Brief article

Two-year-olds can begin to acquire verb meanings in socially impoverished contexts

Sudha Arunachalam *

Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215, United States

ABSTRACT

By two years of age, toddlers are adept at recruiting social, observational, and linguistic cues to discover the meanings of words. Here, we ask how they fare in impoverished contexts in which linguistic cues are provided, but no social or visual information is available. Novel verbs are presented in a stream of syntactically informative sentences, but the sentences are not embedded in a social context, and no visual access to the verb's referent is provided until the test phase. The results provide insight into how toddlers may benefit from overhearing contexts in which they are not directly attending to the ambient speech, and in which no conversational context, visual referent, or child-directed conversation is available.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Learning new words is a critical task of early childhood. It is typically thought of as a social process in which the child apprehends the intentions of an interlocutor and/or engages in joint attention with the interlocutor to the word's referent (e.g., Baldwin, 1993; Baldwin & Moses, 2001; Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009; Tomasello, 2000; Tomasello & Barton, 1994; Tomasello & Farrar, 1986). But many potential learning situations do not present such rich social or visual information. Imagine a toddler playing quietly in the living room while her parents talk in the kitchen. The toddler may overhear many words for which she does not have meanings, but in this situation she lacks visual access to the referents of those words, and the speech she is hearing is not child-directed and is not embedded in a child-centered conversation or interaction. Can toddlers benefit from such socially impoverished learning situations by positing at least basic meanings for some of the new words they hear?

Recent studies have begun to address this question by closing in on the kinds of information, at a minimum, that toddlers require to establish a new word's meaning. For example, by age two, toddlers can acquire the meaning of a novel noun by watching two adults interact with and talk about its referent, even if they are not directly spoken to and do not themselves interact with the object (e.g., Akhtar, Jipson, & Callanan, 2001; Gampe, Liebal, & Tomasello, 2012; O'Doherty et al., 2011; Scofield, Williams, & Behrend, 2007; Shneidman, Buresh, Shimpi, Knight-Schwarz, & Woodward, 2009). They can also acquire novel nouns introduced in adult-directed speech rather than child-directed speech (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011).

For verbs, too, there is evidence of learning in the absence of joint attention. Shown a videotaped interaction between two adults who are animatedly engaged in conversation, and using a novel verb in informative linguistic contexts, toddlers can assign the verb a broad meaning, even though the event it describes is not visually available (e.g., Arunachalam, Escovar, Hansen, & Waxman, 2012; Arunachalam & Waxman, 2010; Scott & Fisher, 2009; Yuan & Fisher, 2009). The critical information provided to toddlers in these studies is syntactic. Toddlers hear the novel verb in syntactic contexts that provide information about the verb's meaning, and can use this information to set

* Address: Department of Speech, Language, and Hearing Sciences, 635 Commonwealth Ave., Rm. 323, Boston, MA 02215, United States. Tel.: +1 617 353 7491.

E-mail address: sarunach@bu.edu

down a basic lexical representation for the verb, an ability known as syntactic bootstrapping (Gleitman, 1990; Landau & Gleitman, 1988). In Arunachalam and Waxman (2010), for example (following on results from Naigles, 1990; see also Hirsh-Pasek, Golinkoff, & Naigles (1996)), toddlers heard adults conversing using novel verbs in either transitive sentences (e.g., *The girl lorped the boy*) or intransitive sentences (e.g., *The girl and the boy lorped*). These two sentence types describe fundamentally different kinds of relations between the event participants: transitive verbs typically describe causative events, in which one actor acts on another (e.g., a girl pushes a boy), and intransitive verbs typically describe non-causative events (e.g., a girl and a boy wave). Even though in Arunachalam & Waxman's task toddlers were only overhearing this syntactic information, and were not ostensibly shown the verb's referent, they nevertheless mapped novel transitive verbs, but not intransitive verbs, to causative events.

But is social context, such as an animated conversation between two adults, *required* to trigger toddlers' abilities to use linguistic context to acquire verb meanings? For some other kinds of learning from auditory stimuli, social context is not required. Statistical learning studies demonstrate that even young infants can glean patterns from auditory input presented in socially impoverished contexts in which the infant hears ambient auditory streams (e.g., Saffran et al., 1996; Marcus et al., 1999). However, these patterns are devoid of meaning. Identifying coherent forms from an auditory stream may differ fundamentally from assigning those forms a referential meaning (Naigles, 2002), with the latter requiring social context. In fact, toddlers have difficulty assigning meaning to isolated word forms if they do not perceive them as referential (Fennell & Waxman, 2010).

