From Meaning to Inference:

Evidence for the distinction between lexical semantics and scalar implicature in online processing and development

Yi Ting Huang

Jesse Snedeker

Harvard University

Address Correspondence to:

Yi Ting Huang

Laboratory for Developmental Studies

Shannon Hall, 25 Francis Avenue

Harvard University

Cambridge, MA 02138, USA

huang@wjh.harvard.edu

Acknowledgments – This work benefited from conversations with members of the Cambridge Conference on Experimental Pragmatics, Lab for Developmental Studies, and the MIT-Harvard number reading group. We are grateful to Ayo Adigun and Hila Katz for their assistance in testing and data coding.

Abstract

While most theorists agree on the need to distinguish between semantic and pragmatic interpretation, there is little consensus about the location of this boundary and the relationship between these processes during comprehension. Past research has demonstrated that pragmatic processing is rapid, often beginning before phrasal completion (Frisson & Pickering, 1999; Sedivy et al., 1999). But this work leaves open the question of whether the pragmatic interpretation is preceded by some degree of semantic interpretation, as most linguistic theories would predict. These current experiments address this question by examining a relatively well-understood test case from linguistics. Horn (1989) noted that scalar quantifiers like *some* have two distinct readings corresponding to lexical semantics (SOME AND POSSIBLY ALL) and pragmatic inference (SOME BUT NOT ALL). A visual world task was used to track the online interpretation of *some, all, two* and *three* in adults and five year olds. The context provided two potential referents for the quantified noun phrase (e.g., a girl with 2 of 4 socks and one with 3 of 3 soccer balls). For *all, two* and *three* the reference of the quantified noun phrase was disambiguated by the semantics of the quantifier and participants were able to converge on the target shortly after quantifier onset. For *some*, however, determining the referent required a pragmatic implicature. On these trials looks to the target were substantially delayed, demonstrating a lag between semantic processing and the calculation of the implicature. Nevertheless, adults showed a preference for the target prior to the disambiguation point, demonstrating that scalar inferences can occur as an utterance unfolds. Children, however, failed to calculate the implicature, converging on the target only after it was lexically disambiguated.

Over the history of psycholinguistics, questions concerning the semantic-pragmatic interface have generated many experimental studies and theoretical accounts of the language processing. While many aspects of utterances are tightly linked to word meaning and syntactic structure, other facets are clearly added by context-sensitive, inferential interpretative processes. The fact that utterances are interpreted at both a semantic and pragmatic level has led many to question how these meanings are integrated over the course of real-time processing. Past research has generally pursued these issues by exploring how contextual information influences reaction times for sentence comprehension and has demonstrated that pragmatic processing is extremely rapid, often beginning before phrasal completion (Frisson & Pickering, 1999; Sedivy et al., 1999). However, these studies leave open the question of whether the pragmatic interpretation is ever preceded by some degree of semantic interpretation, as most linguistic theories would predict.

In order to reconcile these seemingly contradicting accounts, it is necessary to isolate semantic versus pragmatic meaning in situations where the procedural mechanism is well understood. Here, one domain that may serve as an effective test case for many of these issues is the interpretation of scalar quantifiers. Linguists have long noted that terms like *some* have two distinct interpretations (Horn, 1989; Gadzar, 1979). Typically, a sentence like (1) will be taken to imply that Henry ate some, but not all, of the ice cream (the upper-bounded reading).

(1) Henry: I ate some of the ice cream.

However, on occasion *some* can be used in a context that does not exclude the total set. Thus (1) differs from the lower-bounded reading in (2) where Karl asserts that Leif ate both *some* and *all* of the lutefisk.[1]

(2) Eva: Did anyone eat some of the lutefisk?

Karl: Yeah, Leif ate some. In fact, he ate all of it.

Gricean theorists have argued that weak scalars like *some* are semantically compatible with stronger terms like *all* as in (2). Interpretations that exclude the stronger scalar as in (1) require a pragmatic inference called a scalar implicature.

