

RUNNING HEAD: DISFLUENCY EFFECTS IN COMPREHENSION

Disfluency effects in comprehension:
how new information can become accessible

Jennifer E. Arnold & Michael K. Tanenhaus
University of Rochester

Address all correspondence to:

Jennifer Arnold
University of Rochester
Dept. of Brain and Cognitive Sciences
Meliora Hall 495
RC Box 270268
Rochester, NY 14627
585-275-7187
jarnold@bcs.rochester.edu

Arnold, J. E. & Tanenhaus, M. K. (In press). Disfluency effects in comprehension: how new information can become accessible. In Gibson, E., and Perlmutter, N. (Eds) *The processing and acquisition of reference*, MIT Press.

ABSTRACT

Natural speech production often includes disfluencies, such as mid-utterance hesitations, filler words (“um”, “uh”), repeated or repaired words, and pronouncing “the” as /thiy/ (rhyming with “tree”) or “a” as /ey/ (rhyming with “say”) (Fox Tree & Clark, 1997). Yet laboratory research typically ignores disfluency. We describe an on-going research program that demonstrates how disfluency information affects on-line reference comprehension. In two eyetracking experiments, listeners followed instructions like “Put the grapes below the candle. Now put {the / thiy uh} candle...” Their eye movements at the word “candle” in the second instruction revealed a preference for given objects when the instruction was fluent, and for new objects when the instruction was disfluent. An offline experiment supported the idea that the fluency or disfluency in the article affect listeners expectations about what the speaker is referring to. These results are discussed in terms of how disfluency affects the accessibility of referents during reference comprehension.

Spontaneous speech is rarely fluent, resulting in mid-utterance hesitations, filler words (“um”, “uh”), repeated or repaired words, and pronouncing “the” as /thiy/ (rhyming with “tree”) or “a” as /ey/ (rhyming with “say”) (Fox Tree & Clark, 1997). It is often assumed that these phenomena are “performance errors” (Chomsky, 1965), and are merely a hindrance to uncovering the linguistic properties of the input. However, disfluency occurs systematically with respect to other features of the discourse, situation, and speaker’s intentions. This means that the presence, type and location of a disfluency carries potentially useful information. In this chapter we describe ongoing research into the question of how speaker disfluency might affect on-line reference comprehension. Our research is motivated by an interest in understanding the basic processes underlying reference comprehension. In particular, we believe that disfluency can provide a window onto how and why comprehenders perceive some things in a discourse as more accessible than others.

Most researchers agree that reference comprehension is affected by the accessibility of potential referents (e.g., Almor, 2000; Brennan et al. 1987; Dahan et al., 2002; Gernsbacher, 1990; McDonald & MacWhinney, 1995; Sanford & Garrod, 1981). Accessibility is generally assumed to be related to focus of attention, where speakers and listeners focus on some things in the discourse situation more than others. However, it seems necessary to assume that discourse participants coordinate the accessibility of entities jointly, at least partially. A speaker may have focused all her attention on an entity, but if that entity is not available to her addressee, then it is not felicitous to use referring expressions that are normally reserved for highly accessible entities, such as pronouns or deaccented noun phrases.

How, then, do interlocutors coordinate accessibility? The traditional approach to this question is to appeal to the discourse record, especially linguistic, textual, properties (e.g., Ariel, 1990; Chafe, 1976; 1994; Givón, 1983; Gundel et al., 1993; Prince, 1992). The way in which different entities have been treated in the textual history of the discourse is information that is available to all discourse participants. At the most basic level, entities that are **discourse-old** or **given** are considered more accessible than those that are **discourse-new** (Chafe, 1976, 1994; Gundel, 1988; Prince, 1992).

While there are many definitions of these terms, given information is often considered that which is known to the discourse participants, while new information is that which is not known. However, there are different domains over which “known” can be calculated. Prince (1992) distinguishes between information that is known to the hearer (“hearer-old”) and that which is both hearer old and has also been mentioned in the discourse (“discourse-old”). Some studies limit “given” status to only that which is both known and highly accessible. There are also different modes in which information can become given, e.g. through linguistic mention, visual presentation, and inferring something through association with something else that has been mentioned or presented (e.g., Prince, 1992; Haviland & Clark, 1974; Hankamer and Sag; 1976). Nevertheless, many studies operationalize givenness in terms of whether something has been referred to linguistically.

Among given entities, discourse properties also make some entities more accessible than others,. For example a first-mentioned entity in subject position is more accessible than entities mentioned later in non-subject positions (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Gernsbacher, 1990; Gordon, Grosz, & Gilliom,

1993; Hudson-D’Zmura & Tanenhaus, 1998). Other factors shown to be important are the thematic roles that entities have played (Arnold, 2001; Stevenson, Crawley, & Kleinman, 1993), the implicit causality of verbs (e.g., Garvey & Caramazza, 1974, Ehrlich, 1980; Garnham, Traxler, Oakhill, & Gernsbacher., 1996; McDonald & MacWhinney, 1995); focus constructions (Almor, 1999; Arnold, 1998), recency of mention (e.g., Givón, 1983), and parallelism (e.g., Chambers & Smyth, 1998).