Nevertheless, we suggest that assigning basic aspects of meaning from syntactic cues is different from other referential mapping tasks, and that it does not require social context. Because syntactic cues convey aspects of meaning in themselves without identification of a specific visual referent in the world, cues available in social interaction, such as eye gaze, may not be necessary. In fact, we hypothesize that the socially impoverished conditions typically used in statistical learning studies will be sufficient. After all, syntax provides distributional information about where different kinds of words can and cannot appear in a sentence, and it stands to reason that humans' skills at discerning patterns in auditory input will serve them well in discerning that an unfamiliar transitive verb is both preceded and followed by a noun phrase. Syntactic information may thus provide a different inroad to acquisition, one that does not require the rich social contexts that other kinds of mapping tasks do.

Therefore, in the current study we asked whether toddlers could acquire the meaning of a novel verb from severely socially impoverished situations: we did not provide child-directed speech, conversational context, interaction with or eavesdropping on people speaking (neither live nor videotaped), or visual access to the event the verb describes. The only information provided was linguistic: toddlers heard the novel verbs in informative sentences (either transitive or intransitive). These sentences were presented in a maximally non-social context: they were recorded in adult-directed speech and were not embedded in a discourse, and streamed ambiently from a speaker while the child engaged in an unrelated activity (Lany & Saffran, 2011). At test, the toddlers saw two candidate referents for the novel verb: a causative event and a synchronous event, and were asked, e.g., "Find lorping." Our goal was to determine whether toddlers, on hearing the transitive or intransitive sentences, could use their syntactic content alone to map the novel verbs to meaning, even though no social or visual information was available.

2. Methods

2.1. Participants

Forty typically-developing toddlers (age range: 25.0–29.9 months, mean: 27.2) were included in the final sample. All were recruited from Boston, MA and surrounding communities, and were acquiring English as their native language, hearing other languages less than 30% of the time. Caregivers completed the MacArthur-Bates Communicative Development Inventory Short Form Level II A (Fenson et al., 2000). Toddlers' production vocabulary ranged from 38 to 100 words (mean 75 words) and did not differ between conditions. An additional eight toddlers were excluded from analysis due to inattentiveness and one due to parental interference.

2.2. Materials

Participants participated in four trials, each consisting of a Familiarization and a Test phase.

2.2.1. Visual stimuli

During Familiarization, the visual stimuli consisted of animated shapes moving silently on the screen, and during Test, they consisted of digitized video recordings of live actors performing actions.

2.2.2. Auditory stimuli

During Familiarization, the auditory stimuli consisted of 27 sentences playing from the monitor's speakers. These sentences were produced by a female native speaker of American English with list-reading prosody, and included a variety of noun phrase participants in the verb's argument positions as well as a variety of tenses. The stream played as a series of unconnected sentences with approximately 1 s between them. Speech was recorded in a sound-attenuated booth. During Test, toddlers heard two attention-getting phrases (*Look! Wow!*) and the test query (e.g., *Where's lorping? Find lorping!*), also recorded by a female native speaker of American English in a sound-attenuated booth, and synched with the visual stimuli.

2.3. Apparatus and procedure

Toddlers played with toys while the caregiver signed a consent form and completed the MacArthur-Bates checklist. The toddler and caregiver were then brought into the

testing room, where the toddler sat in a car seat, 20 in from a widescreen (24-in.) corneal reflection eye-tracker monitor (Tobii T60XL), which records gaze at a rate of 60 frames/s. The caregiver sat behind the toddler and was instructed not to speak during the session.

Toddlers first participated in two warm-up trials in which they viewed silent videos of familiar characters (e.g., Elmo and Big Bird) and were asked, e.g., "Where's Big Bird?" These trials were designed to accustom toddlers to seeing two different scenes simultaneously and to being asked a question that should draw their attention to one of them. No novel words were introduced during the warm-up trials.

Next, toddlers participated in four test trials. Each featured a different novel verb, and comprised a Familiarization Phase and a Test Phase. In the Familiarization Phase, toddlers heard a stream of either transitive or intransitive sentences, depending on their random assignment to condition (see Fig. 1). While the sentences played, toddlers viewed a silent animation of colored shapes moving on the screen, or, if they appeared uninterested in the animation, they were given a shape sorter to play with during this time. (Toddlers played primarily with the toy and ignored the animation on 8% of trials; performance on these trials did not differ from those on which toddlers looked at the animation.) Toddlers were encouraged to remain quiet but their attention was not explicitly directed to the ambient sentences.