The fact that scalar quantifiers can be interpreted in both of these ways creates an ideal situation for an account of semantic and pragmatic processing. Here is a case where the meaning assigned at each level of interpretation corresponds to different quantities within an array. At the semantic level, the lexical meaning of *some* is compatible the total quantity within a given set (SOME AND POSSIBLY ALL) while at the pragmatic level, *some* is interpreted to exclude the total set (SOME BUT NOT ALL). Research on adult and children's interpretation of scalar terms seems to provide support the existence of these dual interpretations. However, while adults consistently favor the upper-bounded readings, children prefer the lower-bounded interpretations for a variety of scalar terms. For example, Noveck (2001) asked children and adults to evaluate statements like "x might be y" in contexts where "x must be y" was true. He found that while adults overwhelmingly rejected the weaker modal, 7- to 9-year-olds treated the statement to be logically compatible the stronger statement. Similarly, Papafragou and Musolino (2003)

[1] An infamous Norwegian dish made of fish soaked in lye.

found that five-year-olds, but not adults, were content to accept weak scalar predicates like *started* in situations where the stronger scalar term applied (i.e. *finished*).

All together these studies demonstrate that while scalar implicatures characterizes default interpretations among adults, they occur far less reliably among children, who are less pragmatically sophisticated. These results suggest that when children initially acquire scalar terms, their interpretations solely reflect the semantic content. Only later do they develop sensitivity to the pragmatic inference necessary to generate a scalar implicature. Here is a case where the theoretical distinction between semantics and pragmatics aspects of meaning plays itself out in the observable behavior of different populations. This relation can also be explored by investigating how these interpretations arise in the course of real-time processing. That is, how are semantic and pragmatic representations integrated over the course of comprehension?

In order to address this question, recent studies have sought to understand generation of scalar implicatures using measures of reaction time (see also Katsos et al, 2005). Bott & Noveck (2004) compared reaction time judgments from participants asked to evaluate the truth of underinformative statements like “Some elephants are mammals.” They found that participants who spontaneously adopted an implicature interpretation (i.e. judged the statement to be false) took longer than participants who adopted a lower-bounded interpretation (i.e. judged the statement to be true). However, there are many limitations to the generality of these results. First, since these procedures required participants to make overt truth judgments on written sentences, they might induce a more strategic processing than normal conversational discourse. Second, measures of sentence final reaction times are opaque to the underlying stages of processing. These delays could reflect the overall difficulty of judging false statements without specifically attributing them to the generation of scalar implicatures. These possibilities are supported by recent research finding no differences in RT when these interpretations were compared within-participants (Feeney et al., 2004).

One way to circumvent these problems is to use a procedure that could obtain an indirect measure of comprehension during the time-course of interpretation. The visual-world eye-tracking paradigm has been used extensively in psycholinguistic research to yield a sensitive, time-locked measure of linguistic processing (Eberhard et al., 1995). Participants are presented with spoken instructions, asking them to manipulate objects within a visual reference world, while their eye-movements to those objects are measured. This procedure has at least two advantages for exploring semantic and pragmatic interpretations. First, it allows spoken language to be used and provides a dependent measure that is tightly linked to interpretation. Second, measuring eye movements over the course of comprehension permits a dynamic look at how interpretations unfolds prior to the influence of any overt strategic judgments.

