Developmental Science

Developmental Science (2011), pp 1–10 DOI: 10.1111/j.1467-7687.2011.01049.x

PAPER

Toddlers use speech disfluencies to predict speakers' referential intentions

Celeste Kidd,[1] Katherine S. White[2] and Richard N. Aslin[1,3]

1. *Brain & Cognitive Sciences, University of Rochester, USA*
2. *Department of Psychology, University of Waterloo, Canada*
3. *Center for Visual Sciences, University of Rochester, USA*

Abstract

The ability to infer the referential intentions of speakers is a crucial part of learning a language. Previous research has uncovered various contextual and social cues that children may use to do this. Here we provide the first evidence that children also use speech disfluencies to infer speaker intention. Disfluencies (e.g. filled pauses 'uh' and 'um') occur in predictable locations, such as before infrequent or discourse-new words. We conducted an eye-tracking study to investigate whether young children can make use of this distributional information in order to predict a speaker's intended referent. Our results reveal that young children (ages 2;4 to 2;8) reliably attend to speech disfluencies early in lexical development and are able to use disfluencies in online comprehension to infer speaker intention in advance of object labeling. Our results from two groups of younger children (ages 1;8 to 2;2 and 1;4 to 1;8) suggest that this ability emerges around age 2.

Introduction

Inferring a speaker's intention is crucial to successful language learning. For example, mapping a spoken word to the appropriate object in the world requires understanding to which object the speaker intends to refer (e.g. Preissler & Carey, 2005). Though some labeling contexts are unambiguous (e.g. holding a cookie and saying '*cookie*'), most contexts involve multi-word utterances and multiple objects in the child's visual field, making the mapping problem a difficult one.

Previous work has explored various extra-linguistic cues learners can use for determining speaker intention. Social cues include joint visual attention, pointing, and eye gaze (e.g. Baldwin, 1991; Butterworth & Cochran, 1980; Southgate, van Maanen & Csibra, 2007; Yu, Ballard & Aslin, 2005). There are also contextual cues, such as object presence, object–word co-occurrence statistics (e.g. Smith & Yu, 2008), and discourse context (Frank, Goodman, Tenenbaum & Fernald, 2009). In addition to these externally available cues, young children appear to use certain heuristics that facilitate rapid lexical development. One heuristic of particular relevance is the principle of contrast (e.g. Bolinger, 1977; Clark, 1987, 1990; Markman, 1990; Markman, Wasow & Hansen, 2003). Experimental evidence suggests that young word learners make use of the fact that words tend to contrast in meaning, and thus exhibit a bias for a novel referent when encountering a novel word. Use of the principle of contrast for inferring a novel word's referent has been observed in learners as young as 15 months of age (Halberda, 2003).

Here we investigate a previously unexplored cue for inferring speaker intention: speech disfluencies. Disfluencies (e.g. filled pauses '*uh*' and '*um*') occur in highly predictable locations – for example, before unfamiliar or infrequent words, and before words that have not been previously mentioned in the discourse. Since disfluencies occur *before* an object is labeled, they could enable children to anticipate upcoming referents. Thus, speech disfluencies could enable a young word learner to narrow the pool of possible referents that she considers in a given discourse context. Anticipating the referent could also facilitate processing by enhancing the speed of spoken word recognition, and by allowing cognitive resources to be more quickly reassigned to new learning material following the label (Marchman & Fernald, 2008).

Disfluencies are a reliable property of speech between adults. Fox Tree (1995) estimated that about six disfluencies occur per 100 words, excluding pauses (which are not necessarily disfluencies). Shriberg (1996) estimated that disfluencies occur on average every seven to 15 words in conversation between adults.[1] The rate of disfluency varies as a function of several factors, including the speakers' familiarity with one another, utterance length,

[1] The rate varied depending upon which corpus (SWBD or AMEX) was used for the analysis.

Address for correspondence: Celeste Kidd, Brain & Cognitive Sciences, University of Rochester, Meliora Hall RC Box 270268, Rochester, NY 14627, USA; e-mail: ckidd@bcs.rochester.edu

and speech rate (Shriberg, 1996). Disfluencies include pauses, repeated words, lengthened syllables, abandoned phrases, inserted filler phrases, and speech errors. We focus here on the most common type of disfluency, the filled pause – '*uh*' and '*um*' in English (Shriberg, 1996). This type of disfluency is characteristic of planning problems, such as the lexical retrieval difficulties associated with producing infrequent and discourse-new words (Arnold & Tanenhaus, in press; Clark & Fox Tree, 2002). Consider the following example of a filled pause from the Sachs corpus in CHILDES (MacWhinney, 2000):

(1) CHILD: Telephone?

MOTHER: No, that wasn't the telephone, honey. That was *the, uh, timer*.

The filled pause, '*uh*', occurs before '*timer*', a word that is infrequent and previously unmentioned in the discourse. Low frequency and discourse-new lexical items like this require more processing time due to the delay involved in lexical retrieval. Disfluencies before these hard-to-retrieve words function to provide the speaker with time to retrieve the word while simultaneously signaling to the listener that the speaker is having difficulty (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997).