During the Test Phase, two videos were presented simultaneously, side-by-side, of actors performing actions. One depicted two actors engaged in a causative action (e.g., boy spins girl), and the other depicted the same two actors engaged in synchronous actions (e.g., boy and girl each wave). The Test Phase included three subphases. First was Baseline, a 24 s inspection period during which toddlers heard "Look! Wow!" and had an opportunity to inspect the test scenes, both of which were novel to them.[1] Second was Central Fixation, during which the scenes disappeared, a yellow star appeared in the center of the screen, and toddlers heard a test query containing the novel verb, e.g., "Where's *lorping*? Finally, during Response, the two test scenes reappeared for 24 s, and the test query was repeated: "Do you see *lorping*? Find *lorping*!" Crucially, the test query provided no information that could help toddlers distinguish the two scenes; to succeed, they had to rely on the sentences they had heard earlier. Toddlers' eye gaze was recorded and provided an indication of their interpretation of the novel verb.

2.4. Analysis

If toddlers gleaned the novel verb's syntactic properties from the stream of sentences they heard, despite the impoverished social context, then at Test, those who had heard transitive sentences should prefer the causative event compared to those who had heard intransitive sentences. Our analysis therefore asks if there is a difference between the syntactic conditions in looking to the causative scene during the Response window.[2] However, because these are novel verbs that the toddlers are still in the process of acquiring, we do not expect this difference between conditions to arise immediately after the test query. Although toddlers look to targets remarkably quickly when *familiar* words are queried (Swingley, Pinto, & Fernald, 1999), for novel words, and particularly novel verbs, toddlers require longer. Using a very similar paradigm to the current study, Arunachalam et al. (2012) found that 21-month-olds required approximately 1.5 s from the onset of the response window to show a difference between conditions in a novel verb learning task. Anticipating slightly faster processing in 27-month-olds than their younger counterparts (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998), we expected a difference between conditions to emerge after 1 s. We therefore divided the analysis window into two time periods: 0–1 s, during which we expect no difference between conditions, and 1–2.5 s.

For each toddler on each trial, we calculated the proportion of frames on which toddlers looked to the causative scene out of all frames excluding track loss (that is, including looks to locations other than either test scene). We then aggregated the proportion data into 50 ms bins (because of the non-independence of gaze direction on consecutive frames, see Barr, 2008), transformed the binned data using an empirical-logit function (Barr, 2008), and fit the transformed data using a multilevel logistic regression model. We included three fixed effects: Syntactic Condition (Transitive vs. Intransitive), Time (binned at 50 ms), and Time Window (0–1 s, 1–2.5 s). Critically, we are interested in the interaction between Condition and Time Window, predicting that toddlers in the Transitive condition will be more likely to look to the causative scene in the 1–2.5 s window than those in the Intransitive condition.

3. Results

As predicted, within 1 s of the onset of the test scenes, looking patterns in the two conditions diverged (Fig. 2). In the window from 1 to 2.5 s after test onset, toddlers in the Transitive condition preferred the causative event (mean proportion of looks to the causative scene = 0.59, SD = 0.20) compared to those in the Intransitive condition (mean looks = 0.44, SD = 0.23) (Cohen's *d* = 0.70). The multilevel logistic regression analysis supports this interpretation. The critical interaction term (Condition × Time Window) reveals that the odds of looking to the causative scene from 1 to 2.5 s are 2.4 times (0.87 logits) higher for

---

[1] This period also allowed us to determine whether toddlers showed a preference for one of the scenes before any novel words were introduced. They did not: In the Intransitive condition they looked to the causative action scene 46% of the time and the synchronous action scene 47% of the time, and in the Transitive condition they looked to the causative action scene 44% of the time and the synchronous action scene 50% of the time.

[2] Though the Response window lasted 24 s, this length is an artifact of a different experimental procedure in which toddlers were asked to point to the screen at test, and were given ample time to do so. The same timing was maintained in this study so that, in principle, data from the two paradigms could be compared. In the current study, we only analyze data from the first few seconds, as is typical for eye-tracking studies with toddlers (see, e.g., Fernald, Zangl, Portillo, & Marchman, 2008).