The expectancy hypothesis

The discourse-history approach to accessibility has isolated a variety of factors that affect reference comprehension, but it does not explain why this diverse set of factors should affect the focus of attention of discourse participants. A potential answer to this question can be found in what we’ve termed the “expectancy hypothesis”. The expectancy hypothesis was first proposed by Arnold (1998; 2001), who used corpus studies to show that all of these textual characteristics are correlated with a higher likelihood for a particular entity to be mentioned in the following utterance. For example, if an entity is mentioned recently, and in particular in the position of grammatical subject of the preceding clause, it has a higher likelihood of occurring in the incoming clause than a non-recently mentioned or non-subject entity. That is, it has a higher expectancy. The same is true for recently vs. not-recently mentioned entities, entities that just played the thematic role of goal, compared to theme, entities in focus position of clefts, and entities in parallel syntactic position to the noun phrase that is currently being encountered. The expectancy hypothesis suggests that the textual history of the discourse can be a strong determinant of referent accessibility, insofar as it makes

some entities more likely to be mentioned in the upcoming discourse. If the comprehender perceives that an entity is likely to be mentioned, it is likely that the speaker is focusing attention on this entity. The expectancy hypothesis is based on the idea that accessibility is linked to the comprehender's assessment of what is considered most important to the task at hand, at each moment in time. This would need to be a dynamic assessment, as well as multi-dimensional, since different things can be "important" at the same time for different reasons. In this sense, the expectancy hypothesis focuses on what the textual approach originally intended to capture: the cognitive status of discourse entities (e.g., Gundel et al., 1993). Even though the above-mentioned features are meant to affect accessibility as a cognitive phenomenon, many studies focus instead on the effects of the linguistic features themselves. In fact, Grosz and Sidner (1986) explicitly define the attentional state component of their model as an "abstraction of the participants' focus of attention...a property of the discourse itself, not of the discourse participants" (p. 179).

The expectancy hypothesis contrasts with the discourse-history view of accessibility in that it suggests that other, non-textual, factors should also affect referent accessibility, if they affect the likelihood of that entity being mentioned. Here we test this idea by looking at how disfluency affects reference comprehension. While the discourse-history account suggests that given information is always more accessible than new information, the investigation of disfluency effects in comprehension offers an opportunity to investigate conditions under which new entities may become relatively more accessible, perhaps even more accessible than given entities.

Disfluency tends to occur when speakers are having trouble during language production (Clark and Fox Tree, 2002; Clark & Wasow, 1998; Fox Tree & Clark, 1997). There are many aspects of production the speaker could have trouble with, including deciding what to say, selecting a lemma, generating the phonetic plan, or executing it (Levelt, 1989). All of these processes are more difficult for new entities than given ones. This means that disfluent phrases like “thee uh...” may signal comprehenders that the speaker is about to refer to something new – i.e., something that has not already been mentioned.

The research described in this paper builds directly upon Dahan, Chambers, & Tanenhaus’s (2002) findings that comprehenders are biased toward given referents when interpreting fluent definite noun phrases, with different preferences for accented and deaccented noun phrases. Listeners were asked to follow instructions such as “Put the candy above the diamond. Now put the candy/CANDY...” The target noun phrase was the theme of the second instruction, and it was either accented or deaccented. The visual display included four objects, two of which were cohort competitors, for example a candy and a candle. These objects had names with overlapping initial segments, which meant that the target noun was temporarily ambiguous during the initial part of the word (“cand...”). Eye movement data revealed that for deaccented nouns, participants quickly converged on the target, which was the given and most highly accessible entity from the preceding instruction, and experienced little competition from the unmentioned cohort. For accented nouns, participants initially fixated more on the cohort object, revealing a preference for accented noun phrases to refer to an entity that was discourse-new, but visually present, in comparison with a given and highly accessible one. A subsequent

experiment revealed that comprehenders prefer to link accented nouns with referents that have been mentioned, but are not highly accessible, as in “Put the necklace above the candy. Now put the CANDY...”, as opposed to referents that have not been mentioned at all. That is, even though all of the objects were visually available to the experiment participants, both accented and deaccented nouns are preferentially associated with entities that are linguistically given. This research joins other studies showing that reference comprehension is facilitated for given, especially linguistically given information (e.g., Haviland & Clark, 1974; Clark and Haviland, 1977).

Like Dahan et al. (2002), we present listeners with natural-sounding instructions in the context of concrete, real-world referents, and use free head eyetracking to observe the time course of listeners’ hypotheses about what the speaker is referring to. The use of this methodology forms part of an additional goal of our research, to study language comprehension in naturalistic situations, investigating speech characteristics that are normally absent from psycholinguistic experiments. \ While most laboratory research ignores “messy” speech characteristics like disfluency, overlapping speech, or sentence fragments, these features are characteristic of natural language use. In these ways, this research represents an attempt to shed light on the issues around which this volume is organized.

Following Dahan et al.’s methodology, we also use situations with cohort competitors, which introduces a temporary ambiguity of the referent of the target word. Research with the eyetracking visual world paradigm has shown that comprehenders tend to look at the entities they are considering as referents of temporarily ambiguous expressions, revealing a fine-grained picture of the competitor set for particular

expressions and discourse conditions (Allopenna et al., 1998; Dahan et al., 2002; Tanenhaus et al., 1996). Although these cohort competitor effects seem subtle, there is a growing body of evidence that a variety of manipulations can change the magnitude of competitor effects, including target and competitor frequency (Dahan, Magnuson, & Tanenhaus, 2001), and accent (Dahan et al., 2002). Here we investigate the relative proportion of looks to the referents of cohort competitors (e.g., a picture of a candle or camel) to investigate the relative availability of each entity for fluent and disfluent referring expressions.