One important note about these types of filled-pause disfluencies concerns the determiner that precedes them. The word 'the' has two alternative pronunciations: '*thuh*' (i.e. rhymes with '*duh*') and '*thee*' (i.e. rhymes with '*bee*'). The full, unreduced form '*thee*' is far more likely to be produced in conjunction with other evidence of processing difficulties, such as before delays (unfilled pauses) and fillers (e.g. '*thee, uh*', '*thee, you know, thee*'), and during repeats ('*thee, thee* –') (Clark & Wasow, 1998; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). This cannot be said of the more common, reduced form, '*thuh*'. Although '*thuh*' is more common overall in spontaneous speech, it is far less likely to precede an intermediate suspension of speech. Fox Tree and Clark (1997) reported that in their analysis, 81% of the instances of '*thee*' were followed by a suspension of speech, compared to only 7% of a matched sample of instances of '*thuh*'. Thus, the unreduced form, '*thee*', is highly predictive of a subsequent disfluency. As a result, upon hearing an unreduced form, listeners might assume that a retrieval-induced disfluency is forthcoming.

Indeed, previous research with adults suggests that disfluencies facilitate online sentence comprehension: In a series of eye-tracking experiments, adults showed a bias to look at discourse-new or unfamiliar objects when labels were preceded by the types of filled-pause disfluency discussed above (Arnold, Tanenhaus, Altmann & Fagnano, 2004; Arnold, Fagnano & Tanenhaus, 2003; Arnold, Hudson Kam & Tanenhaus, 2007).

Here we ask whether toddlers are able to detect and use disfluencies during online spoken word recognition. In particular, we explore whether young children predict that words preceded by disfluencies will refer to unfamiliar or discourse-new referents. There are two reasons why we might expect children to use disfluencies predictively in this manner. First, there is growing evidence that determiners play an important role in children's language processing: Young children recognize words more rapidly when preceded by an appropriate and informative function word (Kedar, Casasola & Lust, 2006; Lew-Williams & Fernald, 2007; Zangl & Fernald, 2007), and can even use function words to identify the lexical category of unfamiliar labels (Bernal, Millotte & Christophe, 2007). Second, children are sensitive to the prosody of disfluent speech (Soderstrom & Morgan, 2007). Thus, we test whether children, like adults, use indicators of processing difficulties, such as lengthening of the determiner and a subsequent filled pause ('*thee uh*'), in order to anticipate likely referents for an upcoming noun. Given the demands of word learning, sensitivity to the informativeness of disfluencies could be highly advantageous for the young word learner.

Experiment 1

Methods

Participants

Sixteen parents volunteered their toddlers for the study. Parents were recruited through mailings, posters, and web ads. The toddlers ranged in age from 2;4 to 2;8 (M = 2;6), had no reported hearing deficits, and were from monolingual, English-speaking homes. Participants received either $10 or a toy as compensation.

Stimuli

The stimuli consisted of 32 non-animate objects and their labels.[2] Half of these items were familiar objects (e.g. '*ball*') whose labels were the earliest acquired words listed in the MacArthur-Bates Communicative Development Inventories (Dale & Fenson, 1996). The other half were novel objects, matched by adult judgments to the familiar items in brightness and visual complexity. A novel word was created for each novel object. The set of novel words matched the familiar words in syllable lengths, word onsets, and stress patterns (see Appendix 1 for word lists). The items were divided into 16 familiar–novel object pairs, such that each pair contained one familiar and one novel item.

Procedure

Each child was seated on a parent's lap with the child's eyes approximately 63 cm from the 17-inch LCD monitor of a Tobii 1750 eye-tracker. The auditory and visual

[2] Visual stimuli are available for review and download at http://babylab.bcs.rochester.edu/stim/disfluency/.

stimuli were presented from a host Macintosh computer using PsyScope X software (Cohen, MacWhinney, Flatt & Provost, 1993). Calibration of the eye-tracker was performed using Clearview software. The calibration involved the child fixating a shrinking dot located successively at one of five different screen locations. The parent wore headphones playing music to mask the auditory stimuli and was asked to direct their gaze downward during the experiment to prevent influencing their child's behavior. The experiment consisted of 16 trials, each featuring a unique familiar–novel object pair. Each trial was initiated only when the child attended to a small, animated attention-getter (a video of a laughing baby) presented in the center of the Tobii display.

On each trial, the objects from one of the 16 familiar–novel object pairs were presented side-by-side three times in succession. Figure 1 outlines the timecourse of each trial. The objects' locations within a given trial were fixed. During the first two presentations, the familiar object was labeled, first with the carrier phrase '*I see the X!*', then with the phrase '*Oooh! What a nice X!*' During the first two presentations, objects appeared on the screen 500 ms before the carrier phrase, and remained on the screen until 2 seconds after the onset of the familiar target word. The first two presentations were separated by a 1-second pause, during which the screen was blank. During the third presentation, children were instructed to look at either the familiar/mentioned object or the novel/unmentioned object, and the instruction was either produced fluently or contained a filled pause (i.e. '*thee uh*'). The disfluency was preceded by the full, unreduced pronunciation of the determiner, '*thee*', because this form occurs more commonly before suspensions of speech (Fox Tree & Clark, 1997) and could thus be considered most natural. Similarly, '*uh*' was chosen as the filled pause because it is more common than '*um*' in natural speech. Table 1 displays the phrase used for each of these four different trial types. Disfluencies were equally likely to precede familiar and novel targets, thus preventing children from learning any relationship between disfluencies and target familiarity over the course of the experiment. During the third presentation, objects did not appear until 2 seconds before the target object was labeled (which corresponded to the period of disfluency during disfluent trials). On this third presentation, objects remained on the screen for 3 seconds after the onset of the target label.

