RUNNING HEAD: Women and men have different biases

# Women and men have different discourse biases for pronoun interpretation

Jennifer E. Arnold

Department of Psychology, University of North Carolina at Chapel Hill

Address all correspondence to: Jennifer Arnold, UNC Chapel Hill, Dept. of Psychology, Davie Hall #337B, CB #3270, Chapel Hill, NC 27599-3270, Email: jarnold@email.unc.edu

Abstract

Two experiments examine how men and women interpret pronouns in discourse. Adults are known to show a strong "first-mention bias": When two characters are mentioned (*Matthew played with James…),* comprehenders tend to interpret subsequent pronouns as coreferential with the first of the two characters, and to find pronouns more natural than names for reference to the first character. However, this bias is not absolute. Experiment 1 demonstrates a stronger first-mention bias for women than men in their naturalness ratings for short stories. Experiment 2 monitors eye movements during story comprehension, and finds that women are more likely than men to consider the first-mentioned character as the pronoun referent. These findings reveal the first known sex difference in reference processing, and reinforce the view that reference processing is driven by more than the discourse context alone.

Women and men are clearly different, and it is popularly assumed that they use language differently. Yet much is still unknown about how the cognitive mechanisms underlying language use differ, particularly in the domain of discourse processing. This paper examines one of the core aspects of language: how we refer and understand references. I focus on third-person pronouns (*he, she*), which require people to draw inferences about their meaning based on the discourse and real-world context. We know that adults rapidly interpret pronouns, based on a variety of cues from the discourse context. The question examined in this paper is whether men and women differ in their biases about what makes a pronoun appropriate for a particular referent.

Pronouns are often ambiguous, even in context. Consider this story: *Aurora had lunch with Fiona. She made a big mess.* Adults prefer to interpret *she* as coreferential with the first character (Aurora), which in English is usually also the grammatical subject (for a review see Arnold, 2008). Similarly, speakers will tend to use a pronoun more for the first than second character. This first-mention / subject bias is not absolute, however, and can be modulated by other factors, such as the semantics of the verb in the first sentence, and how it relates to the second sentence (e.g., Garvey & Caramazza, 1974; Kehler, 2002). For example, if the second character (*William*) is the goal of an action (*Michael throw the ball to William. He…)*, he is somewhat better as the referent of a subsequent pronoun, especially if the utterance describes the subsequent event (Arnold, 2001; Rohde, Kehler, & Elman, 2007; Stevenson, Crawley, & Kleinman, 1993).

The traditional approach to explaining pronoun production and comprehension focuses on the role of discourse features like the first-mention and verb biases, which make some entities more prominent in the discourse than others. Speakers prefer pronouns for reference to

prominent entities (e.g., Chafe, 1976; Gundel, Hedberg, and Zacharski, 1994), and comprehenders preferentially assign pronouns to prominent referents (Gordon, Grosz, & Gilliom, 1993; Jarvikivi, van Gompel, Hyona, & Bertram, 2005). An extension to this view comes from recent evidence that the effect of the discourse context is mediated by other cognitive factors. Speakers are less likely to use pronouns if they are under cognitive load (e.g., if they are being disfluent; Arnold & Griffin, 2007; Arnold, Bennetto, & Diehl, 2008). Listeners' initial biases are driven by both the discourse context and their own attention (Arnold & Lao, under review). The current paper examines another way in which internal cognitive biases impact processing, by identifying how women and men may differ in their use of the discourse context.

Two experiments test whether men and women differ in their sensitivity to the known first-mention bias. Known sex differences in language make few predictions about this question. Women have been reported to use pronouns more than men, especially first-person pronouns (*I, we*; e.g., Newman, Groom, Handleman & Pennebaker, 2007), but the discourse properties of the referents were not measured. Men use acoustically reduced forms more than women (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009), which might predict more pronoun use for men. But men also are more disfluent than women (Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, and Gildea, 2003), which might predict fewer pronouns for men. However, neither of these findings suggest differential sensitivity to the discourse context. Neither is it clear how referential processing would be impacted by sex differences in sociolinguistic style (Eckert & McConnell-Ginet, 2003; Tannen, 2001), phonological processing (Irwin, Whalen, & Fowler, 2006; Majeres, 1999) or language localization in the brain (e.g., Jaeger et al., 1998).

## EXPERIMENTS 1: OFFLINE PRONOUN PREFERENCES

Experiment 1 tested the first-mention bias by asking participants to rate two-sentence stories for naturalness.[1] In critical items, the second sentence began with either a pronoun or a name, which referred to either the first or second character from the first sentence. The N1 bias is predicted to impact pronoun ratings more than names, such that stories are rated higher if they have pronouns with N1 referents than N2 referents. The critical question is whether this pattern differs for men and women.