In the following experiments, we investigated how processing of scalar terms unfold over the course of on-line speech comprehension. We presented participants with visual displays containing items belonging to two girls—one with SOME-BUT-NOT-ALL of the socks and another with ALL of the soccer balls—and recorded their eye-movements when asked to *"Point to the girl that has some of the socks."* These critical trials contained a period of semantic ambiguity at the onset of the quantifier where the referent of a lower-bounded reading of some is compatible with both characters (i.e. *"…some of the soc-"*). Performance was compared to control trials containing quantifiers with lexically encoded upper-bounds, i.e. a strong scalar term (*all*) and number words (*two/three*). Since these terms do not require a pragmatic inference to specify exact quantities, these control trials do not have the same temporary semantic ambiguity as *some*.[2]

If semantic meaning is processed prior to pragmatic inferences, we would predict quick resolution of the target character in *two, three*, and *all* trials but slower resolution in *some* trials. In addition, if pragmatic implicatures occur rapidly during on-line speech comprehension, we would predict that resolution of the target character would occur prior to the completion of the phrase when the referent is lexically disambiguated (i.e. "…-*ks*").

Experiment 1

Methods

Participants

Twenty students at Harvard University took part and received course credit for their participation. All students were native monolingual English speakers.

Procedure

Participants sat in front of an inclined podium divided into four quadrants, each containing a shelf where pictures could be placed (i.e. upper left, upper right, lower left, and lower right). A camera at the center of the display recorded their face during the task. For each trial, the experimenter would place unlabeled characters on each shelf and then

[2] The semantics of number words has been an area of contention within theoretical linguistics. While some have claimed that numbers pattern like other lower-bounded scalars (Horn, 1989; Gadzar, 1979; Levinson, 2000), others have argued that numbers have an exact semantics (Horn, 1992; Koenig, 1991). See Huang et al. (submitted) for a description of this debate.

act out a scripted story where different objects were distributed among these four characters. Next participants heard prerecorded commands instructing them to select one of the characters (e.g. “Point to the girl with some of the socks”). Their selection was recorded by a second camera located behind them.

Materials

Participants received 16 randomized trials that varied across two factors. Quantifier type contrasted true scalars with number words across two levels of informational strength. Weak quantifiers (e.g. *two* and *some*) referred to terms that were logically compatible with stronger members of their respective scales (e.g. *three* and *all*).

Each trial contained four types of cards (see figure 1): (1) Targets matched both gender and object cues (e.g. girl with socks), (2) Gender Distracters matched gender but not object cues (e.g. girl with soccer balls), (3) Object Distracters matched item but not object cues (e.g. boy with socks), and (4) Irrelevant Distracters matched neither cue (e.g. boy with soccer balls). Targets were arranged horizontally adjacent to Object Distracter and vertically adjacent to Gender Distracters. Two sets of objects were distributed among horizontal pairs in a TWO/TWO and ZERO/THREE configuration for scalar trials and a TWO/TWO and ONE/THREE configuration for number trials.[3] Finally, presentation of materials was counterbalanced by creating four lists such that each list contained four items in each condition and each item was rotated through the four conditions.

[3] An earlier version of this experiment held constant the set of objects across all trial types. While this resulted in similar overall effects, it also produced a slight delay in the disambiguation of *three*. This occurred because the partitive construction of count phrases necessarily picks out a set from within a larger array (i.e. “three of the socks”). While the configuration for *two* trials naturally supports this construction, we fulfilled it for *three* trials by adding an extra Object Distracter. Since this item was always assigned to the opposite gender, it did not directly affect looks to the target during critical periods.

Coding

Trained research assistants coded videotapes of the participant's actions and eye movements. Each recorded trial began at the onset of these instructions and ended with completion of the corresponding actions. Eye movements were coded by a research assistant who was blind to the location of each object using frame-by-frame viewing of the participant's face on SONY digital videotapes. Each change in direction of gaze was coded as towards one of the quadrants, at the center, or missing due to looks away from the display or blinking. These missing frames were excluded from analysis; however they only accounted for approximately 3% of all coded frames. This method of eye-tracking has been validated with high inter-coder reliability and significantly correlates with data obtained from previous studies using head mounted eye tracking (Snedeker & Trueswell, 2004).

Results

We examined the proportion of subjects' eye-movements towards the target character over two divisions of time. Our first analysis examined a coarse-grain measure of subjects' use of the quantifier during the five time windows listed in Table 1.

