Anticipatory Eye Movements Mediated by Word-Order Constraints

Paul E. Engelhardt (paul@eyelab.msu.edu)
Department of Psychology, Michigan State University
East Lansing, MI 48824

Ming Xiang (mxiang@fas.harvard.edu)
Linguistics Department, Harvard University
Cambridge, MA 02138

Fernanda Ferreira (fernanda.ferreira@ed.ac.uk)
Department of Psychology, University of Edinburgh
Edinburgh, UK EH8 9JZ

Abstract

Previous work using the visual world paradigm has demonstrated that listeners can rapidly combine visual and linguistic inputs to make complex inferences about upcoming linguistic information. In this study, we show that this inferential capacity extends to more complex linguistic constraints. In English, pre-nominal modifiers follow particular ordering constraints (e.g. *two red balls, vs. *red two balls*). Using a visual world task, we show that listeners are immediately sensitive to this ordering constraint, such that after processing the fragment *point to the red...,* participants were more likely to look at a red object belonging to a contrastive set distinguished by color compared to red objects from a set distinguished by number. These findings support the hypothesis that the processing system can use linear order constraints to make predictive inferences when establishing reference online.

Keywords: adjective ordering; linear order; referential communication; anticipatory eye movements.

Introduction

A question of increasing importance for psycholinguistic research is how the language system interacts with other cognitive systems, and in the past few years there has been a growing interest in the interaction of vision and language (for review see Ferreira & Tanenhaus, 2007). Much of the previous work has gone into determining what effect visual context has on online language processing. In a very influential study, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) proposed that for a sentence such as *put the apple on the towel in the box,* the temporary structural ambiguity of the prepositional phrase (*on the towel*) can be resolved more quickly when there are two apples in a visual display. This suggests that people are sensitive to multiple-referents, and can immediately treat *on the towel* as a modifier when provided with an appropriate context.

Along similar lines, Sedivy, Tanenhaus, Chambers, and Carlson (1999) investigated the role of contrastive sets on the interpretation of adjective modifiers. When participants were asked to *pick up the tall glass*, the probability of fixating the target (i.e. *the tall glass*) started to increase even before the onset of the noun when the context contained a contrasting object (i.e. *a short glass*). The same result was not obtained when the context did not have a contrasting object. Altmann and Kamide (1999) observed a similar anticipatory effect based on thematic constraints. For example, participants were faster to fixate an object such as *a cake* when hearing *the boy will eat...* versus when hearing *the boy will move....* Together these studies show that the processing system can rapidly integrate visual context and linguistic information when resolving structural ambiguity or establishing reference online (see also DeLong, Urback, & Kutas, 2005).

In the current study, we attempt to determine whether listeners can anticipate reference from a more subtle cue in the linguistic input, namely the ordering constraint on pre-nominal number and color modifiers (Bierwisch, 1987; Danks & Schwenk, 1972; Edwards & Chambers, in press; Ferris, 1993; Martin, 1969; Siegel, 1980). For example, upon hearing *point to the orange...,* the only prediction that can be made with respect to a well-formed utterance is that there must be a noun upcoming. However, a parser that is sensitive to ordering constraints may be able to determine that number information is unlikely to be relevant for identifying the head noun.[1]

It is not obvious that the absence of a modifier directly contributes to the processing of an utterance. Presumably, when the comprehension system builds an interpretation based on partial input, it should focus on what is likely to be in the utterance, rather than on what was not in the

---

[1] Several possibilities exist in which a number modifier can follow a color modifier, such as *point to the red group of six squares* or *point to the red- uh- six squares* . We assume that these alternatives are overall less frequent than the preferred *number-color-noun* ordering.

utterance. Determining whether the parser can make use of the absence of a modifier, what we will refer to as *negative evidence*, is the focus of this study.

Experiment 1

To begin, we assessed the relative ordering strength between different types of adjectives (i.e. number, size, & color), using an offline magnitude estimation paradigm. Magnitude estimation has been used as a tool for assessing subtle differences between different types of sentences (Sprouse, 2007). The basic idea is that participants are given a reference sentence that is ungrammatical, and they must judge test sentences compared to the reference sentence. The reference sentence for this experiment was *Mary figured out what her mother wondered whether she was hiding.* Example instructions that were compared to the reference sentence are shown in Table 1.