[INSERT TABLE 1 ABOUT HERE]

### Method

*Participants*

The analysis included 80 participants, who received course credit. An additional 12 women and 12 men were excluded: 1 was a non-native speakers of English, 1 reported a cognitive disorder, 1 had no response variation, 9 missed too many comprehension questions, and 4 had unusual ratings for the critical catch fillers. 8 participants were excluded to even out the lists, always choosing the last participant(s) run on a particular list.

*Experiment Design and Stimuli*

Stimuli consisted of two-sentence stories. In the 40 experimental stimuli, the first sentence mentioned two human characters (see Table 1). The two characters always had the same gender (half female), and their names were matched by length in syllables. The first

[1] An additional study was conducted in which participants read the stories and made an explicit forced choice between a pronoun and name. The pattern of results was similar to the one reported here for the rating experiment.

sentence described an event in which the verb was expected to create a bias toward either the first or second NP, based on previous research (e.g., Arnold, 2001; Kehler, 2002). The N1-biased verbs included 8 Goal-Source verbs like *grabbed, accepted*; 9 Stimulus-Experiencer verbs like *inspired, amazed;* and 8 Joint Action predicates like *went sledding with, played basketball with*). The N2-biased verbs included 8 Source-Goal verbs like *brought to, sent to,* and 7 Experiencer-Stimulus verbs like *criticized, trusted*).

The second sentence began with a reference to one of the two characters, using either a pronoun or a name. Thus, the design was 2x2 (N1 vs. N2; pronoun vs. name). These items were organized into four lists (with forwards and backwards orders) in a Latin square design.

An additional 20 items were stories about a single character (e.g., *Paige listened to a book on CD. {Paige/She} fell asleep and missed the ending*). Reference form (pro vs. name) was manipulated for these items such that each item appeared on two lists with each form. However, this condition did not include the critical order and verbtype manipulations, and therefore is excluded from further discussion for simplicity.[2]

In the 40 fillers, two different-gender characters were mentioned in the first sentence, either as a conjoined subject (e.g., *Ethan and Rachel danced at the concert*,) or in subject and nonsubject positions, using a variety of event types (e.g., *Audrey invited Zack to the prom*). The second sentence in the fillers began with a gender-disambiguated pronoun referring to N1 or N2, or a word referring to something else.

Two measures assessed whether participants followed instructions: 1) 10 critical fillers had pragmatically awkward continuations, and 3 had grammatical errors. The four participants who rated these more highly on average than the other fillers were excluded from analysis. 2) 12

[2] The average ratings in this condition for women were 4.6 (names) and 5.5 (pronouns); for men they were 4.9 (names) and 5.4 (pronouns).

items were followed by a comprehension question; participants with more than 4 wrong were excluded.

*Procedure*

The experiment was conducted using the Psychology Pool online software, such that participants could complete the experiment on the web at a location of their choosing.

**Results and Discussion**

Ratings were generally high for stories with names, whereas pronoun items received higher ratings when the pronoun referred to N1. Of particular interest was the fact that women's ratings for pronoun and name conditions were more extreme than men's (see Figure 1). Women had very low ratings for items with pronouns referring to N2 (3.8), whereas men found these items relatively more natural (4.3). Women even had a slight preference for N1 pronouns (5.35) over N1 names (5.07), whereas men's ratings were virtually the same for N1 pronouns (5.26) and N1 names (5.20).

[FIGURE 1 ABOUT HERE]

The results were submitted to analyses of variance,[3] with both participants and items as random effects. The critical finding was a 3-way interaction between sex, referent (N1 vs. N2), and referential form (pro vs. name; $F1(1,78) = 4.316$, $p = .041$; $F2(1,38) = 6.987$, $p = .012$). This occurred in addition to the interaction between referent and referential form ($F1(1,78) = 54.931$,

[3] All analyses were also conducted with multilevel model logistic regressions, which revealed the same critical effects.

$p < .001$); $F2(1,38) = 86.221$, $p < .001$), a main effect of referent ($F1(1,78) = 79.962$, $p < .001$; $F2(1,38) = 80.945$, $p < .0001$), and a main effect of referential form ($F1(1,78) = 35.717$, $p < .001$); $F2(1,38) = 79.204$, $p < .001$). There was a main effect of gender, but only in the items analysis ($F1(1,78) = .423$, $p = .517$; $F2(1,38) = 6.673$, $p = .014$). There were no effects or interactions with verb bias, except for a main effect of verb bias in the subjects analysis only ($F1(1,78) = 7.931$, $p = .006$; $F2(1,38) = 1.109$, $p = .299$).

There was no indication that participants differed on any other measure of their performance. There were no sex differences on comprehension question accuracy ($p$'s $> .80$), and no difference in male/female ratings across different filler types ($p$'s $> .11$).