Critically, at the onset of the third presentation, one of the objects was both novel and previously unmentioned in the discourse. Because it was unclear whether toddlers would be sensitive to either of these factors, we jointly manipulated novelty and discourse-new status in order to maximize the chance of observing an effect (i.e. to determine whether toddlers can use disfluencies predictively in any capacity).

Window of analysis

If children do use disfluencies predictively, we would expect to see more looks to the novel/unmentioned object during the period of disfluency. In the disfluent trials, the earliest sign of the disfluency is at the determiner – '*thee*' in disfluent trials versus '*the*' in fluent speech. Thus, the determiner was chosen as the onset of the window of analysis in disfluent trials. Because our focus was on anticipatory looking to the target, the window of analysis ended at the onset of the target word, 2 seconds later. Young children require an estimated 270 ms to program and initiate saccades in response to a stimulus (Canfield, Smith, Brezsnyak & Snow, 1997). To compensate for this stimulus–response latency, we shifted the window of analysis forward by 250 ms. Using this shifted 2-second window, we compared children's anticipatory fixations across disfluent and fluent trials.

Due to the nature of disfluencies, the disfluent utterance is longer; consequently, the linguistic material in the window of analysis varied across fluent and disfluent trials (see Figure 1 and timecourse plots in Figures 2 and 3). To compensate for this difference, the command 'Look!' was repeated in all trials. Thus, in all trials, children had been instructed to look at the screen before the third presentation of the object pair and the onset of the window of analysis. The first 'Look!' instruction was successful in directing children's attention to the screen: on 88.4% of trials, children were looking at the screen immediately prior to the onset of the window of analysis, with no significant difference between fluent and disfluent trials.

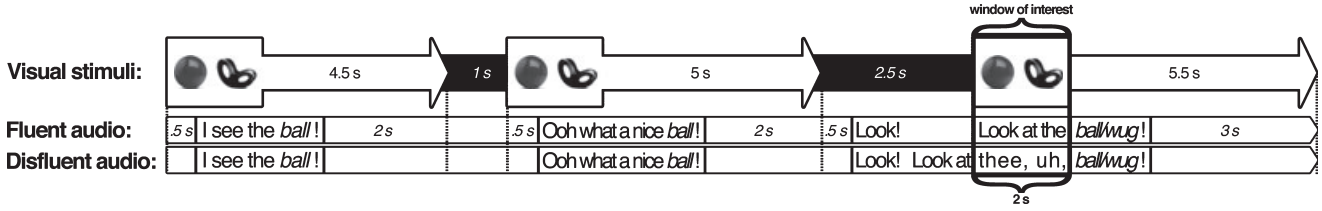

Figure 1 *Each of the 16 novel–familiar object pairs was presented three times in succession. The familiar object was always labeled during the first two presentations. On the third presentation, children were instructed to look either at the familiar or novel object with either a fluent or disfluent command. The window of analysis used was the 2 seconds before the onset of the final object label (the period of disfluency in disfluent trials).*

Table 1 *Trial type examples*

	Familiar target	Novel target
Fluent	Look! Look at the *ball*!	Look! Look at the *wug*!
Disfluent	Look! Look at thee, uh, *ball*!	Look! Look at thee, uh, *wug*!

Results

To ensure that children looked reliably at the appropriate object after it was named, we first calculated for each trial type (fluent, disfluent) the proportion of time the child looked at the target object during the 2-second period *after* the target was labeled.[3] On trials in which the target was familiar, the mean proportion of looking to the target during this window was 0.77. A Wilcoxon signed-rank test found this value to be significantly different from chance ($V = 131$, $p < .0003$), suggesting that children reliably mapped familiar words to familiar objects. During trials in which the target was novel, the mean proportion of looking to the target was 0.74, which was also significantly different from chance ($V = 132$, $p < .0002$). This result suggests that children used the principle of contrast to infer that the novel label referred to the novel object. Finally, the proportions with which children fixated the target did not differ for novel-target and familiar-target trials ($V = 78$, $p > .63$). Taken together, these results suggest that children consistently arrived at the target object, regardless of the trial type.

Next, we calculated the proportion of looks to the novel object at each time point during the critical 2-second window of analysis *before* the onset of the target word (the period of disfluency in disfluent trials). Figure 2 shows the resulting timecourse plot for trials in which the target was novel. As predicted by our hypothesis, children looked more towards the novel object during the 2 seconds before the onset of the target word when there was a disfluency present (i.e. in disfluent trials). This suggests that the disfluency served as a cue that led children to expect that the upcoming referent would be novel/unmentioned. Figure 3 shows the timecourse plot for trials in which the target was the familiar object. In these trials also, children looked more towards the novel/unmentioned object during the pre-target window of analysis in the disfluent trials than in the fluent ones. This again suggests that the disfluency prompted children to anticipate that the novel/unmentioned referent would be labeled, though in these familiar-target trials, that expectation was ultimately violated. Both Figures 2 and 3 also show that, after the onset of the target word, children correctly identified the target picture.

The timecourse plots suggest that children were sensitive to the presence of the disfluency and were biased to interpret that disfluency as signaling that the upcoming word would refer to the novel/previously unmentioned referent. To test that hypothesis, we compared looks to the novel/unmentioned object across fluent and disfluent trials in the 2-second window of analysis before the onset of the target word. During disfluent trials, children looked at the novel object for 1158 ms. During fluent trials, children looked at the novel object for 893 ms. A Wilcoxon signed-rank test found this difference to be highly significant ($V = 125$, $p < .002$). This result suggests that children are sensitive to disfluencies and use them predictively to infer that an upcoming referent is likely to be novel and/or previously unmentioned.

