Learning the Meaning of "*Um*": Toddlers' developing use of speech disfluencies as cues to speakers' referential intentions

Celeste Kidd (*ckidd@bcs.rochester.edu*)

University of Rochester

Katherine S. White (*white@uwaterloo.ca*)

University of Waterloo

&

Richard N. Aslin (*aslin@cvs.rochester.edu*)

University of Rochester

Kidd, C., White, K.S., & Aslin, R.N. 2011. Learning the Meaning of "Um": Toddlers' developing use of speech disfluencies as cues to speakers' referential intentions. In *Experience, Variation, and Generalization: Learning a First Language (Trends in Language Acquisition Research)*, I. Arnon & E.V. Clark (eds.), 91-106. Amsterdam: John Benjamins.

Abstract

Previous research has uncovered various contextual and social cues that children may use to infer speakers' communicative intentions (e.g., joint visual attention, pointing). We review evidence from eye-tracking studies that suggests that by 2;6 years of age, children use another previously unexplored cue to infer speakers' communicative intentions: speech disfluencies. Disfluencies (e.g., "uh" and "um") often occur before unfamiliar, infrequent, and discourse-new words. Thus, disfluencies provide information about a speaker's intended referent. Further, children use the presence of a disfluency before an object label to anticipate a novel, discourse-new referent. These results demonstrate that children go beyond their input, acquiring the generalization that disfluencies precede not just specific words, but rather categories of words that are difficult to produce.

Keywords: language acquisition; speech disfluencies; lexical development; eye-tracking; attention.

Learning and cue use

There are countless ways that human infants might discover facts about the world, but statistical learning is certainly one of the most powerful tools they have. Previous empirical work has demonstrated that infants track and use the statistical properties of their environments in a diverse array of learning tasks pertaining to sounds, words, people, shapes, and objects (Fiser & Aslin, 2002; Maye, Weiss, & Aslin, 2008; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Yu & Ballard, 2007). For example, segmenting words from a continuous speech stream involves tracking the transitional probabilities between words (Saffran, Aslin, & Newport, 1996), and learning word meanings requires tracking cross-situational co-occurrence statistics of objects and spoken words (Smith & Yu, 2008).

Statistical learning begins even before birth, as infants track the prosodic and metrical properties of their auditory environments from inside the womb (e.g., DeCasper & Fifer, 1980). Shortly after birth, newborns can recognize their native language, mother's voice, and even specific stories and songs (e.g., Byers-Heinlein, Burns, & Werker, 2010; DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Moon, Cooper, & Fifer, 1993). By six months of age, infants can track the distribution of acoustic properties of speech sounds, suggesting that this statistical learning mechanism plays an important role in the formation of phonological categories (Maye, Werker, & Gerken, 2002; White, Peperkamp, Kirk, & Morgan, 2008).

Although tracking raw low-level statistics may be a powerful force for solving many learning problems, it is likely insufficient for some. For example, though early word learning could theoretically be done by tracking the co-occurrence statistics of physically present objects and spoken words, an infant's limited attention and memory make this solution intractable. Given infants' cognitive limitations, other cues or biases are needed to limit the pool of possible hypotheses.

For language learning, one way in which infants limit the number of raw statistics they track is by considering a speaker's communicative intentions. In word learning, for example, understanding what a speaker intends to refer to greatly simplifies the otherwise computationally difficult process of mapping spoken words to physical referents (Baldwin, 1991; Bloom, 2002; Frank, Goodman, & Tenenbaum, 2009; Tomasello, 2003). The abilities necessary for sensitivity to speaker intentions first emerge between 9 and 12 months of age (Tomasello, 1995). These abilities include holding joint attention (Bakeman & Adamson, 1984), following gestures (Corkum & Moore, 1995), understanding intentionality (Tomasello, 1998; Tomasello, Kruger, & Ratner, 1993), and understanding goals (Kuhlmeier, Wynn, & Bloom, 2003). Therefore, it is possible that very young children, still in the early stages of lexical development, have access to this type of implicit reasoning during online comprehension.

In fact, understanding communicative intentions underlies learners' use of many extra-linguistic cues for determining the referents of spoken words. These include social cues, such as joint visual attention, pointing, and eye gaze (e.g., Baldwin, 1991, Butterworth & Cochran, 1980; Southgate, van Maanen, & Csibra, 2007; Yu, Ballard & Aslin, 2005), and discourse context (Frank, Goodman, Tenenbaum, & Fernald, 2009). Understanding communicative intentions is also essential to more sophisticated types of pragmatic inference (Behne, Carpenter, & Tomasello, 2005). One example is the *principle of contrast*, by which word learners presume that a speaker intends to refer to a novel object if they use a novel word form. Thus, young word learners will exhibit a preference for associating novel words with novel objects over objects with known names (e.g., Bolinger, 1977; Clark, 1987; Clark, 1990; Markman, 1990; Markman, Wasow, & Hansen, 2003). Use of this bias has been observed in learners as young as 15 months

(Halberda, 2003).

