Chapter 22

THE PROCESSING OF FILLED PAUSE DISFLUENCIES IN THE VISUAL WORLD

KARL G. D. BAILEY

Andrews University, USA

FERNANDA FERREIRA

University of Edinburgh, UK

Eye Movements: A Window on Mind and Brain
Edited by R. P. G. van Gompel, M. H. Fischer, W. S. Murray and R. L. Hill

Elsevier AMS Ch22-I044980 Job code: EMAW 14-2-2007 1:11p.m. Page:488 Trimsize:165×240MM

Abstract

One type of spontaneous speech disfluency is the filled pause, in which a filler (e.g. *uh*) interrupts production of an utterance. We report a visual world experiment in which participants' eye movements were monitored while they responded to ambiguous utterances containing filled pauses by manipulating objects placed in front of them. Participants' eye movements and actions suggested that filled pauses informed resolution of the current referential ambiguity, but did not affect the final parse. We suggest that filled pauses may inform the resolution of whatever ambiguity is most salient in a given situation.

Basal Fonts:Times Margins:Top:4.6pc Gutter:4.6pc Font Size:10/12 Text Width:30pc Depth:43 Lines

The most common type of overt interruption of fluent speech, or disfluency, is the filled pause (Bortfield, Leon, Bloom, Schober, & Brennan, 2001). Speakers produce filled pauses (e.g. *uh* or *um*) for a variety of reasons, such as to discourage interruptions or to gain additional time to plan utterances (Schacter, Christenfeld, Ravina, & Bilous, 1991). While speakers may benefit from producing filled pauses because they gain planning time, listeners may also use the presence of filled pauses to inform language comprehension (Bailey & Ferreira, 2003; Brennan & Schober, 2001; Brennan & Williams, 1995; Clark & Fox Tree, 2002). Thus, given the prevalence of filled pauses, and the use of such pauses by listeners, a complete model of language comprehension should account for how these disfluencies are handled.

In order to construct such a model of disfluency processing, it is necessary to describe and test possible hypotheses about how disfluencies might affect language comprehension. Evidence that supports one such hypothesis, cueing of upcoming structure, comes from a series of experiments involving grammaticality judgments (Bailey & Ferreira, 2003). This hypothesis is built on the observation that filled pauses occur in a particular distribution with respect to syntactic (Clark & Wasow, 1998), semantic (Schacter et al., 1991) or pragmatic (Smith & Clark, 1993) structure. In the case of syntactic structure, filled pauses (and other disfluencies, such as repetitions) are most likely to occur immediately prior to the onset of a complex syntactic constituent (Clark & Wasow, 1998; Ford, 1982; Hawkins, 1971; Shriberg, 1996;). Filled pauses are also likely after the initial word in a complex constituent, especially after function words (Clark & Wasow, 1998). Thus, the cueing hypothesis assumes that listeners might be able to use the presence of a recent filled pause to predict that an ambiguous structure should be resolved in favor of a more complex analysis (Bailey & Ferreira, 2003). In a garden path utterance like [1] below, the filled pause might act as a “good” cue, because it correctly predicts the ultimate structure of the utterance; in [2], the filled pause might be a “bad” cue, because it leads the listener to predict the onset of a new constituent.

[1] While the man hunted the uh uh *deer ran* into the woods.

[2] While the man hunted the *deer* uh uh *ran* into the woods.

Grammaticality judgments supported this cueing hypothesis: [1] was judged grammatical more often than [2], which suggested that [1] is easier to process (Bailey & Ferreira, 2003). However, [1] and [2] confound “good” and “bad” cues with the presence of delay between the ambiguous head noun and the disambiguating verb. This type of delay has led to the same pattern of results in utterances with lexical modifiers (i.e., prenominal adjectives and relative clauses) in place of the disfluencies in [1] and [2] (Ferreira & Henderson, 1991). To avoid this confound, Bailey and Ferreira (2003) tested whether filled pauses that did not introduce delays between temporarily ambiguous head nouns and disambiguating verbs might also affect grammaticality judgments of spoken utterances depending on their location. Disfluencies were placed in two different locations in coordination ambiguity utterances prior to the onset of the temporarily ambiguous head noun. The “good” cue location in [3] below was consistent with the ultimate sentence coordination structure, while the “bad” cue in [4] was consistent with an noun phrase coordination structure (based on the assumption that listeners take a disfluency to be indicative of an upcoming complex constituent).

[3] Sandra bumped into the busboy and the uh uh waiter told her to be careful.

[4] Sandra bumped into the uh uh busboy and the waiter told her to be careful.

Participants were more likely to judge an utterance with a "good" cue disfluency (as in [3]) as grammatical than an utterance with a "bad" cue (as in [4]). This pattern of results was replicated with environmental noises replacing the disfluencies, but, importantly, not with adjectives, suggesting that it is the presence of a non-propositional interruption that is the cue, not the form of that interruption.