A possible alternative explanation for this result is that children simply paid more attention overall to the display (both objects) during disfluencies. To further examine whether disfluencies cause a selective increase in looking to the novel/unmentioned object, we compared the average proportion of total looking time to the novel object during the same temporal window of analysis. The proportion of time children looked at the novel object was 0.68 in the disfluent trials, as opposed to 0.54 in the fluent trials. A Wilcoxon signed-rank test found this difference to be significant ($V = 20$, $p < .01$). Further, the proportion of looking time to the novel object was significantly above chance in the disfluent trials ($V = 132$, $p < .0003$), whereas in the fluent trials, children's looking to the two objects did not differ significantly from chance ($V = 87$, $p = .34$). These results demonstrate that disfluencies cause a selective increase in attention to novel and/or unmentioned objects, suggesting that children use disfluencies online to create expectations about the speaker's intended referent.

Experiment 2

The results of Experiment 1 demonstrated that by 2;6, young children can use disfluencies predictively to anticipate the visual referent of a forthcoming word. In Experiment 2 we repeated the experimental procedure with two groups of younger children to explore at what age this ability emerges. We chose 16 months as a lower limit because 15 months is the youngest age at which children have been demonstrated to use the principle of contrast to map novel words to novel referents. Given the nature of our task, an ability to use the principle of

[3] Restricting the familiar labels to the earliest words children acquire necessarily resulted in labels of differing durations. The target labels varied in duration, from 640 ms to 1170 ms; however, there were no mean differences in label duration across familiar and novel labels, nor across those that followed fluent versus disfluent commands. The 2-second window for target-word recognition reported here began 750 ms after the onset of the target label for all targets, with an additional 250-millisecond shift to compensate for the stimulus–response latency required to program and execute a saccade. At this point, we presume that children have obtained enough acoustic information in order to identify the appropriate referent, even if the label was longer than 750 ms (e.g. '*banana*') and had not been fully uttered.

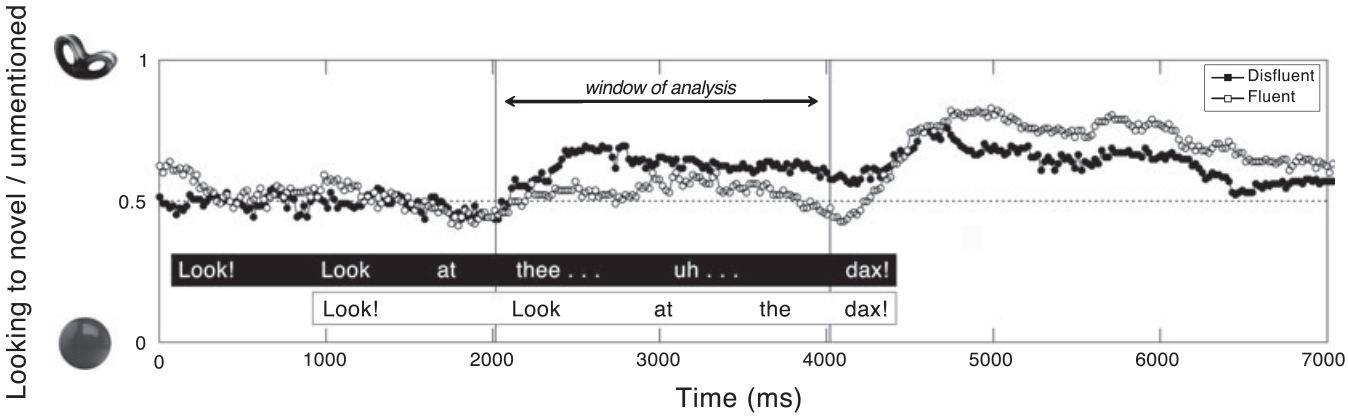

Figure 2 *Proportion of looks to the novel / unmentioned object over the course of the third picture presentation for trials with novel / unmentioned targets (shifted by 250 ms to compensate for saccade latency). During the 2-second window of analysis (the period of disfluency in disfluent trials) children's proportional looking to the novel / unmentioned object was higher overall (*p < *.007). After the target is labeled (just after 4000 ms), looks increase to the target (the novel / unmentioned object in these trials).*

