the initial part of the referring expression (the

---

[1] A list of the 160 critical items, along with the raw data are available at (*OSF link to be added upon publication*). Images are available upon request from the 2nd author.

"early-competitor"), one matched the latter half of the referring expression (the "late-competitor"), one matched a feature of each of the competitors but not the target (the "non-competitor"), and two were fillers. The auditory instructions were pre-recorded by a female research assistant at a steady speaking rate, on average, 2.9 words per second. Once the participant clicked on the named target image, the screen advanced to the next trial (which featured a new set of 6 images, and corresponding audio instruction).
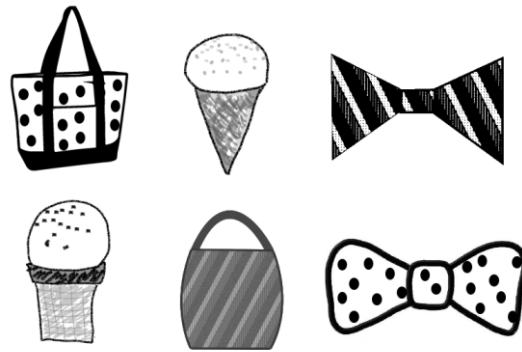

Figure 1: Illustration of an example picture grid for the test sentence "Click on the dotted bag" or "Click on the bag with dots".

Following the study phase, participants completed a series of ten arithmetic questions. This task took about five minutes and was intended to serve as a filled delay in order to bring performance on the recognition memory test off ceiling, as memory for images tends to be excellent (Shepard, 1967).

Following the math questions, participants completed a 2-alternative forced-choice (2AFC) recognition memory test for the images they had viewed in the first phase of the task. The 2AFC test was comprised of 160 trials. On each trial, participants viewed a pair of images on screen, one of which was "old" and had been seen in the first phase of the experiment, and the other which was "new". The new image was always an alternative version of the same picture that the participant had not seen (Figure 2). Participants were instructed to select the "old" image before proceeding onto the next trial (they were not given feedback on this

response). Participants then clicked a button in the bottom right-hand corner of the screen to proceed to the next trial. The test trials were presented to participants in a set random order, and the "old" item was equally likely to be on the left vs. right side of the screen.
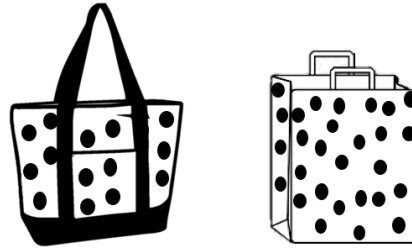

Figure 2: Illustration of an example 2AFC memory test trial; participants were instructed to select the "old" item that had been seen in the study phase.

Sixteen lists were used to counterbalance which of the 4 critical images within a set was the target image (e.g., the dotted/striped bag/bowtie), whether the critical expression was pre- or post-nominally modified, and which version of the critical images was seen in the study phase. Participants were randomly assigned to complete the trials on one of the sixteen lists.

Predictions

If temporary ambiguity among candidate meanings drives the previously reported memory boost for non-referenced items in the referential context (Yoon, et al., 2016), both early and late competitors will be recognized better than items that matched the competitors but not the expression. If activation patterns that privilege early-competitors in spoken word recognition (Allopenna, et al., 1998, Magnuson, et al., 2003) also shape memory when processing referential phrases, early competitors will be remembered better than late competitors. Note, the use of both pre-nominal and post-nominal modification allows us to separate effects of early vs. late activation, from referential form. For pre-nominally modified expressions, interpretation of "*the dotted bag*" should produce better memory for the dotted bowtie than the striped bag. By

9

contrast, for post-nominally modified expressions, "*the bag with dots"*, we predict better memory for the bag with stripes.

Alternatively, memory for non-referenced items in the referential context may be determined by semantic and form-based relationships among the objects. After all, items that do not match the properties of the referent shape referential processes (Fernald, et al., 2010; Grodner & Sedivy, 2011; Olson, 1970), and items semantically-related to the referent become activated during referential processing (Yee & Sedivy, 2006; Yee, et al., 2011). If so, contextual encoding may be unrelated to temporary ambiguity, and instead reflect activation of items meaningfully related to the referent.

