DO CHILDREN COMPUTE SOME OR MOST SCALAR IMPLICATURES? – EVIDENCE FROM GERMAN

CORINNA KOCH, PETRA SCHULZ AND NAPOLEON KATSOS

1. Introduction

Although in recent years much research has focused on the acquisition of scalar implicatures (SIs), up to now SIs have not been studied in German-speaking children, and investigations have mostly focussed on positive scales. In contrast to logically necessary inferences like entailments, SIs are pragmatic inferences that are computed based on principles of conversation (Grice, 1989). Consider the following example:

(1) A and B are at the swimming pool.
A: Look, some of the children are in the pool.

From A's statement B will likely infer that 'Not all of the children are in the pool'. However, the logical meaning of a quantifier like *some* should only license that 'Some and maybe all of the children are in the pool' (Gamut, 1991). How does the extra inference come about? It is assumed that speaker and hearer adhere to the cooperative principle and the maxims as first formulated by Grice (1989), i.e. B assumes that A would use the most informative proposition he could offer. If he used a proposition with *some* instead of the stronger proposition with *all*, this must be for good reason and therefore it is implicated that the proposition with *all* is not the case. This inference is called a scalar implicature, because it is based on the observation that quantifiers, numbers, and other expressions form scales of informativity such as <*all, most, some*> (Horn, 1976), where propositions with *all* are more informative than *most*, which in turn are more informative than *some* (with informativity measured in terms of licensing entailments). Based on these scales an implicature will

be generated by the hearer if a proposition with a weak term of a scale is used by the speaker, as long as the proposition with the stronger term is relevant and could have been offered. In other words, that the use of a proposition with *some* in a context where *all* could be known and would have been relevant suggests that the speaker does not have enough evidence to use a stronger expression. In that case, the hearer can infer that the stronger expression does not hold, although the semantics of the weaker term does not exclude the meaning of the stronger term. Thus, if a weak scalar expression such as *some* is used, the two readings in (2) are possible:

(2) Readings of *some*
a. pragmatic reading: 'some, but not all'
b. logical reading: 'some, and possibly all'

Importantly, if a weaker term such as *some* is used in a context where the stronger term does not hold, the pragmatic reading and the logical reading lead to the same result in evaluating the utterance, - it is true. Imagine, however, a context for (1), in which all of the children present are in the pool. In this case A uses a weaker term although a stronger one holds, and thus A's utterance in (1) would be called underinformative and it also conveys an infelicitous SI. Only in this context, where all of the children are in the pool, the pragmatic reading and the logical reading yield different results: Whereas under a logical reading, the utterance is felicitous, under a pragmatic reading it is infelicitous.

With regard to the computation of a SI, the quantifier *most* behaves like *some*. Thus, from A's statement in example (3), B will most likely infer the SI 'Not all of the children are in the pool' rather than the logical reading 'Most and maybe all of the children are in the pool'. Note, however, that unlike *some*, *most* relates to the number of elements present and can be paraphrased as 'more than half' (Gamut, 1991).

(3) A and B are at the swimming pool.
A: Look, most of the children are in the pool.

If a proposition with *some* or *most* and a negation is used, this leads to a reversal of the scale, where *none* entails *most not* and *some not*, just like propositions with *all* entail those with *most* and *some*. If (4a) and (4b) are uttered in a situation where no child is in the pool, they are under-informative. In this case, since A does not use the stronger quantifier *none* or *all not*, this stronger statement has to be denied, resulting in the SI 'It is

not the case that no children are in the pool' which amounts to 'Some children are in the pool'. This is wrong in the given situation, so the statements (4a) and (4b) should be rejected when a SI is derived in this context.

(4) A and B are at the swimming pool. No children are in the pool
a. A: Look, some of the children are not in the pool.
b. A: Look, most of the children are not in the pool.

The present study investigated to what extent five-year-old German-speaking children compute scalar implicatures by studying whether they rejected underinformative utterances in a sentence evaluation task. We were especially interested in whether the computation of SIs differs depending on the quantifiers used and whether negated quantifiers and therefore reversed scales influence the children's ability to compute SIs.

2. Previous Studies

Focusing on SIs triggered by the quantifier *some*, the acquisition of SIs has been studied in children learning English (Noveck, 2001; Katsos, 2009; Verbuk, 2007), French (Pouscoulous et al., 2008), Greek (Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004), and Italian (Guasti et al., 2005). Depending on the paradigm used, it has been found that children up to age 10 do not compute SIs as often as adults, but prefer the logical interpretation of underinformative utterances, even though they seem to know the core meaning of the quantifiers used.

This finding was first reported by Noveck (2001) using a sentence evaluation task to test the computation of SIs in 8- and 10-year-old English-speaking children. The subjects had to evaluate statements against their world knowledge. Underinformative statements such as *Some giraffes have long necks* were rejected in 11 % of the cases in the children-group, compared to 59 % rejections in the adult group.