All time windows began and ended 200 ms after the relevant marker in the speech stream to account for the time it would take to program saccadic eye-movements (Eberhard et al., 1995). For each trial, we summed the total number of looks to the target character and gender distracter within each of these intervals and calculated the proportion of looks to the target over looks to both. This score ranged from zero (exclusive looks to the gender distracter) to one (exclusive looks to the target character). Looks to the Object and Irrelevant Distracters were infrequent after onset of the gender cue and were not included in the analysis. Each time window was analyzed with subjects and items ANOVAs with quantifier type (number vs. scalar), quantifier strength (weak vs. strong), and list/item group manipulated between subjects and between items

During the Baseline Phase, the proportion of looks to the target character initially remained around chance across all terms (see figure 2). There was no main effect of quantifier type ($F1(1, 16) = 1.30$, $p > .10$; $F2(1, 15) = 0.75$, $p > .10$), quantifier strength ($F1(1, 16) = 3.13$, $p > .05$; $F2(1, 15) = 2.00$, $p > .10$), or interaction ($F1(1, 16) = 4.14$, $p > .05$; $F2(1, 15) = 3.31$, $p > .05$). This continued through the following Gender Phase, where again there was no main effect of quantifier type ($F1(1, 16) = 1.52$, $p > .10$; $F2(1, 15) = 1.55$, $p > .10$), quantifier strength ($F1(1, 16) = 4.33$, $p > .05$; $F2(1, 15) = 4.13$, $p > .05$), or interaction ($F1(1, 16) = 0.01$, $p > .10$; $F2(1, 15) = 0.01$, $p > .10$).

However during the Quantifier Phase, fixations to the Target Object increased when participants heard *two* (66%), *three* (72%), and *all* (72%) but not when they heard *some* (45%), see Figure 3. During this period, there were main effects of quantifier type ($F1(1, 16) = 5.16$, $p < .05$; $F2(1, 15) = 6.39$, $p < .05$) and quantifier strength ($F1(1, 16) = 16.86$, $p < .01$; $F2(1, 15) = 18.29$, $p < .01$), and also critically a significant interaction between both variables ($F1(1, 16) = 6.58$, $p < .05$; $F2(1, 15) = 5.25$, $p < .05$). This quickly disappeared by the DISAMBIGUATION PHRASE where there was a main effect of quantifier strength ($F1(1, 16) = 15.65$, $p < .01$; $F2(1, 15) = 23.66$, $p < .01$) but not quantifier type ($F1(1, 16) = 3.19$ $p > .05$; $F2(1, 15) = 3.20$, $p > .05$) or interaction ($F1(1, 16) = 0.73$, $p > .10$; $F2(1, 15) = 0.63$, $p > .10$). Finally, during the End Phase, total fixations closed in unsurprisingly on the target leading to no differences across type ($F1(1, 16) = 0.24$, $p > .10$; $F2(1, 15) = 0.07$, $p > .10$), strength ($F1(1, 16) = 0.78$, $p > .10$; $F2(1, 15) = 0.48$, $p > .10$), or interaction ($F1(1, 16) = 0.32$, $p > .10$; $F2(1, 15) = 0.29$, $p > .10$).

Additional analyses of 200 ms intervals confirm the differences in time it took subjects to reliably fixate on the target character across the four terms. Approximately 400 ms following the onset of the quantifier, the proportion of looks to the target on the *two*, *three*, and *all* trials were significantly greater than chance, $t(19) = 4.77, p < .001$; $t(19) = 4.20, p < .001$; $t(19) = 2.82, p < .05$. Preference for the *some* trials were slower and the proportion was not significantly above chance until approximately 800 ms following the onset of the quantifier, $t(19) = 2.24, p < .05$. This pattern of differential fixations across terms led to a significant quantifier type by strength interaction approximately 400 ms after the onset of the quantifiers ($F(1, 16) = 6.78, p < .05$). During this period, there were also main effects of quantifier type ($F(1, 16) = 10.80, p < .01$) and strength ($F(1, 16) = 8.93, p < .01$).

