

Relevance of Polarity for the Online Interpretation of Scalar Terms

Daniele Panizza
University of Trento

Yi Ting Huang
*University of
North Carolina*

Gennaro Chierchia & Jesse Snedeker
Harvard University

1. A Puzzle on the Interpretation of Scalar Quantifiers and Numerals

Numeral determiners (such as *two*, *three* etc.) and scalar quantifiers (such as *some*) are known to have different interpretations. Under one interpretation, *two* and *some* have an upper bounded¹ or strong meaning, where *two* means 'two and not more' (or 'exactly two') and *some* means 'some but not all'. Alternately these words can be interpreted with a lower bounded or weak meaning, where *two* means 'at least two' (or 'two or more') and *some* means 'at least some' (or 'some and possibly all'). For example, consider the following dialogue in which speaker (A) interrogates speaker (B) about the quantity of cupcakes eaten by Alan.

- (1) (Speaker A) a. How many of the cupcakes did Alan eat?
(Speaker B) b. He ate some (of the cupcakes).
(Speaker B) c. He ate two (of the cupcakes).

(1a,b) are false - or seriously misleading - if uttered in a situation where Alan ate all or three cupcakes. Thus (1) exemplifies a communicative situation in which *some* and *two* are given upper bounded interpretations. Contrast this with the dialog in (2).

- (2) (Speaker A) a. Who ate some/two cupcakes?
(Speaker B) b. Alan did (eat some/two of the cupcakes).

Here, (2b) appears to be true - or at least felicitous - if uttered in a situation where Alan ate all/three cupcakes. The sentence in (2) is an example of a situation where *two* and *some* are likely to be assigned a lower bounded meaning. The key point here is that the same sentence with a different background generates different interpretations of *two* and *some*. How do these interpretations come about?

In the next paragraph we will review some previous accounts that address this old puzzle by attributing scalar implicature to strengthening. Then, we will consider some recent experimental data which raise questions about this account of the interpretation of numeral and scalar quantifiers. Finally, we will introduce an experimental study designed to address these questions and we will discuss the results in light of different linguistic models of the meaning of scalar items.

¹ See Horn (1972), who first introduced this terminology.

2. The Theory of Scalar Implicatures and Interactions with Quantifiers and Numerals

Perhaps the most influential account of these phenomena stems from Grice's seminal work on the pragmatics of human communication and linguistic behavior (Grice, 1967). This has been further developed by Neogricean theorists (Horn, 1972, 2004; Levinson, 2000; Gazdar, 1979, among others). Under this framework the literal (i.e. basic) meaning of *some* is the lower bounded one while the upper bounded reading arises through a Scalar Implicature (henceforth SI). The SI is a pragmatic inference prompted by a particular class of linguistic items which activate alternative items which can be arranged in an ordinal relationship along a scale (scalar items). *Some*, for instance, activates the alternative quantity *all*. The scalar term and its alternatives form an asymmetrical entailment scale (<some, all>) where each element entails (i.e. logically implicates) the elements to the left, and is entailed by elements to the right. This property holds for other scalar domains such as connectives (<or, and>), modals (<may, must>), adverbs (<sometimes, always>) and so on (see Levinson, 2000, for an overview).

The pragmatic process that gives rise to SIs is grounded in the Gricean Maxim of Quantity ("make your contribution as informative as is required, and don't make it more informative than is required"). Here is a sketch of how it works. When a speaker utters a sentence containing a weak element of an asymmetric entailing scale (such as *some* in 3a), the hearer infers that the speaker was not in the position to utter a parallel sentence containing a stronger element of the scale (i.e. *all* in 3b).

- (3) a. Alan ate *some* of the cupcakes
 b. Alan ate *all* of the cupcakes
 c. Alan ate *some* of the cupcakes, but he did *not* eat *all* of them.

Hence from the Maxim of Quantity - and the assumption that the speaker is cooperative - the hearer infers that (3b) does not hold. As a result, the meaning conveyed by the utterance is the conjunction of what has been said (3a) and what is implicated, namely the negation of (3b), which ends up being (3c).

The Neogricean account of scalar implicature can be extended to numerals. On this hypothesis numbers have 'weak' meanings which are subject to the same entailment relations that hold between other scalar items. A numeral like *two* is a weak element of an asymmetric entailment scale, namely, the number scale <1, 2, 3, 4...>, hence it should also generate SIs². According to this proposal, the upper bounded or 'exact' meaning of numerals is attributable to a SI which arises from the same processes that provide the upper bound for other scalar terms such as *some* (compare 3a-c and 4a-c).

² this was first noted by Horn (1972, 1989), who argued, however, in a more recent paper (Horn, 2004) that numerals are ambiguous among three possible readings, namely 'at least N', 'at most N', and 'exactly N'. Cf. also Carston (1998) for a similar proposal.

- (4) a. Alan ate *two* of the cupcakes
- b. Alan ate *three* of the cupcakes
- c. Alan ate *two* of the cupcakes, but he did *not* eat *three* (or more).

3. Polarity and SIs

Curiously, the interpretation of scalar terms is sensitive to the polarity of the context in which they occur. Intuitively, in Downward Entailing (DE) contexts the lower bounded reading of scalar items seems to be more salient, whereas in Upward Entailing contexts (UE) the upper bounded meaning seems to be more readily available. A DE context can be defined as a linguistic environment that licenses inferences from a set to its proper subset. UE contexts, on the other hand, license inferences from sets to supersets³. The following examples illustrate this apparent generalization.

- (5) a. Alan missed *some/two* exercises and he will flunk out.
- b. Alan missed *some/two* exercises but not *all/more* and he will flunk out.
- (6) a. If Alan missed *some/two* exercises, he will flunk out.
- b. If Alan missed *some/two* exercises but not *all/more*, he will flunk out.