This line of research promises to not only address questions about reference comprehension and accessibility, within a natural setting, but also to provide a rich testing ground for other questions about language processing. Experiments 2 and 3 (described below) begin to explore the different cues that lead to the perception of disfluency, in particular the relationship between traditional manifestations of disfluency and the prosodic characteristics of the utterance. Ongoing research, described briefly below, asks what kind of mechanism underlies the effects of disfluency, and in particular whether comprehenders make attributions about the cause of speaker disfluency. We hope that these studies will shed light on the more general question of whether language comprehension takes into account estimations of the production processes and intentions of the speaker.

THE SYSTEMATIC DISTRIBUTION OF DISFLUENCY IN PRODUCTION

The reason that disfluency provides a potential cue to comprehenders is that it reflects normal production processes, and therefore occurs systematically with respect to

features of the discourse and speech situation. Research on the production of disfluency has suggested that speakers are more likely to be disfluent when they are having trouble with some aspect of language production, and need an extra planning time (e.g., Clark & Wasow, 1998; Fox Tree & Clark, 1997). Clark and colleagues argue that disfluency results from the tension between the time needed to plan upcoming speech (the “formulation imperative”), and the need to avoid long delays or silences, which may signal that the speaker is no longer participating in the conversation (the “temporal imperative”; Clark, 1996; Clark & Wasow, 1998).

There are many ways in which production difficulty changes the speech signal. Many of these are identified as forms of disfluency, including disfluent words “um” or “uh” (e.g., Clark and Fox Tree, 2002), repeats or repairs (e.g., Clark and Wasow, 1998), or pronouncing “the” as “thee” (rhyming with “tree”; Fox Tree and Clark, 1997). Notably, any particular problem in production may result in many such changes to the speech signal, which leads to the tendency for multiple disfluent productions to co-occur (e.g., Fox Tree and Clark, 1997). Production difficulty also affects the prosodic characteristics of speech. It is well documented that words tend to be pronounced with longer durations in situations of production difficulty, or when speakers need an extra moment to plan (Bell, et al., 2003; Gregory, Joshi, & Sedivy, 2003; Griffin, 2003). Mid-utterance pauses and hesitations can also constitute a form of disfluency, although they are often ignored due to the difficulty in distinguishing disfluent from nondisfluent pauses (Fox Tree, 1995). But even the placement of non-disfluent prosodic breaks is influenced by the difficulty of just-completed and upcoming constituents. Watson (2002; Watson and Gibson, 2003) found that the likelihood of a speaker producing an

intonational phrase boundary depends on the complexity of the following constituent, perhaps because speakers need this break to plan upcoming material.

Since disfluent productions tend to occur when the speaker is having difficulty, it stands to reason that they would occur more in some situations than others. Research on language production shows that indeed this is the case. For instance, disfluent speech is most likely to occur at the onsets of complex syntactic constituents (Clark & Wasow, 1998) and intonational phrases (Clark & Fox Tree, 2002), suggesting that these are the units over which speakers do much of their planning. The likelihood of disfluency is also higher at the onset of syntactically complex constituents, suggesting that the difficulty of formulating a constituent is related to its complexity (Clark and Wasow, 1998).

Speakers also are more likely to be disfluent when referring to a difficult-to-name object, for example new entities that have not been mentioned in the discourse. There are many aspects of production that the speaker might have trouble with, including deciding what to say, selecting a lemma, generating the phonetic plan, or executing it (Levelt, 1989), all of these processes are likely to be more difficult for new entities than given ones.

A corpus analysis confirms that speakers are more likely to be disfluent when they are producing a noun phrase that refers to something new. We analyzed data that were collected for a different experiment (Arnold, Wasow, Losongco, and Ginstrom, 2000), in which speakers gave instructions to addressees (both naïve participants) about giving objects to animals, e.g. “Give the corn to the pink duck”. The objects were physically present on a table in front of both participants, and the animals were represented by pictures glued onto boxes. Cue cards led the addressee to begin each trial

with a question about either the animals or the objects, leading one or the other to be “given”. The objects were either simple ones that were generally described with short phrases (e.g., “the scissors”), or sets of objects that required longer descriptions (e.g., “the cup with blue and red spots”).

The speakers’ instructions were transcribed, and the noun phrases referring to each referent (animal and object) were coded for a) whether the object was given or new within the trial, b) whether the NP was “long” (3 or more words) or “short” (1-2 words); and c) whether the speaker was disfluent during or immediately before the noun phrase. “Disfluent speech” included the presence of “um” or “uh”, repeated or repaired words, “thiy” for “the” or “ay” for “a”, and comments like “huh”, or “I mean”.

Analyses showed that a noun phrase was more likely to be disfluent when it was new (21%) than when it was given (16%). Disfluencies were also more common for long NPs than short ones, and for theme NPs (objects) than goal NPs (animals). However, neither of these factors accounted for the correlation between disfluency and new referents. ANOVAs showed main effects of both given/ new status ($F(1,45) = 19.489, p < 0.001$; $F(1,23) = 5.496, p < 0.05$) and NP length ($F(1,45) = 16.801, p < 0.001$), $F(1,23) = 8.245, p < 0.01$), but no interaction (p 's > 0.5). A separate ANOVA considered given/new status with respect to NP type (theme or goal), and revealed main effects of given/new status: ($F(1,45) = 15.210, p < .001$) and NP type: ($F(1,45) = 16.821, p < .001$), but no interaction ($p > .3$).¹

¹ The items analysis could not be performed because of missing cells.