Table 1: Example instructions for Magnitude Estimation

(1) The boy ate the big red apple.	Size-color
(2) ?The boy ate the red big apple.	Color-Size
(3) The boy ate the two red apples.	Number-color
(4) *The boy ate the red two apples.	Color-number
(5) The boy ate the two big apples.	Number-size
(6) *The boy ate the big two apples.	Size-number

The often reported intuition is that size and color is more flexible than number-color and number-size. If the ordering constraint is equally robust, then participants should rate the reversed orders (i.e. 2, 4, & 6) similarly across the pairings.

Methods

Participants

Forty-two students from Michigan State University participated for course credit.

Materials

Twenty-four sentences modified from Kremmerer et al. (2006) were created that each contained two modifiers. For each sentence, the normal and reversed orders were created. These stimuli were placed into two lists. Each list contained 24 critical trials and 72 filler trials. Filler trials consisted of half grammatical and half ungrammatical sentences. Each critical instruction appeared once in each list, and lists were counterbalanced so that subjects received an equal number of trials in each condition.

Design and Procedure

The design was 2 x 3. Adjective order was either normal or reversed, and adjective pair refers to the three possible pairings between number, size, and color (i.e. number-size, number-color, and size-color). Participants were required to complete a short magnitude estimation training session, in

which they rated the lengths of several lines. After training, participants completed three practice trials, and 96 experimental trials. For each trial, participants read both sentences, and typed their estimate of the difference in acceptability using the keyboard. In the instructions, participants were told that the reference sentence had a value of 100. Therefore, ratings of greater than 100 are expected for grammatical sentences. Values of less than 100 are expected for ungrammatical sentences. Analyses were conducted with both participants (F1) and items (F2) as random effects.

Results

Results showed two significant main effects (adjective order $F1(2, 41) = 88.20$, $p < .01$, $F2(2, 23) = 12.18$, $p < .01$, & adjective pair $F1(2, 41) = 24.44$, $p < .01$, $F2(2, 23) = 8.95$, $p < .01$), and a significant interaction $F1(2, 82) = 10.60$, $p < .01$, $F2(2, 29) = 6.66$, $p < .01$ (see Figure 1). These results show that the preferred order resulted in higher ratings than did the reversed order. The reversed order of size and color seems to be the primary factor driving the main effect of adjective pair and the interaction, as this condition resulted in significantly higher acceptability, than did the other reversed orders. This pattern indicates more flexibility with size and color pair. We began our investigation of the use of negative evidence using the strongest ordering constraint (i.e. number & color).
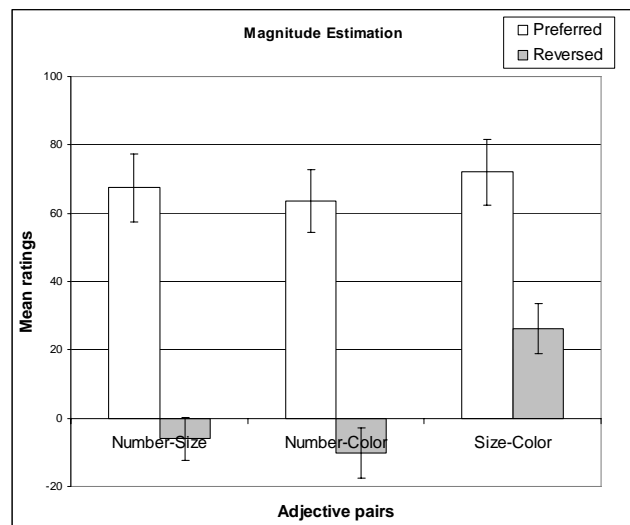

Figure 1: Results from the magnitude estimation study.

Experiment 2

In Figure 2, by the number attribute *two*, there are three sets of objects that fit the description. Similarly, by the color attribute *orange*, there are also three sets of compatible objects. Therefore, when a participant hears an instruction such as *point to the orange…*, the instruction would seem to be ambiguous, since there are three orange sets in the