Some and *two* in (5a) are embedded in a UE context, and they seem to be interpreted with their upper bounded reading (as paraphrased in 5b). In (6a), on the other hand, they occur in the antecedent of a conditional, which is a DE context, and their meaning is more compatible with the lower bounded reading (namely, even if Alan missed all/three or more exercises he will flunk out), and thus they do not appear to get the interpretation spelled out in (6b). The systematicity of these intuitions have received different explanations under different theories of implicature⁴. Panizza, Chierchia and Clifton (in press) conducted two experiments to test whether the interpretation and processing of numerals is affected by polarity. In the first experiment they presented subjects with sentences such as (7a,b) and asked them to report their intuitions about the meaning of the numeral.

- (7) a. John has *two* cars in the garage and he will park a motorcycle in the courtyard.

³ Examples of DE contexts are the antecedent of conditionals, the restriction of universal quantifiers, the scope of negation or negative predicates, questions etc. (see Chierchia, 2004 for an overview).

⁴ Gazdar (1979), proposed that SIs are blocked under negation, while Horn maintained that under negation weak items (e.g. *some*) become strong (*not some*, which means *no*) and thus they do not generate any SI. Chierchia (2004) put forward an algorithm in which the polarity is checked locally (i.e. within the scope of the scalar trigger) and if the SI, generated locally, does not lead to a stronger statement (strength condition) it gets canceled. This theory has the advantage of predicting that weak scalar items should not generate SIs when they occur under the scope of any DE operator.

- b. If John has *two* cars in the garage, he will park a motorcycle in the courtyard.

Subjects were asked to select one of two options ("the speaker is talking about exactly two cars" or "the speaker is talking about at least two cars"). The results showed that when the numeral was embedded in a DE context (as in 7b) it received an upper bounded reading less often than when it occurred in a UE context (as in 7a, 49% vs. 78% upper bounded interpretations).

Second, in an eye-tracking experiment, subjects were given both these sentences and versions of the same sentences in which the interpretation of the numeral (*two*) was strongly biased to the upper bounded reading by material in the second clause (8a and b).

- (8) a. John has *two* cars in the garage and he will park a third car in the courtyard.
- b. If John has *two* cars in the garage, he will park a third car in the courtyard.

The authors found that, when the continuation was neutral, participants made more regressive eye movements toward the numeral (*two*) embedded in a UE context (7a) than in a DE one (7b). However, when the continuation was biased to an exact, upper bounded reading, the numeral in a DE context (8b) received more regressions than one embedded in a UE context (8a).

This gave rise to a significant interaction in the second-pass reading times between *polarity* (UE vs. DE) and *type of continuation* (neutral vs. biased). This interaction was explained in the following terms. When participants were reading the first clause, they were more likely to give the numeral an upper bounded interpretation in UE contexts than in DE contexts. When the continuation was neutral there was no reason to revise this interpretation. However, when the continuation presupposed that the numeral had an exact, upper bounded interpretation (as in 8a,b), this conflicted with the lower-bounded interpretation favored by the DE contexts (as in 8b). As a result, subjects often had to return to the first clause to re-read and reanalyze the numeral. Thus the findings from these two experiments demonstrate that the polarity affects both the off-line interpretation and on-line processing of numerals, in a fashion which suggests that they are subject to scalar strengthening.

4. Divergences between Scalar Quantifiers and Numerals

In the previous paragraphs we have seen that there are compelling reasons for thinking that scalar quantifiers like *some* and numerals are subject to the same mechanism of scalar strengthening. However, there is also evidence both in the linguistic and in the psycholinguistic literature that *some* and numerals behave differently with respect to their interpretation. Horn (1989), for instance, pointed

out that numerals display a different behavior when embedded under negation. Let us consider the following sentences.

- (9) John didn't grade **SOME** of his assignments. He graded **ALL** of them.
(10) John didn't meet Mary **OR** Sue. He met both of them.

In the examples above the strong meaning of *some* and *or* - which is argued to be strengthened towards the exclusive reading (A or B but not both) through SI - is explicitly denied, and the continuation asserts that the truth of the stronger item of the scale holds (namely *all* and *and* are the stronger scale-mates of *some* and *or* respectively). This phenomenon, known as *implicature cancelation*, has been argued to involve a special type of negation called metalinguistic negation (Horn, 1989). In fact in (9) and (10), *some* and *or* are typed in boldface to mark the focal phonological contour accompanying this words when their corresponding SI is canceled under negation. If these sentences are produced with no phonological marking of the weak scalar term, they are ungrammatical or strongly deviant. In contrast, the upper bounded meaning of a numeral is denied more easily and does not require focal stress (11).

- (11) John did not grade three papers. He graded five.

Another piece of evidence that numerals and *some* achieve their upper bound in different ways comes from a recent experimental study by Huang and Snedeker (2009) using the visual-world paradigm. Subjects were presented with scenarios in which a pair of boys and a pair of girls received two different sets of objects (e.g. socks and soccer balls). Participants listened to instructions which asked them to select one of the characters while their eye movements to the display were monitored. The critical trials displayed a girl who had two out of the four socks that were present in the scenario, and another girl with three out of the three soccerballs. Participants were told to "Pick up the girl with some/two of the socks". Notice that there is a temporary ambiguity at the onset of 'some': on its lower bounded interpretation (i.e. 'some and maybe all') the instruction could potentially refer to either, the girl with a subset of socks or the girl with the total set of soccer balls. Similarly, if numbers are truly scalar terms, this ambiguity would also be present for the *two* trials; both the girl with two socks and the girl with three soccer balls have 'at least two' objects and hence, in the absence of a scalar implicature, both are potential referents. In addition, Huang and Snedeker included sentences in which the referent was disambiguated by the semantic lower bound of the numeral or scalar quantifier (i.e. "Pick up the girl that has three/all of the soccer balls"). Under these circumstances, participants rapidly used the meaning of the quantifier to disambiguate reference, developing a preference for the correct character 200-400ms after quantifier onset. In contrast, when they heard the sentences with *some*, subjects did not shift their gaze to the correct referent until about 800 ms after the quantifier (at roughly the same time the noun was disambiguated). Critically, on the two trials, disambiguation was as rapid as it was for *all* and *three*, suggesting that the upper bound for numbers was

available as quickly as their semantically encoded lower bound. The authors interpreted the results in terms of a fast, immediate, access to the 'exact' reading of *two*, in contrast with a slower, delayed, access to the upper bounded meaning of *some*. Numerals are thereby argued to have an 'exact' semantics, while *some* receives its upper bounded meaning through a SI, which is costly in term of processing.