However, the results of Bailey and Ferreira (2003), while promising, are based on offline judgments following the end of the utterance. In other words, the grammaticality judgment task makes it possible to see that filled pauses have had an effect consistent with the cueing hypothesis by the time the utterance is finished, but it is not possible to chart the time course of that effect, nor to observe when processing of the disfluency takes place.

A recently rediscovered methodology that allows spoken language comprehension to be monitored on a moment by moment basis is the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm (henceforth, the VWP), participants listen to utterances while viewing a concurrent array of clip art images on a computer screen (e.g. Altmann & Kamide, 1999) or while interacting with a set of objects within reach (e.g. Tanenhaus et al., 1995). The objects or images which make up the constrained visual world and the relationships between them serve as a context for a concurrent referring utterance (Tanenhaus et al., 1995). Inferences about language comprehension are drawn from listeners' eye movement patterns: The eyes are naturally directed to objects that are related to concurrent language processing (Cooper, 1974). In the VWP, utterances can be presented without distortion and it is not necessary to instruct listeners to look at objects which are related to concurrent speech.

Two particular patterns of eye movements have been used to draw inferences about comprehension: anticipatory and confirmatory eye movements. Anticipatory eye movements (Altmann & Kamide, 1999, 2004; Kamide, Altmann, & Haywood, 2003) are saccades launched to objects before they are directly referenced by the utterance. Confirmatory eye movements (e.g. Spivey, Tanenhaus, Eberhard, & Sedivy, 2002) are made in response to a direct reference to an object and can include fixations on possible referents of a constituent (Tanenhaus et al., 1995) or on disconfirmed competitors (Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Kamide et al., 2003). The presence of confirmatory eye movements is most easily seen in the probability of fixating a given object because participants may simply continue to fixate an object that they were already looking at due to an anticipatory eye movement launched prior to direct reference.

The cueing hypothesis would predict that eye movements during a filled pause should reflect a more complex parse of material currently being processed, and that saccades would be launched to objects consistent with that analysis. Confirmatory eye movements during a later ambiguous referring expression would then identify which of a set of possible parses had been selected and the time course of that selection (as the probability of fixating a given object rises and falls).

In this chapter, we will present an experiment that directly tests whether a cueing mechanism can modulate the interpretation of a fully ambiguous utterance in the presence of a fully ambiguous visual world. As described earlier, the position of a disfluency can affect the probability that an utterance is judged grammatical (Bailey & Ferreira, 2003), suggesting that disfluencies may cue the parser to expect a certain structure. The strongest form of the cueing hypothesis, then, predicts that a fully ambiguous utterance will immediately be interpreted differently based solely on the location of a disfluent interruption. We do not find evidence for this strong hypothesis, but do find support for a weaker form, in which the disfluent interval introduced by the filled pause may allow the parser to further process any existing ambiguities. Depending on the demands of the task, final interpretations may or may not be affected by the disfluency cue. Nevertheless, we suggest that filled pauses provide a unique window on sentence processing in general, because they show what ambiguities are relevant at that point in the utterance.

1. Experiment

In order to test whether filled pauses can change the interpretation of an otherwise fully ambiguous utterance, the concurrent visual world must not constrain the interpretation of that utterance. Previous studies using otherwise fully ambiguous prepositional phrase ambiguities (Spivey et al., 2002; Tanenhaus et al., 1995) used visual worlds that constrained the interpretation of utterances such as [5] below. The objects in these displays required participants to arrive at the same semantic representation as in the disambiguated utterance [6], and disallowed the interpretation in [7].

[5] Put the apple on the towel in the box.
[6] Put the apple that's on the towel in the box.
[7] Put the apple on the towel that's in the box.

Two different constrained display types have been used (Spivey et al., 2002). The first, referred to as the one-referent display, contained a target object (e.g. an apple on a towel), a distractor object (e.g. a frog on a mitten), a goal location (e.g. a box), and a distractor location (e.g. a towel). The second, two-referent, display was identical to the first, except that the distractor object matched the target object in part (e.g. an apple on a mitten). Note that in both displays, the only possible action in response to [5] is for the apple that is on the towel to be placed in the empty box because there is no towel inside a box.

