Use of Referential Context in Children's Language Processing

Yi Ting Huang (huang@wjh.harvard.edu)
Department of Psychology, 33 Kirkland Street
Cambridge, MA 02138 USA

Jesse Snedeker (snedeker@wjh.harvard.edu)
Department of Psychology, 33 Kirkland Street
Cambridge, MA 02138 USA

Abstract

Adult language comprehension is rapid, incremental, and opportunistic—making use of multiple cues from both linguistic and extra-linguistic sources (MacDonald et al., 1994). In contrast, recent research on early parsing has demonstrated that while children readily use bottom-up lexical cues about the meanings of words, they typically fail to use top-down contextual cues from the referential scene (Trueswell et al., 1999). This study examines whether children also fail to use this kind of referential information to interpret scalar adjectives like *big* and *tall*. Using the eye-tracking paradigm, five-year-olds were asked to *"Point to the big coin"*. We varied both the number of referents for the noun (1 or 2) and the polarity of the adjective. We found that children used the linguistic meaning of these terms to reliably differentiate the referents of *big* and *small* approximately 400 ms after adjective onset. Critically, children, like adults, were able to use referential contrast in interpreting scalar adjectives, looking to the Target more often in trials containing 2 referents than those with 1 referent from 600 ms after adjective onset. Altogether these results indicate that adjectives rapidly influence early reference resolution through both their meanings and their referential implications.

Keywords: Language processing; Referential contrast; Scalar adjectives; Development

Introduction

How is spoken language integrated with a visual scene in real-time comprehension? When visiting the Dunkin' Donuts, we coordinate with the person behind the counter to get the bagel we want ("No, not that one. The one on the row below it") or when pointing out an important someone, we add descriptions to differentiate him from the crowd ("The cute one with the red hair"). Thus interpreting these utterances requires a fundamental ability to make a mapping between language and the referential world. In adults, this process generally operates quite efficiently – mature language comprehension has been characterized as rapid, incremental, and opportunistic in its use of linguistic and extra-linguistic information (MacDonald, Pearlmutter, & Seidenberg, 1994).

Nevertheless, the question of how these abilities arise during development remains less clear. Unlike adults, recent research has found that children may be less skilled in recruiting these referential cues during on-line interpretation. For example, Trueswell and colleagues (1999, following Tanenhaus et al., 1995) presented adults and five-year-olds with instructions like (1) in the presence of two referential scenes.

(1) Put the frog on the napkin in the box

In the One-referent context, they saw a display featuring two different locations (e.g., a box and a napkin) and two different referents (e.g., a frog on a napkin and a horse). In the Two-referent context, the different referents were replaced with two of the same kind (e.g., a frog on a napkin and a second lone frog). These displays favored distinct syntactic analyses of the ambiguous phrase "*on the napkin.*" In the Two-referent context, the presence of two potential referents in the scene means that additional modification is necessary to make the definite determiner felicitous. Thus we might expect that, listeners who are sensitive to this effect of referential context would favor the Modifier interpretation. In contrast, in the One-referent condition, a modifier would be overinformative and thus favors an initial misinterpretation of the ambiguous phrase as a Goal.

Adults conformed to these predictions. In the One-referent context, their early incorrect fixations to the empty napkin suggest an initial bias towards the Goal interpretation. Critically, in the Two-referent context, adults demonstrated an overwhelming preference for the Modifier interpretation. In their actions, they correctly selected the frog that was on the napkin while in their eye-movements, they made limited fixations to the empty napkin. In contrast, five-year-olds in both the One-referent and Two-referent contexts had a strong preference for the Goal interpretation, looking to the empty napkin shortly after the critical prepositional phrase. This misanalysis was also reflected in their actions—children often selected a frog at random and systematically move it onto the napkin before moving it into the box.

Overall this striking difference between adults' and children's performance presents a real puzzle. On the one hand, it is possible that children's insensitivity to referential contrast is due to an inability to use this information to make linguistic inferences. Yet we know that over the course of early word learning, even savvy young toddlers are able to exploit salient contrasts among multiple referents to map novel labels in a scene (Clark, 1990; Markman, 1989). Another possibility is that children are simply unable to employ multiple cues or non-lexical cues during parsing. However, this contradicts recent evidence demonstrating early use of both lexical and prosodic information in ambiguity resolution (Snedeker & Yuan,

2008). Finally, one can imagine that children simply do not encode referential contrast or lack the grammatical resources to create the required NP attachment. However, this hypothesis is ruled out by results showing that children successfully produce restrictive modifiers to differentiate between contrastive referents (Hurewitz et al., 2000).