Note that the linguistic and experimental facts we have just reviewed pose several new puzzles about the meaning of scalar quantifiers and numerals. First, if numerals have an 'exact' semantics, how is their 'at least' interpretation derived? Second, how do we explain the apparent similarity of behavior of numerals and scalar quantifiers in UE vs. DE contexts? Finally, how can we reconcile the experimental findings by Panizza et al. (in press) vs. Huang and Snedeker (2009)? To address these questions we conducted an experimental study on numerals and scalar quantifiers that we will describe in the next paragraph. In this study we employed the visual world paradigm, which can provide precise information about how people's interpretation of words and sentences evolves over time. Our experimental design went beyond the previous studies in two critical respects. First, in contrast with Huang and Snedeker (2009), we manipulated the polarity of the context containing the quantifiers. Second, in contrast with Panizza et al. (in press), we tested scalar quantifiers in addition to numerals, so we could directly compare them.

5. An Experimental Study on Scalar Quantifiers and Numerals.

5.1 Procedure

The aim of this study is to test the off-line interpretation and on-line processing of numerals and scalar quantifiers with respect to the polarity of the context in which they are embedded. To do this, we employed the visual world paradigm, where subjects hear sentences while they are attending to a visual scenario and their eye movements are recorded. In our experiment, which builds on Huang and Snedeker (2009), we tested 64 undergraduate students in the following procedure. Subjects were told that they would be playing a game involving a cook and a teacher. The cook was looking for a child to help in the kitchen. On each trial, the teacher (who did not appear in the visual scenario) would tell a story about the four children. Then the cook, who could not hear the story, would ask the participant to help find one of the children. If the subject could help the cook find the boy or the girl that he was looking for, then s/he should do so. But if there was no such person, then the subject could reject the sentence by clicking on a marked box.

5.2 Materials and Design

Half of the subjects were presented with 32 sentences in which the quantifier (either a numeral or a scalar quantifier) appeared in a UE context. These were declarative clauses (such as 12). The other half of the subjects heard 32 sentences

in which the quantifier appeared in the antecedent of a conditional (13) and hence appeared in a DE context. Thus the *polarity* of the embedding context was manipulated between-subjects.

- (12) I know what's going on.
A boy has some/all/two/three of the paperclips. Point to him.
- (13) I don't know what's going on but..
If a boy has some/all/three of the paperclips, then point to him.

Notice that UE and DE conditions differ minimally from each other. In the former ones the Cook thinks he knows what's going on while in the latter conditions he does not. The only further difference in the experimental sentences is the presence of "if" in the DE conditions, at the beginning of the sentence, and the presence of "then" at the beginning of the second one. All the other words, as well as all the scenarios, were exactly the same. The *type* (scalar vs. numeral, e.g., *some* vs. *two*) and the *strength* (weak vs. strong, e.g., *some* vs. *all*) of the quantifier was manipulated within-subjects, which means that each subject heard 8 sentences for each of the four quantifiers that we used (*two*, *three*, *some* and *all*).

As anticipated earlier, we are primarily interested in three questions: (1) what are the mechanisms by which upper bounded interpretations emerge over time, (2) are these mechanisms different for numeral and scalar quantifiers, and (3) how these mechanisms are influenced by the polarity of the context. To pursue this goal, half of our trials were ones in which the visual context provided a good referent for upper bounded interpretation of the quantifier and this object was ultimately the one referred to. These scenarios are similar to those used by Huang and Snedeker (2009). In this study, we will call these conditions the *go* trials. Here participants are always expected to select the same referent (to "go"); the only question is how quickly they are able to do this.

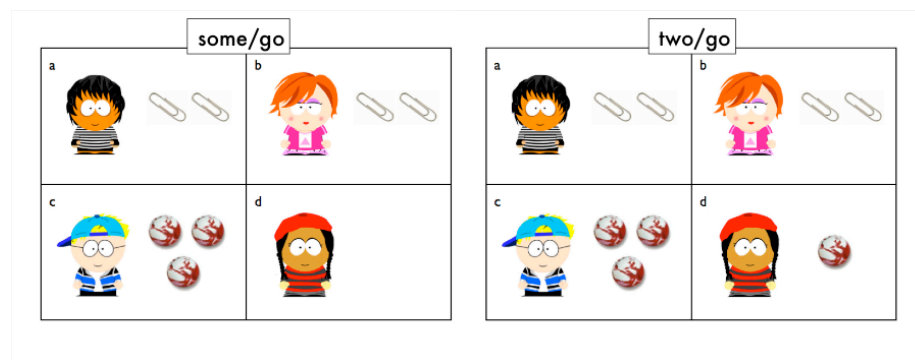


Figure 1. Displays used for the *some* and *two* go trials.

For example, after hearing the story while looking at the scene in Fig. 1, the subjects were asked to point to the boy that has some/two of the paperclips (see examples 12 and 13), who is the dark-haired boy in the top-left quadrant. Critically, however, in the *some* and *two* conditions unambiguous identification of the target is not possible until either the upper bound is accessed or the subjects hear the disambiguation of the noun (i.e. "...clips"). Thus a reliable preference for

the target boy (relative to the other boy) after the quantifier (some/two of the paper...) but *before* the disambiguation point (...clips) would suggest calculation or retrieval of the upper bound. In contrast in the *all* and *three* conditions, the referent is semantically disambiguated by the quantifier (e.g. in Fig.1, participants would be asked to point at the boy with all/three of the paperweights). These trials demonstrate how quickly semantic information can be used to restrict reference in this task and serve as a useful point of reference for interpreting behavior in the *some* and *two* conditions .