In order to modify these displays so that they did not constrain the interpretation of our utterances, we replaced the distractor location described above (a towel by itself) with a modified goal (e.g. a towel in a box as opposed to the unmodified goal, an empty box; see Figure 1). Thus, in the one-referent display, it would be possible to place an apple that is on a towel into an empty box (a modified theme interpretation) in response to [5] or an apple onto a towel that is in a box (a modified goal interpretation). In the case of the two-referent display, of course, the display still constrains the interpretation of the utterance because of the presence of a second apple. The modified goal interpretation in [7] is possible, but only if the listener violates syntactic or discourse constraints and

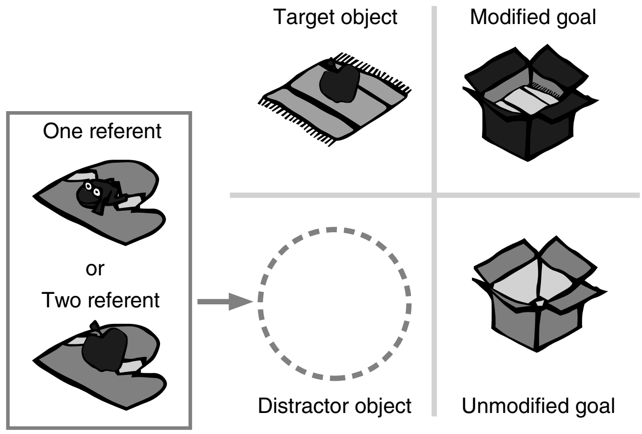

Figure 1. Fully ambiguous displays used in experiment. Only the distractor object differed between the one referent and the two referent displays.

uses the phrase “on the towel” twice: once to identify which apple to pick up (the apple that’s on the towel) and once to identify where to place the apple (the towel that’s in the box). These interpretations are unlicensed because a single constituent cannot play more than one role in a sentence; nevertheless, we have observed that participants occasionally behave as if that is the interpretation that they have obtained, perhaps because they have engaged in “good-enough processing” (Ferreira, Bailey, & Ferraro, 2002).

According to the strong form of the cueing hypothesis, disfluencies placed before one of the two possible modified noun phrases in [5] should bias the parser to prefer the corresponding structure. Specifically, a filled pause placed as shown in [8] is predicted to yield a modified theme interpretation similar to [6], whereas a filled pause placed as in [9] is predicted to yield modified goal interpretations similar to [7].

[8] Put the uh uh apple on the towel in the box.

[9] Put the apple on the uh uh towel in the box.

For the strong form of the cueing hypothesis to be supported, these interpretations should be seen both immediately (in saccades to appropriate goal objects during filled pauses), as the utterance unfolds (in fixations during the ambiguous noun *towel*), and in the overall interpretation of the utterance (the participants’ actions).

1.1. Materials and methods

1.1.1. Participants

Sixteen participants from the Michigan State University community participated in this experiment in exchange for credit in an introductory psychology course or payment ($7.00). All participants were native speakers of English, and had normal hearing and normal or corrected to normal vision.

1.1.2. Materials

Twenty-four prepositional phrase ambiguity utterances were constructed for this experiment using nouns from a set of 12 possible target objects and 12 possible goal objects. Utterances were recorded and digitized at 10 kHz using the Computerized Speech Laboratory (Kay Elemetrics), and then converted to wav format. Each utterance was recorded in two ways: once as an utterance with two disfluencies, as in [10] below, and once as a fluent utterance with two instances of *that's*, as in [11].

[10] Put the uh uh apple on the uh uh towel in the box.

[11] Put the apple that's on the towel that's in the box.

Utterances like [8] and [9] were created from [10], and like [6] and [7] were created from [11] by excising the appropriate disfluency or word. Participants are relatively insensitive to the removal of disfluencies from utterances (Brennan & Schober, 2001; Fox Tree, 1995, 2001; Fox Tree & Schrock, 1999;) and thus this procedure was used to control the prosody of the various utterances. The removal of a single disfluency or word from an utterance did not result in utterances that participants found odd or strange. In the experiment, each participant heard only one version of any given utterance.

Forty-eight filler utterances were also recorded and grouped with the 24 critical utterances into trials of 3 utterances each. A further 72 utterances were recorded to create 24 trials composed of only fillers. The types and proportions of syntactic structures used in the filler utterances and the interleaving of filler and critical trials were identical to those used in Spivey et al. (2002). Filled pauses occurred on half of filler trials and were placed at a variety of different locations within the sentences.

Displays consisted of a 2×2 grid (see Figure 1), and objects were set up according to the description provided by Spivey et al. (2002), so that depending on the height and posture of a given participant, the center of each object (or set of objects) was separated by 10–15° of visual angle from the center of each of its adjacent neighbors (note that previous studies did not report the angular distance between objects). In experimental trials, the possible theme referents (the target and distractor objects) were always on the left, and were each placed equally in both the proximal and distal positions across trials. The possible goal referents (modified and unmodified) were always on the right, and likewise were each placed equally in both the proximal and distal positions across trials. The possible theme and goal referents for filler utterances were equally likely to occur in any of the four positions, and any object on the table could be referenced as a target object in filler utterances. In all, 48 displays were created, one for each set of three utterances. Of the 24 critical displays seen by any participant, 12 were two-referent displays and 12 were one-referent displays. A new random ordering of trials adhering to the interleaving requirements was created for every fourth participant in this experiment.

1.1.3. Apparatus

The eyetracker used in this experiment was an ISCAN model ETL-500 head-mounted eyetracker (ISCAN Incorporated) with eye and scene cameras located on a visor.

