The Pragmatics of Number

Anna Papafragou (anna4@linc.cis.upenn.edu)
Institute for Research in Cognitive Science, 3401 Walnut Street
Philadelphia, PA 19104 USA

Julien Musolino (musolino@indiana.edu)
Department of Speech and Hearing Sciences, 200 S. Jordan Avenue
Bloomington, IN 47405 USA

Abstract

In terms of their semantic and pragmatic properties, number expressions (*one, two, three*…) have standardly been considered similar to quantifiers (*some, a few, all*). For instance, both kinds of expression form a scale: typically, an assertion containing a weaker member of the scale (*Some/Two of the dwarfs loved Snow White*) can be used to implicate that the stronger term of the scale doesn't apply (*Not all/No more than two of the dwarfs loved Snow White*). We report here results from two experiments with young speakers of Modern Greek which support the opposite conclusion: namely, that number terms and quantifiers behave differently in terms of the scalar inferences they support. We discuss implications of these findings for linguistic theories of the semantics/pragmatics of numerals, as well as for developmental theories of the acquisition of number words.

Introduction

In terms of their semantic and pragmatic properties, number expressions (*one, two, three*…) have standardly been considered as scalar expressions similar to quantifiers (*some, a few, all*). Semantically, both numerals and quantifiers have been assigned an 'at least' meaning (Horn, 1972; Grice, 1989): on this 'minimalist' analysis, *two* means *at least two* and *some* means *some (and possibly all).* Pragmatically, both numerals and quantifiers can be used to give rise to so-called *scalar implicatures.* Such implicatures arise when a speaker uses a weak member of the numerical or quantificational scale in order to implicate that the stronger term of the scale does not hold. For instance, an utterance such as (1) is typically used to implicate (2):

(1) Some/Two of the dwarfs loved Snow White.
(2) Not all/No more than two of the dwarfs loved Snow White.

The derivation of scalar implicatures is generally assumed to follow Gricean lines: for instance, if the speaker knew that the more informative statement with *all* (or a higher numeral) were true and relevant, other things being equal, s/he would have preferred to use it. The fact that s/he didn't offers grounds for assuming that such a more informative statement isn't true.

More recently, several objections have been raised to the view that the scalar semantic/pragmatic profile of numerals is identical to that of quantifiers (Carston, 1985; 1998; Horn, 1992). First, it has been observed that cardinals, but not 'inexact' quantifiers, can feature in contextually induced reversals of scale: in (3), *three* is used to communicate *at most three*:

(3) These houses are big enough for families with three kids.

But it is not possible to use *some* in a similar way (e.g. to implicate *at most some*). Second, number terms are regularly used with an 'exact' interpretation in mathematical statements (*Two plus three makes five*), a fact which is hard to reconcile with an 'at least' semantics for numerals (unless we assume that cardinals are ambiguous). Third, the scalar properties of numerals disappear under incorporation: a four-sided figure has exactly (not at least) four sides. Finally, approximation seems to work differently with numerals: *I have $300* is more likely to receive an 'at least' interpretation than its unrounded counterpart *I have $300.17*. For these and related reasons, it has been proposed that cardinals are, in fact, distinct from other scalar expressions. According to these proposals, numerals do not have an 'at least' semantics upper-bounded by a scalar implicature; rather, they are best analyzed as underspecified among an 'at least', 'exact' and 'at most' reading. Pragmatic considerations then are used to determine which reading is more appropriate in a specific context.

There is by now a vast linguistic literature which attempts to adjudicate between the 'minimalist' proposal and alternative theories for number terms (for reviews, see Carston, 1998; Levinson, 2000). The outcome of this debate is important, since theories of scalar predication are a valuable source of insights about how semantic information and contextual cues co-ordinate with each other and contribute to utterance meaning.

