Huang & Snedeker (2009)

Mike Frank
Psych/Ling 236
4/16/12
General Rationale

1. Are generalized contextual inferences
   – E.g. scalar implicatures like “some” -> “some but not all”
     computed automatically?

2. Is there a processing cost for going beyond the literal meaning of utterances?
   (Implying a semantics/pragmatics distinction reflected in processing)
   Or are implicatures processed automatically (cf. Levinson, 2000)
Previous work on a sem/prag border

• Evidence from acquisition, e.g.
  – Noveck (2001): children accept “x might be true” as a description when “x must be true”
  – Papafragou & Musolino (2003): children accept “some of the horses jumped over the fence” when all did.

• Evidence from processing, e.g.
  – Bott & Noveck (2004): judgments slower for “some but not all” than “some and all”
  – Breheny et al. (2006): reading time for “some” higher when it is bounded by contexts
Rationale for current study

• Previous studies rely on context manipulations that are hard to match because they include linguistic material

• Eye-tracking methods improve on RT
  – Provide a more implicit measure of processing
  – Give timecourse information about potential processing slowdowns

• Matched comparison to numerals
Trial types (E1 and E2)

some/two

all/three

“Point to the girl that has some of the socks”
Experiment 1 data

Note: no “some but not all” reading until phonetic material disambiguates!
E1 issues/rationale for E2

• General slowness of “three” contexts might be due to weirdness of “three of the X” when 3 = all of the x
  – Two and three trials: boy with none now has one
  – Quantifier trials: unchanged
In the coarse-grained time windows, there were no reliable effects of Quantifier Scale or Strength during the Baseline and Gender phases (all $F$'s < 4.00, all $p$'s > .05). This changed during the Quantifier phase where fixations to the Target character increased when participants heard two (66%), three (72%), and all (72%) but not when they heard some (45%). During this period, there were both main effects of Quantifier Scale ($F_1(1,16) = 5.16, p < .05; F_2(1,15) = 6.39, p < .05$) and Quantifier Strength ($F_1(1,16) = 16.86, p < .01; F_2(1,15) = 18.29, p < .01$). Critically, there was also the predicted significant interaction between these variables ($F_1(1,16) = 6.58, p < .05; F_2(1,15) = 5.25, p < .05$). Planned comparisons within the levels of Quantifier Strength revealed that looks to the Target in the some trials were significantly lower than in the two trials ($t_1(19) = 3.22, p < .01; t_2(15) = 3.11, p < .01$) but there was no reliable difference between the all and three trials ($t_1(19) = 0.01, p > .50; t_2(15) = 0.04, p > .50$). Comparisons within the Quantifier Scales revealed that there was no difference between two and three trials ($t_1(19) = 1.08, p > .20; t_2(15) = 0.98, p > .30$) but a reliable difference between the some and all trials ($t_1(19) = 3.93, p < .01; t_2(15) = 4.15, p < .01$).

However, unlike Experiment 1, this pattern quickly disappeared after the onset of the final phoneme (see Fig. 6). Fixations to the Target character during the Disambiguation phase increased for all trial types (82% for the two trials, 91% for the three trials, 86% for the all trials, and 71% for the some trials). In this region there was a significant effect of Quantifier Strength ($F_1(1,16) = 15.65, p < .01; F_2(1,15) = 23.66, p < .01$) but no effect of Quantifier Scale ($F_1(1,16) = 3.19, p > .05; F_2(1,15) = 3.20, p > .05$) and no interaction ($F_1(1,16) = 0.73, p > .10; F_2(1,15) = 0.63, p > .10$). Finally, during the End phase, total fixations closed in unsurprisingly on the Target leading to no effect of Quantifier Scale ($F_1(1,16) = 0.24, p > .10; F_2(1,15) = 0.07, p > .10$) and Strength ($F_1(1,16) = 0.78, p > .10; F_2(1,15) = 0.48, p > .10$), and no interaction between them ($F_1(1,16) = 0.32, p > .10; F_2(1,15) = 0.29, p > .10$).

Additional analyses of 200 ms intervals following the quantifier onset confirmed the difference in time it took participants to reliably fixate on the Target character across the four terms. Table 3 displays the proportion of looks to the Target for each quantifier type during each of these time windows.
Switch/stay to the Target were significantly lower during the some trials than during the all, three, and two trials (all $p$'s < .05).

A similar pattern emerged in the Target initial trials (Table 5). The pattern of fixations across the four trial types began to differ about 200 ms after the onset of the quantifiers ($F_{1}(1,16) = 4.59, p < .05; F_{2}(1,15) = 7.73, p < .05$). Following the onset of two, three, and all, participants adhered to their initial looks to the Target, presumably because the Distractor was inconsistent with the semantics of these terms. Planned comparisons again revealed no differences in switches to the Target for the number words (two vs. three) or for the strong terms (three vs. all, all $p$'s > .15). In contrast, following the onset of some, participants often abandoned their initial looks to the Target, suggesting that a scalar implicature was not initially available to restrict looks to the subset. As a result, the proportions of switches off the Target were significantly higher during the some trials than during the all, three, and two trials (all $p$'s < .05).