Here, we explore one additional cue for inferring speaker intentions: speech disfluencies. Speech disfluencies (e.g., "*uh*" and "*um*") provide an especially interesting cue because their occurrence is not a particularly salient feature of a child's word-learning experience—unlike such cues as pointing or object presence. One simple null hypothesis is that children, with their limited computational abilities, might only attend to the most salient or informative cues. Indeed, many computational systems simplify learning and inference tasks by only focusing on highly informative cues, or by modeling data in artificially low dimensionality spaces. However, it is possible that the range of data that enters into children's learning computations is considerably more broad, and not limited to the most salient cues. We provide evidence that children are capable of using disfluencies predictively in online comprehension, suggesting that children's abilities to recognize and integrate cues for language learning and processing are extraordinarily sophisticated, even from a young age. Our results show that children can use disfluencies online as a way to thin the pool of potential referents they consider during comprehension.

Disfluencies as a cue to speaker intention

Speech disfluencies include pauses, repeated words, lengthened syllables, abandoned phrases, inserted filler phrases, and speech errors. Disfluencies are numerous in speech between adults. Fox Tree (1995) estimated that about 6 disfluencies occur per 100 words, excluding pauses (which are not necessarily disfluencies). Shriberg (1996) estimated that disfluencies occur on average every 7 to 15 words in conversations between adults. Most disfluencies appear to result from processing difficulties relating to speech production, such as difficulty retrieving a word or planning an utterance. Notably, disfluencies occur in highly predictable locations—before difficult linguistic material.

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 particularly common before infrequent and discourse-new words (Arnold & Tanenhaus, 2011). Consider the following example of a filled pause from the Brent corpus in CHILDES (MacWhinney, 2000):

(1) MOTHER: *Lemme see about these baked potatoes. I mean these <u>uh... yams</u>.*

In Example (1), the speaker erroneously referred to *baked potatoes*, a higher frequency vegetable than *yams*. The speaker then began to correct herself and produced a filled pause while she searched her mental lexicon for the intended word form, *yams*. Lower frequency lexemes like *yams*—as well as lexemes that are new to the discourse—require more processing time due to the delay involved in lexical retrieval. Disfluencies before these hard-to-retrieve words provide speakers with time necessary to locate the correct word forms (Clark & Fox Tree, 2002). Speakers may produce some types of disfluencies for the purpose of announcing their processing difficulties to listeners (Fox Tree & Clark, 1997). Thus, filled pauses may be intentional communicative acts that convey meaning to listeners (e.g., "*Hold on… I'm thinking*" or "*I'm having trouble accessing a difficult word*"). Disfluencies could also simply be artifacts of the processing system. That is, it could be easier for speakers to keep their vocal folds vibrating and tongues held at-the-ready during delays in speech production. Regardless of whether or not speakers explicitly *intend* to communicate their processing difficulties, speech disfluencies provide the listener with a reliable indicator that the speaker is trying to describe something that is difficult (Arnold, Hudson Kam, & Tanenhaus, 2007).

Previous research suggests that adult listeners can use disfluency cues in online comprehension: adults show a bias to look at discourse-new or uncommon referents when labels are preceded by a disfluency (Arnold et al., 2004; Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Hudson Kam, & Tanenhaus, 2007). That is, adults appear to have extracted the generalization that disfluencies are reliable predictors of difficult-to-process material and are able to apply this knowledge in novel contexts. Interestingly, adults show no such bias when provided with an alternative explanation for why the speaker is disfluent. 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 in labeling even familiar objects. Taken together, these results show that adults engage the following causal inference: novel and discourse-new referents cause increased processing difficulty in normal speakers, but not in patients with object agnosia (who have general difficulties with lexical retrieval). Thus, even if the bias that adults exhibit during comprehension is driven by a low-level association between disfluencies and novel/unmentioned referents, it is at least moderated by high-level explicit inference about the cause of the speakers' difficulty.

If children have access to either the low-level association or high-level understanding about what causes speakers difficulty, they could use disfluencies to anticipate upcoming referents. This could be useful for a number of different reasons. First, it might enable young word learners to narrow the pool of possible referents that they consider given the discourse context. Second, it could enhance learning by informing toddlers that the subsequent linguistic material is likely to provide new information. Third, it could make spoken word recognition more rapid by enabling toddlers to anticipate an upcoming referent. Finally, anticipating the referent could enhance toddlers' processing speed, allowing cognitive resources to be more quickly reassigned to learning new material that follows the label (Marchman & Fernald, 2008). Given the demands of word learning, sensitivity to the information contained in disfluencies could be even more advantageous for the young word learner than for the adult.