In this paper, we present the results from two developmental studies which compare the semantic-pragmatic properties of both cardinals and quantifiers. Our goal is to use the scalar behavior of numerically modified and quantified phrases in child language to shed light on the theoretical debate surrounding these predicates. To preview our discussion, we find that number terms and quantifiers behave differently in child language in terms of the scalar inferences they support. We take these results to be incompatible with 'minimalist' semantic accounts of numerals. Thus we show that developmental data offer an additional piece of evidence for the different status of numerals and quantifiers within semantic/pragmatic theories.

Our experiments build on previous cross-linguistic studies of the acquisition of scalar predicates (for English, see Chierchia, Crain, Guasti, Gualmini & Meroni, 2001; Gualmini, Crain, Meroni, Chierchia & Guasti, 2001; Musolino & Lidz, in press; for French, see Noveck, 2001). Even though they were not concerned with the pragmatics of number terms, these studies have shown that preschoolers have difficulty with the pragmatics of other scalar expressions such as quantifiers (even though they seem to know the semantics of such quantifiers). In our experiments, we turn to Modern Greek for further evidence. Since the scalar inferences associated with numbers and quantifiers apply universally, we should expect to see cross-linguistic similarities in the acquisition of the pragmatics of scalar predicates.

Experiment 1

Methods

Participants. Participants were a group of 20 Greek-speaking 5-year-olds between the ages of 4;11 and 5;11 (mean 5;3) and a group of 20 adult native speakers of Greek. The children who participated in this study were recruited from daycares in the Athens area. The adults were all undergraduate students at the University of Athens.

Procedure and Materials. In this experiment, we asked children (and adults) to offer pragmatic judgements on sentences containing either the numerical scale <*three, two*> or the quantifier scale <*all, some*>.[1] We used a slightly modified version of the Truth Value Judgment Task (Crain & Thornton, 1998). The TVJT typically involves two experimenters. The first experimenter acts out short stories in front of the subjects using small toys and props. The second experimenter plays the role of a puppet (in this case Minnie) who watches the stories alongside the subjects. At the end of the story, the puppet is asked to say what happened in the story. In our version, instead of asking subjects if the puppet is 'right' or 'wrong' (as in the original TVJT), we then asked whether the puppet 'answered well' (i.e., *Apantise kala;*, 'Did-(she)-answer well?'). This modification was made since we were interested in felicity, not truth. Finally, the subjects were asked to justify their answers by explaining why they thought that Minnie answered well or not.

The children were tested individually in a quiet room away from the class. Adult subjects were shown a videotaped version of the stories witnessed by the children, including the warm-up stories. They were given a score sheet and were instructed to indicate, for each story, whether Minnie had 'answered well' or not. They were also asked to provide a brief justification for their answers.

For each scale, subjects were asked to judge four statements like the ones in (4-5):

(4) Meriki apo tus dinosavrus efagan dedra.
'some of the dinosaurs ate trees'
(5) Dio apo tus dinosavrus efagan dedra.
'two of the dinosaurs ate trees'

In each case, these utterances were used to describe situations which satisfied the truth conditions of utterances containing stronger terms on the respective scales, i.e., *all, three.* The critical stories were identical for both scales. For instance, for both (4) and (5), a group of three dinosaurs went to get something to eat. After contemplating other options, all three dinosaurs ended up eating trees. In this context, assuming an 'at least' semantics for the scalar predicates, both utterances in (4) and (5) express a true but pragmatically infelicitous proposition.

Before the main phase of the experiment, each child received two 'warm-up' stories, one designed to elicit a 'Yes' answer and the other a 'No' answer. Furthermore, in addition to the critical statements, and for each scale, subjects were asked to judge four control statements like the ones in (6-7):

(6) Donald cleaned some of the cars/airplanes.
(7) Donald cleaned two of the cars/airplanes.