One potential insight into this mystery comes from a prominent psycholinguistics model called the Constraint-Based Lexicalist theory (MacDonald et al., 1994; Trueswell & Tanenhaus, 1994). This account was initially advanced to explain comprehension in adults but has recently been extended to capture the developmental process in children (Trueswell & Gleitman, 2004). It posits that at the heart of the processing system is a linguistic architecture that represents input at many different levels (prosodic, syntactic, semantic, phonological, discourse) and a processing mechanism that is highly attuned to the regularities of individual lexical items. Processing at each level of representation makes use of constraints from the other levels as well as the stored lexical information to resolve ambiguity and make predictions about material that has yet to come.

Critically, this theory claims that children begin organizing language in this way as they acquire it.[1] However, such a system provides a clear role for experience even after the acquisition of the representational systems is largely complete: the links between the various representations must be formed and may change over the course of development. The rapidity with which these links form will crucially depend upon the strength of the correlation between the two phenomena. In the case of syntactic ambiguity, prior research has shown that referential contrast often fails to be a reliable predictor of the underlying structure of the sentence (Brown-Schmidt et al., 2005). This would perhaps explain why adults but not children were sensitive to this cue in earlier studies (Trueswell et al., 1999, see also Hurewitz et al., 2000; Snedeker & Trueswell, 2004).

Nonetheless, one way to increase to the reliability of referential cues is to look at a phenomenon where the number of referents is a more reliable predictor of intended meaning. Here we turn to a class of terms scalar adjectives like *big* and *tall* (Bierwisch, 1987; Kennedy, 1999). The lexical semantics of these words specify a scale along which entities can be compared (e.g., size or height) and a pole along that scale (e.g., greater in this dimension). Nevertheless to extend and interpret these adjectives, children must also set a standard of comparison that can depend upon both the noun which is being modified (*big car* versus *big coin)* and a contextually defined comparison class (*He is tall in Japan but not in Sweden*). Our choice of standard can dramatically alter the interpretation of the term and this flexibility of interpretation opens the question of how this comparison class is determined during real-time comprehension.

In adults, this class is often related to the set of potential referents is in the visual scene (Sedivy et al., 1999). For example, listeners were quicker to comprehend instructions like (2) in the presence of another contrasting referent of the same category (e.g. a short glass).

(2) Pick up the tall glass

This demonstrates that adults in this task can quickly use information about comparison classes to make predictions about upcoming referent. Like in the case of syntactic ambiguity, these contrast effects have been explained in terms of a communicative expectation that listeners have, namely that speakers will make their contributions as informative as is required (Grice, 1975). Thus when there is only one potential referent available, speakers will use a bare NP. However, when there are two potential referents in the display, they will generate an additional modifier to disambiguate.

In the following experiments, we explored whether the presence of a contrast set in the visual contexts influences children's real-time comprehension of scalar adjectives. Using the visual-world eye-tracking paradigm, five-year-olds were asked to "*Point to the mouse that has the big coin"* when presented with displays containing pairs of items which varied in size and belonged either to the same kind or to different kinds (e.g., mouse with BIG COIN vs. SMALL COIN or SMALL BUTTON). To assess the rapidity of linguistic processing, we examined the point at which looks to the Target (big object) exceeded looks to the Distractor (small object). To assess children's sensitivity to referential cues, we compared Target looks in the presence of within-category contrast (coin vs. coin) to Target looks in the presence of between-category contrast (coin vs. button).[2]

Critically, if children's real-time interpretations are sensitive to referential contrast, then we would predict facilitation of Target resolution when the contrast set is determined by the presence of a same category item. If, however, children have general difficulties incorporating this cue, then we would expect no differences between within- and between-category conditions.