DOES DISFLUENCY AFFECT COMPREHENSION?

Given that disfluency correlates with reference to new objects, we wanted to know whether comprehenders make use of this information during comprehension. We tested this idea in two on-line experiments (Exps. 1 and 2), and one offline experiment (Experiment 3). In experiments 1 and 2, participants followed fluent and disfluent instructions (e.g., “Now put thiy, uh...” vs. “Now put the...”; see Table 1) to move objects on a computer screen (Fig. 1), while their eye movements were recorded with an head-mounted eyetracker. On each trial the scene contained two cohort competitor objects (e.g. candle, camel) and two distracters (e.g., grapes, salt shaker). The names of the distracters had no phonetic overlap with those of the cohorts. Participants heard one of the four pairs of instructions resulting from the cross of the context and target phrases in Table 1. The context instruction established the target as either given or new, and the second (critical) instruction was either disfluent or fluent.

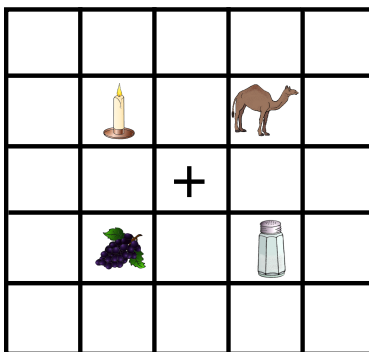


Figure 1. Sample visual display for all experiments

TABLE 1: Sample instructions .

GIVEN CONTEXT INSTRUCTION:	Put the grapes below the candle.
NEW CONTEXT INSTRUCTION:	Put the grapes below the camel.
FLUENT TARGET INSTRUCTION:	Now put <u>the CANDLE</u> below the salt shaker
DISFLUENT TARGET INSTRUCTION:	Now put <u>thiy, uh, CANDLE</u> below the salt shaker.

All target words were accented, which meant that the fluent NPs were predicted to lead to a bias toward the given but less accessible entity from the context sentence, i.e. the second-mentioned object (cf. Dahan et al., 2002). By contrast, if comprehenders are sensitive to the correlation between disfluency and reference to new objects, the disfluent NP (which was also accented) should lead to a bias toward the discourse-new entity.

One of the challenges of studying features of naturally-occurring language, like disfluency, is to maintain experimental control and still present participants with natural-sounding stimuli. While disfluency is a frequent part of language, it may sound out-of-place in a laboratory setting, where scripted, fluent speech is the norm. We overcame this problem by telling participants that the instructions had been generated by another subject in the context of the same visual scene, while in fact they had been recorded by the experimenter. It was emphasized that the speaker was shown **what** to say by graphic cues, but had to come up with her own words. A post-experiment questionnaire

confirmed that most participants believed the story, and those that expressed any doubts about the natural production of the stimuli were excluded from the analyses (less than 6%).

Experiment 1

Sixteen participants' eye movements were monitored as they viewed scenes like Figure 1, and followed instructions like those in Table 1. Each of the 16 experimental items were rotated through the four conditions resulting from the cross of the context and target instructions, and combined with 32 fillers. The target item (camel vs. candle) was also manipulated as a control variable, resulting in 8 lists, with both forward and backward versions. All filler items contained cohorts; half began like the target items, but did not mention either cohort in the second utterance. The other half mentioned no cohort in the first utterance. Half the fillers contained disfluencies of various types and in various locations. The visual stimuli were versions of the Snodgrass and Vanderwart (1980) pictures, colored and normed for frequency, visual complexity, and familiarity (Rossion and Purtois, 2001). These dimensions were counterbalanced across items, so on average the properties were the same for the target and competitor (c.f. Dahan et al., 2001). The location of the cohorts was also counterbalanced across items.

The instructions were recorded to achieve a natural-sounding fluent or disfluent instruction. As in naturally-occurring speech, there were a number of features that differentiated the fluent from disfluent items: 1) Disfluent “thiy” vs. fluent “thuh”, 2) presence vs. absence of “uh”, 3) longer durations for “Now” and “Put” in the disfluent condition, and 4) a higher pitch excursion on “Now” in the disfluent condition. The

prosodic characteristics of the disfluent “Now put” resulted in the impression that the speaker was thinking, henceforth termed “thinking prosody”.

In condition (a), the fluent, accented NP provided an initial bias toward the given but nonfocused object, so we expected faster looks to the target in the given condition, and a bigger cohort effect in the new condition (replicating Dahan et al., 2002). If the disfluency shifts attention to discourse-new entities, in condition (b) participants should show increased looks to the target in when it is new, and a bigger cohort effect when it is given.

The results supported these predictions. Starting 200 msec after the onset of the head noun (e.g., “candle”), there was an interaction between disfluency and referent, such that there were more looks to the competitor in the disfluent/given and fluent/new conditions. That is, the disfluent condition gave rise to a preference for the new cohort, and the fluent condition to a preference for the given cohort. These results show that the previously established bias toward given entities holds only for fluent referring expressions. Disfluent expressions disrupt this pattern, instead facilitating reference resolution for discourse-new objects (see Arnold, Fagnano, & Tanenhaus, 2003; for details).