Participants were able to view 103° of visual angle horizontally and 64° vertically. No part of the object display was occluded at any time by the visor. Eye position was sampled at 30 Hz and merged with scene video data. Eye position in this merged video was later hand-coded relative to Regions of Interest (henceforth, ROIs) frame by frame, starting with the onset of each critical utterance and ending with the movement of an object to a new location.

1.1.4. Procedure

After a participant was introduced to the objects and apparatus, and had provided informed consent, the eyetracker was placed on the participant's head and adjusted. Depending on the height of each participant, participants either stood or were seated at a table. Participants' eye positions were calibrated to the scene by recording pupil and corneal reflection locations while they looked at nine predetermined targets. The sentence comprehension task was introduced to the participant via three practice utterances involving the movement of a single object from one location to another. The practice utterances did not contain any lexical or syntactic ambiguities.

Immediately before beginning a trial, the experimenter set up the appropriate objects in front of the participant, which allowed 20–30 s of view time prior to the onset of the first utterance in the trial (as in Spivey et al., 2002 and Trueswell, Sekerina, Hill, & Logrip, 1999). Participants were instructed to respond as quickly as possible prior to practice trials, but were not reminded thereafter. In addition, no "Look at the center cross" instruction was given prior to the start of each trial (cf. Spivey et al., 2002), as pilot studies indicated that participants tended to perseverate in fixating the center cross when this instruction was given.

Design. The four utterance types (theme and goal disfluencies, and theme and goal modifiers) were combined with the two displays (one and two referent) to create eight unique conditions for this experiment. Three trials in each condition were presented to each participant, for a total of twenty-four critical trials. Each display occurred in each condition an equal number of times over the course of the experiment.

1.2. Results and discussion

The analysis of eye tracking data presented here differs somewhat from previous studies using this version of the VWP. These studies (Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999) calculated probabilities of fixating particular objects at each sampling interval during arbitrary time segments that did not take into account variations in word length across individual utterances. Probabilities in this study, on the other hand, were calculated separately for each ROI and each word in each utterance and then averaged (see Altmann & Kamide, 2004, who described this procedure). These probabilities were then arcsine-transformed (Winer, 1971) and submitted to a 2 (cue location: theme or goal) by 2 (cue type: disfluency or modifier) by 2 (number of possible theme referents: one or two) ANOVA. In addition, behavioral responses to disfluent instructions in the current experiment were classified as either modified goal (e.g. towel in box) directed or unmodified goal (e.g. empty box), and were submitted to a 2 (number of referents) by 2 (location of disfluency) ANOVA. Unambiguous controls were not included in the behavioral analysis as participants moved an object to the appropriate goal on over 90% of trials.

Participants were more likely to move a target object to the unmodified goal ($F_{1,15} = 23.6$, $p < 0.001$) in the two-referent display (70.8% of trials with a theme disfluency; 64.6% with a goal disfluency) than in the one-referent display (37.5% of trials with a theme disfluency; 35.4% with a goal disfluency). The location of the disfluency had no effect ($F < 1$) on participants' actions and there was no significant interaction between the number of referents and disfluency location ($F < 1$). The effect of number of referents on the final interpretation of the utterance is not surprising, as the two-referent display should have constrained the interpretation of the utterance (due to the presence of two apples), while the one-referent display should not have. However, the lack of effect of disfluency location on the final interpretation of the utterances, even in the one-referent display, is evidence against the strong form of the cueing hypothesis, and suggests that disfluencies were not interpreted as strong predictors of the syntactic parse.

Eye movement patterns, on the other hand, did support a form of the cueing hypothesis. Figure 2 shows graphs representing the probability of fixating and launching a saccade to each ROI for each condition in the experiment. Gray polygons represent the probability of fixation on, and black lines the corresponding probability of launching a saccade to that ROI for each word. Each point on the polygons and line graphs corresponds to a single word in each utterance. Content words, disambiguating function words, and disfluencies are indicated above the fixation polygons. The eight conditions form rows, while the four ROIs form columns.

An effect of number of referents is presented in Figure 2; the different display types elicited different patterns of fixation, especially on the distractor and the modified goal. Consistent with previous studies (Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999; the incorrect goal in previous studies corresponds to our modified goal), there is a significant increase in the probability of fixation on the modified goal in the one-referent display relative to the two-referent display during the word *towel* ($F1_{1,15} = 36.2$, $p < 0.001$; $F2_{1,23} = 30.6$, $p < 0.001$). This difference was found for all utterance types, including theme modifiers, which should rule out the modified goal as a possible referent of *towel* because of the preceding *that's*. Main effects of cue location ($F1_{1,15} = 6.29$, $p < 0.03$; $F2_{1,23} = 5.12$, $p < 0.04$), and cue type ($F1_{1,15} = 7.64$, $p < 0.02$; $F2_{1,23} = 7.22$, $p < 0.02$) were also present. The effect of cue type was due to an increased proportion of looks to the modified goal in the disfluency conditions, which would be expected if the language comprehension system treated those utterances as more ambiguous than either modifier condition. A significant interaction between number of referents and cue location ($F1_{1,15} = 6.75$, $p < 0.02$; $F2_{1,23} = 7.99$, $p < 0.02$) was present, but interactions between cue type and number of referents ($F < 1$), between cue type and location ($F1_{1,15} = 3.54$, $p > 0.05$; $F2_{1,23} = 3.61$, $p > 0.05$), and between all three variables ($F < 1$) were nonsignificant. This pattern (Figure 3) is consistent with the prediction that