FAMILIARIZATION	TEST		
	BASELINE	CENTRAL STAR	RESPONSE
(toddler watches an irrelevant animation or plays quietly with a shape sorter)	Causative Synchronous		Causative Synchronous
Transitive condition The girl wants to *lorp* the teacher. Mommy *lorped* the train. ...	Look! Wow!	Where's *lorping*?	Find *lorping*!
Intransitive condition The girl and the teacher want to *lorp*. Mommy and the train *lorped*. …			

Fig. 1. Representative set of stimuli from one trial (of four). Toddlers viewed the dynamic video scenes on an eye-tracker monitor. The only difference between conditions is whether toddlers heard transitive or intransitive sentences during Familiarization.

toddlers in the Transitive than the Intransitive condition. The beta coefficients are reported in Table 1.

4. Discussion

These results reveal that neither observational nor social information is required for toddlers to posit a basic representation for a novel verb. For verbs, at least, rich linguistic information alone can be sufficient for acquiring at least some aspects of meaning. Of course we do not claim that toddlers do not use observational and social information to acquire the meanings of words when they

Fig. 2. Toddlers' looks to the causative scene over time from the onset of the Response window after the test query (e.g., *Where's lorping?*), separated by condition. As predicted, toddlers in the Transitive condition reliably preferred the causative scene compared to those in the Intransitive condition from 1 to 2.5 s from Response onset.

are available. It is well documented that toddlers recruit and integrate information from multiple sources, including visual, social, and linguistic, to discover word meanings, including verb meanings (e.g., Bloom, 2000; Hollich et al., 2000). Rather, we argue that when these sources of information are absent, toddlers nevertheless posit rudimentary word meanings that can be built upon in future encounters.

Our results therefore provide insight into how toddlers may be able to benefit from overhearing contexts in which they are not directly attending to the ambient speech, and in which no visual referent, discourse context, or child-directed conversation is available. Puzzlingly, however, several naturalistic studies have found that the amount of speech toddlers overhear does not predict their vocabulary size, while amount of speech directed to the toddler does (Shneidman & Goldin-Meadow, 2012; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2012; Weisleder & Fernald, in press). This suggests that toddlers may not be using their experimentally demonstrated abilities to benefit from overhearing in daily life. Why might this be? Shneidman et al. (2012) suggest that in experimental studies, toddlers' focus of attention is quite restricted compared to naturalistic home settings, and that therefore toddlers were able to focus on the overheard speech. However, the current study differs in that toddlers' attention was divided; unlike previous experimental studies in which toddlers have watched and
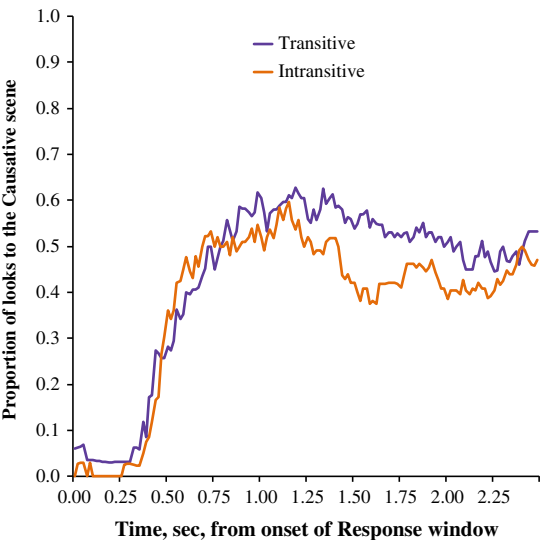
Table 1
Fixed effects from multi-level logistic regression model of proportion of time spent looking at causative scene (empirical logit transformed).

Effect	Estimate	S.E.	t-Value
Subject (Intercept)	−2.28	0.14	−16.16[*]
Syntactic condition (Trans vs. Intrans)	0.02	0.20	0.11
Time (in 50 ms bins)	0.15	0.15	0.98
Time window (0–1 s vs. 1–2.5 s)	1.68	0.11	14.76[*]
Condition × time window interaction	0.87	0.21	4.25[*]

[*] $p < 0.05$ (on normal distribution).

overheard an interaction in which adults were labeling a novel object (e.g., Akhtar et al., 2001; Gampe et al., 2012; O'Doherty et al., 2011; Scofield et al., 2007; Shneidman et al., 2009), here toddlers were presented with auditory information while watching an unrelated animation or playing with a toy. The attention hypothesis therefore would not appear to capture toddlers' success in our study.