Analysis and Results

Accuracy during the 2AFC memory test (Figure 3) was analyzed using a mixed-effects logistic regression analysis with the glmer function in lme4 (Bates, et al., 2018). We used the buildmer function (Voeten, 2020) to identify a parsimonious random effects structure (see Matuschek, et al. 2017). This analysis indicated a random intercept model was a good fit. Whether the expression was pre-nominally or post-nominally modified was included as a mean-centered fixed effect. The match between the expression and candidate referents was coded using Helmert contrasts: The first contrast compared memory for Targets vs. non-targets. The second compared memory for Early- and Late-competitors vs. non-competitors. The third compared memory for Early- and Late-competitors.
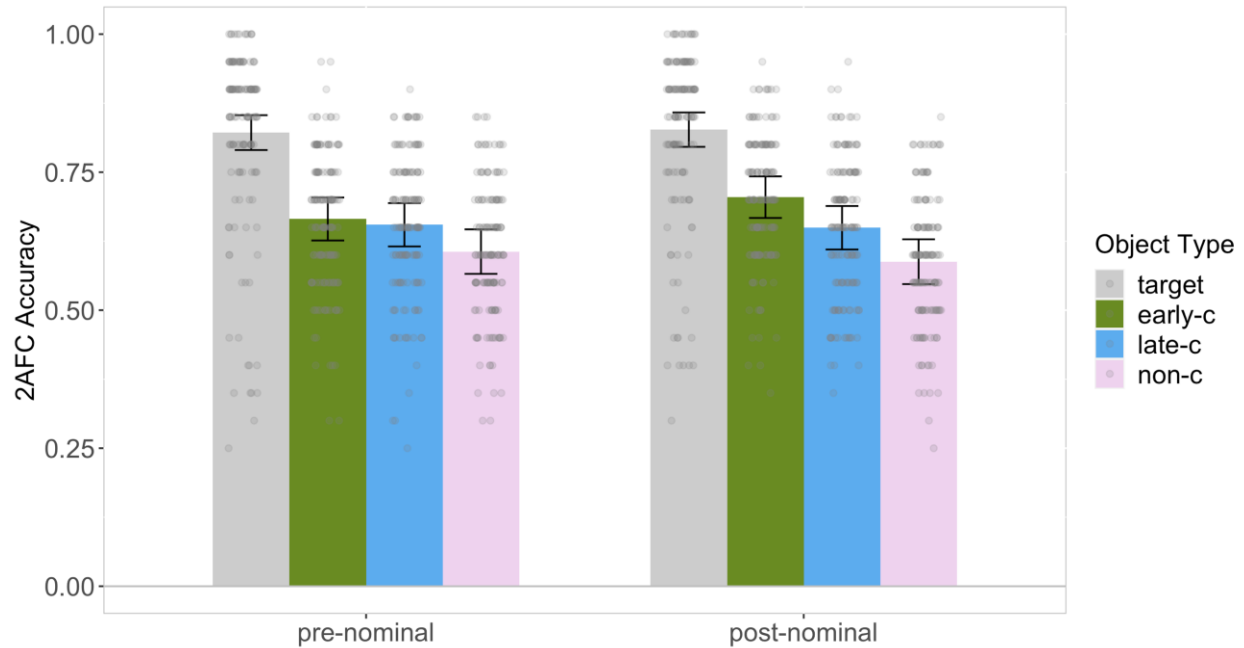
Figure 3: Experiment 1, 2AFC accuracy across conditions. Error bars indicate by-participant standard error of the mean. Individual points indicate by-participant condition means.