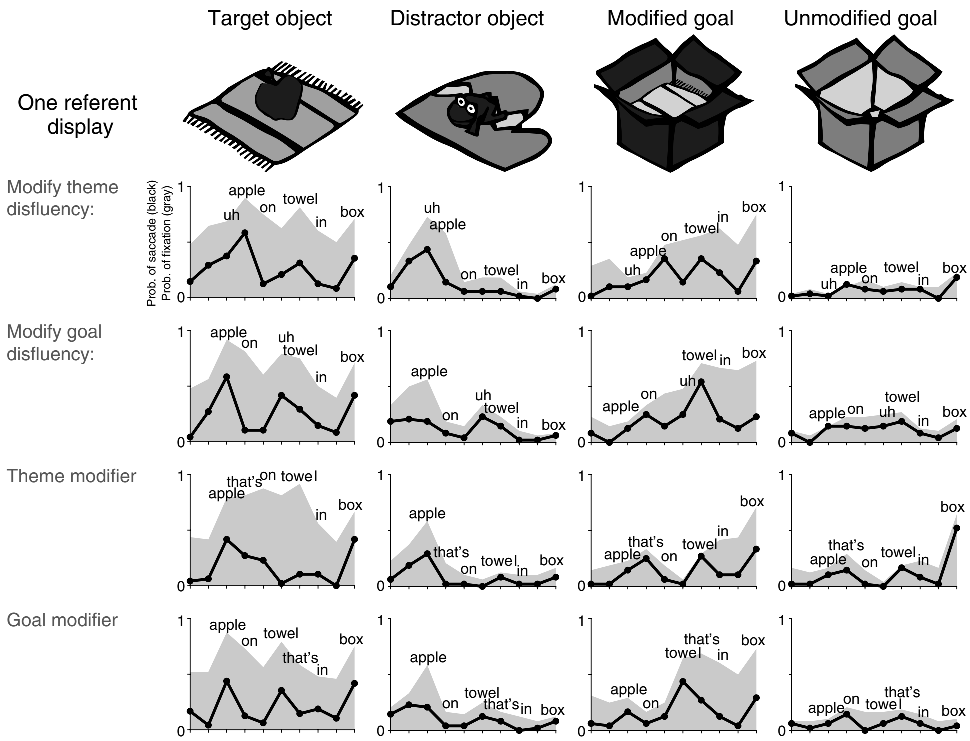

Figure 2. Probability of fixation on or saccade launch to regions of interest for each word in each utterance and display condition in Experiment 1. Gray polygons represent the probability of fixation; black line represent probability of saccade. The locations of content words, disambiguating function words, and disfluencies are indicated above the fixation graphs. The y-axis of each graph represents probability (0–1) and the x-axis the course of the utterance (one word per tick).

theme disfluencies and modifiers should elicit fewer looks to the modified goal (being consistent instead with a modified theme) than the corresponding modify goal utterances, but only in the one-referent display, where the identity of the theme has already been ascertained and the eye movement system is not engaged in deciding between the target and distractor objects. However, separate analyses of the disfluent conditions found only an effect of number of referents ($F1_{1,15} = 29.8$, $p < 0.001$; $F2_{1,23} = 24.9$, $p < 0.001$). The effect of cue location ($F < 1$) and the interaction between number of referents and cue location ($F1_{1,15} = 2.69$, $p > 0.1$; $F2_{1,23} = 2.29$, $p > 0.1$) were not significant, suggesting that the modifier conditions were carrying the overall interaction between the number of referents and cue location. This would suggest that disfluencies were not interpreted by the parser in the same ways as modifiers.

Similar patterns (Figure 4) are also present in the saccade data to the modified goal during *towel*, consistent with confirmatory saccades as the source of fixation patterns.