Discussion

We found that adults were able to integrate semantic and pragmatic interpretations of quantifiers over the course of real-time speech comprehension. However, the speed at which they did so were greatly affected by the term they heard and in particular, we found a reliable preference for the target during the Quantifier Phase for *two*, *three*, and *all* trials but a delay for the *some* trials. This suggests that the referent is rapidly disambiguated when the upper-boundary of the term is semantically specified. In the case of a weak scalar quantifier, the lower-bounded semantics initially left the referent ambiguous. However, later in the ambiguous period, the preference for the target on the *some* trials was reliably above chance. This indicates that participants were not relying solely on the disambiguation from the final phoneme to find the target but were in fact arriving at an upper-bounded interpretation by generating the scalar implicature during the course of real-time processing.

In Experiment 2, we investigated early semantic and pragmatic processing by performing a parallel experiment on children. Recent research on the development of on-line comprehension has taken advantage of the eye-tracking paradigm's ability to provide an implicit measure of children's interpretation rather than requiring explicit judgments over utterances. If children's acquisition of words and structures are initially guided by the understanding of speaker's intent (Tomasello, 1998), we might expect that they would be more inclined to interpret words pragmatically or might initially misinterpret the upper-bound as part of the word's meaning. In contrast, studies using explicit judgment tasks suggest that children are more literal than adults (Noveck, 2001; Papafragou & Musolino, 2003).

Experiment 2

Methods

Participants

Twenty-four five-year-olds (mean age 5;6) were recruited from the database of the Laboratory for Developmental Studies at Harvard University. All children were native monolingual English speakers.

Procedures, Materials, and Coding

These components were identical to Experiment 1.

Results

Children's performance was analyzed using the same procedure as Experiment 1. During the Baseline Phase, there was a strong bias to look at cards with greater quantity, see Figure 4. This led to a main effect of quantifier strength ($F1(1, 20) = 25.19$, $p < .01$; $F2(1, 15) = 13.85$, $p < .01$) despite no effect of type ($F1(1, 20) = 0.84$, $p > .10$; $F2(1, 15) = 0.96$, $p > .10$) or interaction ($F1(1, 20) = 0.21$, $p > .10$; $F2(1, 15) = 0.12$, $p > .10$). This continued through the following

Gender Phase, where again there was a main effect of quantifier strength ($F1(1, 20) = 10.66$, $p < .01$; $F2(1, 15) = 8.63$, $p < .05$) but no effect of type ($F1(1, 20) = 3.55$, $p > .05$; $F2(1, 15) = 3.26$, $p > .05$) or interaction ($F1(1, 20) = 0.75$, $p > .10$; $F2(1, 15) = 0.18$, $p > .10$). This did not change during the Quantifier Phase, leading to a main effect of quantifier strength ($F1(1, 20) = 18.25$, $p < .01$; $F2(1, 15) = 9.15$, $p < .01$) but no effect of type ($F1(1, 20) = 3.37$, $p > .05$; $F2(1, 15) = 4.04$, $p > .05$) or interaction ($F1(1, 20) = 2.42$, $p > .10$; $F2(1, 15) = 0.61$, $p > .10$).