Even if children are sensitive to associations between particular words and disfluencies in their input, they must go one step further in order to use disfluencies effectively in comprehension: they must generalize beyond these raw lexical associations present in their input. For instance, noticing that "*yams*" is preceded by a disfluency (as in Example 1) is of little help to a young word learner: disfluencies precede many words other than "*yams*", and thus the raw lexical association offers little predictive power by itself. The predictive power afforded by the presence of a disfluency relies on children making the generalization that disfluencies precede not just specific words, but rather categories of words that are difficult to produce (those which are unfamiliar and not previously mentioned).

In the remainder of this chapter, we will first examine children's linguistic input and provide evidence that speech disfluencies are a reliable feature of speech to children. Next, we will investigate whether young children have been able to generalize from lexically specific instances of disfluency in their input to learn that disfluencies precede referents that are difficult to name. We ask whether toddlers are able to detect and use disfluencies during online comprehension to infer a speaker's intended referent when that referent is entirely novel (and thus could not have been predicted with a lexically specific disfluency-label association). We provide evidence from a series of eye-tracking studies that by around 2 years of age, young children have learned to interpret disfluencies as indicating that the speaker is trying to refer to difficult referents (objects that are unfamiliar and discourse-new).

How fluent is speech to young children?

Since fluency is widely regarded as a hallmark of child-directed speech, it may initially seem odd to consider the possibility that young children could learn that disfluencies signal a difficult-to-label referent. The features of child-directed speech, such as slow speed and short utterances, are in fact correlates of fluency; on this view, it would seem that children would not be able to learn about disfluencies until later in development.

However, children may still be able to learn from the abundance of disfluencies present in speech between adults. Indeed, there is some experimental evidence that demonstrates that young children (1;10) attend to disfluencies in speech between adults (Soderstrom & Morgan, 2007). Moreover, as the frequency with which disfluencies occur in child-directed speech is not known, it is important to consider the possibility that child-directed speech may itself provide enough examples for learning. We explored this possibility by analyzing the rate of child-directed filled-pause disfluencies in CHILDES (MacWhinney, 2000).

For this analysis, we used only English-language CHILDES transcripts involving two participants—the target child and an adult caretaker. This restriction ensured that the adult-produced disfluencies extracted from the transcripts were directed to the target child, and not to an older sibling or another adult. After extracting filled-pause disfluencies from the remaining transcripts, we computed the log mean probability of a filled-pause disfluency spoken to children at each age.

Evidence from CHILDES of disfluencies in child-directed speech

Figure 1 displays the regularity with which filled pauses occurred in speech to young children. At age 2, filled pauses occurred at an approximate rate of 1 every 1,000 words. Further, the plot demonstrates that children heard filled pauses more frequently as they got older (Spearman's rank correlation, *rho* = 0.85, $p < 0.002$), in accord with the fact that caretakers tend to use longer, more complicated utterances with older children.

Insert Figure 1 about here

Although this rate is much lower than that estimated by Shriberg (1996) for speech between adults (1 filled pause every 50 words), comparing these analyses in terms of absolute rate is problematic. In our CHILDES analysis, only *transcribed* filled pauses could be detected[1]. Since these 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. Our analysis therefore yields an underestimate of the true disfluency rate. Disfluencies, then, are a reliable feature of speech to children and they become increasingly frequent with age. Thus, children could learn not only from the disfluencies in speech between adults, but also potentially from disfluencies in child-directed speech.

Our analysis, in conjunction with work by Soderstom and others on disfluencies in speech between adults, suggests that young children are exposed to disfluencies in their input. Next, we examine whether young children have used these specific instances in their input to form more generalized knowledge about speech disfluencies.

General knowledge of speaker difficulty from lexically specific examples in the input

If children have learned that disfluencies indicate speaker difficulty, they could potentially use them predictively during comprehension to identify a difficult-to-label referent, even when that referent is novel. In the case of difficult-to-label referents with known labels—for example, "*alligator*", which though infrequent is known by most 2-year-olds—children could have learned a lexically specific rule indicating that "*alligator*" is likely to be preceded by a

[1] A review of the audio corresponding to a sample of our data—100 filled pauses with recordings available in CHILDES—confirmed that 98 of the 100 transcribed filled pauses were true disfluencies. The remaining two comprised a partially transcribed "*uh huh*", and a verbally produced sound effect intended to simulate a thud.

disfluency. With novel referents (e.g., "*wug*"), however, children cannot rely on a specific lexical association; rather, children must have acquired general knowledge that disfluencies precede difficult-to-produce labels. We provide empirical evidence from a series of eye-tracking studies that suggests that children make this generalization at around two years of age. Further information on the experimental work described below is available in Kidd, White, & Aslin (2011).

Experimental evidence of young children's use of disfluencies in comprehension

To test whether children, like adults, are biased to look towards discourse-new or novel referents when they hear disfluencies, we used a simplified version of the Visual World paradigm employed by Arnold and colleagues (Arnold et al., 2004; Arnold et al., 2003; Arnold, Hudson Kam, & Tanenhaus, 2007). Unlike the four-object displays used in Arnold's adult eye-tracking studies, ours featured only two objects at a time: one highly familiar object (e.g., "*ball*"), and one novel object (e.g. *"wug*"). The experiment consisted of 16 trials, each featuring a unique familiar-novel object pair.