The results (Table 2), revealed a significant intercept (b = .88, p<.0001), reflecting the finding that participants were more accurate than not at recognizing previously-viewed images. Accuracy was higher for previously-referenced Targets than non-target context images (*b* = -0.99, *p*<.0001). Accuracy was higher for competitors (Early and Late) than non-competitors (*b* = .33, *p*<.0001). Further, early-competitors were remembered better than late-competitors (*b* = .16, *p*<.0001). Both competition effects significantly interacted with modification type. Exploration of these interactions revealed that for pre-nominal modifiers, competitors were remembered significantly better than non-competitors (*b* = .24, *p*<.0001), however memory for early and late competitors was not significantly different (*b* = .05, *p*=.40). In contrast, for post-nominal modifiers, there was a strong competition effect (*b* = .41, *p*<.0001), and better memory for early- vs. late-competitors (*b* = .27, *p*<.0001).

Fixed Effects	Estimate	SE	z-value	p-value
(Intercept)	0.881	0.043	20.389	<.0001
Targets (-.75) vs. Early, Late, and Non-competitors (+.25)	-0.993	0.038	-25.836	<.0001
Early and Late Competitors (+.333) vs. Non-competitors (-.666)	0.328	0.034	9.704	<.0001
Early (+.5) vs Late (-.5) competitors	0.157	0.040	3.913	<.0001
Pre (+.5) vs. Post (-.5) modification	-0.032	0.030	-1.068	0.286
Target*PrePost	0.012	0.077	0.159	0.874
Competitor*PrePost	-0.169	0.068	-2.495	0.013
EarlyLate*PrePost	-0.219	0.080	-2.736	0.006
Random Effects	Variance	Std.Dev.		
Item	0.089	0.298		
Participant	0.160	0.400		

Table 2. Experiment 1: Results of logistic mixed effects model of 2AFC accuracy. 23680 observations, 160 Items, 148 Participants.

Discussion

Consistent with the hypothesis that temporary ambiguity among candidate meanings drives memory for items in the referential context, we observed a memory boost for competitors that temporarily matched the referring expression, over those that did not. In addition, early competitors were remembered better than late competitors, suggesting the enhanced activation of early-competitors in spoken language processing (Allopenna, et al., 1998; Magnuson, et al. 2003), shapes memory for the referential context. A significant interaction with utterance form revealed that this early vs. late competitor memory boost emerged only when expressions were post-nominally modified.

One explanation is that in English, post-nominal modification is infrequent (Brown-Schmidt & Konopka, 2008; 2011). Thus, post-nominal modifiers may have been perceived as marked. Indeed, post-nominal modification is associated with increased referent salience (Karimi, Diaz, & Ferreira, 2019). This enhanced salience may result in a primary distinctiveness effect (von Restorff, 1933), conferring better memory much like auditory oddballs are

remembered better (Fabiani, Karis, & Donchin, 1986). If so, post-nominal modification may have enhanced the relative salience of early- vs. late-competitors. Another possibility is that post-nominal modification boosted memory for early-competitors because the early-competitor was activated first *and* the same type of object as the target, forming a contrast set (e.g., bag with stripes/dots; Sedivy et al., 1999). In contrast, for pre-nominal modification, while the early competitor was activated early, the late competitor formed a contrast set with the target (e.g., striped/dotted bag). If so, these different factors may have resulted in a similar memory boost for early- and late-competitors.

Finally, we can entertain a simpler explanation based on timing. Post-nominal modification resulted in a longer average period of initial ambiguity in our stimuli (e.g., "bag with", 800ms) compared to pre-nominal modifiers (e.g., "dotted", 610ms). Prior work examining *reading* of temporarily ambiguous sentences shows that readers are more likely to maintain misinterpretations of sentences when the period of ambiguity is longer (Christianson, et al., 2001). Similarly, Karimi, et al. (2020) report that reading times are faster following modified vs. non-modified noun phrases, suggesting it is the amount of *time* per se the reader thinks about a noun that enhances attention to and subsequent retrieval of it. A longer period of temporal ambiguity may also increase the chance of a fixation to the competitor, thereby increasing memory for it (Loftus, 1972), a point we return to in the General Discussion.