In sum, women had a greater preference for stories using pronouns to refer to N1, the first-mentioned / subject referent. By contrast, men were more willing to accept pronouns for N2 than women. That is, the expected order-of-mention effect was stronger for women than men. There was no evidence that verbtype modulated either the sex difference, or the preference for pronouns over names.

## EXPERIMENT 2: ON-LINE PRONOUN COMPREHENSION

Experiment 1 demonstrated that men and women have different judgments about the naturalness of pronouns, in a task where they can spend plenty of time thinking about it. If these biases influence everyday language use, we would also expect differences during "on-line", moment-by-moment processing, as listeners seek out potential referents for pronouns. Experiment 2 addressed this question by monitoring participants' eye movements as they heard a story and decided if it matched a picture. The stories were about four characters: Doggy (male), Birdy (male), Bunny (female), and Kitty (female). In critical items, the first sentence mentioned two characters of the same gender, e.g. *Birdy picked apples with Doggy near the farmhouse*. The second (target) sentence in the story was always linguistically ambiguous, e.g. *he was wearing a hat to protect himself from the sun*. Both characters were consistent with the verb (i.e., both were wearing something), but only one matched the critical characteristic (wearing a hat). Therefore, participants' direction of gaze immediately following the pronoun revealed their initial preferences to consider each character as the referent (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000). Final responses indicated their willingness to accept the pronoun with the intended referent.

### Method

*Participants*

Fifty-one undergraduates participated for course credit. One was excluded because of track loss, one for poor accuracy on critical filler trials, one because she participated in

Experiment 1, and one who was familiar with my research. The analysis included 23 women and 24 men.

*Experiment Design and Stimuli*

*Eyetracking Task*. Participants' eye movements were monitored while they viewed a picture and heard a story (Figure 2). Using a paradigm adapted from Arnold, et al. (2000), each story introduced 2 same-sex characters in the first sentence, and referred to one with a pronoun in the second sentence. The visual stimulus manipulated whether the pronoun referred to the first- or second-mentioned (N1 vs. N2); e.g. in Figure 2 only one character has a hat.

[FIGURE 2 ABOUT HERE]

There were 20 experimental items in two conditions (N1 vs. N2). Eight of the verbs were N1-bias (joint action); twelve were N2-bias (dative). These were psuedorandomly combined with 8 fillers in two lists, with a forwards and backwards version. Right/left location of target and N1 were counterbalanced across items.

Participants judged whether the story matched the picture, responding by clicking on a button on screen after the story was over. All experimental stimuli matched the picture under the intended referential assignment. Of the 8 fillers, 6 had stories that were not intended to match the picture for non-pronoun reasons (e.g., *Doggy went boating one summer…,* when the picture showed Doggy snorkeling). There were two practice items: one that matched and one that did not. The practice and filler items had 1-3 characters, different structures and no ambiguous pronouns.

*Participant Matching Tasks*

Two measures ensured that the sample of males and females were similar in terms of general cognitive abilities: 1) Participants were asked to report on a questionnaire their SAT scores,[4] since SAT scores correlate with other measures of working memory (Engle, Kane, & Tuholski, 1999). 2) Working memory was measured with the kiosk-ready version of the operation span task (Unsworth, Heitz, Schrock, & Engle, 2005). Participants read math problems out loud (e.g., *(1*2) + 1 = ?*) and mentally solved them, while simultaneously remembering a letter (e.g., L). Their score was the sum of all perfectly recalled sets.

*Apparatus and Procedure*

An Eyelink II monitored participants' eye movements; the visual and auditory stimuli were presented on a PC running the ExBuilder software (Longhurst, 2006). Fixations were sampled every 4 msec and later collapsed into 16-msec samples (McMurray, 2002). Participants then completed a verbal questionnaire about the study, and the operation span task.

**Results and Discussion**

*Eyetracking Task*. Eye movement data are presented in terms of **looks**, where a look is defined as a fixation grouped together with the prior saccade (e.g., McMurray, Tanenhaus, & Aslin, 2009). Figure 3 illustrates proportion looks over time to both target and competitor characters, immediately following the pronoun. There was strong evidence for the N1 advantage: in the N1 condition there were more target looks than competitor looks, but in the N2

[4] A standardized aptitude test required for application to college in the U.S.

condition there were more competitor looks than target looks. Moreover, women showed a stronger N1 advantage than men, although this contrast emerged later than the contrast between the N1 and N2 conditions. Both men and women looked at the target more in the N2-biased items, for both N1 and N2 targets.

Statistical analyses were performed by calculating the percentage of looks to target and competitor over two time windows: 400-800 msec and 800-1200 msec after the pronoun (following Arnold et al, 2000). These were submitted to analyses of variance, with both participants and items as random effects[5].