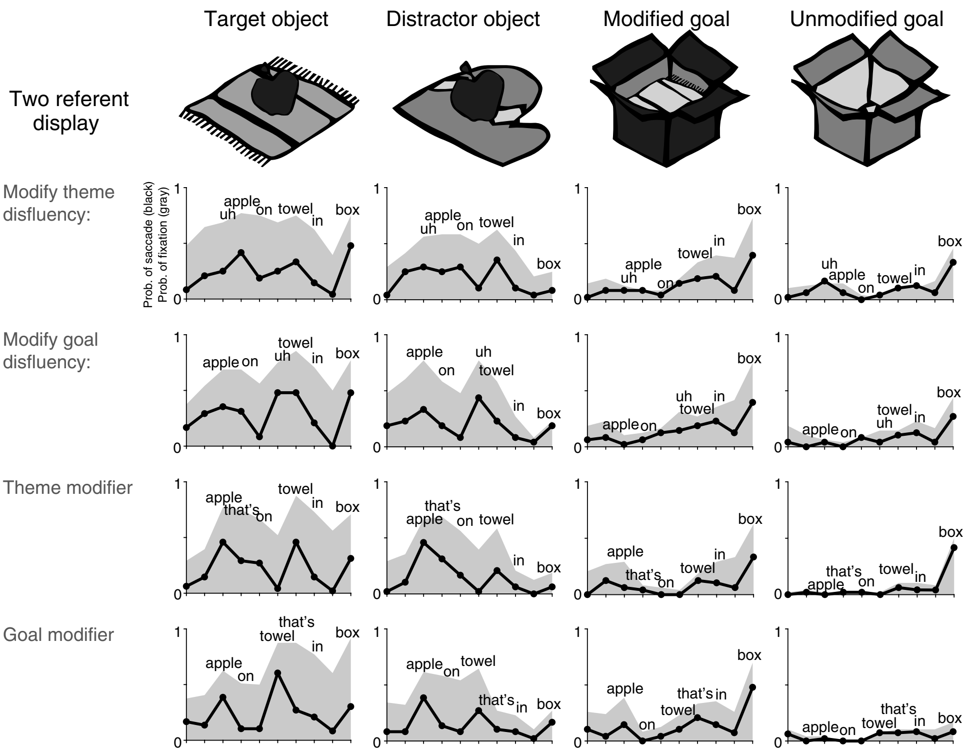

Figure 2. (*continued*)

Main effects of number of theme referents ($F1_{1,15} = 39.6$, $p < 0.001$; $F2_{1,23} = 36.8$, $p < 0.001$), cue location ($F1_{1,15} = 4.68$, $p < 0.05$; $F2_{1,23} = 2.63$, $p > 0.1$), and cue type ($F1_{1,15} = 5.03$, $p < 0.05$; $F2_{1,23} = 3.76$, $p > 0.05$) are again present (although the latter two effects are significant only by participants), as well as a marginal interaction between number of referents and cue location ($F1_{1,15} = 3.18$, $p = 0.095$; $F2_{1,23} = 4.17$, $p = 0.053$). All other interactions were nonsignificant ($F < 1$). Anticipatory saccades, however, may also have contributed to the probability of fixating the modified goal during *towel*, as saccades were also launched to this object during the word *on*, suggesting that these saccades may have been launched based on the expected arguments of the verb *put* (which requires both a theme and a goal when used imperatively; cf. Altmann & Kamide, 1999, 2004).

Separate analyses of disfluent conditions again revealed a main effect of number of referents ($F1_{1,15} = 14.4$, $p < 0.01$; $F2_{1,23} = 15.4$, $p < 0.01$), but only a marginal effect of cue location by participants ($F1_{1,15} = 3.23$, $p < 0.1$; $F2_{1,23} = 1.56$, $p > 0.1$) was present. The interaction between number of referents and cue location was nonsignificant ($F1 = 1.23$; $F2 = 1.87$). The marginal effect of cue location tentatively suggests that disfluencies may have some immediate effect on the parser; however, it is clear that the display itself had a much greater impact on eye-movement patterns.

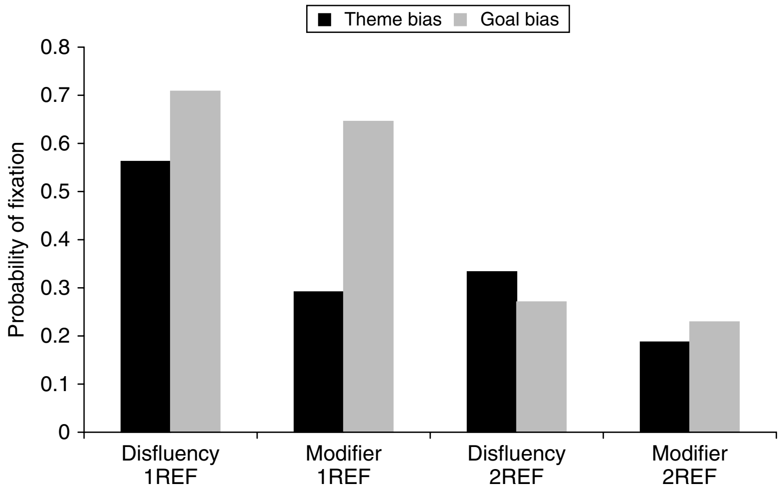

Figure 3. Probability of fixation on the modified goal (e.g. towel in a box) during the word *towel* for each of the eight utterance and display conditions. 1REF and 2REF refer to the number of possible theme referents in the display; theme bias and goal bias refer to the locations of the disfluencies; disfluency and modifier refer to cue types.