If referential activation during the period of temporal ambiguity is responsible for the competitor-memory boost, intentionally lengthening the ambiguity should exaggerate this memory advantage for early- over late-competitors. We test this hypothesis in Experiment 2.

Experiment 2

Methods

This experiment was preregistered on the Open Science Framework (https://osf.io/pvyrw). The experimental design was similar to Experiment 1; changes are detailed below.

Participants

The sample size was determined based on a-priori simulation-based power analyses using the simr package in R (Peter et al., 2019). That analysis revealed that a planned sample size of 128 would result in over 90% power to detect the effect of early vs. late competitors that was observed in Experiment 1.

Participants were recruited through Amazon Mechanical Turk (mTurk). As in Experiment 1, participation was restricted to persons who were native English speakers, located in the United States, at least 18 years old, and who had over 500 approved HITS and at least a 95% acceptance rate on MTurk. Data from participants who reported themselves to be non-native English speakers were excluded prior to analysis. The final sample size submitted to analysis was 128 participants; eighty-four reported their gender as male and forty-four reported their gender as female, and the average age was 39 years (SD=11).

Materials

To create the manipulations of interest we used the same 40 sets of 4 critical images as in Experiment 1 (totaling 160 critical images). Recall that in Experiment 1, we counterbalanced across 16 lists which of the 4 critical images in the set was the target, whether the critical expression was pre- or post-nominally modified, and which version of each image participants viewed in the study phase. Because Experiment 2 added a manipulation of speech rate, it was necessary to simplify the counterbalancing scheme to avoid an unwieldy number of lists and audio files to record. Thus, for each item set only two of the four images were used as targets across the different experimental lists. The two possible target items were selected in a way that allowed us to counterbalance, across lists, which image was the early-competitor and which

image was the late-competitor. As before, in each item set, one feature matched two of the referents, and the other feature matched the other two referents, as illustrated in Table 3.

	Target	Early competitor	Late competitor	Non-competitor
Target 1 pre-nominal: *Click on the dotted bag*	dotted bag	dotted bowtie	striped bag	striped bowtie
Target 1 post-nominal: *Click on the bag with dots*	bag with dots	bag with stripes	bowtie with dots	bowtie with stripes
Target 2 pre-nominal: *Click on the striped bowtie*	striped bowtie	striped bag	dotted bowtie	dotted bag
Target 2 post-nominal: *Click on the bowtie with stripes*	bowtie with stripes	bowtie with dots	bag with stripes	bag with dots

Table 3. Experiment 2: Illustration of experimental manipulation with example item set. The 4 critical images in the pre- and post-nominal conditions are the same; the labels are provided in pre- or post-nominal form for explanatory purposes. Across experimental lists, we counterbalanced the target image, whether it was referenced using pre- or post-nominal modification, the speech rate, and which image versions were viewed during the study phase.

The images used in Experiment 2 were the same as those used in Experiment 1. As before, one version of each of the 160 critical items was shown to participants during the study trials. During the study phase of the task, participants were shown 40 image sets, one at a time in a 3 x 2 array. Including the critical and filler images, across the 40 study trials, participants viewed a total of 240 images. The 40 study trials appeared in a different random order for each participant. During each trial, participants pushed a button to play the instruction indicating which image to select of the six images on the screen. Participants had to select the correct image before continuing on to the next trial. Selecting an incorrect image generated an error response and instructions for participants to "try again". The form of the critical noun phrase was either pre-nominally modified, e.g., *Click on the dotted bag*, or post-nominally modified, e.g., *Click on the bag with dots*. The auditory instructions were pre-recorded by the first author at two different speaking rates using an online metronome tool to keep time. For each participant, half the trials featured the slower rate (on average, 1.52 words per second), and the other half of

trials featured the faster rate (3.19 words per second). As in Experiment 1, the trial order was randomized.

As in Experiment 1, participants next completed a series of math questions, followed by a surprise 2AFC memory test (the test was identical to Experiment 1). We created sixteen experimental lists (see Table 3) to counterbalance the target item utterance form (pre- vs. post-nominal modification), speech rate (slow vs. fast), and which image set was viewed at study (version one vs. version two). Each participant completed the trials on a single list.