*Target Looks Analyses.* From 400-800, the critical interaction between condition (N1 vs. N2) and gender only emerged in the items analysis, where it narrowly missed significance ($F1(1,45) = 1.80$, $p = .187$; $F2(1,18) = 4.33$, $p = .052$). There was also a main effect of referent ($F1(1,45) = 49.83$, $p < .001$; $F2(1,18) = 19.01$, $p < .001$) and a main effect of verb ($F1(1,45) = 18.928$, $p < .001$); $F2(1,18) = 10.98$, $p = .004$).

From 800-1200, the critical finding was an interaction between gender and referent ($F1(1,45) = 14.22$, $p > .001$; $F2(1,18) = 39.19$, $p < .001$). There was also a main effect of referent ($F(1,45) = 102.44$, $p < .001$; $F2(1,18) = 58.84$, $p < .001$), a main effect of verb bias ($F1(1,45) = 5.32$, $p = .026$, $F2(1,45) = 5.32$, $p = .033$), and a marginal verb x gender interaction in the items analysis only ($F1(1,45) = 2.22$, $p = .143$; $F2(1,18) = 3.17$, $p = .092$).

*Competitor Looks Analyses.* The analyses with looks to competitor as the dependent variable mirrored the target looks analyses. From 400-800 the referent x gender interaction was only significant by items ($F1(1,45) = 2.22$, $p = .1434$; $F2(1,18)$, and there was a main effect of referent ($F1(1,45 ) = 58.71$, $p < .001$; $F2(1,18) = 22.25$, $p < .001$). From 800-1200 the critical

[5] All analyses were also conducted with multilevel models (Barr, 2008), and revealed the same critical effects.

referent x gender interaction was significant ($F1(1,45) = 9.45$, $p = .004$; $F2(1,18) = 17.57$, $p < .001$), there was a main effect of referent ($F's > 60$), and a main effect of gender in the subjects analysis only ($F1(1,45) = 5.37$, $F2(1,18) = 2.59$).

[FIGURE 3 ABOUT HERE]

*Final Response*. After hearing each story, participants indicated whether the story matched the picture. For Mismatch responses, they provided a verbal explanation. We only counted Mismatch responses that were related to the pronoun (e.g., *Doggy was wearing the hat, not Birdy*). Mismatch responses that were irrelevant to pronoun resolution (e.g., *It's not sunny in the picture)* were counted as Matches. The proportion of Match responses in each condition was submitted to analyses of variance.

[TABLE 2 ABOUT HERE]

Both men and women provided Match responses to nearly all N1 target items ($M = 96\%$), and less often to N2 target items ($M = 20\%$), which resulted in a significant main effect of referent ($F1(1,45) = 473.6$, $F2(1,18) = 788.53$, $p's < .001$). Men were more likely to accept the N2 target (24%) than women (16%), but this emerged only a marginal main effect of gender in the items analysis only ($F1(1,45) = 1.925$, $p = .172$; $F2(1,18) = 3.617$, $p = .073$), and a marginal interaction between gender and referent in the items analysis only ($F1(1,45) = 1.559$, $p = .218$; $F2(1,18) = 3.546$, $p = .076$). Both sexes accepted N2 targets more for verbs with an N2-bias

(23%) than the N1-bias (17%), resulting in a condition x verbtype interaction, but only in the subjects analysis ($F1(1,45) = 4.51$, $p = .039$; $F2(1,18) = 2.449$, $p = .135$).

*Control Tasks.* There was no difference between men and women's performance on the operation span task ($F(1,45) = 1.89$, $p=.18$), or for the 45 participants who reported their total SAT score ($F(1,43)=.02$, $p=.88$). Women and men did not differ in their ability to correctly reject the filler items that mismatched the picture for nonreferential reasons (female $M = 91\%$; male $M = 89\%$). There was also no evidence that metalinguistic awareness differed across groups. The postexperiment questionnaire probed their awareness of the manipulation by asking increasingly leading questions. On the first two general questions (*What did you think the task was about?* and *Was there anything difficult about the task?*), 17% of women and 25% of men spontaneously mentioned pronouns.

In sum, women displayed a stronger first-mention bias than men during online spoken language comprehension. By contrast, men and women exhibited equal sensitivity to the verb bias.

## GENERAL DISCUSSION

In both a naturalness rating study (Experiment 1) and an on-line pronoun comprehension task (Experiment 2), women exhibited a greater preference for pronouns to refer to first-mentioned characters, and a greater dispreference for pronouns to refer to second-mentioned characters. This is the first observation that women and men have different reference processing biases, and it has potentially wide-reaching implications.