These results demonstrate that disfluency affects on-line reference resolution. But they raise the question of what features of disfluency contribute to the effect, since our fluent and disfluent conditions differed on multiple dimensions. Of particular interest is the different pitch contour on the words “Now put”. Pitch movement can signal different patterns of accenting, which have also been linked to information status (e.g., Dahan et al., 2002; Terken & Hirschberg, 1994). Experiment 2 investigated the role of pitch in this

effect.

Experiment 2

We used the same materials and methods as in Experiment 1, except that the pitch contour on “Now put” was manipulated explicitly, resulting in a 2 (given vs. new) x 2 (fluent vs. disfluent) x 2 (large pitch excursion vs. small) design. Target identity (e.g. camel vs. candle) was not manipulated because there were no main effects or interactions with this variable in Experiment 1. The given/new conditions were created by cross-splicing each target sentence with one of the two context sentences, as in Experiment 1. Fluency and pitch were manipulated by creating a single “Now put” for each condition, and cross-splicing these into each item (disfluent/large excursion, disfluent/small excursion, fluent/large excursion, fluent/small excursion). These stimuli were further controlled by cross-splicing a single target word into each condition of an item.

Acoustic analyses with *Praat* (Boersma and Weenink, 2002) confirmed that the “large” conditions had greater pitch ranges than the “small” conditions” (disfluent/large: 120 Hz; disfluent/small: 21 Hz; fluent/large: 108 Hz.; fluent/small, 38 Hz). Duration was maintained as a feature of disfluency, since it tends to co-occur with other forms of disfluency (Gregory, et al., 2003; Bell, et al., 2003; Shriberg, 2001). That is, the “Now put” segments in the disfluent conditions were longer (disfluent/large: 965 msec; disfluent/small: 1026 msec) than in the fluent conditions (fluent/large: 642 msec; fluent/small: 531 msec). The amplitude of “Now” and “put” was roughly the same for all conditions.

The results replicated those for Experiment 1. Again there was an interaction

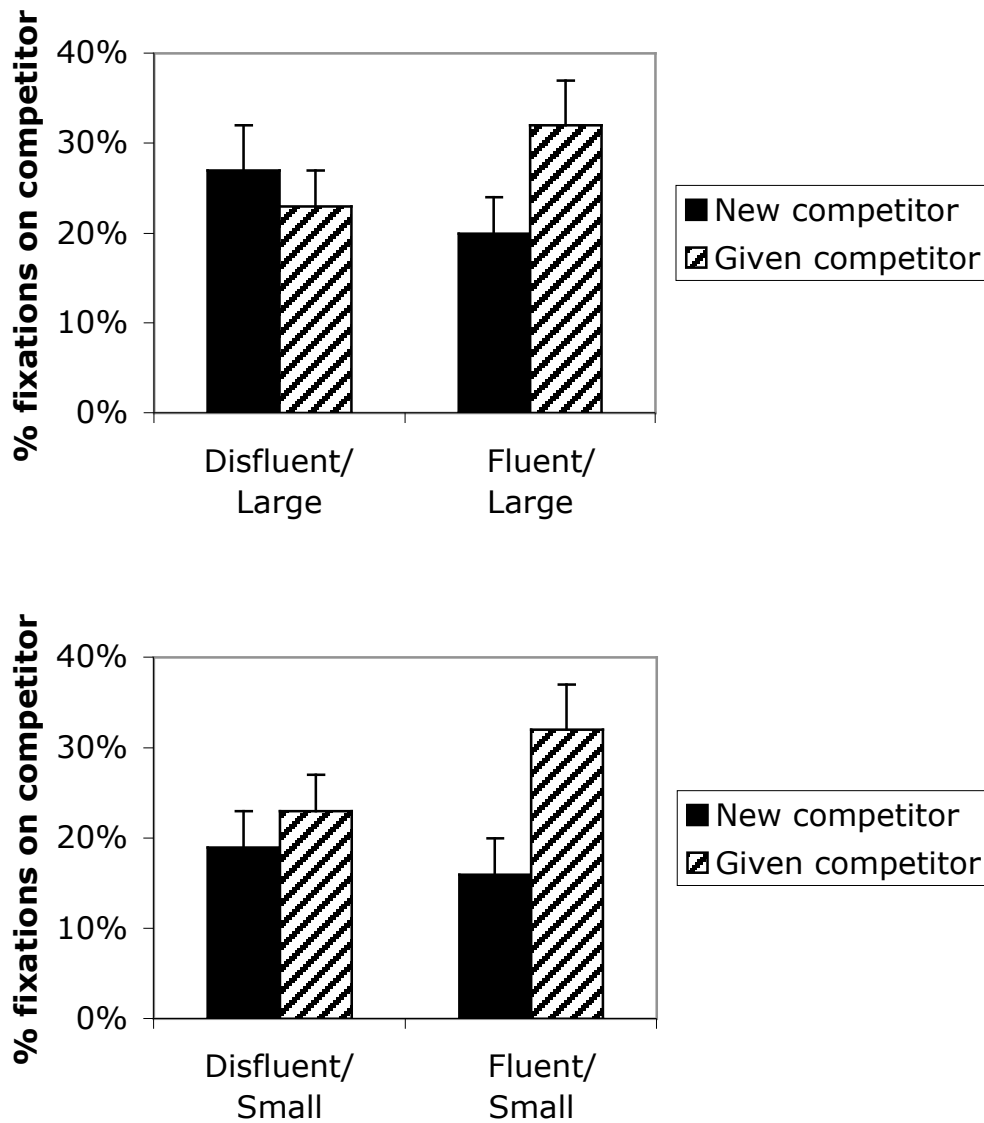
between disfluency and referent beginning 200 msec after the onset of the target noun. Participants fixated on the competitor more in the fluent/new condition than the fluent/given condition, but this bias disappeared for the disfluent stimuli. There were no significant effects of the pitch manipulation or interactions between it and other variables. (see Arnold, Tanenhaus, Altmann, & Fagnano, 2003, for details).