We tested 16 toddlers from monolingual, English-speaking homes in each of three age groups: ages 1;4 to 1;8 (*M* = 1;6), 1;8 – 2;2 (*M* = 2;0), and 2;4 – 2;8 (*M = 2;6*). Each child was seated on a parent's lap in front of a 17-inch LCD monitor of a Tobii 1750 eye-tracker. Parents wore headphones playing masking music to prevent parental influence on the children's behavior throughout the experiment.

On each trial, the objects from one of the 16 familiar-novel object pairs were presented side-by-side three times in succession. The objects' locations within a trial were fixed. Figure 2 outlines the timecourse of each trial. 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 third presentation, children were instructed to look at either the familiar/mentioned object or the novel/unmentioned object. The instruction was either produced fluently or contained a filled pause (i.e., "*theeee uhhh*"). Disfluencies were equally likely to precede familiar and novel targets, preventing children from learning any relationship between disfluencies and target familiarity during the experiment.

Insert Figure 2 about here

Critically, at the onset of the third presentation, one of the objects was both novel and previously unmentioned by the speaker. Because it was unclear whether toddlers would be sensitive to either of these factors, we jointly manipulated novelty and discourse status in order to determine whether toddlers can use disfluencies predictively in any capacity.

If children 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 et al., 1997). We shifted the window of analysis forward by 250 ms to compensate for this stimulus-response latency.

We calculated the proportion of looks to the novel object at each time point during the critical 2-second window of analysis to examine whether disfluencies caused preferential looking to the novel/unmentioned object. The results for all three age groups are plotted in Figure 3. We found more looks to the novel/unmentioned object during the period of disfluency only in the oldest age group we tested (2;4 – 2;8, *M* = 2;6), though we observed a marginal difference in looking behavior in the expected direction for the next oldest age (1;8 – 2;2, M = 2;0).

Insert Figure 3 about here

The proportion of time children looked at the novel object in the 2;6-year-old group was 0.68 in disfluent trials, as opposed to 0.54 in fluent trials. This difference was significant ($V = 20$, $p < 0.01$). Further, the proportion of looking time to the novel/unmentioned object was significantly above chance in the disfluent trials ($V = 132$, $p < 0.0003$), but not in the fluent ones ($V = 87$, $p = 0.34$). These results demonstrate that disfluencies cause a selective increase in attention to 'difficult' referents, suggesting that 2;6-year-olds use disfluencies online to compute expectations about the speaker's referential intentions.

The difference for the 2;0-year-old group—0.57 in disfluent trials, compared to 0.49 in fluent trials—reached marginal significance in the expected direction ($V = 33$, $p = 0.07$). Neither the mean for disfluent nor fluent trials differed significantly from chance ($V = 99$, $p = 0.12$ for disfluent; $V = 62$, $p = 0.78$ for fluent).

Finally, for the 1;6-year-old-group, there was no difference in proportional looking for disfluent and fluent trials: the mean was 0.49 for both trial types ($V = 62$, $p = 0.93$). Neither mean differed significantly from chance ($V = 49$, $p = 0.56$ for disfluent trials; $V = 55$, $p = 0.80$ for fluent ones).

Insert Figure 4 about here

A regression analysis of individual children's difference scores in all three age groups indicated that age was a significant predictor of sensitivity to disfluency ($\beta = 0.07$, $t = 2.62$, $p < 0.02$). Age accounted for 13.3% of the variance in differences scores ($r^2 = 0.133$). Difference scores were computed by subtracting a child's proportion of looking to the novel/unmentioned object in the window of analysis during *fluent* trials from that in their *disfluent* trials. The difference scores are plotted (with age binned into 2-month groups) along with the best-fitting line representing the correlation between age and difference score in Figure 4. This pattern suggests that young children's ability to use disfluencies predictively emerges around 2 years of age.

Discussion

Contemporary theories typically 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 research 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. This highlights young learners' ability to glean general knowledge from lexically specific associations. Moreover, disfluencies are not an especially salient or informative cue for speaker intention, which implies that young learners track many environmental statistics, and not just the most informative ones.

An alternative possibility: Learning by doing?

We have not yet addressed the possibility that children could learn to interpret the disfluencies of others though their own disfluent productions, but this prospect is certainly worth mention. Consider the following child-produced disfluencies from CHILDES:

(2) CHILD (2;5): *This, uh, this…*
MOTHER: *What? Show me.*
CHILD: *This, uh, this—uh, this.*
MOTHER: *Show me.*
CHILD: *This. Uh, this! This! This!*

MOTHER: *What is it? What is he looking at ?*

(3) CHILD (3:2): *I wanna get some candy from— from Scotty for Valentine.*

In Examples (2) and (3), children produced disfluencies in association with presumed lexical retrieval difficulties. These and other examples of child-produced disfluencies suggest that at least some children produce filled-pause disfluencies before difficult words and linguistic material. Thus, children may also be able to learn about the classes of referents that trigger processing difficulties through their own production experiences.