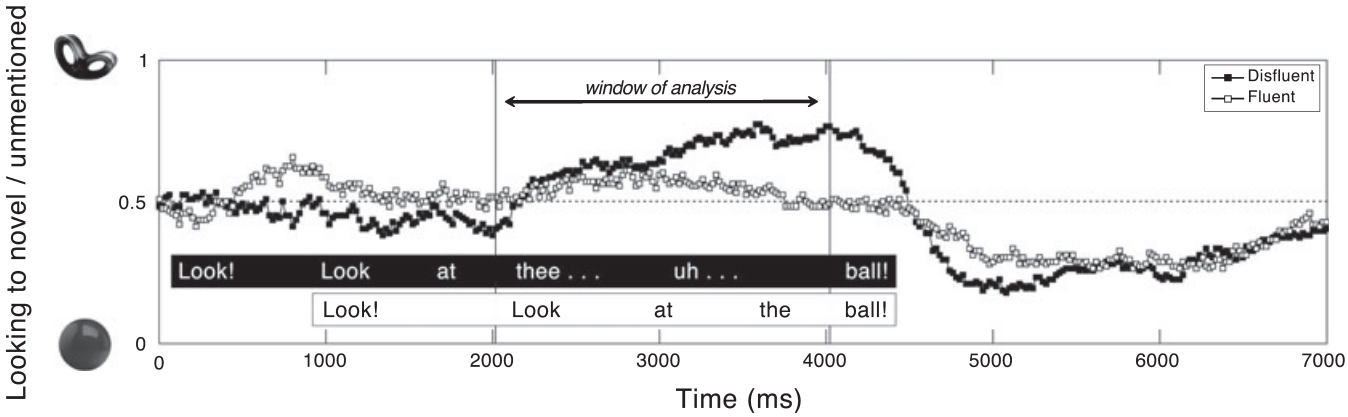

Figure 3 *Proportion of looks to the novel / unmentioned object over the course of the third picture presentation for trials with familiar / previously mentioned targets (shifted by 250 ms to compensate for saccade latency). During the 2-second window of analysis, children's proportional looking to the novel / unmentioned object was higher overall (*p < *.006). After the target is labeled, looks increase to the target (the familiar / previously mentioned object).*

contrast to map novel words may be a prerequisite for observing an effect of disfluencies. The design was identical to Experiment 1.

Methods

Participants

Thirty-two parents volunteered their toddlers. Sixteen of the toddlers ranged in age from 1;8 to 2;2 (M = 2;0), and 16 ranged from 1;4 to 1;8 (M = 1;6). None of the toddlers had known hearing deficits, and all were from monolingual, English-speaking homes. Participants received either $10 or a toy as compensation.

Stimuli and procedures

These were identical to Experiment 1.

Results

Again, to ensure that children looked reliably at the appropriate object after it was named, we calculated the proportion of time the child looked at the target item during the 2 seconds *after* the target was labeled.[4] In the 2;0-year-old group, the mean proportion of looking to the target during this window was 0.62 for familiar target labels – significantly higher than chance by a Wilcoxon signed-rank test (V = 122, p < .003). For novel labels, mean proportional looking to the target was also significantly higher than chance at 0.64 (V = 118, p < .007). These results suggest that the children in the 2;0-year-old

[4] The window for target-word recognition reported here was identical to that used for the 2;6-year-old group reported in Study 1: it began 750 ms after the onset of the target label for all targets, with an additional 250-millisecond shift to compensate for the stimulus–response latency required to program and execute a saccade.

group reliably arrived at the appropriate target after labeling, regardless of whether the target was familiar or novel. The pattern was different for the 1;6-year-old group, however. Among these younger children, the mean proportion of looking to the target was 0.55 for familiar target labels – a value *not* significantly different from chance ($V = 78$, $p = .33$). The mean proportion of looking to the target *was* significantly different for *novel* target labels, at 0.63 ($V = 116$, $p < .0004$). The latter result suggests that 1;6-year-old children used the principle of contrast to identify the novel objects given novel labels. This finding is in line with previous work that has demonstrated that children can use the principle of contrast starting at 15 months of age (Halberda, 2003). However 1;6-year-old children's apparent failure in the case of the familiar labels is somewhat surprising, since all familiar labels used in this study were selected from the earliest words that children acquire.

One possible explanation is that the 1;6-year-old group was less inclined to look at the familiar object during the third redundant labeling event due to boredom. A review of these infants' fixations from earlier in the trial revealed that the 1;6-year-old group had reliably attended to the familiar object following each of the first two labeling events (in which the familiar object was labeled). Prior to labeling, no bias existed for either object: Mean proportion of looking at the familiar object *before* labeling was 0.50 during the first presentation and 0.51 during the second, values not significantly different from chance ($V = 59$, $p = .98$ and $V = 79$, $p = .56$). Mean proportion of looking to the familiar object *after* labeling,[5] however, was 0.58 and 0.61, respectively, for the first and second presentations, both of which were significantly different from chance ($V = 103$, $p < .01$; $V = 109$, $p < .004$).

Taken together with the results of Experiment 1, these results appear to reflect a change in attentional capacities, processing speed, and accuracy in line with previous work on the developmental trajectory of spoken word recognition and attention (Columbo, 2001; Fernald, Zangl, Portillo & Marchman, 2008). Critically, these results also show that the two younger groups of children readily identified the labeled objects.

Next, we calculated the proportion of looks to the novel object at each time point during the critical 2-second window of analysis, for the fluent and disfluent trials, to examine whether disfluencies cause preferential looking to the novel object *before* the object has been labeled. For the 1;6-year-old-group, there was no difference in proportional looking for disfluent and fluent trials ($V = 62$, $p = .93$); the mean was 0.49 in both trial types. Neither the mean for disfluent nor fluent trials differed significantly from chance ($V = 49$, $p = .56$ for disfluent trials; $V = 55$, $p = .80$ for fluent ones). The difference for the 2;0-year-old group – 0.57 in disfluent trials compared to 0.49 in fluent trials – reached only marginal significance in the expected direction ($V = 33$, $p = .07$). Neither the mean for disfluent nor fluent trials differed significantly from chance ($V = 99$, $p = .12$ for disfluent; $V = 62$, $p = .78$ for fluent). These data are plotted along with those from Experiment 1 in Figure 4.