The other half of the trials constituted the *no-go* conditions. In these trials, the scenario did not include a referent that was consistent with the strengthened or upper bounded interpretation of the quantifier. In the case of the strong quantifier conditions (sentences with *all* and *three*), there was simply no person who matched the description. In the case of the weak quantifier conditions (sentences with *some* and *two*), there was a referent that matched the description under its lower bounded interpretation. For example in Fig. 2 subjects were asked to look for the boy with some/two paperclips. If they interpreted the quantifier as having no upper bound, they could choose the dark-haired boy in the top-left corner, who has all/three paperclips. However, if they had adopted an upper bounded interpretation of *some* and *two* (i.e. 'some but not all' and 'two but not more', respectively) they would have to reject the statement.

In the UE conditions, the rejection option was "the cook is wrong" while in the DE conditions it was "there is none". Thus the *all* and *three* trials in both the *go/no-go* conditions and UE/DE conditions, allowed us to verify that subjects understood the experimental sentences, were able to use them to guide eye movements and could perform the task correctly.



Figure 2. Displays used for the *some* and *two* no-go trials

5.3 Experimental Predictions

This experiment permitted us to address two issues about the interpretation of scalar quantifiers and numerals. First, the participants' decisions in the *no-go* trials allowed us to assess how polarity affects the ultimate interpretation of utterances containing *some* and *two*. On these trials participants who had a lower bounded interpretation of the utterance would presumably select the character who matched this description but violated the strengthened meaning. In contrast, those who had an upper bounded interpretation of the utterance would be expected to select the rejection box (indicating that there was no referent for the

phrase). If the upper bounded meaning of numerals and scalar quantifiers is controlled by a SI-sensitive mechanism, SIs should be discouraged in DE contexts but encouraged in the UE contexts. This should lead to higher acceptance rate in DE conditions relative to the UE conditions.

Our analyses of online processing focus on the *go* trials where the target matches the utterance both under both the weak and the strong interpretation of the quantifier. Here, we may draw the following predictions. First, irrespective of polarity, if the basic meaning of *two* includes a semantic upper bound and the meaning of *some* does not, then we would expect rapid disambiguation for *two* in the *go* trials, but slower disambiguation for *some*. Second, if the polarity of the context influences the probability of calculating the implicature, then we should see differences between the UE and DE conditions for *some*. Furthermore if the mechanisms underlie polarity effects are the same for numeral and scalar quantifier, then we would expect to see similar differences between UE and DE contexts for both the *some* and *two* trials. For example, since UE contexts are more likely to give rise to strengthened readings, we might expect more rapid disambiguation for both *some* and *two* in the UE contexts, and prolonged competition between referents in the DE contexts.

Recall that Huang and Snedeker found that the upper bound of *two* restricted reference as rapidly as the lower bound of *three* in a UE context. One possible interpretation of these findings is that SIs for numbers are incredibly rapid because the scale is over practiced and highly salient (see Barner & Bachrach, in press). If this is the case, we might expect *two* to pattern differently from strong quantifiers in DE contexts in which SI's are typically absent. In contrast if the rapid disambiguation of *two*, reflects its lexical properties this early disambiguation should persist regardless of polarity.

5.4 Off-line results

Unsurprisingly, in the *go* trials participants' accepted virtually all of the sentences, while in the *no-go* trials for *all* and *three*, they consistently rejected the sentences. These results suggest that participants understood the task, performed it correctly, and were able to understand both the UE and DE sentences.

The *no-go* trials for *some* and *two* were the primary focus of our offline analyses. Here we found the predicted effect of polarity: participants accepted the target more often in the DE conditions than in the UE ones, both for *some* (74% UE vs. 93% DE) and for *two* trials (11% UE vs. 40% DE). Thus the polarity of the context in which the quantifier is embedded influences both *some* and *two*, and does so to roughly the same degree (no significant interaction was found between *polarity* and *type of quantifier*). Nevertheless, *two* was strengthened very often whereas *some* received an upper bounded interpretation far less frequently. This asymmetry is consistent with the hypothesis that there are lexical differences between numerals and scalar quantifiers. For example, *two*, unlike *some*, could have a lexically encoded upper bound (an 'exact' semantics). We explore this possibility further in our analyses of participants' eye movements.

5.5 On-line results (broad time-window analysis)

Our analysis of fixation patterns focused on the *go* trials. For each time window, we computed target preference: the time spent fixating the target (the referent of the description) as a proportion of the total time spent in fixating both the target and the distractor (the person of the same gender as the target). When target preference equals 1, it means that the subject is looking only at the target. When it equals 0, then the subject is only looking at the distractor. The early part of the sentence (e.g. "a boy has...") allows the participants to focus in on target and distractor character but provides no additional information, thus the target proportion should hover near 0.5 in this time window. After quantifier onset, we would expect target proportion to increase sharply for the strong quantifiers. As we noted earlier, for *some* and *two*, the latency before this rise in target preference indexes the amount of time required to interpret the quantifier with an upper bounded meaning.

Analyses were conducted on the target preference scores for two broad time-windows of 1.1 seconds each corresponding to the *quantifier* ("some/two of the paper...") and the *disambiguation* ("...clips"). First, we found that the target preference scores for the quantifier region were well above chance for *three* and *all*, demonstrating that our participants were able to rapidly use the meaning of the quantifiers to restrict reference. Second, target preference scores in the quantifier window were also well above chance for *two*, both in the UE and DE conditions, suggesting that participants had very fast access to the 'exact' or upper bounded meaning of *two*. In contrast, target preference scores for *some* remained at chance in this time window, rising only after participants heard the disambiguating noun. This suggests that subjects did not compute a SI to restrict reference prior to the disambiguation. These results replicate what was found by Huang and Snedeker (2009) and suggest that *some* and *two* diverge strongly with respect to their initial interpretation, with *two* having an immediate upper bounded interpretation.