The purpose of these controls was to ensure that subjects, and in particular children, could accept or reject the puppet's statements when appropriate and, more importantly, that they could do so when these statements involved felicitous uses of terms like *some* and *two*. For each control statement, the experimenter had a choice between two versions: one that was a correct description of the story and would therefore elicit a 'Yes' response and one that was an incorrect

[1] For ease of exposition, we provide English glosses throughout. The Greek terms are <*tris, dio*> and <*oli, meriki*> respectively.

description of the story and would therefore elicit a 'No' response. The experimenter selected the version of the control statement (correct or incorrect description) based on the child's response of the preceding critical statement. If the child had rejected the puppet's statement on the previous critical trial, the experimenter selected the version of the control statement that would elicit a 'Yes' response, and vice-versa. This step was taken to keep a balance between 'Yes' and 'No' responses.

Subjects (5-year-olds and adults) were randomly assigned to one of two conditions, determined by scale type (i.e., <*all, some*>, <*three, two*>) which gave rise to a 2X2 design with age and scale type as between subject factors and 10 subjects per cell. The age range and mean ages for the 10 children assigned to each scale condition, i.e. <*all, some*> and <*three, two*>, are 5;0 to 5;11 (mean 5;4) and 4;11 to 5;10 (mean, 5;3) respectively. In each condition, subjects received four critical trials and four control trials administered in a pseudo-random order. Within each condition, order of presentation was counterbalanced between subjects.

Results

Beginning with test trials, we found that adult subjects overwhelmingly rejected the puppet's statements in each of the two conditions, i.e. 92.5% of the time in the <*all, some*> and 100% of the time in the <*three, two*> condition. Statistical analysis revealed no reliable difference between these rejection rates ($t(18) = 1.96$, $p = 0.06$). By contrast, we found that while 5-year-olds rejected the puppet's statements in the case of <*three, two*> 65% of the time, they did so reliably less often in the case of <*all, some*> i.e., 12.5% of the time ($t(18) = 3.47$, $p < 0.01$) .The proportions of 'No' responses were entered into an analysis of variance (ANOVA) with two factors: age (5-year-olds vs. adults) and scale type (<*all, some*> vs. <*three, two*>). The analysis revealed a main effect of age ($F(1,36) = 54.41$, $p < 0.0001$), a main effect of scale type ($F(1,36) = 14.81$, $p < 0.001$) and a reliable interaction between age and scale type ($F(1,36) = 8.33$, $p < 0.01$).

Recall that subjects in this study were also asked to provide justifications for their answers. Adults in 98% of the justifications they offered for rejecting a statement made reference to the stronger term of the scale, as expected. That is, they said that the puppet was wrong that some or two of the dinosaurs ate a tree because ALL or THREE of them had eaten a tree. Children's justifications for rejecting a numerically modified statement always invoked the pragmatically more appropriate stronger numeral. The same is true for the few cases in which a quantified statement with *some* was rejected by children.

On control trials, adults gave correct responses 100% of the time in the <*all, some*> condition and 80% of the time in the <*three, two*> condition. No reliable difference was found among these means ($t(18) = 1.92$, $p = 0.07$). On the same items, children gave correct responses 90% of the time for <*all, some*> and 95% of the time for <*three, two*>. Again, no reliable differences among the means were found ($t(18) = .77$, $p = 0.44$).

Discussion

Two main conclusions emerge from the first experiment. First, children are much less likely to compute scalar implicatures than adults. This finding comports well with previous research showing that preschoolers cross-linguistically have difficulties understanding scalar inferences, especially those associated with quantifiers (Chierchia et al., 2001; Gualmini et al. 2001; Musolino & Lidz, in press; Noveck, 2001). Second, and more crucially for present purposes, 5-year-old children are more successful in drawing scalar inferences triggered by numerals than by quantifiers. This finding is even more remarkable given that our critical trials with *some* and *two* used identical scenarios and props. This result is unexpected given standard 'minimalist' semantic accounts, since it points to a difference in status between numerals and inexact quantifiers such as *some*.