Predictions

Experiment 2 tests the hypothesis that the amount of time a referential expression is consistent with a non-referenced item in the referential context determines how well it will be remembered. On this lexical activation-time hypothesis, we predict better memory for competitors and an enhanced advantage for early over late competitors in the slow speech condition. Alternatively, if the asymmetry between pre- and post-nominally modified expressions in Experiment 1 was due to the fact that post-nominal constructions are less frequent or because the noun is mentioned first, we would expect a similar pattern of findings as Experiment 1, with no effect of speech rate.

Analysis and Results

The data were analyzed in a mixed-effects model (Figure 4). The buildmer function (Voeten, 2020) indicated a model with intercepts by participants and items, and a random by-participants slope for the Speed effect was a good fit to the data. Referential form (pre-nominal vs. post-nominal modification), and audio speed (slow vs. fast) were included as mean-centered fixed effects. The match between the expression and candidate referents was coded using Helmert contrasts.

Fixed Effects	Estimate	SE	z-value	p-value
(Intercept)	**0.857**	**0.047**	**18.337**	**<.0001**
Targets (-.75) vs. Early, Late, and Non-competitors (+.25)	**-1.005**	**0.051**	**-19.527**	**<.0001**
Early and Late Competitors (+.333) vs. Non-competitors (-.666)	**0.358**	**0.059**	**6.099**	**<.0001**
Early (+.5) vs. Late (-.5) competitors	**0.247**	**0.044**	**5.674**	**<.0001**
Slow (+.5) vs. Fast (-.5) speed	**0.149**	**0.073**	**2.045**	**0.041**
Pre (+.5) vs. Post (-.5) modification	-0.007	0.032	-0.227	0.820
Target*Speed	0.097	0.083	1.174	0.241
Competition*Speed	-0.006	0.074	-0.088	0.930
EarlyLate*Speed	**0.266**	**0.087**	**3.048**	**0.002**
Target*PrePost	-0.031	0.082	-0.374	0.709
Competition*PrePost	-0.048	0.073	-0.657	0.512
EarlyLate*PrePost	**-0.177**	**0.087**	**-2.038**	**0.042**
Speed*PrePost	0.064	0.064	0.997	0.319
Target*Speed*PrePost	-0.252	0.165	-1.534	0.125
Competition*Speed*PrePost	-0.064	0.147	-0.437	0.662
EarlyLate*Speed*PrePost	-0.169	0.174	-0.968	0.333
Random Effects	Variance	Std.Dev.	Correl.	
Item (intercept)	0.084	0.289		
Participant (intercept)	0.178	0.422		
Speed (slope)	0.543	0.737	-0.240	

Table 4. Experiment 2: Results of logistic mixed effects model of 2AFC accuracy. 20480 observations, 160 Items, 128 Participants.

A significant intercept (b = .86, p<.0001), indicated that participants were more accurate than not (Table 4). Accuracy was higher for Targets than non-targets (*b* = -1.01, *p*<.0001), and for competitors than non-competitors (*b* = .36, *p*<.0001). Early-competitors were better remembered than late-competitors (*b* = .25, *p*<.0001). Referential form (pre vs. post) interacted with competitor type (*b* = -.18, *p*=.04): the memory advantage for early- over late-competitors was larger for post-nominal (*b* = .33, *p* <.0001) than pre-nominal modifiers (*b* = .16, *p*=.010).

Unlike Experiment 1, the memory advantage for competitors over non-competitors was not significantly different between pre-nominally and post-nominally modified phrases (*b* = -.05, *p*=.512).