First, this finding suggests a functional difference in the way that men and women understand pronouns. In both experiments, women had a greater preference for pronouns to refer

to N1 characters than men. In experiment 2 this preference emerged as a stronger consideration of the N1 character following the pronoun. This suggests that women (compared to men) find pronouns easier to understand when they refer to N1, whereas men (compared to women) find pronouns relatively easier to understand when they refer to N2. This does not mean that women are "better" than men in their understanding of pronouns. However, it is more common for pronouns to refer to N1 than N2 (Arnold, 1998), which means that processing is facilitated for women in the more frequent case. In some situations, men and women may even differ in their final interpretation of the pronoun. In experiment 2, there was a marginal trend for men to say that the story and picture matched with N2 pronouns more than women. This effect may become stronger with a task where responses in the N1 condition are not at ceiling, or where responses need to be made more quickly.

Second, the results reported here demonstrate that discourse processing is an area of language processing in which we may find further sex differences. Despite the popular assumption that women and men use language differently, surprisingly few language processing differences have been reported. Moreover, those behavioral processing differences that have been observed have come from tasks where participants responded to words and sentences out of context. The current study goes beyond existing findings by demonstrating that sex differences occur in tasks where information must be integrated across two utterances.

The current findings also raise numerous questions about the scope and mechanism of the observed sex difference. Although future research is needed, we can use existing data to speculate about why women have a greater order-of-mention effect.

The first question is whether a greater first-mention bias for women is the result of any known sex differences in other language processes. Of most relevance are studies on word and

sentence processing, which have suggested a female superiority in either syntactic or semantic processing. For example, Osterhout, Bersick & Mobey (1997) measured event related potentials (ERPs) as participants read sentences, and found that women showed a larger reaction to gender violations (*The woman prepared himself…)* than men. This emerged as a greater P600 event-related potential, which is often interpreted as a syntactic-violation component. Daltrozzo, Wioland, & Kotchoubey (2007) used ERPs to studied the effects of semantic priming in both word pairs (e.g., *day ~ night*) and sentence contexts (e.g., *Because he disobeyed, his mother gave him a smack/house).* They found greater contrast between the incongruent and congruent conditions for women than men in the N400 component, but a greater contrast in the sentence contexts for men than women in the LPC (Late Positive Complex) component. They concluded that women may have more automatic language processing, and that men may be more greatly influenced by the context. Ullman and colleagues have suggested that a female advantage in episodic memory results in an ability to produce morphologically complex words as wholes rather than using rules more (e.g., retrieving *climbed* as a whole rather than generating it from *climb + ed*; Hartshorne & Ullman, 2006; Ullman et al., 2002).

In sum, some findings suggest that women may process semantic and syntactic information more quickly than men, leading to suggestions that women are generally better at language processing than men (but see also MacCauley 1978). However, the current results cannot be explained as the result of verbal ability. Women were more sensitive to the order of mention bias than men, but this is not the only determinant of pronoun understanding. Indeed, pronouns can be felicitously used to refer to second-mentioned entities, as indicated by the fact that participants in both experiments were sometimes willing to accept this interpretation of the pronoun. Moreover, if the current finding were the result of a general verbal ability, we would

expect to see stronger effects of all discourse biases, both order of mention and verb biases. However, we found that men and women responded equally to the verb bias, and the sex difference only occurred for the order-of-mention effect.

Another possibility we must consider is whether our sample of women was simply more cognitive able or more motivated than our sample of men. However, the available evidence is not consistent with this conclusion. In experiment 1, women and men had similar accuracy and ratings for fillers. In experiment 2, there were no sex differences on the operation span task and SAT scores. There was also no evidence that men and women differed in their metalinguistic awareness of the task, as measured by a post-experiment questionnaire. More generally, the participants in both experiments were mostly drawn from the fall 2008 psychology 101 course at UNC Chapel Hill, in which men and women did not differ in their final course grades ($t(190) = 1.48$, $p=.14$; Elizabeth Jordan, p. c.).

In sum, the most plausible interpretation is that the observed sex difference results from some aspect of language processing, and not just overall cognitive or linguistic ability. The next question is what that difference might be. The order-of-mention bias represents one way in which the discourse context constrains the understanding of subsequent utterances. This constraint is mediated by a nonlinguistic mental representation of linguistic information, which includes the entities and events of the discourse (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). The details of the preceding discourse – for example, who was mentioned first – results in some information in the mental model being more accessible than other information. This discourse accessibility is then utilized when comprehenders interpret a subsequent pronoun, in combination with constraints like the semantic properties of the pronoun (e.g., gender/number).

One possibility is that women have a greater tendency than men to utilize discourse accessibility when interpreting referential expressions. That is, perhaps women use the preceding discourse to develop stronger expectations about upcoming referential expressions more than men. If so, we would predict that women rely on discourse information more than men in other tasks as well. For example, women may have a stronger preference than men for given over new referents. Another possibility is that women specifically pay more attention to structural prominence when building their representation of the discourse, whereas men instead weight other constraints that are known to be relevant, like recency or semantic content (Hobbs, 1979). If so, we would predict that women might use other kinds of structural information for pronoun resolution more than men (for example, parallelism, e.g. Chambers & Smyth, 1998).