display. However, if participants are sensitive to the number-color ordering asymmetry, then they should be more likely to fixate the correct target (i.e. the orange triangles) before hearing the noun. This is because number information is needed to distinguish the sets of objects with squares. In other words, beginning a noun phrase with a color modifier should signal that the target is unlikely to be from a set of objects discriminated by number.
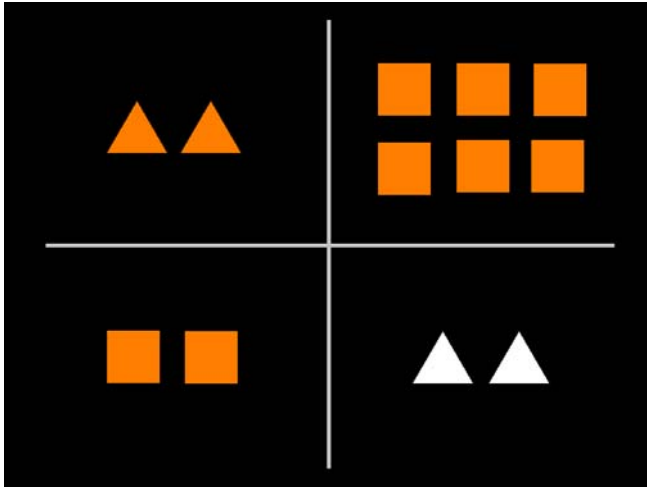

Figure 2: Example stimulus for Experiment 1.

In contrast, for an instruction such as *point to the two...* , ordering constraints cannot be used to rule out any of the three potential candidates because a color adjective is still an option (e.g. *the two orange triangles*). Therefore, our critical comparison is between the two types of modifiers: We expect more looks to the correct target during the modifier with an ambiguous color modifier compared to an ambiguous number modifier. We also included two unambiguous instructions (see Table 2). In these conditions, the target can be identified at the modifier. Therefore, we should see no difference between these two instructions.

Methods

Participants
Forty students from Michigan State University participated for course credit. All were native speakers of English and had normal or corrected-to-normal vision.

Materials

A total of 99 visual displays were created (three practice, 24 experimental, & 72 filler). For all displays, the number, color and shape of objects was randomly determined (see Table 3). A complete list of stimulus materials is available at (http://eyelab.msu.edu/pubs/exf.pdf/). All filler trials contained a single modifier instruction, and each uniquely identified one quadrant of the display. For a viewing distance of 60 cm, the visual array of objects on average

subtended 22° of visual angle horizontally and 17° vertically.

Table 2: Example instructions for Figure 1

(1) Point to the orange triangles. (color/ambiguous)
(2) Point to the two squares. (number/ambiguous)
(3) Point to the white triangles. (color/unambiguous)
(4) Point to the six squares. (number/unambiguous)

Utterances were digitally recorded by a native speaker of English, and converted to .wav format. Four lists were created so that for each experimental display, all four possible instructions were created. Each visual display appeared once in each list, and instructions were rotated in a Latin Square design. The mean duration of number modifiers was 406 ms, and the mean length of color modifiers was 413 ms, across all four critical conditions. This difference was not significant $t(47) = .395$, $p = .69$. The mean lengths of the head nouns comparing number and color modified instructions was 749 ms and 747 ms respectively. This difference was also not significant $t(47) = .088$, $p = .93$.

Design and Procedure

The design had two variables: *modifier type* and *ambiguity* (see Table 2). The participants' task was to point to a target as quickly as possible. Each trial began with a fixation cross. Participants were instructed to look at the cross and to press the space bar to start the trial. The visual display appeared for 1000 ms before the instruction began. Primary data analysis occurred in two time segments corresponding to the modifier and noun. These segments were identified by aligning utterances on a word-by-word and trial-by-trial basis. The modifier and noun time windows were offset (for all analyses) by 200 ms to account for the time required to initiate a saccade.

Table 3: Modifiers and nouns tested.

Numbers	Colors	Heads
two	blue	circle
five	gray	diamond
six	green	heart
	orange	moon
	pink	square
	purple	star
	red	triangle
	white	yellow

The dependent measure was the proportion of trials with a fixation to each of the four regions of the display. The probability of fixation was calculated by determining the

number of trials that contained a fixation in one of the regions. The number of trials with a fixation was divided by the total number of trials per condition (i.e. 6). (Note that we excluded fixations that crossed either the modifier or noun onsets.) Therefore, the dependent measure could also be described as the proportion of trials in which a saccade was launched to each of regions during the modifier or noun. Hand-coding of the data was performed from the onset of the instruction to the time at which the participant pointed to the target. The criterion for fixations was three consecutive frames with little to no movement of the fixation cross.

Results

The following results are presented sequentially in three main segments. First, we report the probability of fixating the objects prior to the modifier word (i.e. during *point to the*). This analysis establishes that participants are not biased to attend to any particular region of the display prior to the critical material. The second time segment corresponds to the modifier, and this is the region of most interest for our main hypotheses. Finally, we report the results from the noun region. The general prediction is that early in the utterance participants should be randomly sampling the four regions of the display. However, as the instruction unfolds, attention will be increasing restricted to the object named in the instruction.