Finally, consistent with the lexical activation-time hypothesis, memory was better when speech was slow vs. fast (*b* = .15, *p*=.041). Critically, competitor type (early vs. late) interacted with speech rate (*b* = .27, *p*=.002): a significant memory advantage for early- over late-competitors was present for slow speech (*b* = .38, *p* <.0001), but not fast speech (*b* = .11, *p*=.066).
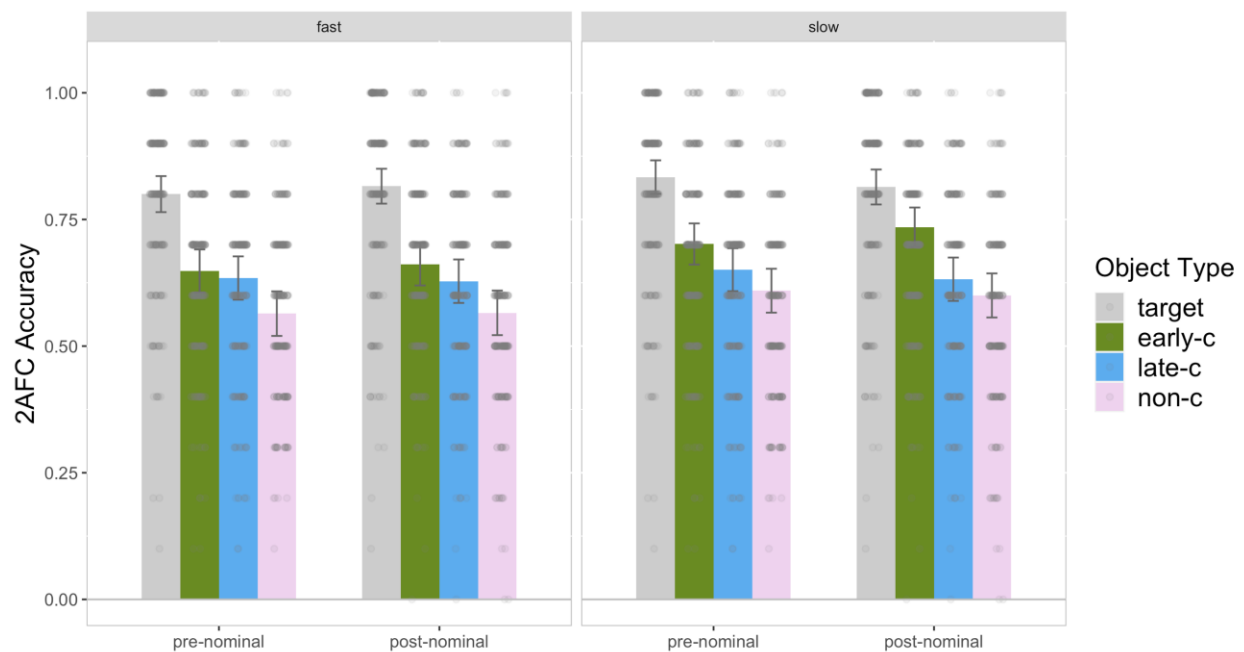

Figure 4: Experiment 2, 2AFC accuracy by condition. Error bars indicate by-participant standard error. Individual points indicate by-participant condition means.

General Discussion

Theories of language processing observe that language use and understanding are shaped by both the immediate referential context and memory for past contexts (Clark & Wilkes-

Gibbs, 1986; Tanenhaus, et al. 2000; Yoon, et al. 2021). Converging evidence from empirical studies and computational models of spoken language understanding demonstrate that language is interpreted incrementally, with listeners continuously mapping the unfolding speech stream to multiple candidate meanings (Allopenna, et al., 1998; Eberhard, et al., 1995; Sedivy, et al., 1999). Somewhat surprisingly, then, little is known about the consequences of incremental processing on enduring memory for these candidate meanings.

The present research demonstrates that when interpreting a referring expression, temporary ambiguity among candidate referents makes those candidate referents more memorable. Consistent with evidence of earlier and stronger activation of early competitors in spoken word recognition (Allopenna, et al., 1998; Creel, et al., 2008; Magnuson, et al. 2003; Sedivy, 2003), we observed better memory for early- than late-competitors when expressions were post-nominally modified, and when speech was slow. These findings offer support for the activation-time hypothesis of the link between language processing and memory for items in the referential context. Much in the same way that misinterpretations of ambiguous sentences are more likely to be retained when the period of ambiguity is long (Christianson, et al., 2001; Ferreira, Lau, & Bailey, 2004; Lau & Ferreira, 2005), our findings show that the longer the period of time a referring expression is consistent with a candidate referent, the better the memory for the referent.