Finally, during the DISAMBIGUATION PHRASE, fixations to the Target Object increased for *two* (60%), *three* (73%), and *all* (74%) but not when they heard *some* (36%), see Figure 5. This resulted in a main effects of quantifier type ($F1(1, 20) = 7.31$, $p < .05$; $F2(1, 15) = 2.73$, $p > .10$) and quantifier strength ($F1(1, 20) = 28.64$, $p < .01$; $F2(1, 15) = 21.79$, $p < .001$), and also critically a significant interaction between both variables ($F1(1, 20) = 10.33$, $p < .01$; $F2(1, 15) = 3.89$, $p > .05$). This quickly disappeared by the End Phase where total fixations closed in on the target leading to no differences across type ($F1(1, 20) = 1.74$, $p > .10$; $F2(1, 15) = 2.15$, $p > .10$), strength ($F1(1, 20) = 3.08$, $p > .05$; $F2(1, 15) = 6.38$, $p < .05$), or interaction ($F1(1, 20) = 0.60$, $p > .10$; $F2(1, 15) = 0.40$, $p > .10$).

Additional analyses confirm that approximately 800 ms following the onset of the quantifier, differential fixations across terms led to a significant quantifier type by strength interaction ($F(1, 20) = 11.30$, $p < .01$). During this period, there were also main effects of quantifier type ($F(1, 20) = 8.00$, $p < .01$) and strength ($F1(1, 20) = 34.08$, $p < .001$).

Discussion

We found that children, like adults, demonstrated earlier disambiguation for *two*, *three* , and *all* and a later disambiguation for *some*. The fact that lexically upper-bounded quantifiers patterned differently from a pragmatically specified one suggests that children, like adults, privilege initial semantic analysis of utterances. However, this interaction occurred at a later time window relative to adults and likely reflects children's difficulty in overcoming initial bias to fixate on cards with greater quantities. All together, these results suggest that children, perhaps more so than adults, rely heavily on the logical meaning when interpreting utterances (Noveck, 2001; Papafragou & Musolino, 2003). Further investigation will be needed to explore whether children ever perform these pragmatic inferences over the course of real-time speech comprehension.

General Discussion

This study explores the real-time interaction between semantic and pragmatic meaning by investigating interpretations of scalar terms. In Experiment 1 and 2, we found that semantic meaning is activated prior to inferential procedures but that these pragmatic inferences occur quickly over the course of on-line speech comprehension. These findings add to a growing literature demonstrating delays of pragmatic interpretations of scalar quantifiers relative to semantic ones (Bott & Noveck, 2004; Katsos, et al., 2005). However, we extend this work by demonstrating that information provided by distinct systems (as defined by work in theoretical linguistics) becomes available at different times during processing.

There is an apparent tension between our results and previous findings demonstrating rapid assimilation of extra-linguistic cues. Sedivy et al. (1999) demonstrated that participants who heard "*Pick up the tall cup*" identified the correct target faster in the presence of an adjectival contrast (e.g. short cup), suggesting that incremental semantic interpretations make almost immediate use of contextual information. Our results suggest that despite the speed at which this occurs, pragmatic interpretation is preceded by some degree of semantic interpretation. Ultimately, both studies are consistent with a model of language that is characterized by representational modularity in a massively interactive system (Trueswell & Tanenhaus, 1994). Consequently, while processing at one level does not need to be completed before processing at the next level begins, the fact that semantic representations mediate between phonological form and pragmatic interpretation, requires that they have some priority during interpretation.

Finally, our data also bear upon the current debates in linguistic theory on the mechanisms underlying generation of pragmatics interpretations (Bott & Noveck, 2004; Katsos et al., 2005; Feeney et al., 2004). One hypothesis is that pragmatic interpretations, like scalar implicatures, generally occur by default unless otherwise cancelled by the context (Levinson, 2000). Others argue that all pragmatic interpretations, including implicatures, are constructed with reference to the global situation, rejecting notions of default inferences (Sperber & Wilson, 1986). While our study was not specifically designed to distinguish these two accounts, features of our data are compatible with predictions that each might make about language processing. The quickness of scalar implicatures in adults suggests that these inferences may selectively attend to particular informational sources (e.g. linguistic/visual context). Nonetheless, the reliable delay suggests that the automatcity of even the most robust of pragmatic inferences requires some initial processing of lexical semantics.