During the initial region, we found a main effect of ambiguity $F1\ (1, 39) = 27.52$, $p < .05$; $F2\ (1, 23) = 23.22$, $p < .05$, in which there was a higher probability of fixating the targets of the unambiguous instructions. Both targets are unique in terms of number and color. Therefore, the bias to attend to these objects is likely due to their salience. For our purposes however, the main effect of modifier type was not significant $F1\ (1, 39) = .40$, $p > .05$; $F2\ (1, 23) = .45$, $p > .05$, and neither was the interaction $F1\ (1, 39) = 3.97$, $p > .05$; $F2\ (1, 23) = 3.69$, $p > .05$.

At the modifier position, there was a main effect of ambiguity (see Figure 3). The unambiguous instructions had a higher probability of fixation than the ambiguous instructions $F1\ (1, 39) = 94.07$, $p < .05$; $F2\ (1, 23) = 188.36$, $p < .05$. There was also a main effect of modifier type, in which the color adjectives resulted in more fixations than did number adjectives $F1\ (1, 39) = 12.74$, $p < .05$; $F2\ (1, 23) = 8.11$, $p < .05$. The interaction was not significant $F1\ (1, 39) = 1.06$, $p > .05$, $F2\ (1, 23) = .56$, $p > .05$. Ordinarily, one would stop with statistical analyses at this point. However, because of a critical difference between the targets for the ambiguous and unambiguous instructions, we went ahead with simple effects tests. The reason for doing so will be discussed in the subsequent paragraph. The pair-wise comparisons showed a significant difference between the color and number adjectives within the ambiguous conditions $t1(39) = 2.88$, $p < .05$, $t2(23) = 2.37$, $p < .05$, and a significant difference between the color and number adjectives within the unambiguous conditions $t1(39) = 2.38$, $p < .05$, $t2(23) = 2.16$, $p < .05$.[2]

These results show that with the ambiguous instructions, the color modifier produced more looks to the target than the number modifier, consistent with the hypothesis that the processing system can make use of word-order constraints to anticipate the target. However, we also found a color advantage with the unambiguous instructions. This color advantage in the unambiguous conditions is not surprising, as a color "pop-out" effect is well documented in the visual search literature (Itti & Koch, 2000). In the unambiguous color condition, the modifier directly points to the one unique color in the display, and thus, listeners' eye-gaze is attracted by the highly salient color feature.
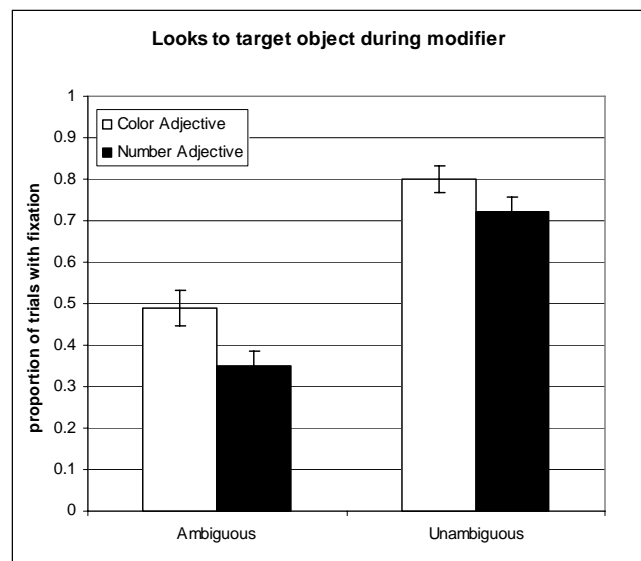

Figure 3: Looks to the target object during the modifier.

One possibility is that there is some inherent processing advantage of color over number. However, with the ambiguous instructions, the ambiguous modifier is compatible with three regions of the display. If there is a processing advantage of color over number, then it should lead to a higher probability of fixating the orange objects, but it should not help listeners to discriminate the target from the relevant distractors (i.e. the other sets of objects that are compatible with the ambiguous modifier). To rule out the possibility of a general processing advantage of color over number, we examined looks to the target versus the relevant distractors. We defined the relevant distractors in the following way. For the ambiguous color condition, the six orange squares differ from the target in number, and

---

[2] Because of the a priori hypothesis concerning ordering restrictions, we conducted one-tailed tests for comparisons involving the ambiguous instructions, and two-tailed tests for the unambiguous instructions.

we refer to it as relevant distractor 1. The two orange squares are relevant distractor 2. For the ambiguous number condition, the two white triangles differ from the target by color, so we refer to it as relevant distractor 1. The two orange triangles are relevant distractor 2.