On the basis of the available evidence on children's performance with scalar inferences, previous literature has concluded that children are generally incapable of deriving scalar implicatures on-line (Chierchia et al., 2001). It might be tempting to interpret our results (at least for *some*) in a similar way. However, there are alternative hypotheses which are worth pursuing. For instance, it is possible that children's failures are not due to an inability to derive the implicatures but to a misunderstanding of the nature of the task. Perhaps children (unlike sophisticated adult communicators) treat this as a truth-value judgement task. Since no special motivation is provided for drawing the scalar inferences, children may be more willing to let the puppet score an appropriate response if she has simply given a true (albeit infelicitous) description of what happened. This inability to correctly assess the experimental demands may therefore make preschoolers in our study appear less pragmatically savvy than they really are (for similar explanations of children's 'failures', see Shipley, 1979; Gelman & Greeno, 1989). Our aim in designing Experiment 2 was to investigate whether a methodologically improved version of the same task might raise children's overall performance with scalar terms. We were also interested in testing whether such a task would yield a similar asymmetry between cardinals and quantifiers.

Experiment 2

Methods

Participants. 20 Greek-speaking children ranging in age between 5;1 and 6;3 (mean 5;6) participated in this experiment. These children were recruited at daycare centers in the same Athens area as the children used in Experiment 1.

Materials and Procedure. Experiment 2 introduces several modifications to the design of our previous study. First, we included a training phase, in which we presented children with four warm-up stories designed to enhance their awareness of the fact that they were being asked to produce pragmatic judgments. Children were told that the puppet, Minnie, sometimes said 'silly things' and that the purpose of the game was to help Minnie 'say things better'. For example, Minnie would be shown a toy dog which she would describe using the truth-conditionally accurate - but pragmatically infelicitous - statement 'This is a little animal with four legs'. The child would then be asked whether 'Minnie answered well' and whether 'we can say it better'. In case the child failed to correct the puppet and provide a better description, the experimenter eventually corrected Minnie and provided the appropriate description, i.e. 'Minnie didn't say that very well. This is a DOG'.

The second change we introduced concerns the test scenarios. As before, subjects witnessed four test stories in which they were asked to judge statements containing the scalar terms *some* and *two* in situations which satisfied the truth conditions of the stronger terms of the respective scales, i.e. *all* and *three*. However, the stories in Experiment 2 were all based on scenarios in which the main character was involved in a contest or a challenge. The main character's performance therefore became the focal point of the stories and at the end, the puppet was asked to comment on how well the character in question had done, 'How did X do?' (*Pos ta pige o X?).* In one of the stories for example, one of the characters claims that he is very good at throwing hoops around a pole and he challenges Mickey to try and do the same with three hoops. Mickey really concentrates hard and he's able to put all the hoops around the pole. At the end of the story, Minnie is asked "How did Mickey do?" and she answers by saying that "Mickey put some of the hoops around the pole". The idea behind this manipulation is to make clear the demands of the communicative situation: given that Mickey's performance is being directly evaluated, only an answer making reference to *all* the hoops would satisfy the expectations of the hearer.

Children heard four test stories and four control stories administered in a pseudo-random order. As before, order of presentation was counterbalanced between subjects within a single condition. Finally, as in Experiment 1, subjects were randomly assigned to either of the two scale conditions. The age range for the 10 children assigned to the quantifier scale condition was 5;1 to 6;2 (mean 5;6). For the number scale condition, the range was 5;4 to 6;3 (mean 5;7).