The effects of *polarity* emerged later, for both the *some* and *two* trials. On the *some* trials, the target proportion during the *disambiguation* window was significantly higher in the DE condition than in the UE one. This result is quite surprising for two reasons. First if the UE context promotes implicature, then we would have expected more growth in target preference in this context, not less. Second, in both UE and DE contexts the target is unambiguous at this point, thus it is unclear why polarity should continue to have any effect on reference restriction. Interestingly, we found precisely the opposite effect for *two*. In this case, participants had developed a rapid target preference during the *quantifier* window with no difference between the UE and DE conditions. However during the *disambiguation* time window, target preference grew more rapidly in the UE condition than the DE condition. The time course of these effects is explored in more detail in the next section.

5.6 On-line results (fine grained time-windows analysis)

In this analysis we divided the presentation time of the sentence (5 seconds) into 50 regions of 100 ms each. This analysis had two goals. First we wanted to rule out the possibility that there were early differences between the UE and DE contexts that had disappeared when averaged in our broad time windows. Second, we wanted to gain a clearer understanding of the timing of the polarity effects that emerged in the disambiguation window. Fig. 3 graphs the target proportion for *some* and *two* sentences in the UE and DE contexts.

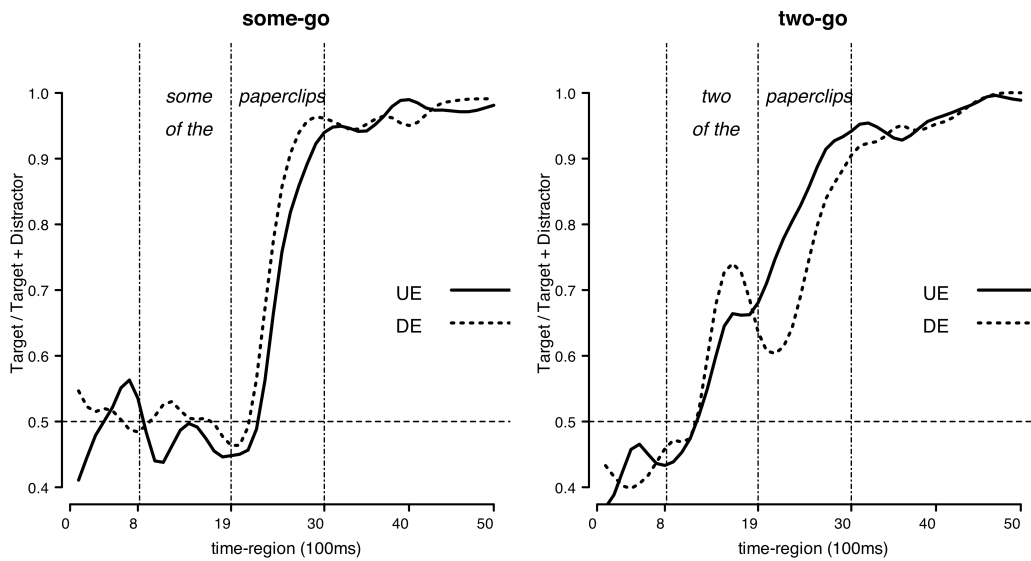


Figure 3: Target Preference in Eye Movements for Some and Two trials

In the *some go* conditions, participants did not disambiguate the target until they got to the disambiguation point and there was no reliable difference between the UE and DE conditions before this point (the differences visible in the graph were not significant). However at the disambiguation point the UE and DE contexts diverge: in the DE contexts the preference for the target (the boy with some of the paperclips) becomes reliable at 400ms after the disambiguation while in the UE condition this preference is delayed by about 200ms.

What could possibly account for this delay? One possibility is that this delay reflects processing costs associated with implicature. Prior studies suggest that SI calculation is often a slow and costly process (Bott and Noveck, 2004). Additional data from Huang and Snedeker (under review) demonstrate that participants in visual world studies may begin calculating the SI for *some* within 1000 ms of quantifier onset. Thus on some trials participants may have been hard at work calculating the implicature at the moment that the disambiguating phonological information arrived, leaving them less able to use this information to disambiguate reference. Because SIs are more common in UE contexts, this should occur more often in the UE condition than in the DE condition resulting in a slower growth in target preference.

In the *two go* trials, on the other hand, the data pattern is quite different. In both UE and DE contexts participants converge on the target shortly after the onset of the quantifier. In the DE trials however there is a sudden decrease in

target preference caused by shifts back to the distractor which begins roughly 800ms after the onset of the quantifier but 300ms before phonological disambiguation. In the UE *two go* trials this pattern is highly attenuated, if not absent.

The polarity effect in the *two go* trials is open to several interpretations. The first and most straightforward possibility is that in the DE context, participants who initially interpreted the number as upper bounded revise this interpretation after they integrate the number into the context and adopt the lower bounded reading which leads them to consider the distractor (i.e. the boy with three paperweights) as a possible referent. On this interpretation, we would conclude that this contextual integration begins about 600-800ms after quantifier onset and well after the 'exact' reading allows participants to map the quantifier to sets in the display. Second, the polarity effect in the DE context could be explained in a way that parallels our explanation for the polarity effect in the *some* condition. Perhaps in the DE contexts, participants are engaged in a costly pragmatic analysis which limits the resources available for identifying the referent of the phrase or making use of the disambiguating phonological information. We discuss this possibility in section 6.2.

Taking stock, these results confirm previous findings on the interpretation of scalar quantifiers and numerals. On the one hand, *some* is different from *two*, in that the former only receives an upper bounded reading after considerable delay and through a process which is apparently effortful. In contrast, *two* is immediately interpreted with an upper-bound that is semantically specified. On the other hand, *some* is similar to *two*, in that they are both affected by the polarity (UE vs. DE) of the context embedding the quantifier.