We compared the looks to the target and the relevant distractors using difference scores. Target advantage 1 was computed by subtracting the proportion of trials with a fixation to relevant distractor 1 from the proportion of trials with a fixation to the target (see Figure 4). Target advantage 2 was computed by subtracting relevant distractor 2 from the target. Target advantage 1 was significantly greater for the color adjective compared to the number adjective $t1(39) = 2.36$, $p < .05$, $t2(23) = 1.56$, $p = .06$. Target advantage 2 was significantly greater in the color condition than in the number condition $t1(39) = 3.63$, $p < .05$, $t2(23) = 2.02$, $p < .05$. These results show that at the modifier, participants are already beginning to discriminate the target from the distractors in the ambiguous color condition, consistent with our hypothesis that listeners can make predictive inferences based on linear ordering constraints. No such evidence was found for the ambiguous number instruction.
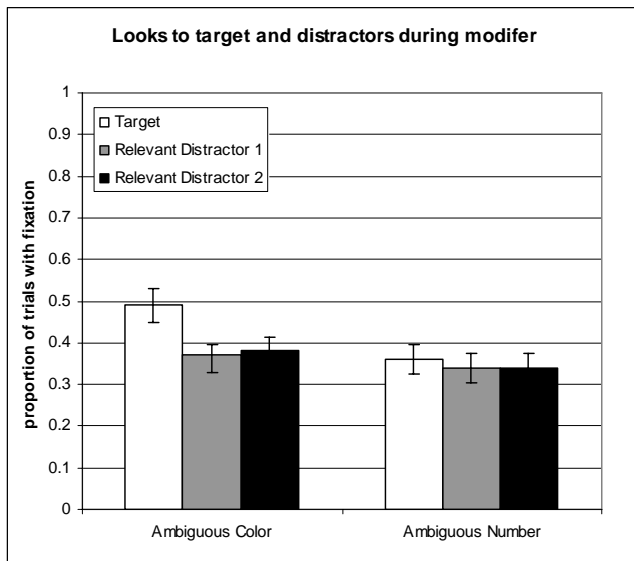

Figure 4: Looks to the target and relevant distractors.

At the head-noun neither of the main effects were significant (adjective type $F1 (1, 39) = .33$, $p > .05$; $F2 (1, 23) = .13$, $p > .05$, & ambiguity $F1 (1, 39) = .29$, $p > .05$; $F2 (1, 23) = 1.13$, $p > .05$), but the interaction was significant by subjects and marginal for items $F1 (1, 39) = 4.12$, $p < .05$; $F2 (1, 23) = 3.20$, $p > .05$ (see Figure 5). The pattern of results for the ambiguous instructions is similar to pattern at the modifier, namely that the color modifier resulted in a marginally higher probability of fixation $t1(39) = 1.91$, $p < .05$, $t2(31) = 1.46$, $p = .075$. The pattern for the unambiguous instructions was in the opposite direction of

the previous time region, and was not significant $t1(39) = -1.00$, $p > .05$, $t2(31) = -1.17$, $p > .05$.

In summary, we found that linear order constraints can be used by the processing system to make anticipatory fixations when confronted with an ambiguous color modifier. In addition, we were able to demonstrate that this effect is unlikely to be due to an inherent processing advantage of color over number. Instead, we conclude that people can make rapid anticipatory eye movements based on the ordering constraints on pre-nominal adjectives.
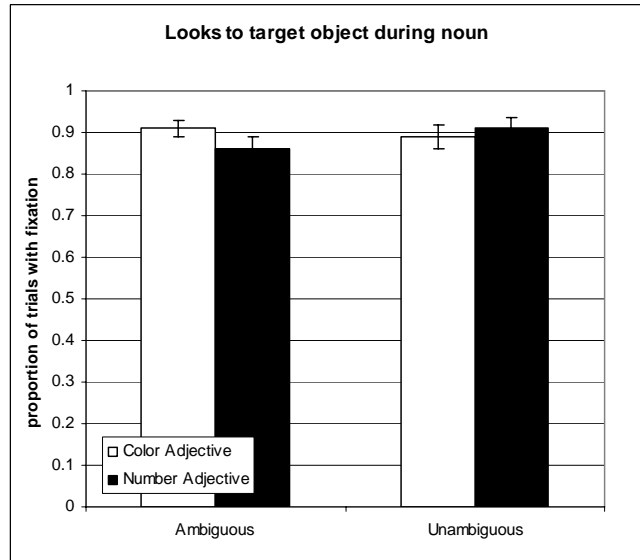

Figure 5: Looks to the target object during the head noun.