Experiment

Methods

Subjects Twenty-seven five-year-olds (mean age 5;4) were recruited from daycares and schools around the Cambridge,

[1] An alternate view would be that children are initially impervious to extra-linguistic information sources like referential context (see Boland & Cutler, 1996 for a parallel theory for adult comprehension). Information from the lexicon may be encapsulated from other sources of information, thus children may not develop an ability to coordinate these cues until a later age.

[2] We also examined a second way in which interpretations of scalar adjectives could be sensitive to the norms within the relevant comparison class. In addition to referential contrast, we explored sensitivity to comparison based on stored knowledge of a particular category, evaluating looks to the Target when it was good size representative of the category (3" coin) relative to when it was not (0.5" coin). However, since they do not directly bear on our current topic, these results will not be discussed.

MA area. Twenty-four students at Harvard University also took part and received course credit for their participation. All participants were native English speakers.

Procedure Participants sat in front of an inclined podium divided into four quadrants, each containing a shelf where objects could be placed. A camera at the center of the display recorded their face during the task. Prior to the study, the experimenter introduced two mice and two frogs each of whom would own the object placed in their quadrant. On each trial, participants heard names for each object ("*coin, coin, ruler, pencil*") followed by prerecorded commands like (3).

(3) Point to the mouse that has the big coin

Materials Participants received 32 randomized trials with two within subject factors. Referential contrast was manipulated by varying whether the critical Distractor was from the same basic level (*within-category*) or from a different one (b*etween-category*). Token fit was manipulated by varying whether the Target was good exemplar of the adjective noun combination (*good token*) or a poor one (*poor token*). For each trial, participants saw three types of objects placed on the shelves (see Figure 1): (1) Targets matched both animal and object size cues (e.g. mouse with big coin), (2) Distractors matched animal but not object size cues (e.g. mouse with small coin or button), and (3) Objects that did not match the animal cues (e.g. frogs with ruler and pencil).[3] The adjective encoded either height (*tall, short*) or size scale (*big, small*) and were counterbalanced for polarity through the creation of eight lists, each containing eight items per condition with each item rotated through the four conditions.

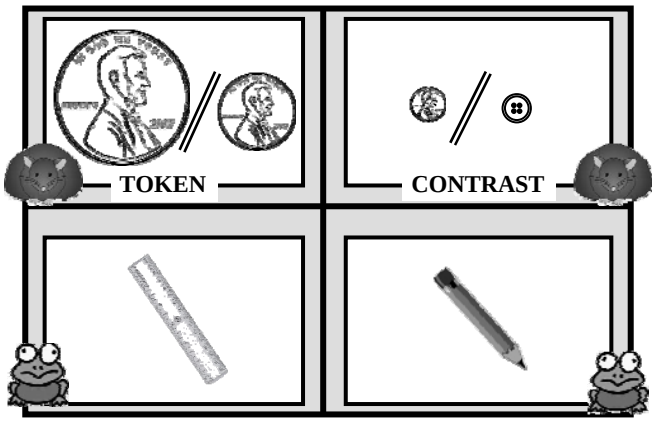

Figure 1: Visual display for the four conditions. Participants saw one token and one contrast object.

Coding Trained research assistants coded videotapes of the participant's actions and eye movements. For details see Snedeker and Trueswell (2004). This method of eye-tracking has been validated with high inter-coder reliability and highly correlates with data obtained from studies using head mounted eye tracking.

[3] The presence of the animal cue functioned to shift attention to the relevant pair across all trials, preventing any baseline difference in looking to this pair across the contrast and no contrast trials. To ensure that membership in a contrast set did not predict the Target, the other animals received contrast sets for half of the trials.

Results

First, we assessed the rapidity of semantic processing, by focusing our attention on performance in the within-category/good token condition. Since these trials use the same display for Targets from the positive pole (*big/tall*) and the negative pole (*small/short*), they allow us to determine the moment at which polarity, which is part of the semantic meaning of these terms, is used to restrict reference. All time points offset by 200 ms to account for the time it would take to program a saccadic eye-movement (Matin, Shao, & Boff, 1993). For each trial, we summed the total number of looks to the *big/tall* object and the *small/short* object and calculated the proportion of looks to the *big/tall* over looks to both. This score ranged from zero (exclusive looks to the *small/short*) to one (exclusive looks to the *big/tall*).