However, Figure 2 shows that the disfluency effect was numerically greatest when a large pitch excursion supported the idea that the speaker was having some kind of trouble. Furthermore, separate analyses of the two pitch conditions revealed that the disfluency by referent interaction was only reliable in the large pitch excursion condition (although it was marginal in the participants analysis, $F(1,31) = 3.633$; $p = .066$; $F(1,8) = 9.255$; $p = .016$), and not in the small pitch excursion condition ($F(1,31) = 2.677$, $p = .112$; $F(1,8) = 2.116$, $p = .184$)².

These results suggest that comprehenders perceive disfluency through multiple cues. When disfluency was signaled through only the presence of “thiy uh” and longer durations on “Now put”, the presence of disfluency removed the given bias that occurs with fluent stimuli. When a large pitch excursion contributed to the impression that the speaker needed a moment to think, the disfluent stimuli lead to an initial bias toward the new cohort. These findings underscore the need to think of disfluency as multifaceted; many features of the signal may reflect a single underlying source of production difficulty. The more the speech signal supports the impression of disfluency, the stronger the effects on processes like reference resolution. In particular, when “thinking prosody”

² The analyses of variance with participant as the random variable also included disfluency and referent (given vs. new) as independent variables. The items analyses also

occurs early on in an utterance, it may foreshadow later production difficulty. This information may build over time, as it is supported by other speech characteristics, and lead listeners to generate expectations about what the speaker is likely to refer to.



Figures 2a and b. % fixations on the competitor from 200-600 msec after the onset of the target noun (e.g., “candle”)

included disfluency, referent, and item group (i.e., those items that rotated together through the lists).

We cannot, however, reduce the disfluency effect to pitch. If that were this case, we would expect similar results in both “large excursion” conditions, whether fluent or disfluent. Instead, in both fluent conditions comprehenders fixated on the competitor more when it was given than when it was new. While pitch contributed to the new bias in disfluent conditions, the most reliable information about disfluency came from the words “thiy uh” and longer word durations.

In sum, when the speaker was fluent, listeners were biased toward given objects. But when listeners perceived that the speaker was being disfluent, new objects became relatively more accessible. There were many cues that may have signaled disfluency, but different pitch excursions on “Now put” was not necessary to observe the contrast between disfluent and fluent conditions.

These results are consistent with the expectancy hypothesis, which suggests that accessibility is influenced by the comprehender’s perception of the likelihood that the speaker will refer to that entity. The results from Experiments 1 and 2 show that disfluency affects the accessibility of entities during reference comprehension,. The correlation between disfluency and reference to new objects further suggests that disfluency may provide information that the speaker is relatively more likely to be referring to something new than something given. We tested this idea more specifically in Experiment 3.

Experiment 3

If disfluency leads comprehenders to perceive a higher likelihood of new entities being mentioned, we should see this reflected in listener’s guesses about what the speaker is likely to say, at different points in the utterance. We tested this idea in an off-line

experiment, where we played instructions from Experiment 2 that were truncated either after “Now put”, or after “Now put the/ thiy uh”. This experiment provided a further test of the information carried by the prosodic characteristics of “Now put” and the other cues to disfluency.

Twenty-four participants were presented with eight of the experimental items in Experiment 2, combined with four fillers. The experimental items occurred in one of two length conditions, where listeners heard either “Now put”, which contained the pitch (and duration) information, but no other cues to disfluency, or “Now put the/thiy uh”, which provided more definitive evidence of fluency or disfluency. Their task was to choose which object on the screen they thought the speaker was about to mention. The filler items were truncated at some point after the onset of the target word, to encourage participants to pay attention.

Results showed that after only “Now put”, participants chose one of the new objects 63% of the time. But after “Now put the/thee uh”, the proportion of new objects chosen rose to 83% for the disfluent stimuli, and dropped to 35% for the fluent stimuli. There were no main effects or interactions with the prosody manipulation, i.e. whether the “Now put” had a large pitch excursion or a flatter pitch..

These results suggest that listener’s expectations are not influenced by the thinking prosody on “Now put” alone. Although “thinking prosody” foreshadows other manifestations of disfluency in one condition, it was not in and of itself enough to create a new bias. However, the occurrence of definitive information about disfluency leads listeners to probabilistically expect upcoming reference to something new. By contrast, definitive information about fluent speech increases the expectancy of the given objects.

DISCUSSION

The results from experiments 1 and 2 provide strong support for the hypothesis that disfluency affects reference comprehension. Listeners showed a preference to fixate the discourse-new cohort when the instruction was disfluent, and the discourse-given cohort when the instruction was fluent. These biases occurred before the target word could be disambiguated on the basis of speech information alone. These results, along with those from experiment 3, suggest that disfluency influences listeners' expectations about what the speaker is referring to. In fact, the difference between fluent and disfluent conditions occurred beginning 200 msec after the onset of the target word. Since it takes people approximately 200 msec to program and launch an eye movement, this is the earliest we could expect to observe any effects from the speech signal. Thus, fluency information is driving reference resolution as early as the phonetic information from the input.