General Discussion

The results from this experiment demonstrate that participants can fixate a target object more quickly when they hear a temporarily ambiguous color modifier when compared to a similarly ambiguous number modifier. Our results can only be explained by assuming that when confronted with an ambiguous modifier, the processing system has the ability to quickly apply knowledge of linguistic constraints to the available visual information to anticipate what the upcoming referent will be.

If people do employ word order constraints to make rapid predictions, then one question that naturally arises is under what circumstances are these predictions generated. Do people always generate predictions whenever the linguistic constraints and the visual displays support such predictions, or is this anticipatory behavior bound by the particular task or experimental situation? One potential explanation of our results is that because all of the stimuli (experimental and filler) contained only one modifier, it is possible that participants learned over the course of the experiment that a single modifier was the only information needed to succeed in the task. Therefore, in making

maximal use of modifier information, they might have employed more predictions, than they would in a more varied context. To investigate this possibility we analyzed the first half and second half of the critical trials for each participant. The results from this analysis showed that the effect was slightly (but not significantly) greater in the first half of trials, which goes against a learning explanation of our results.

For the structures and constraints tested in previous studies (e.g. Altmann & Kamide, 1999; Delong et al., 2005; Sedivy et al., 1999), partial linguistic information provides evidence as to what should be upcoming in the speech stream. In the current study however, when hearing *point to the orange…*, the only necessary element is that there must be a noun at some point following the adjective. This prediction by itself will **not** drive the fixation patterns we observed. Rather what seems to be happening is that participants are able to use the absence of a particular type of modifier (i.e. negative evidence) to infer what information will **not** follow, namely, that after a color modifier it is unlikely that any information about number will be included.

Turning our attention to the mechanism that allows the processing system to successfully link the linguistic input with the visual context, we assume that minimally three steps are required. First, at the (ambiguous) modifier, our participants must narrow down the referents to those that match the description. This is possible only if the parser builds linguistic representations incrementally. The second step is that participants must have access to the ordering constraints between different types of adjective modifiers. If the modifier is a color term, the parser must be aware that number information is unlikely to follow, whereas if the modifier is a number term, the possibility for color modification remains. The final step is for the parser to apply the ordering constraint to the available visual context.

Under this analysis, all our subjects need is a detailed linguistic representation coupled with an active mapping between linguistic and visual representations. Going beyond the previous literature, we have shown that people can use more subtle and indirect linguistic cues to help them succeed in locating an object in a visual array. It remains to be seen whether the same type of anticipatory eye movements would be observed in a task environment with more varied utterances.

Author Note

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73(3),* 247–264.

Altmann, G. T., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action* (pp. 347-386). New York: Psychology Press.

Altmann, G. T., & Steedman M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191-238.

Bierwisch, M. (1987). The semantics of gradation. In: M. Bierwich, & E. Lang (Eds.): *Dimensional Adjectives.* Berlin: Springer Verlag, pp. 71-261.

Danks, J. H., & Schwenk, M. A. (1972). Prenominal adjective order and communication contexts. *Journal of Verbal Learning and Verbal Behavior, 11,* 183-187.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117-1121.

Edwards, J. D., & Chambers, C. G. (in press). 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.* Eds. N. Pearlmutter, and E. Gibson. Cambridge MA: MIT Press.

Ferreira, F. & Tanenhaus, M. K. (2007). Introduction to the special issue on language-vision interactions. *Journal of Memory and Language, 57,* 455-459.

Ferris, C. (1993). *The meaning of syntax: a study of adjectives in English.* London & New York: Longman.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489-1505.

Martin, J. E. (1969). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior, 8,* 697-704.

Quirk R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of English Language.* London: Longman.

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

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

Siegel, M. (1980). *Capturing the Adjective.* New York: Garland.

Sprouse, J. (2007). A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge. Ph.D. thesis University of Maryland.

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