6. Optimizing Informativeness During Processing. Possible Solutions to the Puzzle.

Let us go back to the questions we raised in Section 4: How is the 'at least' interpretation of numerals derived? How do we explain the apparent similarity of behavior of numerals and scalar quantifiers in UE vs. DE contexts? In this section we describe two theoretical models that might account for the results of the present experiment and provide meaningful answers to those questions.

6.1 The grammatical view of SI

The grammatical account is grounded on a model proposed by Chierchia, Spector and Fox (in press), which draws on previous works by Chierchia (2004) and Fox (2005). According to this view, the basic meaning of numerals is adjectival and exact⁵ (14).

⁵ See also Ionin and Matushansky (2004), who maintain that the meaning of simple cardinals is adjectival, and they provide a compositional model of how they combine to generate complex cardinals. Also Landman (2003) defends the view that numerals are exact at the lexical level.

- (14) Basic meaning of numerals
- a. The basic meaning of numerals is adjectival and exact
 - b. Logical type: $\langle e, t \rangle$ (or $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$)
 - c. $\text{three}_{\text{ADJ}}(x) = \text{card}(x) = 3$ (the cardinality of the atoms in x is 3)
 - d. i. "Those are three cats"
 ii. $\lambda x[\text{cat}(x) \wedge \text{three}(x)](\text{those}) = [\text{cat}(\text{those}) \wedge \text{three}(\text{those})]$

In this first step, the meaning of a number phrase like "three cats" is a set of *cats* that is composed by exactly three atoms. This accounts for the immediate access to the 'exact' interpretation of *two* that was found in our experiment, as well as by Huang and Snedeker (2009).

Notice that in cases like (14d.i), numerals are used with a predicative meaning (i.e. they are interpreted as properties) and only the 'exact' meaning is available. If numerals are used as arguments, like in our experiment (e.g. "a boy has three paperclips"), they are interpreted as quantifiers and like other indefinites, they are subject to existential closure (15).

- (15) Existential closure
- a. Indefinites are subject to existential closure (Heim, 1982; Winter, 1989; Chierchia, 2005)
 - b. "Three cats are purring"
 i. $[\text{cat}(x) \wedge \text{three}(x) \wedge \text{purring}(x)]$
 ii. $\exists x[\text{cat}(x) \wedge \text{three}(x) \wedge \text{purring}(x)]$ ($\exists$ -closure)

During this process, the 'at least' interpretation arises at the clausal level, as in (15b.ii). In fact, if there is a set of *four* cats that are purring in a situation s , it follows that there is also a set of *three* cats that are purring in s , which is what (15b.ii) means. This accounts for how 'at least' readings of numerals are derived, in general, and provides a theoretical explanation for the possible 'at least' reading we found in the DE condition of the *two go* trials (namely, the delay arising 600-800ms after the onset of *two*). Finally, SIs are obtained via exhaustification at Logical Form, where numerals gets again an 'exact' reading (16).

- (16) Scalar Implicatures as exhaustifications
- a. SIs are obtained via an operator present at LF that acts like a silent counterpart to 'only.'
 $\text{Exh}(p) = p \wedge \forall q \text{ ALT}(p) [q \rightarrow p \subseteq q]$
 - b. $\text{Exh} [\exists \text{ three cats are purring}]$
 - c. $\text{Exh} [\exists x[\text{cat}(x) \wedge \text{three}(x) \wedge \text{purring}(x)]]$
 - d. $\exists x[\text{cat}(x) \wedge \text{three}(x) \wedge \text{purring}(x)] \wedge \forall q \text{ ALT}(\exists x[\text{cat}(x) \wedge \text{three}(x) \wedge \text{purring}(x)]) [q \rightarrow \exists x[\text{cat}(x) \wedge \text{three}(x) \wedge \text{purring}(x)] \subseteq q]$
 - f. $\exists x[\text{cat}(x) \wedge \text{two}(x) \wedge \text{purring}(x)] \supseteq \exists x[\text{cat}(x) \wedge \text{three}(x) \wedge \text{purring}(x)] \subseteq \exists x[\text{cat}(x) \wedge \text{four}(x) \wedge \text{purring}(x)]$

- g. (c) = Three cats and not more are purring = only THREE cats are purring

The exhaustivity operator in (16a) operates in the following way. It takes as argument a proposition p (the assertion). Then it says that p is true, and each alternative q to p which is true must be entailed by the assertion p ; hence non entailed alternatives must be false. If we apply this to (15b.ii), as in (16c,d), we end up excluding any alternative stronger than p (such as "four cats are purring"), and we get the upper bounded interpretation spelled out in (16g).

This process is subject to a principle of optimization of informativeness as shown in 17. The effect of Optimize Informativeness (OI) is that exhaustification is carried through preferentially when it leads to a stronger meaning, namely in UE contexts.

- (17) Optimize informativeness
- a. [IP[... A....]] A is a scalar term
 - b. [IPExh [... A....]]
 - c. Every else being equal, choose (b) iff it is stronger than (a).
Exh leads to a stronger sentence in UE contexts.

This means that when we interpret a sentence containing a scalar term, its basic meaning (17a, which is 15b.ii in the previous example, viz. "at least three cats are purring") is compared to the same sentence to which an exhaustification operation has been applied (17b, which is 16d in the previous example, that is, "only three cats are purring"). As a result of this comparison, the stronger meaning should be selected - or, at least, favored - against the weaker one. This step explains why upper bounded readings of numerals are generated more often in UE contexts, compared to minimally different DE ones (this study and Panizza et al., in press).