A final implication for the sex difference reported here is that it reinforces role of processing itself on reference interpretation. The speaker's meaning is not purely contained in their words, but rather is constructed by the listener. The ease with which the listener adopts an interpretation, and possibly the interpretation itself, is modulated by individual and group differences in processing.

## REFERENCES

Arnold, J. E. (1998). Reference form and discourse patterns (Doctoral dissertation, Stanford University, 1998). *Dissertation Abstracts International, 59,* 2950.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes, 31,* 137–162.

Arnold, J. E. (2008). Reference Production: Production-internal and Addressee-oriented Processes. *Language and Cognitive Processes*, *23*, 495-527.

Arnold, J. E., & Griffin, Z. (2007). The effect of additional characters on choice of referring expression: Everyone competes. *Journal of Memory and Language, 56,* 521-536.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S, & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time-course of pronoun resolution from eyetracking. *Cognition*, *76,* B13-B26.

Arnold, J.E. & Lao, S. C. (Under review). Effects of non-shared attention on pronoun comprehension. Ms., University of North Carolina at Chapel Hill.

Arnold, J.E., Bennetto, L., & Diehl, J. J. (2009). Reference Production in Young Speakers with and without Autism: Effects of Discourse Status and Processing Constraints. *Cognition*, *110*, 131-146.

Bell, A., Brenier, J. M., Gregory, M., Girand, C. & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Langauge*, *60*, 92-111.

Bell, A., Jurafsky, D. Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*, 1001-1024.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–56). New York: Academic Press.

Chambers, C. G., and R. Smyth. 1998. Structural parallelism and discourse coherence: a test of centering theory. *Journal of Memory and Language, 39*, 593–608.

Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2007). Sex differences in two event-related potentials components related to semantic priming. *Archives of Sexual Behavior*, *36*, 555-568.

Eckert, P. & McConnell-Ginet, S. (2003). *Language and Gender*. Cambridge University Press.

Engle, R.W., Kane, M.J. & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In Miyake, A. & Shah, P. (Eds.) *Models of working memory: Mechanisms of active maintenance and executive control*. London:Cambridge Press.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry, 5,* 459-464.

Gordon, P. C., Grosz, B. J. & Gilliom, L.A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17,* 311-347.

Gundel, J. K., Hedberg, N. & Zacharaski, R. (1993). Cognitive Status and the form of referring expressions. *Language, 69*, 274-307.

Hartshorne, J. K., & Ullman, M.T. (2006). Why girls say ‘holded’ more than boys. *Developmental Science,* 9, 21-32.

Hobbs, Jerry. 1979. Coherence and coreference. *Cognitive Science, 3*, 67–90.

Irwin, J., Whalen, D., & Fowler, C. (2006). A sex difference in visual influence on heard speech. *Perception & Psychophysics*, *68*, 582-592.

Jaeger, J., Lockwood, A., Van Valin, R., Kemmerer, D., Murphy, B., & Wack, D. (1998). Sex differences in brain regions activated by grammatical and reading tasks. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*, *9*, 2803-2807.

Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference and consciousness*. Cambridge, MA: Harvard University Press.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16*, 260-264.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Longhurst, E. (2006). ExBuilder. [Computer program]

Macaulay, R. (1978, June). The myth of female superiority in language. Journal of Child Language, 5(2), 353-363

Majeres, R. L. (1999). Sex differences in phonological processes: speeded matching and word reading. *Journal of Memory and Cognition*, *27*: 246-253.

McMurray, B. (2002). EyelinkAnal, version 1.5. Analysis scripts written in Microsoft Access.

Newman,M. L., Groom, C. J., Handleman, L. D., & Pennebaker, J. W. (2007). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes, 45*, 211-236.

Rohde, H., Kehler, A., & Elman, J. L. Pronoun Interpretation as a Side Effect of Discourse Coherence. In the *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, Nashville, Tennessee, 2007.

Stevenson, R., Crawley R., & Kleinman D. (1994). Thematic roles, focusing and the representation of events. *Language and Cognitive Processes*, *9*, 519–548.

Tannen, D. (2001). *You Just Don't Understand: Women and Men in Conversation*. New paperback edition: Quill, New York.

Unsworth, N., Heitz, R.P., Schrock, J.C., & Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498 - 505.