In conclusion, the results of our study demonstrate a temporal lag between semantic and pragmatic processeing. Using the visual-world eye tracking, we found that listeners quickly restricted interpretations to the correct referent when utterances involved semantically unambiguous terms. However, when presented with a lexically lower-bounded quantifier like *some*, participants' initial interpretations failed to differentiate between quantities that include (POSSIBLY ALL) and exclude the total set (NOT ALL). However, we also found that adults in these cases quickly generated pragmatic implicatures prior to the phonological disambiguation at the completion of the phrase. All together these results provide evidence for a model of real-time processing where semantics of quantifiers are activated prior to inferential procedures but that these pragmatic inferences occur quickly over the course of real-time speech comprehension.

References

Bott, L. & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*. 51, 437-457.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research*, 24, 409-436.

Feeney, A., Scafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inferences by children and adults. *Canadian Journal of Experimental Psychology*, 58, 121-132.

Frisson, S. & Pickering, M. J. (1999). The processing of Metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 1366-138

Katsos, N., Breheny, R., & Williams, J. (2005). Interaction of structural and contextual constraints during the on-line generation of scalar inferences. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*.

Gadzar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. New York: Academic Press.

Horn, L. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

Huang, Y., Snedeker, J., Spelke, E. (submitted). What exactly do number words mean?

Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.

Noveck, I. A. (2001). When children are more logical than adults: experimental investigation of scalar implicatures. *Cognition*, 78, 165-188.

Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, 86, 253-282.

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

Snedeker, J. & Trueswell, J. C. (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.

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Tomasello, M. (1998). *The new psychology of language: Cognitive and functional approaches to language structure.* Mahwah, NJ: Erlbaum.

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

Table 1: Time Windows used in analysis

Phase	Period within instructions
1. Baseline	"POINT TO THE. . ."
2. Gender	"GIRL THAT HAS. . ."
3. Quantifier	"TWO/SOME/THREE/ALL OF THE SOC–"
4. Disambiguation	"–KS."
5. End	TIME TO CARD SELECTION

Figure 1: Visual display for *some* trials

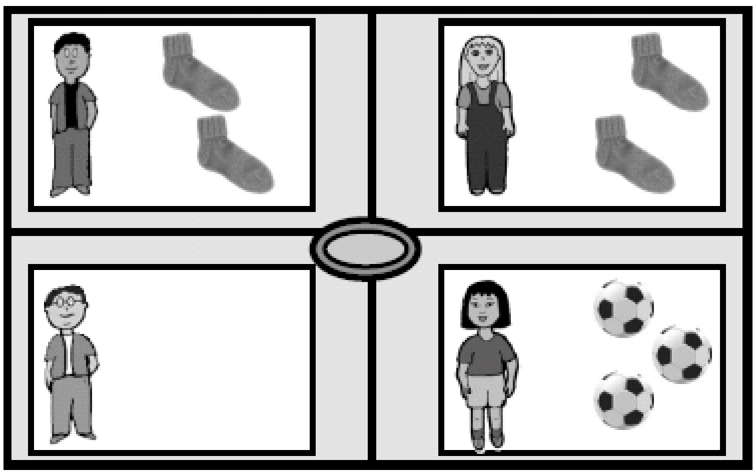

Figure 2: In Experiment 1, the proportion of looks to target during each time window (adults).