To test whether age was a significant predictor of sensitivity to disfluency, difference scores were computed for each child from all three groups of children (across Experiments 1 and 2) and entered into a regression analysis. Difference scores were computed by subtracting each child's mean proportion of looking to the novel object during the 2-second window of analysis (before the target-label onset) in fluent trials from their mean proportion in disfluent trials. Age was binned into 2-month groups. The regression analysis revealed that age was a significant predictor of children's difference scores ($\beta = 0.07$, $t = 2.62$, $p < .02$), and accounted for 13.3% of the difference-score variance ($r^2 = 0.133$). The difference scores are plotted along with the best-fitting line representing the correlation between age and difference score in Figure 5.

These results suggest that young children's ability to use disfluencies predictively emerges over the second year. Further work would be necessary to ascertain whether the ability emerges gradually in individual children, or whether the pattern observed here reflects a more discontinuous pattern (e.g. some 2; 0-year-olds

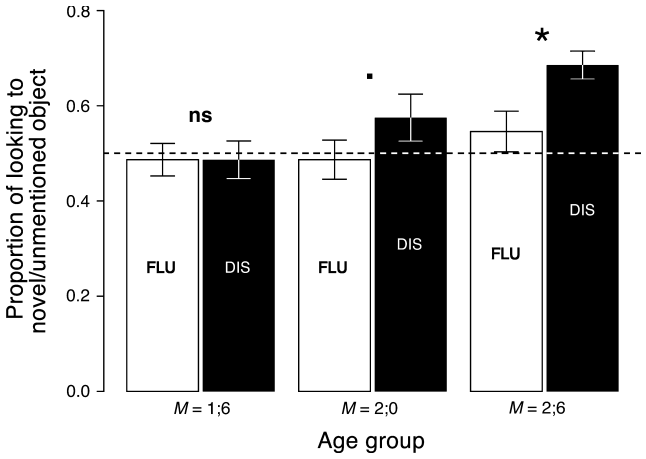

Figure 4 *Proportion of looks to the novel / unmentioned object during the disfluency for all three age groups. For the 1;6-year-old-group, there was no difference in proportion of looking for disfluent and fluent trials (V = 62, p = .93); the mean was 0.49 in both trial types. The dierence for the 2;0-year-old group – 0.57 in disfluent trials, compared to 0.49 in fluent trials – reached only marginal significance in the expected direction (V = 33, p = .07). The significant dierence for the 2;6-year-old group in Experiment 1 is also plotted for comparison – 0.68 in the disfluent trials, as opposed to 0.54 in the fluent trials (V = 20, p < .01).*

[5] The window for target-word recognition reported here was identical to that used for all other target-word recognition analyses: a 2-second period starting 750 ms after the familiar label onset, with an additional 250-millisecond shift to compensate for the stimulus–response latency required to program and execute a saccade.

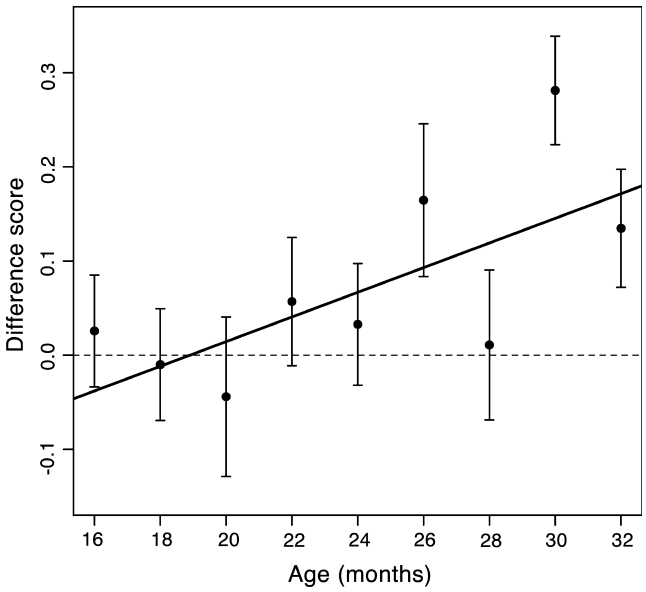

Figure 5 *Children's difference scores (proportion of looks to the novel/unmentioned object before the onset of the target word in disfluent trials minus that in fluent trials) are plotted as a function of age in 2-month-old bins for all three groups of subjects (Experiments 1 and 2). The solid line represents the correlation between children's ages and difference scores. Age accounted for 13.3% of the difference-score variance (r^2 = 0.133). A regression analysis revealed that age was a significant predictor of difference score (β = 0.07, t = 2.62, p < .02).*

reliably use disfluencies to infer reference in advance of labeling, while others do not).

Discussion

Many contemporary theories model word learning as a process of mapping the arbitrary association between sounds and meanings (Siskind, 1996; Smith, 2000; Yu & Ballard, 2007). The results of our experiment demonstrate that young children's ability to match sounds with meanings is considerably more general: they are able to match disfluencies not with a single observable referent, but rather with a broader property of communication that is not directly observed and which relates to the speaker's referential intentions. These results raise several important issues.

First, it is unclear whether novelty or discourse status is primarily responsible for these effects. Adults' interpretation of disfluencies is affected by both of these factors. In our study, the novel objects were both previously unmentioned and novel. Work in progress attempts to uncover which of these – or both – drives the effect.