The hypothesis that temporary activation of candidate referents boosts memory for them leads to specific predictions regarding subsequent language processing and memory. A variety of linguistic and non-linguistic factors activate candidate meanings (Chambers, et al. 2002; Yee & Sedivy, 2006); we predict enhanced memory for activated items that reflects the degree of activation. Gaze at objects is associated with better memory for them (Loftus, 1972) thus trial-by-trial analyses that relate gaze to subsequent memory is expected to reveal better memory for candidate referents that the listener fixated. Yet referential activation is not isomorphic with gaze, as listeners activate non-pictured referential candidates both in studies where the speaker

names objects in visual displays, and when spoken language is not linked to objects in the visual world (Magnuson, et al. 2007; Dahan, et al. 2001; Van Petten, et al., 1999). If temporary activation of candidate referents is the mechanism driving the observed competitor-memory boost, listeners should retain in memory representations of candidate meanings even when they are not fixated or visually presented on-screen. Quantifying the predicted contribution of activation time to the observed memory boost, apart from fixation-driven memory, would likely involve references to absent objects (Saylor & Ganea, 2007), which are readily interpreted despite the absence of a co-present referent in the immediate context.

Lastly, a consequence of enhanced memory for temporarily activated candidate referents is the possibility that they will be *mis*-remembered as having been referenced. Exposure to falsehoods increases people's belief in them even when they contradict prior knowledge (Fazio, et al., 2013). This illusory truth effect is *enhanced* when presentation rates are slowed (Fazio & Marsh, 2008). The prediction, then, is that undesired meanings that are temporarily activated during spoken language processing may be nonetheless retained in memory and later believed. If so, when precision is important, speakers may wish to avoid phrasings that activate problematic alternative meanings (e.g., "peanut" in "*peanut-free sandwich"*, cf. "*sun-butter sandwich"*).

Conclusion

Spoken language is interpreted incrementally. As the words of a referring expression unfold over time, listeners activate multiple candidate meanings based on the items in the referential context. The present research provides insight into the implications of temporary ambiguity in language processing for subsequent memory for what was – and what could have been – said. Two experiments demonstrate a clear link between the amount of time candidate referents in the referential context were temporarily consistent with spoken referential expressions, and later memory. We posit an activation-time hypothesis, which argues that

factors that increase the amount of time items in the referential context are considered, will

increase memorability of those items.

Acknowledgements

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419-439.

Altmann, G., & Steedman, M. (1998). Interaction with context during human sentence processing. *Cognition*, *(3):191-238*.

Bates, D., Mäachler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., . . . Green, P. (2018). Package "lme4": Linear mixed-effects models using 'eigen' and s4. Retrieved from https://cran.r-project.org/web/packages/lme4/lme4.pdf

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

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., & Konopka, A. E. (2008). Little houses and casas pequeñas: message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition, 109*(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

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.

Burt, J. S., McFarlane, K. A., Kelly, S. J., Humphreys, M. S., Weatherall, K., & Burrell, R. G. (2017). Brand name confusion: Subjective and objective measures of orthographic similarity. *Journal of Experimental Psychology: Applied, 23*(3), 320–335. https://doi.org/10.1037/xap0000127

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of memory and language*, *47*(1), 30-49.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology*, *42*(4), 368-407.

Clark, H. H., and Marshall, C. R. 1978. "Reference Diaries." In D. L. Waltz, ed. *Theoretical Issues in Natural Language Processing-2. New York: Association for Computing Machinery*, pp. 57-63.

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

Connine, C. M., Blasko, D., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30, 234–250*.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: the role of talker variation in lexical access. *Cognition*, *106*(2), 633–664.