Insert Figure 5 about here

An analysis akin to our earlier one of child-directed speech reveals that the probability of a filled pause in child-*produced* speech similarly increased with children's age (*rho* = 0.94, $p < 0.001$) (Figure 5). At age 2, the rate was approximately 1 every 230 words (though, again, this is most likely an underestimate). Further work is needed to ascertain whether these production experiences influence children's ability to make causal inferences about other speakers' disfluencies.

The nature of the knowledge

Precisely what children understand about disfluencies is an open question, though the results of our eye-tracking studies suggest that the knowledge goes beyond mere lexical associations. Clark and colleagues have suggested that speech disfluencies signal to the listener that a speaker is having difficulties producing speech (Clark & Fox Tree, 2002). Work with adults shows that high-level reasoning about the cause of speakers' processing difficulties may drive the interpretation of disfluencies or, at the very least, modulate the use of generalized low-level statistical associations between disfluencies and certain types of referents (Arnold et al., 2007).

It is possible that children, too, engage in this type of causal reasoning. Children may reason that disfluencies are the result of processing difficulties, and subsequently seek out the referent that is likely to have caused those 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/unmentioned referents. Thus, children could gain the low-level association by way of a higher-level inference about the speakers' communicative goals and troubles. 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 understanding of the linguistic processing mechanisms involved on the part of the speaker. Disfluencies might simply be associatively linked through experience to novel/unmentioned referents. (Note that our study leaves unresolved which of these factors–novelty or discourse status or both–is relevant.) That is, toddlers could observe that when an adult produces a disfluency, the word that follows is likely to be unknown to the child or previously unmentioned, or the adult's overt behavior is likely to be directed to a novel or unmentioned object. They could then generalize the association to the class of novel/unmentioned referents. Thus, the disfluency could be interpreted to mean "*look at the less familiar or unmentioned referent*". 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 class of referents is quick and direct. Such an account would predict that children could not alter their interpretation of disfluencies on the basis of high-level reasoning about the cause of a speaker's disfluency.

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 et al., 2005), and even very young children are able to infer the intentions and difficulties of others (Warneken & Tomasello, 2006). It is possible that by 2;6 years, children have access to this type of reasoning during online sentence processing. Future research is needed to determine which of these two accounts best describes toddlers' understanding of disfluencies.

Acknowledgments

The first author was supported by a National Science Foundation Graduate Research Fellowship. The research was supported by National Institutes of Health grants to R. Aslin (HD-37082) and to M. Tanenhaus (DC-05071).

We thank Johnny Wen and Steven T. Piantadosi for assistance with Matlab and R; Rochester Baby Lab staff for help running subjects, especially Holly Palmeri, Laura Zimmermann, Alyssa Thatcher, and Suzanne Horwitz; and the following people for helpful conversations and advice: Jennifer Arnold, Inbal Arnon, Alison Austin, Colin Bannard, Eve V. Clark, Herbert H. Clark, Michael S. DeFreitas, Anne Fernald and lab members, Austin Frank, Casey Lew-Williams, Elena Lieven, Elissa L. Newport and lab members, Michael K.Tanenhaus and lab members, and Michael Tomasello and lab members.

References

Arnold, J. E & Tanenhaus, M. K. 2011. Disfluency effects in comprehension: how new information can become accessible. In *Referential Processing in Adults and Children*, E. Gibson & N. Pearlmutter (eds.), 197-217. Cambridge, MA: MIT.

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

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. 2003. Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research* 32, 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, & Cognition* 33, 914-930.

Bakeman, R. & Adamson, L.B. 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development* 55(4): 1278-1289.

Baldwin, D. A. 1991. Infants' contribution to the achievement of joint reference. *Child Development* 63, 975-890.

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.

Bloom, P. 2002. *How Children Learn the Meanings of Words (Learning, Development, and Conceptual Change)*. Cambridge, MA: MIT Press.

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, 253-272.

Byers-Heinlein, K., Burns T. C., & Werker, J. F. 2010. The roots of bilingualism in newborns. *Psychological Science* 3, 343-348.

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 *Mechanisms of language acquisition*, ed. B. MacWhinney (1-33), Hillsdale, NJ: Lawrence Erlbaum.

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, 73-111.

Corkum, V. & Moore, C. 1955. Development of joint visual attention in infants. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origin and role in development*, Hillsdale, NJ: Erlbaum.

DeCasper, A.J. & Fifer, W.P. 1980. Of human bonding: Newborns prefer their mothers' voices. *Science* 208(4448): 1174-1176.

DeCasper, A.J. & Spence, M.J. 1986. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development* 9(2): 133-150.

Fiser, J. & Aslin, R. N. 2002. Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences* 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 & Language* 34, 709-738.