Results

As before our dependent measure was children's *Yes/No* responses to the puppet's statements. We found that the manipulations described above led children to reject the puppet's statements much more often than in Experiment 1. Nevertheless, the difference between scales persisted: children answered correctly 52.5% of the time for the <*all, some*> scale, and 90% of the time for the <*three, two*> scale ($t(18) = 2.39$, $p = 0.02$). We compared the rejection rates from Experiment 1 and Experiment 2 by entering them into a 2 (training vs. no training) by 2 (scale type*)* ANOVA. The analysis revealed a main effect of training ($F(1,36) = 8.92$, $p < 0.01$), a main effect of scale type ($F(1,36) = 17.1$, $p < 0.001$) and no reliable interaction between training and scale type ($F(1,36) = 0.47$, $p = 0.49$). On control items, children gave correct responses 85% of the time in the quantifier condition and 95% of the time in the number condition. No reliable differences between these means were found ($t(18) = 0.89$, $p = 0.38$). Finally, the justifications children offered for their rejections in the overwhelming majority of cases (93%) made explicit reference to the stronger term of the scale (just like adults' justifications in Experiment 1).

Discussion

The results from this Experiment show that children's sensitivity to scalar inferences improves dramatically if children are provided with clear contextual cues about the communicative expectations of the task.[2] This is an important and novel finding in itself (see Papafragou & Musolino, 2001, for discussion). For present purposes, a more pertinent finding is that the asymmetry found in the previous experiment persists: in child language,

[2] Children in Experiment 2 are slightly older than those in Experiment 1. However, there are strong reasons to think that the differences in performance are not due to these small age differences. First, in the <*all, some*> case, the age difference is not reliable (mean 5;4 vs. 5;6, $t(18) = 0.985$, $p= .33$) but the difference in performance persists. Second, and more importantly, previous studies have reported children's difficulties with scalar implicatures well beyond the age of 5;0 (and up to the age of 10;0).

numerically modified phrases give rise to scalar inferences much more readily than quantified phrases.

General Discussion

As we mentioned in the introduction, there are several theoretical reasons for considering cardinals as distinct from quantifiers and other scalar terms, and our experimental data seem to confirm this difference. Our studies demonstrate that, in child language, inexact quantifiers such as *some* are assigned a lower-bounded reading ('at least some', or 'some and possibly all') and the associated scalar inferences are ignored in the absence of strong contextual cues. By contrast, preschoolers typically reject lower-bounded interpretations of numerals and are very attentive to the scalar properties of number terms. Since the test environment was identical in both situations, this difference in interpretive preferences points to a difference in the semantic/pragmatic status of numbers and quantifiers. Specifically, it suggests that, while a minimalist semantics may be plausible for quantifiers and other scalar predicates, cardinals may best be analyzed in terms of either an 'exact' or an underspecified semantics.

An interesting observation which arises from our experiments is that children often used the counting routine as a means of formulating their responses. For instance, when Minnie offered *Two of the dinosaurs ate trees*, several children protested by saying *No, one, two, THREE dinosaurs ate trees* (while at the same time pointing to and counting the dinosaurs one by one). Explicit counting of this sort offers a specific and precise way of verifying statements containing number terms (by placing the referents of the corresponding NPs in a one to one correspondence with objects in the world). Counting games may also encourage an 'exact' interpretation of the numerals. Notice that neither of these steps is available for the inexact *some*.

There is additional evidence that the observed asymmetry between quantifiers such as *some* and numerals is related to the difference between discrete and non-discrete (vague) scalar predicates. Papafragou (2002) tested Greek preschoolers' understanding of the scalar modifier *half* (e.g. *The bear built half of the tower*). This modifier resembles numerals in that it is discrete (it denotes a precise partitioning of a quantity into two equal parts). By means of the same methodology as the second experiment reported above, it was found that young Greek learners rejected lower-bounded interpretations of the modifier *half* in contexts which licensed scalar implicatures. For instance, children overwhelmingly rejected the statement *The bear built half of the tower* in cases where a whole tower had been built. This pattern reinforces the conclusion that discrete quantity modifiers (e.g. *half*, numerals) have distinct properties from inexact quantifiers and other scalars (e.g. *some, a few*).