One possible alternative explanation for our results might be that children interpreted some component of the disfluency as a label for a novel object. If, for example, young children interpreted the onset of the filled pause '*uh*' as the onset of a novel label, we would also expect more looks to the novel object. We believe this possibility is unlikely for a number of reasons. First, work by Soderstrom and Morgan (2007) suggests that by 22 months of age, children are sensitive to prosodic indicators of disfluency. Further, there is evidence to suggest that they would expect an article at this position in the sentence (Kedar *et al*., 2006; Lew-Williams & Fernald, 2007; Zangl & Fernald, 2007). Yet more evidence against this possibility lies in the fact that if our young subjects were initially (in the first few trials) interpreting '*thee uh*' as a label for the novel object, we would expect the effect to decrease over the experiment, as they repeatedly hear the same potential label in the presence of many different novel objects. This is not the case: the effect stays constant over the 16 trials in our data. Perhaps most convincingly, however, is the fact that the youngest group of children demonstrated that they were able to employ contrast to determine the correct referent for novel labels *after* the onset of the target words. This youngest group did not, however, look towards the novel object after hearing the disfluency. We therefore believe it to be likely that these children and the two older groups interpreted the unfilled pause as a disfluency, and not as the onset of a novel object label. Finally, in an analysis of English-language CHILDES transcripts, we found that filled-pause disfluencies were a regular feature of speech to children (see Appendix 2).

A second important issue is what children understand about disfluencies. Clark and colleagues have suggested that speech disfluencies signal to the listener that a speaker is having production difficulties (Clark & Fox Tree, 2002). Furthermore, there is evidence that adults' use of disfluencies is at least moderated by a causal understanding that the disfluencies result from the speaker's processing difficulties: Arnold *et al*. (2007) demonstrated that adult listeners do not use disfluencies predictively if they are told the speaker suffers from object agnosia, a condition characterized by difficulty labeling even familiar objects.

It is possible that children, too, engage in this type of causal reasoning. Children may be aware that disfluencies are the result of processing (specifically, lexical access) difficulties, and therefore look for a referent that is likely to have caused difficulties. While this reasoning almost certainly does not happen consciously, children may nonetheless have learned that disfluencies occur because of speaker difficulty, and that speaker difficulty often arises with novel referents. If so, we might expect children – like adults (Arnold *et al*., 2007) – to alter their interpretations when disfluencies can be attributed to an external cause.

Alternatively, children could show the patterns demonstrated here without any explicit understanding of the linguistic processing mechanisms involved on the part of the speaker. Disfluencies might simply be associatively linked through experience to novel referents. That is, toddlers could learn that when an adult exhibits a disfluency, the word that follows is likely to be unknown to

the child or the adult's overt behavior is likely to be directed to a novel object. Thus, the disfluency could be treated just like words that mean 'look at the novel referent', or, alternatively, 'look at the other thing' (i.e. the thing that was *not* just talked about). This theory does not assume intermediate stages of processing or conceptual reasoning about the internal state of the speaker: the association between a disfluency and a word-referent pair is direct and quick. Such an account would predict that children could not alter their interpretation of disfluencies based on whether they were perceived as internally or externally driven. If true, it is not clear why children younger than 2:6 would not have access to such an associative mechanism, especially because they are confronted with many novel words in the speech of their parents. Yet, as shown in Figure 4, these younger children did not respond in a manner predicted by this associative account.

Both accounts are plausible, given what is known about infants' and young children's capabilities. Infants and children are known to be capable statistical learners (e.g. Fiser & Aslin, 2002; Saffran, Aslin & Newport, 1996), which could enable them to detect correlations between disfluencies and referent novelty in the environment. Young children are also able to engage in pragmatic inference (e.g. Behne, Carpenter & Tomasello, 2005), and even very young children are able to infer the intentions and difficulties of others (Warneken & Tomasello, 2006). Therefore, it is possible that by the middle of the second year children have access to this type of reasoning during online sentence processing.

Conclusions

Together, the results of these studies indicate that young children have learned that disfluencies contain information, attend to disfluencies in speech, and can make use of the information contained in disfluencies online during comprehension in order to infer speaker intention.

Acknowledgements

The first author was supported by a National Science Foundation Graduate Research Fellowship. The research was supported by a National Institutes of Health grant (HD-37082) to R. Aslin and (DC-05071) to M. Tanenhaus. We thank Johnny Wen and Steven T. Piantadosi for their help with the analyses; the Rochester Baby Lab staff and research assistants, especially Holly Palmeri, Laura Zimmermann, Kathryn Schuler, Lindsay Woods, Rosemary Ziemnik, Suzanne Horwitz, and Alyssa Thatcher, for their help recruiting and scheduling subjects; and Michael S. DeFreitas, Susan Wagner-Cook, Mohinish Shukla, Austin Frank, Alison Austin, Patricia Reeder, Sarah Davis, Eve V. Clark, Elena Lieven, Jennifer Arnold, Michael Tanenhaus, Colin Bannard, Anne Pier Salverda, and members of the Tomasello, Fernald, Tanenhaus, and Aslin-Newport labs for their advice, comments, and suggestions. We also thank two anonymous reviewers for their helpful comments.

References

Arnold, J.E., Fagnano, M., & Tanenhaus, M.K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, **32** (1), 25–36.

Arnold, J.E., Hudson Kam, C.L., & Tanenhaus, M K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**, 914–930.

Arnold, J.E., & Tanenhaus, M.K. (in press). Disfluency effects in comprehension: how new information can become accessible. In E. Gibson & N. Pearlmutter (Eds.), *Referential processing in adults and children*. Cambridge, MA: MIT Press.

Arnold, J.E., Tanenhaus, M.K., Altmann, R.J., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science*, **9**, 578–582.