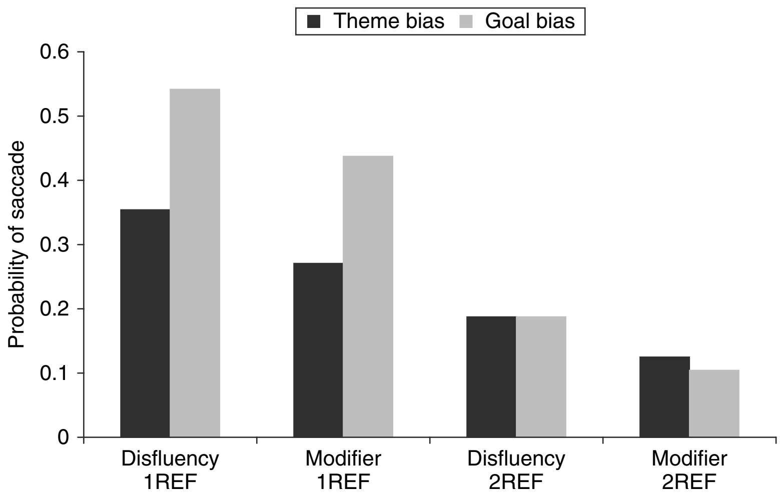

Figure 4. Probability of saccade launch to the modified goal (e.g. towel in a box) during the word *towel* for each of the eight utterance and display conditions. 1REF and 2REF refer to the number of possible theme referents in the display; theme bias and goal bias refer to the locations of the disfluencies; disfluency and modifier refer to cue types.

Disfluencies, then, do not affect the final interpretation of utterances in this study, contrary to what was suggested by previous experiments employing grammaticality judgment tasks (Bailey & Ferreira, 2003), nor do they appear to be strongly biasing the parser during the utterance. The strong form of the cueing hypothesis must therefore be rejected in favor of a hypothesis that can account for both current and previous results. A possible modification might suggest that further processing of the most salient current ambiguities (whether lexical, syntactic, referential, or discourse related) may occur during the disfluency, or that less salient ambiguities may become more salient. As a result, participants' eye movements may reflect the processing of possible resolutions of current ambiguities, and, in cases where the discourse context does not constrain the parse, the final parse may be affected (e.g. Bailey & Ferreira, 2003).

Evidence that ambiguities (not necessarily syntactic) are processed during disfluencies can be seen in the probability of saccade launch to each of the four objects during the filled pauses in the theme and goal disfluency conditions (Figure 5). The probability of launching a saccade (as opposed to probability of fixation) is sensitive to changes in visual attention (and by inference, cognitive operations) during a disfluency (Altmann & Kamide, 2004). As expected, participants are more likely to launch a saccade to the modified goal during the goal disfluency than the theme disfluency, regardless of condition, as indicated by a main effect of cue location (marginal by items; $F1_{1,15} = 5.70$, $p < 0.05$; $F2_{1,23} = 3.21$, $p = 0.086$), a nonsignificant main effect of number of referents, and a nonsignificant interaction between number of referents and cue location. This pattern is consistent with the fact that the goal disfluency occurs later in the utterance, often after the theme has been unambiguously identified.

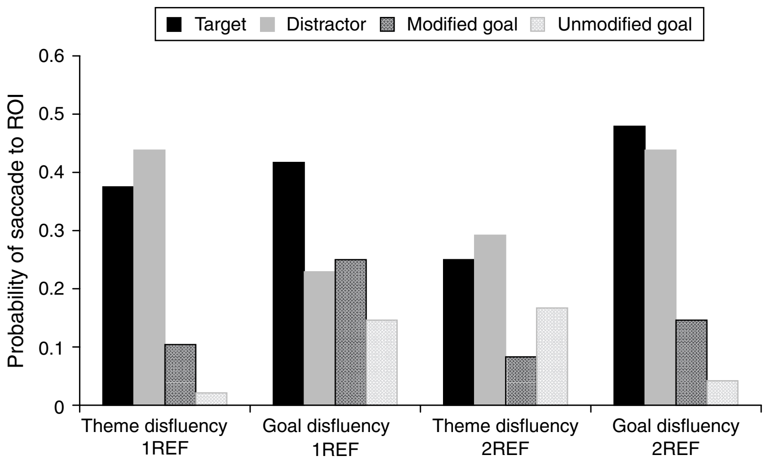

Figure 5. Probability of saccade launch to each of the four regions of interest for each of the disfluent conditions. 1REF and 2REF refer to the number of possible theme referents in the display; theme disfluency and goal disfluency refer to the locations of the disfluencies.