Although our results can be taken as evidence against a minimalist semantics for numerals, they leave open the question of whether number terms in natural language have an exact or an underspecified semantics. It is worth pointing out that, in the considerable developmental literature which looks at children's acquisition of number terms (Carey, 2001; Gelman & Gallistel, 1978; Wynn, 1992; Bloom & Wynn, 1997, among many others), it is usually assumed that children ultimately arrive at an 'exact' semantics for number terms (which is the correct adult meaning). Moreover, according to one influential position, children assign meaning to cardinal expressions in natural language by placing them in a one-to-one correspondence with an innate conceptual 'integer list' (Gelman & Gallistel, 1978). The results reported in this paper, even though not univocally in favor of an 'exact' over an underspecified semantics for numerals, are certainly consistent with these positions.

To conclude: Throughout this paper, we have assumed that aspects of child language can be instructive about the nature of the semantic representations in adults. This position accepts some fundamental continuity in the representational systems of children and adults - here, in the specific domain of number. It thus allows us to bring acquisition data to bear on theoretical debates about the architecture of the semantic-pragmatic system in adults. Even though several interesting questions remain unresolved about both the adult system and its acquisition by young learners, we hope to have shown that an approach which treats numbers as regular scalar predicates alongside quantifiers misses important generalizations about their developmental properties.

Acknowledgments

We wish to thank Lila Gleitman, Henry Gleitman, John Trueswell, and the members of the CHEESE seminar at the University of Pennsylvania for comments and input. Thanks go to the teachers and children at the 3rd and 5th daycares at Vrilissia (Athens), and to Profs. D. Chila-Markopoulou, D. Theofanopoulou-Kontou and S. Hoidas for the experiments conducted at the University of Athens. This research was supported by NSF-STC Grant #SBR-89-20230, and by NIH Grant #1-R01-HD37507-01L.

References

Bloom, P. & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language, 24*, 511-533.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind and Language, 16,* 37-55.

Carston, R. (1990). Quantity maxims and generalized implicature. *UCL Working Papers in Linguistics* 2: 1-31. Reprinted in *Lingua, 96* (1995), 213-244.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications*. Amsterdam: Benjamins.

Chierchia, G., Crain, S., Guasti, M., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. *Proceedings from the 25th Annual BUCLD* (pp. 157-168). Somerville, MA: Cascadilla Press.

Gelman, R. & Gallistel, R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gelman, R. & Greeno, J. (1989). On the nature of competence: Principles for understanding a domain. In L. Resnick (Ed.), *Knowing and learning: Essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.

Grice, P. (1989). *Studies in the ways of words*. Harvard: Harvard University Press.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. (2001). At the semantics/pragmatics interface in child language. *Proceedings of SALT* 11. Ithaca, NY: Cornell University.

Horn, L. (1972). *On the semantic properties of logical operators in English.* Doctoral dissertation, Department of Linguistics, UCLA. Distributed by IULC.

Horn, L. (1992). The said and the unsaid. *Proceedings of SALT* 2 (pp. 163-191). Department of Linguistics, Ohio State University.

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

Musolino, J. & Lidz, J. (in press). Preschool logic: Truth and felicity in the acquisition of quantification. *Proceedings of BUCLD* 26. Somerville, MA: Cascadilla Press.

Noveck, I. (2001). When children are more logical than adults. *Cognition, 78,* 165-188.

Papafragou, A. (2002). Scalar implicatures in language acquisition: Some evidence from Modern Greek. To be presented at the 38th Annual Meeting of the Chicago Linguistics Society, University of Chicago, 25-27 April 2002.

Papafragou, A. & Musolino, J. (2001). *Scalar implicatures: Experiments at the Semantics-Pragmatics Interface*. IRCS Technical Report 01-14. Philadelphia, PA: University of Pennsylvania, Institute for Research in Cognitive Science.

Shipley, E. (1979). The class inclusion task: Question form and distributive comparison. *Journal of Psycholinguistic Research, 8,* 301-331.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24,* 220-251.