Fox Tree, J. J., & Clark, H. H. 1997. Producing “the” as “thee” to signal problems in speaking. *Cognition* 62, 151-167.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. 2009. Using speakers’ referential intentions to model early cross-situational word learning. *Psychological Science* 20, 579-585.

Frank, M. C., Goodman, N. D., Tenenbaum, J. B., & Fernald, A. 2009. Continuity of discourse provides information for word learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (1418-1423). Austin, TX: Cognitive Science Society.

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

Kidd, C., White, K. S., & Aslin, R. N. 2011. Toddlers use speech disfluencies to predict speakers' referential intentions. Developmental Science 14(4): 925–934.

Kuhlmeier, V., Wynn, K. & Bloom, P. 2003. Attribution of dispositional states by 12-month-olds. *Psychological Science* 14(5): 402-408.

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

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 children place on word meanings. *Cognitive Science* 14, 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, 241-275.

Maye, J., Weiss, D. J., & Aslin, R. N. 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science* 11, 122-134.

Maye, J., Werker, J. F., Gerken, L. A. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.

Moon, C., Cooper, R. P, & Fifer, W. P. 1993. Two-day-olds prefer their native language. *Infant behavior & Development* 16(4), 495-500.

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

Saffran, J.R., Johnson, E.K., Aslin, R.N. & Newport, E.L. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1): 27-52.

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

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

Smith, L. B. 2000. Learning how to learn words: An associative crane. In *Becoming a Word Learner: A debate on lexical acquisition,* ed. R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. J. Hollich (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. and 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, 735-740.

Tomasello, M. 1995. Joint attention as social cognition. In *Joint Attention: Its origins and role in development,* ed. C. Moore and P. J. Dunham (103-130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M. 1998. Reference: Intending that others jointly attend. *Pragmatics & Cognition* 6, 229-244.

Tomasello, M. 2003. *Constructing a Language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M., Kruger, A., & Ratner, H. 1993. Cultural learning. *Behavioral & Brain Sciences*, 16, 495-552.

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

White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. 2008. Rapid acquisition of phonological alternations by infants. *Cognition* 107, 238-265.

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.

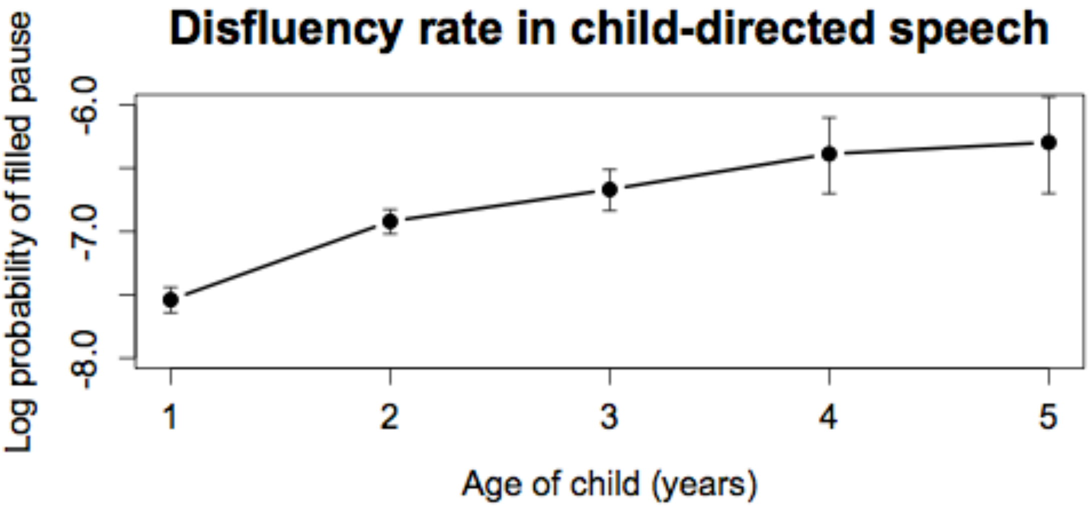

Figure 1: Log probability of a filled pause in child-directed speech in CHILDES, plotted by age with standard errors.

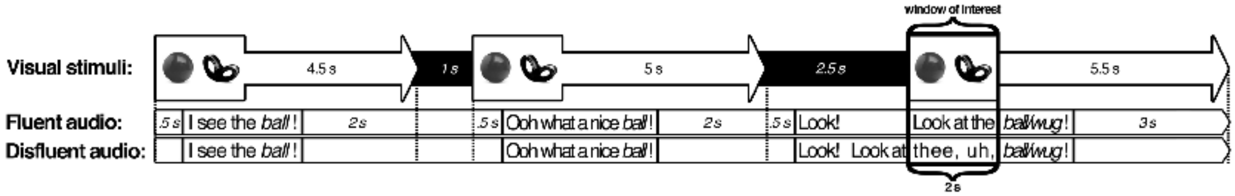

Figure 2: Each of the 16 novel-familiar object pairs was presented three times in succession. The familiar object was always labeled during the first two presentations. On the third presentation, children were instructed to look either at the familiar or novel object with either a fluent or disfluent 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).