Ullman, M. T., Estabrooke, I. V., Steinhauer, K., Brovetto, C., Pancheva, R., Ozawa, K., Mordecai, K., and Maki, P. (2002). Sex differences in the neurocognition of language. *Abstracts, Brain and Language, 83*, 141-143.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Acknowledgements

I am grateful to Kellen Carpenter, Shin-Yi Lao, and Cathleen Sparks for their help designing stimuli and administering the tasks. I am also thankful for Chris Wiesen's patient help with the statistical models. This research partially funded by NSF grant BCS-0745627.

Tables

Table 1. Example stimuli for Experiment 1.

| | |
|---|---|
| N1 referent: | Stanley noticed Lucas across the street.<br>{Stanley/He} waved at Lucas. |
| N2 referent: | Stanley noticed Lucas across the street.<br>{Lucas/He} tried to hide from Stanley. |

Table 2. % Match responses in Experiment 2. Standard error in parentheses.

| | Women | Men |
|---|---|---|
| N1 pronoun, N1-bias verb | 97% (2%) | 98% (1%) |
| N1 pronoun, N2-bias verb | 96% (3%) | 95% (3%) |
| N2 pronoun, N1-bias verb | 13% (4%) | 21% (5%) |
| N2 pronoun, N2-bias verb | 18% (5%) | 28% (6%) |

Figures

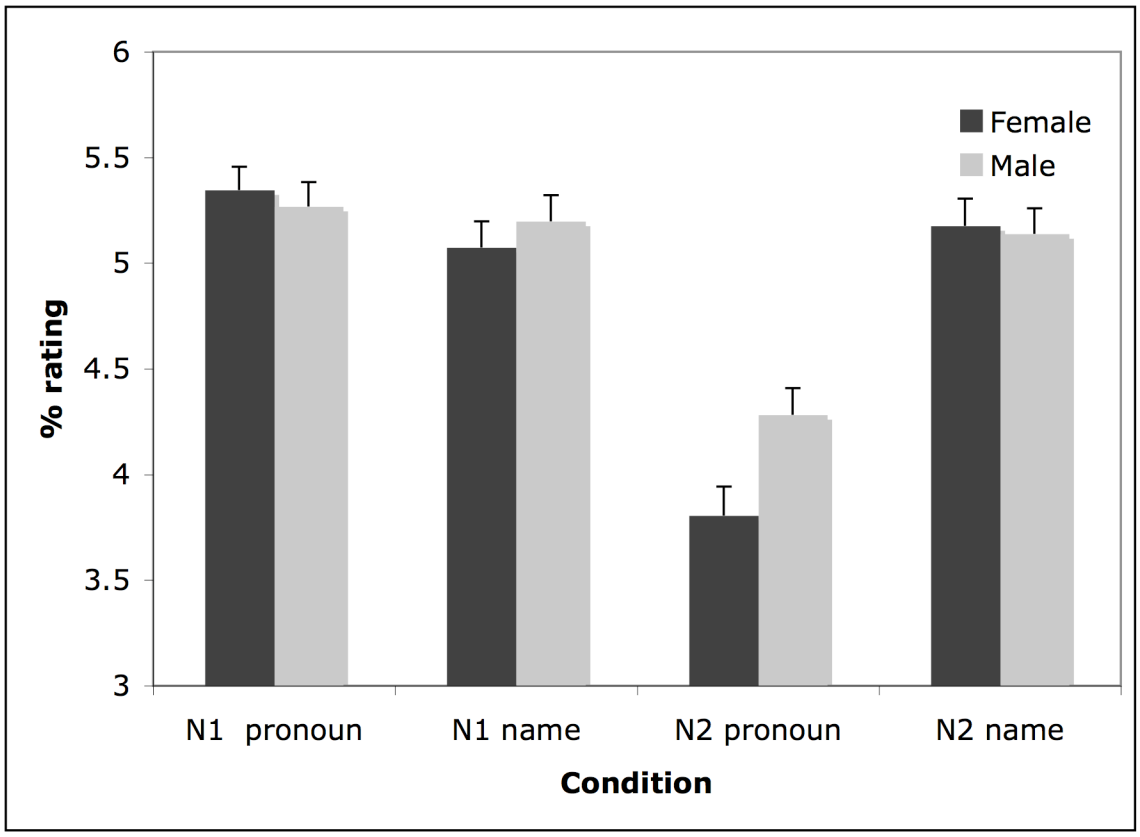
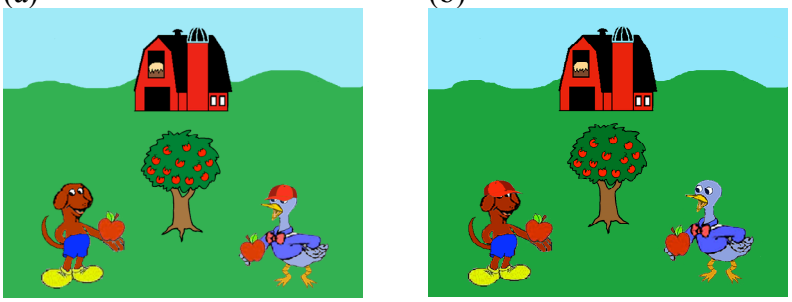



Figure 1: Offline rating experiment results. Average subject means of ratings for items as a function of their referent (N1 vs. N2) and reference form (name vs. pronoun).