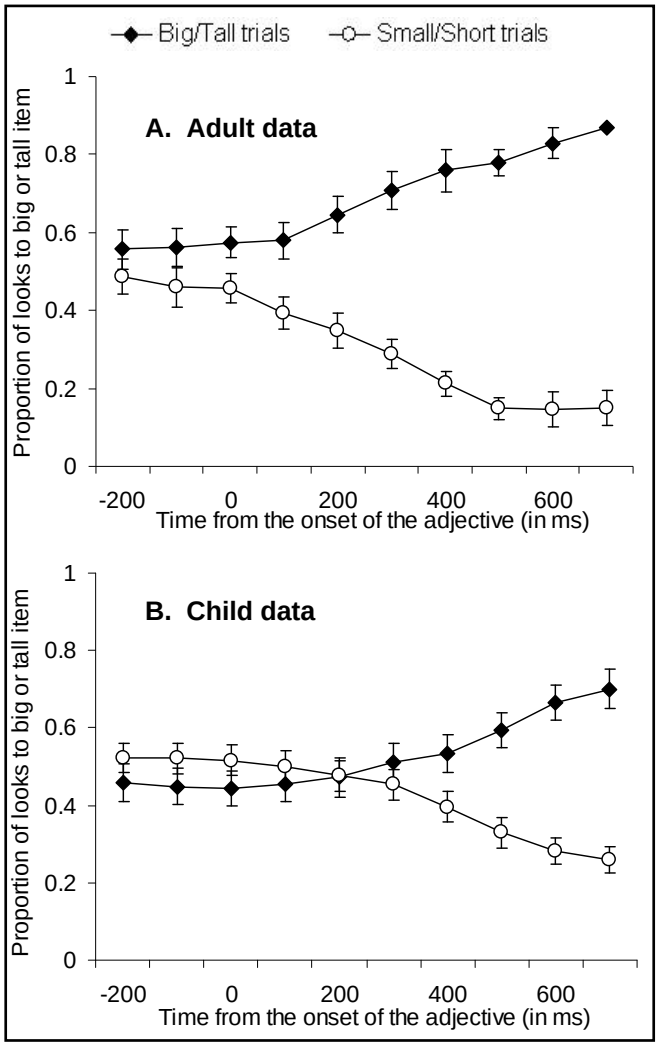

Figure 2: Semantic processing in adults and children

Figure 2 illustrates these results. Adults reliably differentiated the referents of *big/tall* from *small/short* immediately after adjective onset (57% vs. 45%; $t_1(23) = 2.31$, $p < .05$; $t_2(31) = 1.73$, $p < .10$) while children reliably did so during the 400 ms time window (60% vs. 33%; $t_1(26) = 4.08$, $p < .01$; $t_2(31) = 4.57$, $p < .01$). Thus we found that the meaning of scalar adjectives is used in reference resolution at the earliest moments of language processing.

Next, we turned to adult and children's sensitivity to referential contrast. We examined the proportion of eye-movements towards the Target over two divisions of time. Our first analysis examined a coarse-grained measure during four periods of analysis (see Table 1). For each trial, we summed the total number of looks to the Target and Distractor and calculated the proportion of looks to the Target over looks to both. This score ranged from zero (exclusive looks to the Distractor) to one (exclusive looks to the Target). Looks to the other objects were infrequent after onset of the animal cue and were not included in the analysis. Each time window was analyzed with ANOVAs with referential contrast (within- vs. between-category) and token contrast (good vs. poor) as within subject and item factors. List/item group were analyzed between subjects and between items.

Table 1: Time windows used in course-grained analysis

Region	Period within instructions
1. BASELINE	*Point to the*
2. ANIMAL CUE	*mouse that has the*
3. ADJECTIVE	*big*
4. NOUN	*coin*

During the BASELINE REGION, the proportion of looks to the target character initially remained around chance across both within-category trials and between-category trials for both adults (49% vs. 48%; $F_1(1, 16) = 0.14$, $p > .70$; $F_2(1, 31) = 0.29$, $p > .50$) and children (50% vs. 51%; $F_1(1, 20) = 0.73$, $p > .40$; $F_2(1, 31) = 0.79$, $p > .30$). This continued through the following ANIMAL CUE REGION, where again there was no difference between the two displays in both adults (49% vs. 50%; $F_1(1, 16) = 0.55$, $p > .40$; $F_2(1, 31) = 0.29$, $p > .50$) and children (49% vs. 52%; $F1(1, 20) = 2.16$, $p > .15$; $F2(1, 31) = 4.36$, $p < .05$).