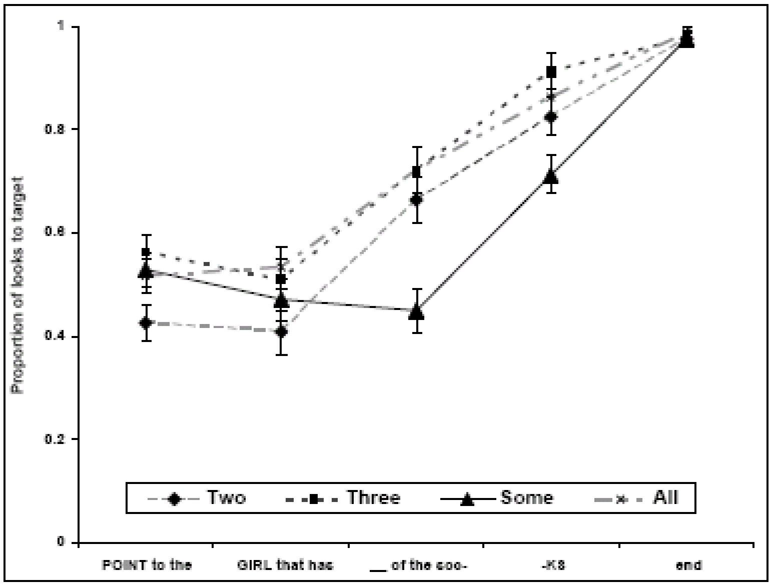

Figure 3: In Experiment 1, the proportion of looks to target during the Quantifier Phase (adults).

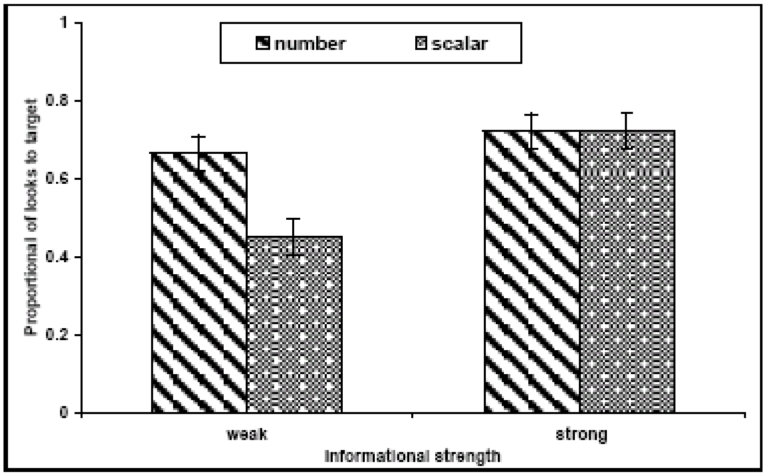

Figure 4: In Experiment 2, the proportion of looks to target during each time window (children).

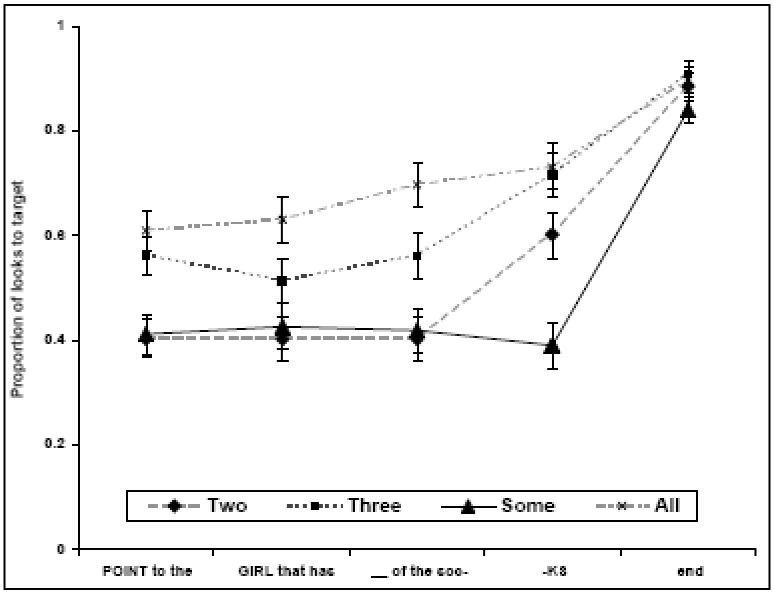

Figure 5: In Experiment 2, the proportion of looks to target during the Disambiguation Phase (children).