Dahan, et al. 2001: Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5-6), 507–534.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *J Psycholinguist Res* 24, 409–436. https://doi.org/10.1007/BF02143160

Edwards, J. D., & Chambers, C. G. (2011). It's not what you said, it's how you said it: How modification conventions influence on-line referential processing. *The processing and acquisition of reference*, 219.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology, 23(3), 298–308*.

Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, *15*(1), 180-185.

Fazio, L. K., Barber, S. J., Rajaram, S., Ornstein, P. A., & Marsh, E. J. (2013). Creating illusions of knowledge: Learning errors that contradict prior knowledge. *Journal of Experimental Psychology: General*, *142*, 1-5.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive psychology*, *60*(3), 190-217.

Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and Tree Adjoining Grammars. *Cognitive Science, 28(5)*, 721–749.

Foraker, S., & McElree, B. (2011). Comprehension of Linguistic Dependencies: Speed-Accuracy Tradeoff Evidence for Direct-Access Retrieval From Memory. *Language and linguistics compass*, *5*(11), 764-783.

Fukumura, K., & Carminati, M. N. (2021). Overspecification and incremental referential processing: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Grodner, D., & Sedivy, J. C. (2011). The Effect of Speaker-Specific Information on Pragmatic Inferences. In: *The processing and acquisition of reference* (Vol. 2327, pp. 239-272). MIT Press.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition, 96*(2), 127–142.

Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream Behavioral and Electrophysiological Consequences of Word Prediction on Recognition Memory. *Frontiers in human neuroscience*, 13, 291. https://doi.org/10.3389/fnhum.2019.00291

Karimi, H., Swaab, T. Y., & Ferreira, F. (2018). Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *Journal of Memory and Language*, *102*, 68-82.

Karimi, H., Diaz, M., & Ferreira, F. (2019). "A cruel king" is not the same as "a king who is cruel": Modifier position affects how words are encoded and retrieved from memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(11), 2010.

Karimi, H., Diaz, M., & Wittenberg, E. (2020). Sheer Time Spent Expecting or Maintaining a Representation Facilitates Subsequent Retrieval during Sentence Processing. *Presentation at CogSci.*

Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, *20*(5), 633-666.

Lieberman, A. M., Borovsky, A., & Mayberry, R. I. (2018). Prediction in a visual language: real-time sentence processing in American Sign Language across development. *Language, cognition and neuroscience*, *33*(4), 387-401.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive psychology, 3(4),* 525-551.

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1400–1416.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202.

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(1), 133–156*.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language, 94*, 305-315.

McKinley, G.L., Brown-Schmidt, S., & Benjamin, A.S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition, 45,* 1281-1294.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language, 60*(1), 65–91.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological review, 77(4),* 257.

Pechmann T. (1989). Incremental speech production and referential overspecification. *Linguistics 27* 89–110. 10.1515/ling.1989.27.1.89

Peter, G., Catriona, M., Phillip, A. (2019). Package "simr": Power Analysis for Generalized Linear Mixed Models by Simulation. Retrieved from https://cran.r-project.org/package=simr

Saylor, M. M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Developmental Psychology*, *43*(3), 696.

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

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

Shepard, R. N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations and linguistic processing. *Journal of psycholinguistic research*, *29*(6), 557–580. https://doi.org/10.1023/a:1026464108329

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 394–417.

Voeten, C.C. (2020). Package "buildmer": Stepwise Elimination and Term Reordering for Mixed-Effects Regression. Retrieved from https://cran-r-project.org/package=buildmer

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1.

Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: Activation of shape and function features during object identification. *Journal of Experimental Psychology: General*, *140*(3), 348.

Yoon, S. O., Benjamin, A. S., and Brown-Schmidt, S. (2016). The historical context in conversation: Lexical differentiation and memory for the discourse history. *Cognition, 154*, 102-117

Yoon, S. O., Benjamin, A. S. & Brown-Schmidt, S. (2021). Referential form and memory for the discourse history. *Cognitive Science*. https://doi.org/10.1111/cogs.12964