Thus the analyses of both the Target initial trials and the Non-Target initial trials confirm that while listeners rapidly use the meanings of all, three and two to restrict reference, they do not use the pragmatic upper bound of some to guide their initial interpretation of the utterance.
Experiment 3 rationale

- Alternative explanation: lexical ambiguity
- Perhaps two words:
  - some = some | all
  - some = some & ¬all
- Might compete, producing slowdown
- We should see this competition independent of context

“Point to the girl with some of the socks”

(one-referent)
Experiment 3 data

This changed during the Quantifier phase: when some was contrasted with none fixations to the Target character increased (87%) but when some was contrasted with all looks to Target remained at chance (45%), resulting in a reliable difference between these conditions, $t_{1(19)} = 6.48, p < .001; t_{2(15)} = 7.58, p < .001$. This difference between 1-referent and 2-referent trials persisted throughout the Disambiguation phase (Target preference 95% and 74%, respectively) ($t_{1(19)} = 5.01, p < .001; t_{2(15)} = 3.77, p < .01$) and into the End phase where the difference in looks was significant by subjects ($t_{1(19)} = 2.36, p < .05$) but not by items ($t_{2(15)} = 1.41, p > .10$). However, since the Target fixations in this final period were near ceiling (99% in the 1-referent trials and 96% in the 2-referent trials), there was limited variability. Thus we followed up on the $t$-test with a non-parametric Wilcoxon signed-rank test and confirmed that the difference in Target preference during the End phase was significant by subjects ($W = 45, Z = 2.61, p < .005$) and marginally significant by items ($W = 19, Z = 1.68, p < .10$).

However, two features of this data complicate the interpretation of these analyses. First, there were systematic preferences for particular quantities prior to the quantifier onset. In the 1-referent display, looks to the Target in the some trials (57%) were significantly higher than in the none trials (40%) during the Gender phase, $t_{1(19)} = 2.16, p < .05; t_{2(15)} = 2.16, p < .05$. In contrast, looks in the 2-referent some trials (48%) were no different from looks in the all trials (51%) during the same period, $t_{1(19)} = 1.48, p > .15; t_{2(15)} = 1.05, p > .30$. This pattern suggests that prior to the onset of the quantifier participants preferred to look at characters with items rather than those with nothing.

Second, we found that looks to the Target in the all trials were slow to rise after the onset of the quantifier. Target preference during the Quantifier phase (55%) was significantly lower than in the comparable trials of Experiments 1 (66%) and 2 (72%), $F_{1(2,57)} = 3.68, p < .05; F_{2(2,45)} = 3.27, p < .05$. This could reflect differences in how the sets are construed across the two types of displays. In the none trials, participants heard instructions that quantify over a single set distributed among the four characters while in the all trials, there were two sets (socks and soccer balls) and the critical term only quantified over one of these sets. Perhaps, on some proportion of the all trials, participants were confused by this and attempted to interpret all as referring to the total set of objects. Since neither the Target nor the Distractor character had "all of the things" this should result in no reliable preference for either character. When these trials were averaged with trials in which participants limited the possible domains of quantification to the two basic level sets, we would expect to see a slight preference for the Target which would emerge more slowly across the trial.
The proportion of switches to the Target and (B) Target initial trials: the proportion of switches off the Target.

Fig. 10.

Types began to differ about 400 ms after the onset of the quantifiers (subjects and items variable. OVs with all four trial types as within subject and item variables, and list/item group as a between subjects variable. Each time window was analyzed with one-way ANOVA, and the results are reported in Table 7.

For Experiment 2, we used this method of analysis to factor out initial perceptual biases. However, it has the added advantage of splitting the data into non-target initial and target initial. In Experiment 3, trials were separated based on fixations prior to the onset of the quantifier (A) Non-target initial trials: the pattern of fixations across the four trial types. The analysis revealed that the proportion of switches to the Target was significantly higher in the non-target initial trials compared to the target initial trials (F(3,48) = 5.88, p < .01; some (one-referent) vs. all; none vs. all; some (two-referent) vs. all).

The proportion of switches to the Target and (B) Target initial trials: the proportion of switches off the Target.

Switch/stay

Proportion of switches to Target

Time following the onset of the quantifier (in ms)

Proportion of switches

Time following the onset of the quantifier (in ms)
Discussion

• Processing cost for making a pragmatic inference that goes beyond the semantics
  – Consistent with previous developmental and processing evidence

• Alternative explanations
  – “Verification conditions”: quantifiers hard?
  – Simultaneous lexical ambiguity for some

• As with Breheny’s results, results constrain theories that posit automaticity of implicature
  – (Even if implicatures are sometimes fast...