However, during the ADJECTIVE REGION, fixations to the Target increased more in the presence of a within-category contrast relative to a between-category item. Adults showed clear effects of referential context making significantly more looks to the Target during the within-category trials compared to the between-category trials (74% vs. 66%; $F_1(1, 16) = 10.73$, $p < .01$; $F_2(1, 31) = 13.16$, $p < .01$). Critically, children also showed more rapid processing of the adjective when the target belonged to a contrast set. Looks to the Target were greater for the within-category trials during this region (60% vs. 53%; $F_1(1, 20) = 7.50$, $p < .05$; $F_2(1, 31) = 8.28$, $p < .01$). In adults, this effect was short-lived and quickly disappeared by the NOUN REGION (90% vs. 89%; $F_1(1, 16) = 0.53$, $p > .40$; $F_2(1, 31) = 0.51$, $p > .40$). In children, however, looks to the Target continued to be greater for the within-category trials compared to the between-category trials (78% vs. 72%; $F_1(1, 20) = 6.00$, $p < .05$; $F_2(1, 31) = 6.26$, $p < .05$).

Finally, we examined fine-grained differences in fixation across the within-category and the between-category trials using 100 ms intervals immediately following the onset of the adjective. This period was of particular interest since it corresponds to the region when the referent of the referent ambiguous between the Target and Distractor.

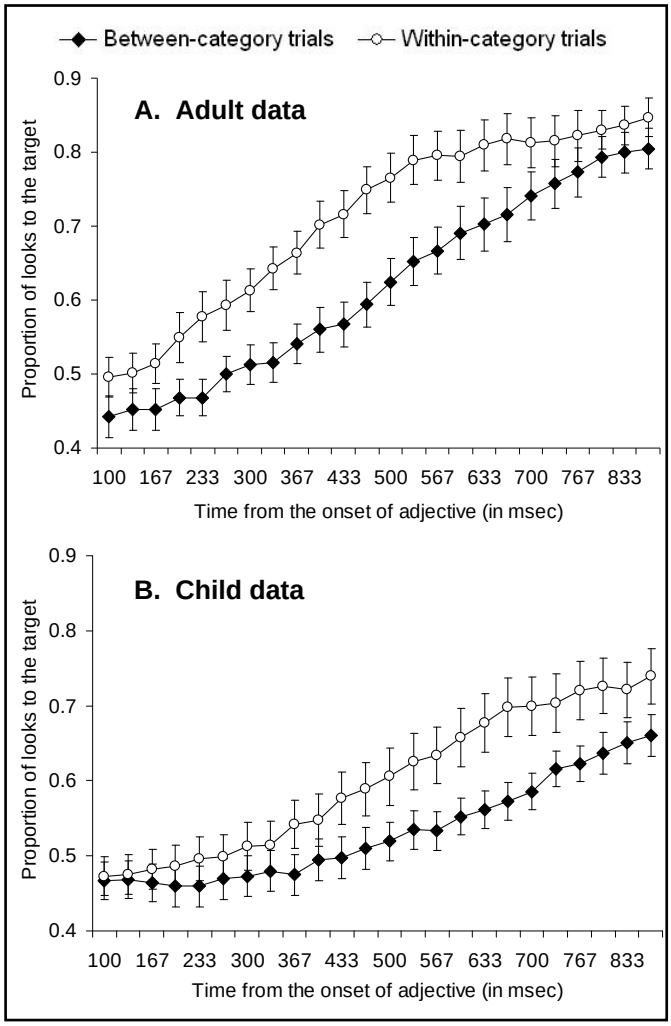

Figure 3: Use of referential contrast in adults and children

These analyses confirmed the differences in time it took participants to reliably fixate on the Target across the two displays (see Figure 3). In adults, a greater Target preference in the within-category trials emerged approximately 200 ms after the onset of the adjective ($F_1(1, 16) = 6.13$, $p < .05$; $F_2(1, 31) = 5.62$, $p < .05$) and continued until the 600 time window ($F_1(1, 16) = 7.24$, $p < .05$; $F_2(1, 31) = 7.53$, $p < .05$). In children, this effect emerged approximately 500 ms after the onset of the adjective ($F_1(1,$