One feature of our data is unexpected given this hypothesis. On the face of it, this account predicts that there will be an early phase in which participants interpret numbers as exact, then a second phase in which the at least reading is generated in both UE and DE contexts (after existential closure), followed by a third phase in which exhaustification occurs giving rise to the differences between UE and DE contexts. We clearly find evidence for the first phase (the rapid disambiguation for *two*) and we see evidence that polarity has the predicted effects on offline judgments and online processing. However our data provide no evidence for the second stage in which at least readings are generated (and equally so) in both UE and DE contexts. The most straightforward explanation for this fact under the present hypothesis is $\exists$ -closure is occurs quite late but is very fast. As a result, the time lag between the insertion of $\exists$ and the subsequent insertion of Exh is just too short to leave a trace in how the relevant sentences are processed.

6.2 The D(omain)-Widening approach to numerals

The Domain-widening account⁶ comes in two parts. The analysis of *some* is the same as in the previous approach: its strengthening is due to a scalar implicature. For explicitness sake, we assume that SIs arise as exhaustifications as described in section 6.1. Numerals, however, work differently. Like before, they start out with an exact adjectival lexical meaning.

- (18) Basic meaning of numerals: same as above (14)
 $\text{three}_{\text{ADJ}} = \lambda x [\text{card}(x) = 3]$

The meaning in (18) is used to build predicative NPs (as in *those are three cats*). However their meaning in argument position (as in *John fed three cats*) is derived in a different manner: by turning numerals into generalized quantifiers that retain their original exact meaning.

- (19) Quantified meaning
 a. $\text{three}_D = \lambda P \lambda Q [\text{three}_{\text{ADJ}} (\cup [P \cap Q]_D)]$
 b. $\text{Three cats are purring} = [\text{three}_{\text{ADJ}} (\cup [\text{CATS} \cap \text{PURR}]_D)]$
 = the cardinality of the sum of the set of things that are cats and purr (in D) is three

What (19b) describes is a situation where given a domain D, the biggest set of things (i.e. the supremum) that are cats and purr has cardinality three. Here the meaning of the constituent quantified by a numeral is still 'exact', for if in the domain there is a group of four individuals that are cats and purr the formula in (19b) ends up being false. Notice that the steps in (18) and (19) accounts for the immediate 'exact' interpretation given to *two* in our experiment.

The 'at least' meaning of numerals, then, is derived through an optional existential closure over the domain variable D as in (20).⁷

- (20) Optional domain widening
 a. $[\dots n_D \dots] \rightarrow \exists_D [\dots n_D \dots]$
 b. $\exists_D [\text{three}_D \text{ cats are purring}]$
 c. $\exists_D [\text{three}_{\text{ADJ}} (\cup [\text{CATS} \cap \text{PURR}]_D)]$
 = there is some domain D such that the set of things that are cats and purr in D is three.
 d. (19b) entails (20c)

(20c) says that there is *at least* a domain satisfying the formula in (19b). From this follow that if we consider different domains containing different groups formed by three purring cats, we will wind up taking into account more than three cats

⁶ Our proposal is directly inspired by Breheny (2007).

⁷ It should be noted that on the current presentation of the domain widening approach this operation is part of grammar: we are constructing two different logical forms for the sentence in question. Pragmatics enters into determining which of the available representations is chosen on a given occasion.

and thus the 'at least' reading arises. Arguably, according to this model, the domain widening would induce the generation of the 'at least' reading that caused our subjects to reconsider the "boy with three objects" when asked to point to the boy with "two objects" in DE contexts.

Finally, like in the previous proposal, this process is subject to a principle of optimizing informativity (21), whereby the existential operator leads to a stronger statement in DE contexts.

- (21) Optimize informativeness
- a. [IP[... A....]] A is a scalar term
 - b. [IP $\exists$ [... A....]]
 - c. Everything else being equal, choose (b) iff it is stronger than (a): $\exists$ leads to a stronger statement in DE contexts

Notice that, so far, both the D-widening and the exhaustivity-based approaches account for most of the results of our experiment, as well as for Huang and Snedeker (2009), even though they radically differ from each other. However, what they do share is that they rely on an operation that is subject to a principle of optimizing the informativeness. The addition of an exhaustivity operator is favored in UE contexts, where it leads to a stronger meaning, whereas the addition of an existential operator over domains is favored in DE contexts.

Here is how the account for the observed eye-movement patterns in the *go* trials would be accounted for under these two theories. On the D-widening approach, there simply is no level of representation or processing in which the UE utterances receive a lower bounded interpretation. Thus it is unsurprising that we do not find any evidence for one. Moreover, the D-widening approach allows us to explain the polarity effects for *some* and *two* in precisely the same way. In the case of *some*, there is an additional operation in the UE conditions (exhaustification) which is resource intensive and slows down reference resolution and use of the disambiguating phonological information. Conversely, in the case of *two*, there is an additional pragmatic operation in the DE conditions (D-widening) which is also resource intensive and slows down the use of disambiguating information.

Let us compare this with how the explanation would go on the grammatical view of section 6.1. These are the candidate logical forms in the case of *some*:

- (22)
- a. i. **Exh** [A boy has some of the paperclips]... (Exh)
 = A boy has some but not all of the paperclips. Point to him.
 - ii. [A boy has some of the paperclips]... (No Exh)
 = A boy has some and maybe all of the paperclips. Point to him.
 - b. i. If **Exh** [a boy has some of the paperclips] then... (Local Exh)
 = If a boy has some but not all of the paperclips, then point to him.
 - ii. **Exh** If [a boy has some of the paperclips] then... (Global Exh)
 = If a boy has some and maybe all of the paperclips, then point to him.
 - iii. If [a boy has some of the paperclips] then... (No Exh)

= If a boy has some and maybe all of the paperclips, then point to him.

In the UE case, Optimize informativeness leads one to prefer (22a.i) over (22a.ii). In the case of DE contexts, OI leads one to choose (22b.iii), where no operator is present. Clearly (22a.i) should be more costly than (22b.iii) since no operator is present in the latter. Let us contrast this with what happens with numerals.