These data further show that under some circumstances, given (mentioned) entities are no more accessible than ones that have not been mentioned, and possibly less so. This suggests that accessibility can not be purely described in terms of how an entity has been treated in the textual history of the discourse. In our stimuli, given entities with an identical discourse history could be either more or less accessible, depending on the fluency of the target instruction.

What mechanism underlies the disfluency effect?

While the above findings show that disfluency affects language comprehension, they raise the question of exactly how they do so. Of particular interest is the question of the degree to which listeners make attributions about why the speaker is being disfluent, and use this information to guide reference comprehension.

On one hand, listeners may use an attributional mechanism. They may use disfluency to infer that the speaker is having trouble with production, and identify possible sources of difficulty. It is more difficult to refer to new objects than given ones, which makes new referents more plausible sources of difficulty than given ones. At a more intuitive level, it would seem strange for a speaker to be disfluent when referring to something that was just mentioned (Do you see the candle? Pick up, thee, uh, candle,”), unless there was some additional reason for the speaker to be distracted. Do comprehenders make these inferences quickly enough to drive the reference comprehension? If so, it would be dramatic evidence that comprehension processes are influenced by representations of the mental processes of the speaker. Such evidence would be relevant to the current theoretical debate over the degree to which language processing is influenced by representations of the mental processes, goals, and perspectives of one’s interlocutor.

On the other hand, the disfluency effects described here are also consistent with a correlational mechanism. Since disfluent NPs are correlated with new referents, one possibility is that listeners store and use this correlational information automatically. When they hear signs of disfluency, it may increase the activation of discourse representations for referents that are evoked in the discourse situation (e.g., visually), but

have not been mentioned yet. At the same time, activation may fall for discourse representations that have been recently mentioned, perhaps proportional to the accessibility of the entity. Critically, this correlational information can be calculated on the basis of information about the referents themselves (e.g., whether they have recently been mentioned), without concern for either why the speaker is being disfluent, or whether new objects are difficult to name.

Preliminary evidence from a second line of research sheds light on this question. If disfluency effects arise from a correlation between disfluent speech and reference to certain types of entities (e.g. new ones), the processing system would need to keep track of correlations for all the types of entities that become more accessible when speakers are disfluent.

In an initial experiment, we investigated whether disfluency also introduces an expectation that the speaker is referring to a novel, complex object (e.g., a funny squiggly shape), compared with a known object (e.g., an ice cream cone). Participants viewed a scene like in Figure. 3, with two novel and two known objects, each in two colors. They heard instructions like “Click on {the/ thiy uh} red ice cream cone,” or “Click on {the/ thiy uh} red funny squiggly shape that looks kind of like a monkey.” We examined fixations during production of the color word, that is, before the target noun was heard. In the fluent condition, where there was no expectation for either a known or novel target, there were increased fixations on both color-matched objects (e.g., the red ice cream and the red squiggly shape). By contrast, disfluency created an expectation for a novel target, resulting in increased fixations on only the novel color-matched object.

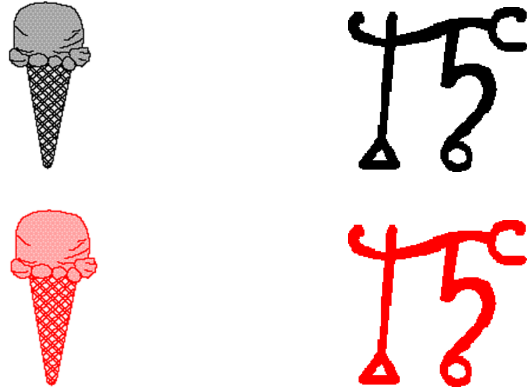


Figure 3. Sample visual display in the novel/known experiments. The top two items here were in black in the actual experiment, the bottom two in red.

.Disfluency therefore facilitates reference comprehension for at least two kinds of entities: those that are new to the discourse and those that are novel. Although these results are also consistent with both attributional and statistical mechanisms, they complicate the kind of statistics that would be needed. Listeners would need to store information about two correlations (disfluency with new referents, and disfluency with novel referents). Alternatively, they might learn that disfluency correlates with “difficult-to-name objects”. This, however, would require making decisions about which objects are difficult to name, thus narrowing the distinction between the attributional and statistical mechanisms.

An ongoing study more directly tests whether an attributional mechanism underlies the effect of disfluency, by asking whether the novelty bias disappears in the context of an alternate plausible cause for disfluency. Although disfluency may be plausibly attributed to difficulty in formulating a description for new or novel objects, it may also be the result of other problems, for example when the speaker is momentarily distracted. We repeated the novelty experiment described above, but this time we inserted a beep at the onset of half the trials. Participants were told that the speaker was

doing a secondary task of pressing a button at the beep. If this beep (and the putative resulting distraction) is a plausible cause of disfluency, we would expect to see an attenuation of the novelty effect. That is, if comprehenders attribute the disfluency to the distraction, they may not have as high an expectation for the speaker to refer to a difficult-to-name object, such as the novel one.