20) = 6.47, $p < .05$; $F_2(1, 31) = 5.64$, $p < .05$) and continued until the 800 ms time window ($F_1(1, 20) = 7.93$, $p < .05$; $F_2(1, 31) = 10.23$, $p < .01$). These findings demonstrate that, like adults, five-year-olds rapidly use referential contrast to interpret adjectives.

Discussion

This study explores the real-time interaction between semantic information and referential contrast in children by investigating their interpretations of scalar adjectives. We found that five-year-olds possess a robust ability to exploit lexical information concerning the semantics of these terms. Like adults, they incorporated the linguistic meanings of these terms to distinguish between opposite poles on the scale at the earliest moments of processing. Critically, we also found that these children exhibited a capacity to make predictive inferences about the up-coming noun based on the set of potential referents in the scene. Like adults, they were quicker to restrict reference in the presence of another contrasting referent of the same category. More broadly speaking, these findings add to a growing literature finding early sensitivity to multiple, non-lexical cues like discourse (Hurewitz et al., in prep) and prosody (Snedeker & Yuan, 2008) in children's on-line interpretations.

Yet these findings creates an apparent tension between our current results and previous findings highlighting children's disproportionate use of lexical information about the meanings of words and failure to use contextual cues from the referential scene (Trueswell et al., 1999; Hurewitz et al., 2000; Snedeker & Trueswell, 2004). Why would children use referential contrast in one situation but not another? Below we explore a couple potential explanations for the apparent discrepancy.

Reliability of cues

One possibility lies in the use of contextual contrast for the purpose of resolving syntactic ambiguities. As mentioned in the Introduction, previous research has shown that that, referential contrast is in fact a less reliable cue to the underlying structure of the sentence (Trueswell et al., 1999; Snedeker & Trueswell, 2004). This is consistent with surprising evidence from production studies showing that speakers would frequently utter a bare definite NP (e.g., "*the frog*") in the presence of multiple potential referents (e.g., many frogs) (Brown-Schmidt et al., 2005). It turns out that rather than producing a restrictive modifier, speakers would instead rely on functionally equivalent information from prior discourse and established goals to avoid ambiguity. This lack of a straightforward link between the number of items in the scene and production of a particular structure explains previous insensitivity to referential contrast in syntactic parsing.

However, in the case of scalar adjectives, there is in fact a tight correlation between the number of referents and production of these terms. Sedivy (2003) found that words like *big* and *tall* were much more likely to be produced in the context of multiple referents from the same category (e.g., several cups) compared to a single referent. This makes referential contrast a particularly valid cue to the use of the adjectives. The fact that we found reliable contrast effects in children's real-time interpretations of these terms provides evidence that increased cue validity contributes to establishing critical links between various representational systems. It also suggests that like adults, children as listeners do indeed abide by sophisticated notions embodied in communicative principle. From an early age, they expect that speakers will avoid redundant modification in presence of a single referent and produce disambiguating modifiers in situations containing multiple referents.

Making predictions in real-time comprehension

A second possible explanation for the discrepancies between current and previous findings lies in a qualitative difference in the kinds of processes that are tapped in the two sets of studies. Most of the prior research including the original Trueswell et al. (1999) experiment focused on tasks that measured syntactic ambiguity resolution. Recall that participants were presented with a post-nominal modifier (e.g., "*on the napkin*") that was consistent with either a Goal or Modifier interpretation. Nevertheless, children's behavior indicates that they were only readily entertaining the Goal one. This could happen for two reasons. First, children's on-line interpretation has been shown to be largely dominated by their knowledge of lexically-specific syntactic/semantic information (Snedeker & Trueswell, 2004). Thus given the strong bias for a destination found in the initial verb *put*, children committed to the need for a goal even before hearing the ambiguous prepositional phrase (i.e. "Put it where? Put it on the napkin"). Furthermore, following the onset of the noun, children may quickly establish a referent in the scene and subsequently fail to see the need for additional referential specificity.[4]