- (23) a. i. **Exh** [A boy has two of the paperclips]... (Exh)
= A boy has exactly two of the paperclips. Point to him.
ii. [A boy has two of the paperclips]... (No Exh)
= A boy has two or more of the paperclips. Point to him.
b. i. If **Exh** [a boy has two of the paperclips] then... (Local Exh)
= If a boy has exactly two of the paperclips, then point to him.
ii. **Exh** If [a boy has two of the paperclips] then... (Global Exh)
= If a boy has two or more of the paperclips, then point to him. If he has less than two of the paperclips, you don't have to point to him.
iii. If [a boy has two of the paperclips] then... (No Exh)
= If a boy has two or more of the paperclips, then point to him.

Here, in UE contexts, OI leads to prefer (23a.i) over (23a.ii); in DE contexts OI leads to choose (23b.ii) over its competitors, for that is the logically strongest interpretation. Thus we see in comparing UE with DE that in both 'winning' cases we have an occurrence of Exh. But in the DE context, we have two sites at which it can be applied. We conjecture that this choice is the source of the delay in the DE contexts with respect to the UE in the case of numerals.

Obviously, further fine-grained experimentation will be needed to tease these two accounts apart. However notice that the logic of our second account reveals a potential weakness of the D-widening approach. The implicature generated by the LF in (23b.ii) is a so called 'indirect implicature' (cf. Chierchia, 2004). It is unclear how such implicatures would be generated on the D-widening approach. To appreciate the phenomenon in its full generality, consider a further example. The sentence in (24a) may suggest the implicature in (24b).

- (24) a. Everyone who has two cars can help us out
b. Not everyone who has one car can help us out

This is due to the fact that in DE context the 'standard' entailment (n entails $n-1$) is reversed (n entails $n+1$). Under the grammatical view of SI, this fact is readily exploited and the inference in (24) is derived straightforwardly by applying the exhaustivity operator globally, as in (23b.ii) – cf. (25a). In fact what the operator (Exh) does in (25a) is to deny the stronger alternatives in (25b), which entail the assertion (24a). Thus the result of this process (25b) is added to the meaning of the utterance, resulting in a strengthened proposition, that is (25d)

- (25) a. O_{ALT} [everyone who has two cars can help us out]
b. $ALT(31a)$ = everyone who has one car., everyone who has three cars...

- c. everyone who has n car can help us out $\subseteq$
everyone who has $n+1$ cars can help us out
- d. everyone who has two or more cars can help us out and not everyone
who has one car can

The existential operator over domain variables, on the other hand, would not be of any help in cases such as (24). The local addition of the $\exists$ -closure operator over domains would yield the two or more reading; the non addition of the operator would lead to the exact reading. There is no way to obtain the reading in (25d), without introducing further apparatus. And it seems difficult to see how such an apparatus could differ from that of grammatical approach to SI.

7. Conclusions

In conclusion, these experiments provide further evidence that numerals have an 'exact' semantics. Unlike *some* they are typically strengthened in both UE and DE contexts, and their upper bounded meaning is immediately available. However numerals can get a weak, lower bounded ('at least') interpretation, which is more salient, or more often derived, in DE contexts. We discussed two accounts under which the 'at least' meaning of numerals is derived through an existential closure operation of a variable ranging either over pluralities (*grammatical view of SI*) or over the domain of quantification (*D-Widening approach*).

While different in some respects, scalar quantifiers and numerals clearly have parallel properties as well. Both are affected by the polarity of the context in which they are embedded. Polarity affects both the temporal dynamics of processing and the ultimate interpretation of the terms, as attested by the off-line and on-line results we discussed. To explain this fact from a psycholinguistic perspective, we must acknowledge that the parser has to be sensitive to a (economy) Principle such as Optimize Informativeness, which - under both accounts - is geared to maximize logical strength. That is, it favors the derivation of a new meaning only in cases where it leads to a stronger statement.

References

- Barner, D., Bachrach, A. (in press). Inference and exact numerical representation in early language development. *Cognitive Psychology*.
- Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51, 437–457.
- Breheny, R. (2008) A new look at the semantic and pragmatics of numerically quantified noun phrases. *Journal of Semantics*, 25, 93-140.
- Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: applications and*

- implications* (pp. 179–236). Amsterdam: Benjamins.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond*. Oxford: Oxford University Press.
- Chierchia, G., Spector, B., & Fox, D. (2008). The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics, to appear in Portner, P., Maienborn, C., & von Stechow, K. (Eds.), *Handbook of Semantics*, Mouton de Gruyter.
- Fox, D. (2007). Free Choice Disjunction and the Theory of Scalar Implicatures. *Presupposition and Implicature in Compositional Semantics*. U. Sauerland and P. Stateva. New York, Palgrave Macmillan: 71-120.
- Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.
- Grice, H. P. (1967), Logic and Conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics, vol. 3: Speech Acts* (pp. 41-58). New York: Academic Press, 1975. Reprinted in H. P. Grice 1989, (pp. 22-40).
- Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Doctoral dissertation, University of Massachusetts, Amherst
- Horn, L. (2004). Implicature. In L. R. Horn and G. Ward (Eds.), *The Handbook of Pragmatics*. Malden, MA: Blackwell.
- Horn, L. R.: (1989). *A Natural History of Negation*. Chicago, Ill.: University of Chicago Press.
- Huang, Y. & Snedeker, J. (2009). On-line Interpretation of Scalar Quantifiers: Insight into the Semantic-Pragmatics Interface. *Cognitive Psychology*, 58, 376-415.
- Huang, Y., Snedeker, J. and Spelke, E. (under review). What exactly do numbers mean?.
- Ionin, T., Matushansky, O. (2006). The composition of complex cardinals. *Journal of Semantics* 23(4), 315-360.
- Landman, F. (2003). The Syntax and Semantics of Noun Phrases (pp. 211-237). *Linguistics Today* 55. John Benjamins: Amsterdam and Philadelphia. Cambridge, MA: MIT Press.
- Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.
- Panizza, D., Chierchia, G., and Clifton, C. (in press). On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language*.