Alternatively, it could be that toddlers in our study succeeded because we provided many (27) exposures to the novel verbs in informative sentence contexts in a concentrated time window. Even still, however, that they succeeded despite their divided attention indicates strong abilities to extract meanings, and not just word forms, from the ambient speech stream. Another possibility is that for some kinds of words, like verbs, for which syntactic information is particularly useful, overhearing may be more successful than for the nouns that make up a larger part of toddlers' early vocabularies. Such subtle advantages for verbs may not have manifested in large naturalistic studies. Future studies must bridge the gap between naturalistic and experimental findings by understanding how toddlers acquire the meaning of any given word from naturalistic overheard input.

Acknowledgments

References

Akhtar, N., Jipson, J., & Callanan, M. (2001). Learning words through overhearing. *Child Development, 72*, 416–430. http://dx.doi.org/10.1111/1467-8624.00287.

Arunachalam, S., Escovar, E., Hansen, M. A., & Waxman, S. R. (2012). Out of sight but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and Cognitive Processes.* http://dx.doi.org/10.1080/01690965.2011.641744.

Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition, 114*, 442–446. http://dx.doi.org/10.1016/j.cognition.2009.10.015.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418. http://dx.doi.org/10.1017/S0305000900008345.

Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 309–329. http://dx.doi.org/10.1111/1467-9507.00168.

Barr, D. J. (2008). Analyzing 'visual world' eye-tracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457–474. http://dx.doi.org/10.1016/j.jml.2007.09.002.

Bloom, P. (2000). *How Children Learn the Meanings of Words.* Cambridge, MA: MIT Press.

Fennell, C., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development, 81*, 1376–1383. doi: 0.1111/j.1467-8624.2010.01479.x.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21*, 95–115.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*, 228–231.

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: Online methods in children's language processing* (pp. 97–135). Amsterdam: John Benjamins.

Gampe, A., Liebal, K., & Tomasello, M. (2012). Eighteen-month-olds learn novel words through overhearing. *First Language, 32*, 285–297. http://dx.doi.org/10.1177/0142723711433584.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55. http://dx.doi.org/10.1207/s15327817la0101_2.

Hirsh-Pasek, K., Golinkoff, R. M., & Naigles, L. (1996). Young children's use of syntactic frames to derive meaning. The origins of grammar, 123–158.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. With Hennon, E., Chung, H. L., Rocroi, C., Brand, R. J., & Brown, E. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. Monographs of the Society for Research in Child Development, 65, (3, Serial No. 262).

Landau, B., & Gleitman, L. R. (1988). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Lany, J., & Saffran, J. R. (2011). Interactions between statistical and semantic information in infant language development. *Developmental Science, 14*, 1207–1219. http://dx.doi.org/10.1111/j.1467-7687.2011.01073.x.

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development, 7*, 209–225. http://dx.doi.org/10.1080/15475441.2011.579839.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*, 357–374. http://dx.doi.org/10.1017/S0305000900013817.

O'Doherty, K., Troseth, G. L., Shimpi, P. M., Goldenberg, E., Akhtar, N., & Saylor, M. M. (2011). Third-Party Social Interaction and Word Learning From Video. *Child Development, 82*, 902–915.

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

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development, 80*, 1360–1375. http://dx.doi.org/10.1111/j.1467-8624.2009.01338.x.

Scofield, J., Williams, A., & Behrend, D. A. (2007). Word learning in the absence of a speaker. *First Language, 27*, 297–311. http://dx.doi.org/10.1177/0142723707077252.

Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes, 24*, 777–803. http://dx.doi.org/10.1080/01690960802573236.

Shneidman, L. A., Buresh, J. S., Shimpi, P. M., Knight-Schwarz, J., & Woodward, A. L. (2009). Social experience, social attention and word learning in an overhearing paradigm. *Language Learning and Development, 5*, 266–281.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2012). What counts as effective input for word learning? *Journal of Child Language, 1*(1), 1–15.

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: how important is directed speech? *Developmental science, 15*(5), 659–673.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*, 73–108. http://dx.doi.org/10.1016/s0010-0277(99)00021-9.

Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology, 30*, 639–650. http://dx.doi.org/10.1037/0012-1649.30.5.639.

Tomasello, M. (2000). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development.* Cambridge University Press.

Tomasello, M., & Farrar, J. (1986). Joint attention and early language. *Child Development, 56*, 1454–1463. http://dx.doi.org/10.2307/1130423.

Weisleder, A., & Fernald, A. (in press). Talking to children matters: Early language experience strengthens processing and builds vocabulary. Psychological Science.

Yuan, S., & Fisher, C. (2009). ''Really? She blicked the baby?'': Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science, 20* (pp. 619–262, doi: 10.1111/j.1467-9280.2009.02341.x).