In contrast, our current study is a predictive task. We presented participants with a pre-nominal modifier ("*big*") and examined whether their predictions about the up-coming noun is facilitated by the elements in the visual scene. Here across all conditions, the linguistic information was always congruent with what was in the display since the Target was consistently a better example of the adjective relative to the Distractor. The critical question was whether the presence of referential contrast that would facilitate this predictive process. Here we found no age-differences in patterns of reference resolution: Children, like adults, readily encoded the contrast and used it to generate predictive inferences about the up-coming noun. More generally, the appearance and disappearance of contrast effects arising from the pre- versus post-nominal position of the modifier suggests that the ordering of information within

[4] Both these alternatives would also be consistent with previous findings in the adult-literature demonstrating a relationship between cognitive processing capacity and the computation of syntactic alternatives (MacDonald et al., 1992). Children may in fact pattern with adults with low working memory span in their decreased propensity to generate multiple interpretations.

a sentence may have critical effects on children's on-line interpretations of utterances.

Conclusion

In conclusion, the results of our study suggest that the same features that characterize an adult model of language comprehension are fundamentally present and operational in the child listener (Trueswell & Gleitman, 2004). Using the visual-world eye tracking, we found that children, like adults, exploit regularities in the referential context to facilitate real-time interpretations of scalar adjectives. Critically, these findings demonstrate that previous failures to incorporate top-down cues cannot solely be due to failure to encode referential contrast or ability to use it during comprehension. All together these results provide evidence for a model of real-time processing where the linguistic architecture is attuned to process information from multiple sources—linguistic as well as extra-linguistic regularities—early in development.

Acknowledgments

We are grateful to Maria Markhelyuk, Jenny Lee, Tricia O'Loughlin, Silvia Chen, and Claire Huang for their assistance in testing and data coding. We also thank the parents and children from the Arlington Children's Center and the Medford elementary schools for their participation. This work is supported by the NSF Grant No. 0623845.

References

Bierwisch, M. (1987). The Semantics of Gradation . In: M. Bierwisch, E. Lang (eds.): *Dimensional Adjectives*. Berlin: SpringerVerlag.

Boland, J. & Cutler, A. (1996). Interaction with autonomy: Multiple output models and the inadequacy of the Great Divide. *Cognition*, 58, 309-320.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. (2005). Reference resolution in the wild: Circumscription of referential domains by naive participants during an interactive problem solving task. In: J. Trueswell & M. Tanenhaus (Eds.) *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge: MIT Press.

Clark, E. (1990). On the pragmatics of contrast. *JCL*, 17, 417-431.

Grice, H. P. (1975). Logic and Conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics, Vol. 3*. New York: Academic Press.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. & Trueswell, J. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *JPR*, 29, 597-626.

Hurewitz, F., Brown-Schmidt, S., Trueswell, J. & Gleitman, L. (in prep). The contribution of conversation goals and verb-preferences to children's syntactic decisions.

Kennedy, C. (1999). Projecting the adjective: The syntax and semantics of gradability and comparison. New York: Garland Publishing.

MacDonald, M., Just, M., & Carpenter, P. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56-98.

MacDonald, M., Pearlmutter, N. & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

Markman, E (1989). *Categorization and naming in children: Problems of induction.* Cambridge: MIT Press.

Matin, E., Shao, K. & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perception & Psychophysics*, 53(4), 372-380.

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

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effect of informativity expectations. *JPR*, 32(1), 3-23.

Snedeker, J. & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49, 238-299.

Snedeker, J. & Yuan, S. (2008). The development of parsing: Children use lexical and prosodic cues in online sentence comprehension. To appear in *JML*.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632.

Trueswell, J. & Tanenhaus, M. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In Clifton and Frazier (Eds), *Perspectives on sentence processing*. Hillsdale, NJ: Lawrence Erlbaum.

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

Trueswell, J. & Gleitman, L (2004). Children's eye movements during listening: evidence for a constraint-based theory of parsing and word learning. In J. Henderson & F. Ferreira (eds.). *Interface of Language, Vision, and Action: Eye Movements and the Visual World.* NY: Psychology Press.