Baldwin, D.A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, **63**, 875–890.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, **8**, 492–499.

Bernal, S., Lidz, J., Millotte, S., & Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language, Learning, and Development*, **3** (4), 325–341.

Bolinger, D.L. (1977). *Meaning and form*. London: Longman.

Butterworth, G., & Cochran, E. (1980). Towards a mechanism of joint visual attention in human infancy. *International Journal of Behavioral Development*, **3** (3), 253–272.

Canfield, R.L., Smith, E.G., Brezsnyak, M.P., & Snow, K.L. (1997). Information processing through the first year of life: a longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development*, **62** (2, Serial No. 250).

Clark, E.V. (1987). The principle of contrast: a constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, E.V. (1990). On the pragmatics of contrast. *Journal of Child Language*, **17**, 417–431.

Clark, H.H., & Fox Tree, J.E. (2002). Using uh and um in spontaneous speaking. *Cognition*, **84** (1), 73–111.

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

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, & Computers*, **25** (2), 257–271.

Columbo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, **52**, 337–367.

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, **28**, 125–127.

Fernald, A., Zangl, R., Portillo, A.L., & Marchman, V.A. (2008). Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children. In I.A. Sekerina, E.M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). Amsterdam: John Benjamins Publishing.

Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA*, **99**, 15822–15826.

Fox Tree, J.J. (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.J., & Clark, H.H. (1997). Producing 'the' as 'thee' to signal problems in speaking. *Cognition*, **62** (2), 151–167.

Frank, M.C., Goodman, N.D., Tenenbaum, J.B., & Fernald, A. (2009). Continuity of discourse provides information for word learning. In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 31st Annual Cognitive Science Society* (pp. 1418–1423). Amsterdam: Cognitive Science Society.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, **87** (1), B23–B34.

Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18- and 24-month-old infants' use of function words to determine reference. *Child Development*, **77** (2), 325–338.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, **18** (3), 193–198.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd edn.). Mahwah, NJ: Lawrence Erlbaum Associates.

Marchman, V.A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, **11**, F9–F16.

Markman, E.M. (1990). Constraints childen place on word meanings. *Cognitive Science*, **14** (1), 57–77.

Markman, E.M., Wasow, J.L., & Hansen, M.B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, **47** (3), 241–275.

Preissler, M.A., & Carey, S. (2005). The role of inferences about referential intent in word learning: evidence from autism. *Cognition*, **97** (1), B13–B23.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928.

Shriberg, E.E. (1996). Disfluencies in SWITCHBOARD. *Proceedings of the International Conference on Spoken Language Processing, Addendum* (pp. 233–251). Philadelphia, PA: Institute of Electrical and Electronics Engineers.

Siskind, J.M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, **61** (1), 39–91.

Smith, L.B. (2000). Learning how to learn words: an associative crane. In R.M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L.B. Smith, A.L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). New York: Oxford University Press.

Smith, L.B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, **106**, 333–338.

Soderstrom, M., & Morgan, J.L. (2007). Twenty-two-month-olds discriminate fluent from disfluent adult-directed speech. *Developmental Science*, **10**, 641–653.

Southgate, V., van Maanen, C., & Csibra, G. (2007). Infant pointing: communication to cooperate or communication to learn? *Child Development*, **78** (3), 735–740.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, **31**, 1301–1303.

Yu, C., & Ballard, D.H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, **70**, 2149–2165.

Yu, C., Ballard, D.H., & Aslin, R.N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, **29**, 961–1005.

Zangl, R., & Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language, Learning, and Development*, **3** (3), 199–231.

Received: 17 June 2010
Accepted: 21 January 2011

Appendix 1

*Familiar-novel object pairs**

Familiar object	Novel object
Cup	Biffle
Stroller	Prag
TV	Semp
Ball	Gorp
Shoe	Mog
Cookie	Blimmick
Bathtub	Tibble
Brush	Toma
Telephone	Bep
Spoon	Perniscle
Bottle	Dax
Car	Spad
Bed	Kib
Book	Wug
Sock	Bleet

* Visual stimuli available for review and download at http://babylab.bcs.rochester.edu/stim/disfluency/.

Appendix 2

Analysis of disfluencies in speech to children in CHILDES

Figure A1 *Log probability of a filled pause in child-directed speech in CHILDES with standard errors.*

An analysis of English-language CHILDES transcripts involving only two participants – the target child and an adult caretaker – revealed that filled pause disfluencies are a regular feature of speech to children. (Other transcripts were excluded to ensure that disfluencies extracted from the adults were from utterances directed to the target child, and not to an older sibling or another adult.) In our analysis, we computed the log mean probability of a filled pause disfluency spoken to children at each age. The resulting plot suggests that at age 2, filled pauses occurred at an approximate rate of 1 every 1000 words. Further, the plot demonstrates that children heard filled pauses more frequently as they got older (Spearman's rank correlation, $rho = 0.85$, $p < .002$), in accord with the fact that caretakers tend to use longer, more complicated utterances with older children.

Though this rate is much lower than that estimated by Shriberg (1996) for speech between adults (1 filled pause every 50 words), our analysis likely grossly underestimates the number of disfluencies in speech to children. Since the parent–child interactions were not transcribed specifically for analyzing speech disfluencies (with the exception of the Soderstrom corpus), the transcribed disfluencies represent only a subset of those that occurred. Regardless, our analysis suggests that disfluencies are a reliable feature of speech to children and become increasingly frequent with age.