The pattern of results for the distractor object is more complex: Only the interaction between number of referents and cue location is significant ($F1_{1,15} = 5.31$, $p < 0.05$, $F2_{1,23} = 5.89$, $p < 0.03$).

Looks to the distractor are more likely in the theme-disfluency one-referent display condition and the goal-disfluency two-referent display condition. These results are key because they indicate why the display exerts such a powerful influence on parsing in the structurally ambiguous disfluency conditions. When a disfluency occurs later in the utterance (in the goal disfluency conditions), the distractor is still a possible candidate theme in the two-referent condition (leading to many looks to the distractor and few to either goal), as opposed to the one-referent condition (leading to fewer looks to the distractor and more looks to either goal in anticipation of the speaker identifying the goal). Moreover, because the theme disfluency occurs after the word *the*, participants could expect immediate resolution of the referential ambiguity in the one-referent condition, while immediate resolution is not expected until the goal disfluency in the two-referent condition. The interaction between number of referents and cue location thus shows that the language comprehension system is sensitive to immediately upcoming ambiguity resolution.

This ambiguity resolution hypothesis, moreover, suggests a mechanism by which we can account not only for the results in this experiment but also for previous syntactic cueing results (Bailey & Ferreira, 2003). Recall that in the current experiment, the number of possible continuations for any phrase that began with *the* was limited by co-present objects (*the* preceded the disfluency in all of the critical items in this study, as well as in the Bailey & Ferreira, 2003, study). Thus, it was easy for the listener to predict the actual object that would be referenced. However, in Bailey and Ferreira's (2003) study, no context was present. As a result, the number of possible lexical continuations at a disfluency was very large (limited only by the context introduced by the initial utterance fragment). On the other hand, the number of possible syntactic continuations was relatively small. Thus, the parser could have used the disfluent delay to consider less preferred structures, rather than possible lexical items. The grammaticality judgment task may have been sensitive to the occasions on which the parser identified the ultimately correct (less preferred) structure during the disfluency. "Bad" disfluencies may have occurred too early (i.e. [4]) to provide enough information to deduce possible structures, or so late (e.g. [2]) that the parser had committed to a single parse. "Good" disfluencies (i.e. [1] and [3]) may have occurred just late enough that less preferred structures could be identified, but not so late that multiple structures were no longer being considered. In essence, then, the nature of the grammaticality judgment task (which focuses participants on syntactic structure, with relatively little context) may have affected the way in which disfluencies were interpreted.

In the current experiment, however, it was possible for the listener to pick out the complete set of possible lexical continuations for any disfluency and interpret the disfluency as referential uncertainty. The final parse may therefore have been driven by the biases of the parser and the constraints of the display only (i.e. cue location did not lead to differences in commitments), leading to a null effect of disfluency location.

The processing done by the language comprehension system during a disfluency, then, may amount to identifying concurrent ambiguities, but the type of ambiguity that receives further processing depends on the partial parse and discourse context, and on the number of alternative continuations to be considered. Moreover, a particular type of ambiguity may be more salient than others in certain experimental settings (e.g. referential ambiguity in the VWP). This suggests that the language comprehension system uses the delay in propositional input and the distributional cues provided by disfluencies in very flexible ways that fit the comprehension goals of the listener.

Finally, these results suggest that filled pauses may provide an opportunity for studying the relative saliency of a variety of ambiguities during processing in different experimental paradigms, in that they provide a natural interruption of propositional input during which otherwise obscured ambiguity resolution processes may continue to run and thus be more easily observed. In fact, identifying the processes at work in the VWP may be especially important, as a model of cognition and language processing in this paradigm is necessary to ground and guide further study. Additional research is also needed to examine the processes that occur during filled pauses, to test the delay hypothesis described in this chapter, and to further understand the relationship between eye-movement patterns and language comprehension processes in the VWP.

References

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

Altmann, G. T. M., & 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: Eye movements and the visual world*. New York: Psychology Press.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language, 49*, 183–200.

Bortfield, H., Leon, S., Bloom, J., Schober, M., & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123–147.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274–296.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383–398.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201–242.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Ferreira, F., & Henderson, J. M. (1991) Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 31*, 725–745.

Ford, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 798–827). MIT Press, Cambridge, MA.

Fox Tree, J. E. (1995) The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709–738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition, 29*(2), 320–326.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language, 40*, 280–295.

Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech, 14*, 277–288.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 113–156.

Schacter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology, 60*, 362–367.

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

Shriberg, E. E. (1996). *Disfluencies in SWITCHBOARD*. Proc. International Conference on Spoken Language Processing, Addendum (pp. 11–14). Philadelphia, PA.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*, 25–38.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002) Effects of visual context in the resolution of temporary syntactic ambiguities in spoken language comprehension. *Cognitive Psychology, 45*, 447–481.

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.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73*, 89–134.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.