In sum, the above results clearly establish that disfluency affects on-line reference comprehension, increasing the accessibility of both new (unmentioned) referents, and novel objects with no common name. These results support the need for language comprehension research to consider features of naturally-occurring discourse, like disfluency, to fully understand the processes of language comprehension (Ferreira, Lau, & Bailey, to appear). The results also suggest that the impression of production difficulty may stem from a variety of speech characteristics, including prosodic ones. Finally, this line of research offers a vehicle for exploring some fundamental questions about the mechanisms of language comprehension. In particular, disfluency research may shed light on questions about whether listeners make attributions about why speakers say things in a particular way, which is related to questions about whether listeners model the thought processes, intentions, and goals of the speaker.

ACKNOWLEDGEMENTS

We are grateful to Rebecca Altmann, Maria Fagnano, and Dana Subik for their help in collecting and coding the data, Katherine Crosswhite for her help with intonational analyses, and Bob McMurray for his help with the setup and analysis of the novelty experiments. We also thank Sarah Brown-Schmidt, Mikhail Masharov, and Duane Watson for helpful discussion about the experimental design. This research was partially supported by NIH grants HD-41522 to J. Arnold and HD-27206 to M. Tanenhaus.

REFERENCES

- Allopenna, P. D., Magnuson, J. S. & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419-439.
- Almor, A. (1999). Noun-Phrase Anaphora and Focus: The Informational Load Hypothesis. *Psychological Review*, 106(4), 748-765.
- Arnold, J. (1998). *Reference form and discourse patterns*. Ph.D. dissertation, Stanford University.
- Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research*, 32(1), 25-36.
- Arnold, J., Wasow, T., Losongco, T. & Ginstrom, R. (2000a). Heaviness vs. Newness: the effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28-55.
- Brennan, S. E., Friedman, M. W. & Pollard, C. J. (1987). A Centering Approach to Pronouns. *Proceedings from the 25th Annual meeting of the Association for Computational Linguistics*. July 6-9, 1987.

- Chafe, W. (1994). *Discourse, Consciousness, and Time*. Chicago: Chicago University Press.
- 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, Inc.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- 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.
- Clark, H. H. and Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse Production and Comprehension, volume 1 of Discourse Processes: Advances in Research and Theory* (pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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.
- Dahan, D., Magnuson, J.S. & Tanenhaus, M.K. (2001). Time course of frequency effects in spoken word recognition: evidence from eye movements. *Cognitive Psychology*, **42**, 317-367.
- Dahan, D., Tanenhaus, M.K. & Chambers, C.G. (in press). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language*. **47**: 292-314
- Ehrlich, K. (1980). Comprehension of pronouns. *Journal of Experimental Psychology*, **32**.247-255.
- Ferreira, F., Lau, E. F., and Bailey, K. G. D. (To appear). Disfluencies, Language Comprehension, and Tree Adjoining Grammars. *Cognitive Science*.
- Fox Tree, J. E. & Clark, H. H. (1997). Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, **62**, 151-167.
- Garnham, A., Traxler, M., Oakhill, J. & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, **35**.517-543.
- Garvey, C. & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, **5**, 459-64.

- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Givón, T. (1983). Topic Continuity in Discourse: An Introduction. In T. Givón (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study* (pp. 1-42). Amsterdam: John Benjamins Publishing.
- Gordon, P. C., Grosz, B. J. & Gilliom, L. A. (1993). Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science*, *17*, 311-347.
- Gregory, M., Joshi, A., and Sedivy, J. (2003, March). *Adjectives and processing effort in production: So, uh, what are we doing during disfluencies?* Paper presented at the 16th annual CUNY conference on human sentence processing. Boston, MA.
- Griffin, Z. M., & Garton, K. L. Procrastination in speaking: *Ordering arguments during speech*. Paper presented at the 16th annual CUNY conference on human sentence processing. Boston, MA.
- Gundel, J. K., Hedberg, N. & Zacharaski, R. (1993). Cognitive Status and the form of referring expressions. *Language*, *69*(2), 274-307.
- Hankamer, J., & Sag, I. (1976). Deep and Surface Anaphora. *Linguistic Inquiry*, *7*, 391-426.
- Haviland, S. E; Clark, H. H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning & Verbal Behavior*, *13*(5), 512-521.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge: MIT Press.
- McDonald, J. & MacWhinney, B. (1995). The Time Course of Anaphor Resolution: Effects of Implicit Verb Causality and Gender. *Journal of Memory and Language*, *34*, 543-566.
- McDonald, J. & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*, 543-566.
- Prince, E. (1992). The ZPG letter: subjects, definiteness, and information-status. In S. Thompson & W. Mann (Eds.), *Discourse description: diverse analyses of a fund raising text* (pp. 295-325). Philadelphia/Amsterdam: John Benjamins B.V.
- Rossion, B. & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition [Abstract]. *Journal of Vision*, *1*(3), 413a, <http://journalofvision.org/1/3/413>, DOI 10.1167/1.3.413. Sag and Hankamer, 197?
- Sanford, A. J. & Garrod, S. C. (1981). *Understanding Written Language*. New York: Wiley.

- Shriberg, E. (2001). To 'errr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31(1).
- Snodgrass, J. G. & M. Vanderwart (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory* 6(2), 174-215.
- Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position. *Language and Speech*, 37, 125-145.
- Watson, D. & Gibson, E. (2003) The relationship between intonational phrasing and syntactic structure in language production. Unpublished manuscript, University of Rochester.
- Watson, D. (2002) Intonational Phrasing in Language Production and Comprehension. M.I.T., Dissertation.