[SET THIS SIDEWAYS ON PAGE (LANDSCAPE VIEW & ENLARGE FIGURE]

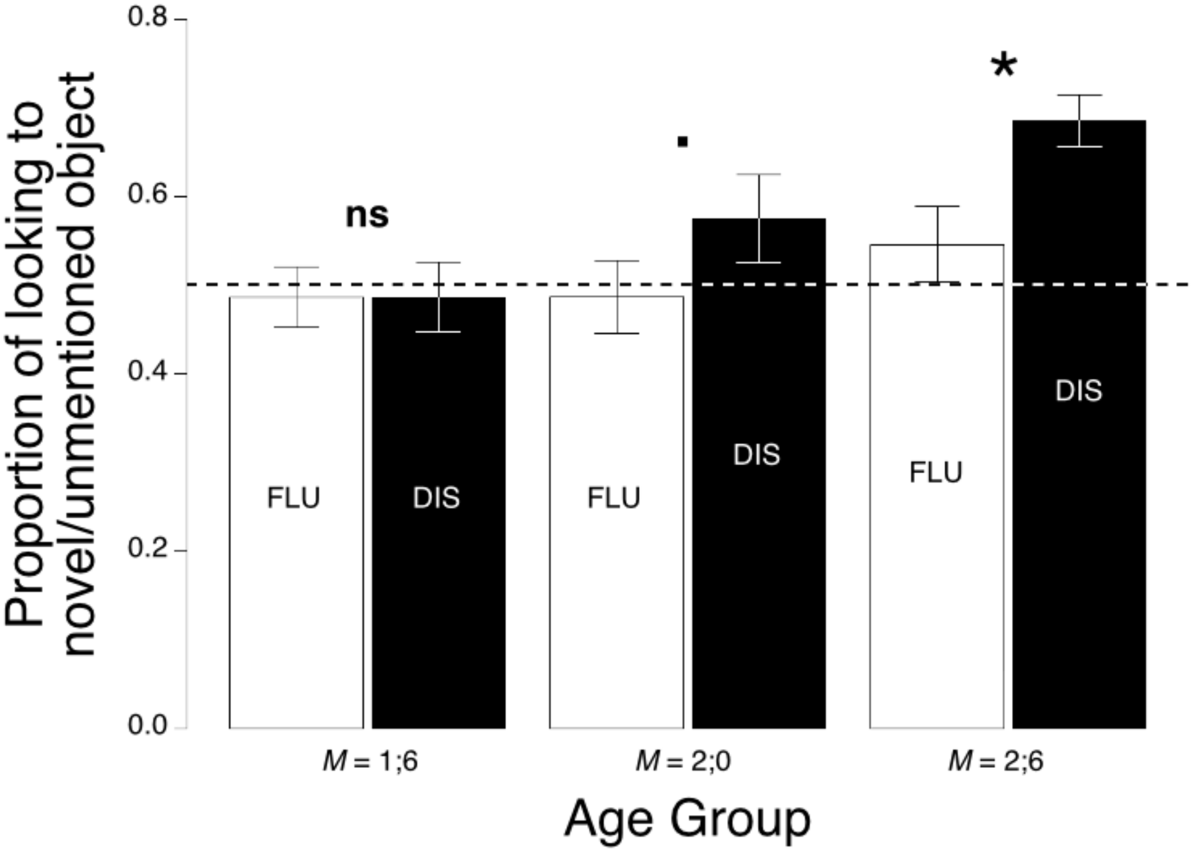

Figure 3: The proportion of looking to the novel object in the two seconds before the onset of the target label (the period of disfluency in disfluent trials) for the three age groups. The error bars denote standard errors.