(a) (b)



Example story:

Doggy picked apples with Birdy near the farmhouse. He was wearing a hat to protect himself from the sun.

Figure 2: Example visual and linguistic stimuli for Experiment 2. Panel a: First-mentioned condition (Birdy is wearing the hat); Panel b: Second-mentioned condition (Doggy is wearing the hat).

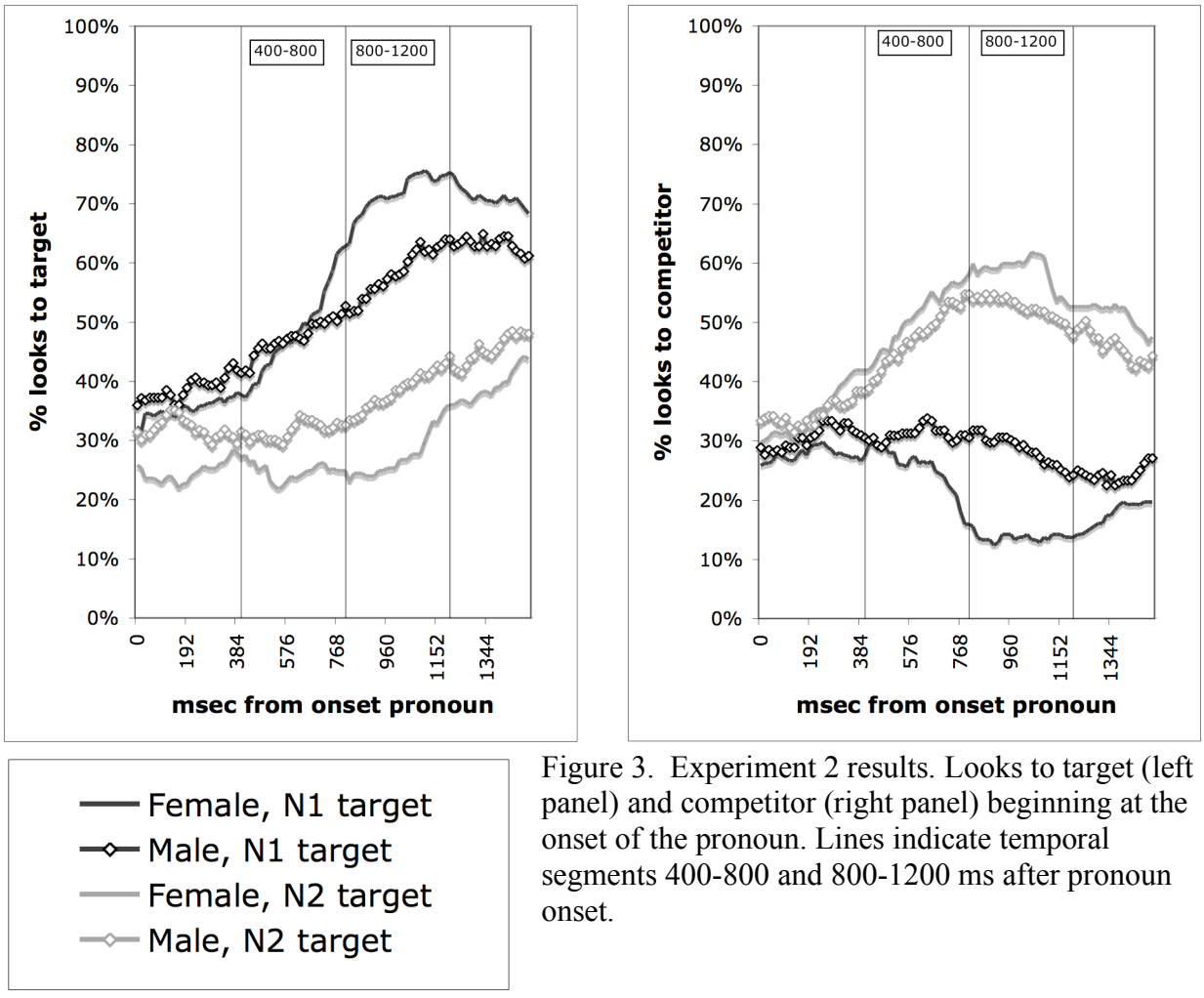



Figure 3. Experiment 2 results. Looks to target (left panel) and competitor (right panel) beginning at the onset of the pronoun. Lines indicate temporal segments 400-800 and 800-1200 ms after pronoun onset.