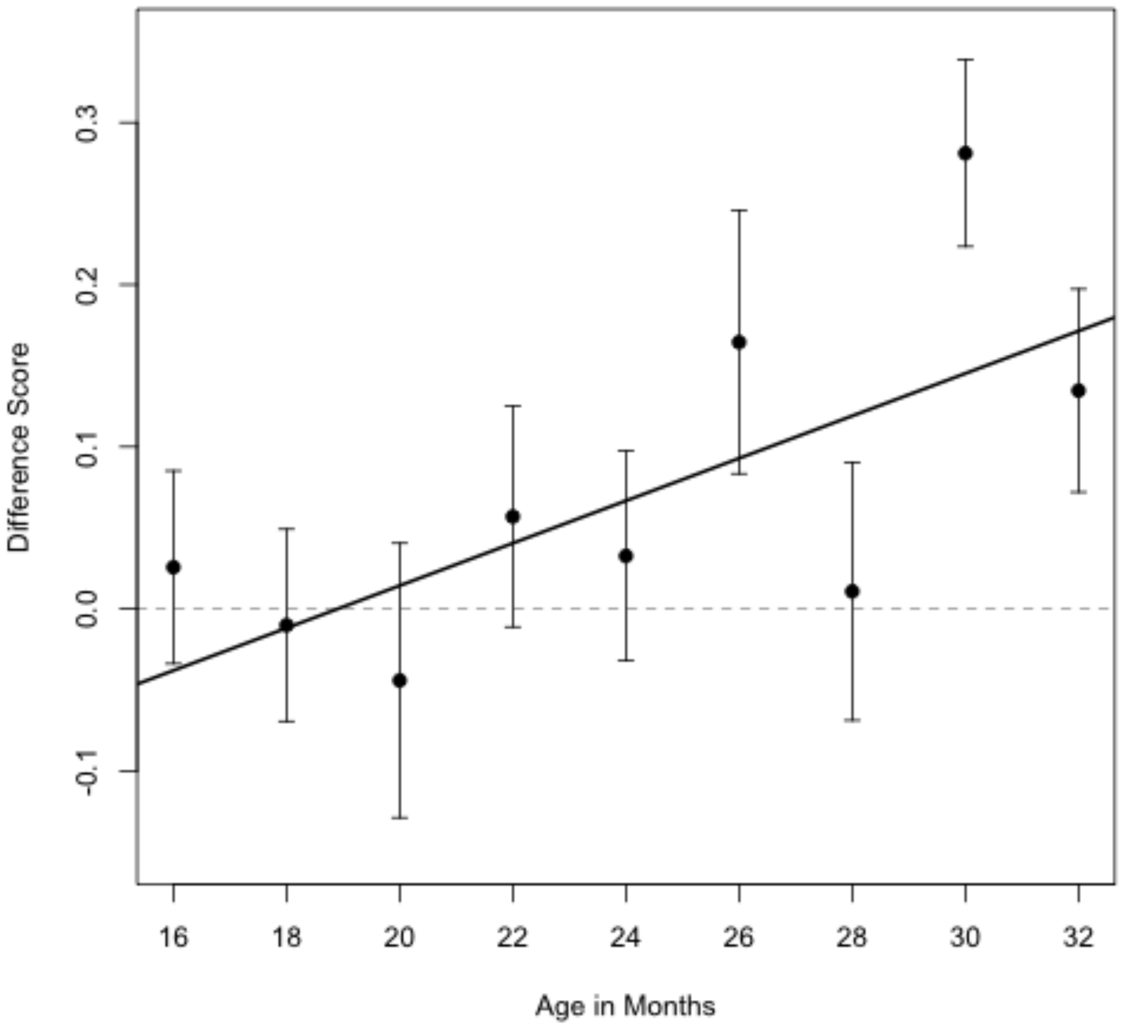

Figure 4: Children's difference scores (proportion novel/unmentioned minus proportion familiar/mentioned in the 2-second window before the onset of the target word) are plotted as a function of age in 2 month bins for all three age groups (Studies 1 and 2). The solid line represents the correlation between 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 ($\beta = 0.07$, $t = 2.62$, $p < 0.02$).

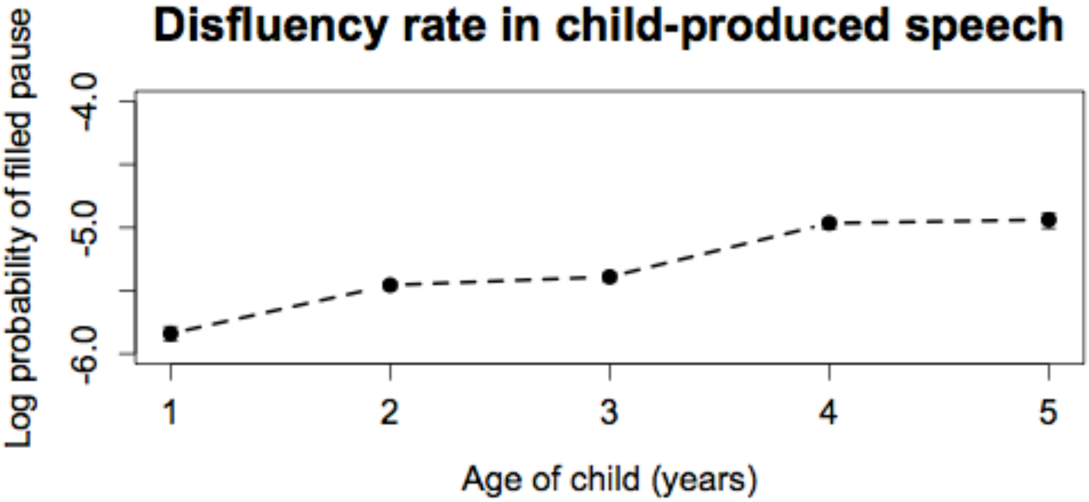

Figure 5: Log probability of a filled pause in child-produced speech in CHILDES, plotted by age with standard errors.