To investigate why logical interpretations were preferred in underinformative conditions by the children, further studies manipulated the task design and demands (Papfragou & Musolino, 2003; Guasti et al., 2005; Puscoulous et al., 2007). Papafragou and Musolino (2003) added a training session focusing on pragmatic interpretations of utterances and created a more natural test situation by providing a visual context to evaluate the sentences. They used a scenario that was acted out in front of a child and against which the underinformative test sentences had to be evaluated. As a result of this focus on pragmatic interpretations performance

on the computation of SIs triggered by *some* improved from 12.5 % to 52.5 % for the 5-year-old children, although this was still lower than in the adult group (92.5 %). In addition, Papafragou and Musolino showed that the rate of generated SIs depended on the type of scale. With numerals such as <three, two> even at age 5 children computed SIs in 65 % of the underinformative contexts without prior training and in 90 % of the cases following pragmatic training. Guasti et al. (2005) demonstrated that the task demands and the task itself may influence the results. Seven-year-old children computed scalar implicatures adult-like if they evaluated sentences that summarized a movie they had seen before and if they did not have to rely on their world knowledge in judging the summaries. The findings of Pouscoulous et al. (2007) support the view that task demands and a focus on pragmatic aspects are crucial for children in order to generate scalar implicatures. Their results indicate that children's performance on scalar implicatures improved if the number of control and distractor items was reduced.

Negated scales, that were also part of the design of the present study, were first investigated in two different experiments by Pouscoulous et al. (2007). In the first experiment 9-year old children had to evaluate sentences against a scenario of boxes in front of them with plastic animals placed in and around the boxes. Negated underinformative test sentences like *Some elephants are not in the boxes* were used in a context in which there were no elephants in the boxes. A SI would lead to the interpretation that some elephants have to be in the boxes, which should then be rejected. In the second experiment 4- to 7-year old children had to act in correspondence to a wish-sentence, i.e. they had to change the content of the boxes so that they match sentences like *I would like some of the boxes to contain a token.* Critical conditions of the negative scale were those where all boxes were empty and the children had to act with respect to the sentence *I would like some of the boxes to not contain a token.* If the children believe *some not* to be compatible with *none* the boxes can be left unchanged, but if a SI is derived some boxes have to be filled with a token. In addition distractor items were removed and a different French translation of the quantifier *some* was chosen. Due to the different task demands and the age of the subjects the results in the two experiments were different (Exp. 1: 30%, Exp. 2: 59% pragmatic answers in 7 year old children); the influence of negation on the task demands remains unclear.

Building on the previous studies we designed a sentence evaluation task to investigate the following questions: (1) To what extent do 5-year-old German-speaking children compute scalar implicatures? (2) Does the rate of scalar implicatures differ depending on the type of quantifier within

the same scale? And (3) how do negative scales influence the computation of scalar implicatures in German?

3. Experiment

3.1 Method

We developed a sentence evaluation task (Katsos, in prep.[1]) including the quantifiers *einige* 'some', *die meisten* 'most', and their negated counterparts *einige nicht* 'some not', *die meisten nicht* 'most not', as well as *alle* 'all', and *keine* 'none'. During a short practice session subjects were familiarized with the task design and the objects used. In the test phase subjects evaluated 38 pre-recorded utterances, ascribed to an animated cartoon character, Wilma, against a visual context. All test items were presented on a notebook as a slideshow. To make the task more natural and therefore easier for the subjects, the following story background was provided: Wilma would like to improve her German and asks the child if she would be willing to practice with her. After the child agrees, Wilma describes the picture on the slide, and the child's task is to decide whether Wilma's description matches the picture or not. Note that the phrasing of the task was not on truth-value judgments (judgment in terms of true, false), but rather on pragmatic felicity (i.e. felicity judgment, cf. Guasti et al. (2005), Papafragou & Musolino (2003)).

Underinformative statements were tested on two different scales, positive <*all, most, some*> and negative <*none, most not, some not*>. In addition to the experiment a standardized language test was conducted to ensure that the children show typical language development.

3.1.1 Participants

Twenty typically developing monolingual German-speaking preschool children (mean age: 5;3; range: 4;9 to 5;9) participated in the study. The children group consisted of 12 male and 8 female subjects, who all showed above performance in the standardized language test SETK 3-5 (Grimm, 2001). In addition 20 adults (mean age: 24;4; range: 22;2 to 27;8) served as a control group.

[1] The task is based on the quantifier semantics, implicature, and scope task developed by the COST Action A33 WG-5, http://cost.zas.gwz-berlin.de/cost/.

3.1.2 Material

All pre-recorded auditory test sentences were of the form: "Here [QUANTIFIER] [OBJECTS] are in the box", as illustrated in the examples (1) and (2) below.

(1) Hier sind einige Äpfel in der Kiste
Here are some apples in the box
'Some of the apples are in the box.'

(2) Hier sind die meisten Enten in der Kiste
Here are the most ducks in the box
'Most of the ducks are in the box.'

Note that the German translation does not contain partitive constructions, because they are not equally acceptable with all of the German quantifiers used.

Visual stimuli showed a box and five identical objects and Wilma at the lower right corner. The pictures varied in the number of objects in and outside of the box, but the total number of objects was always five (cf. Figure 1). All objects were used only once during the test trials.

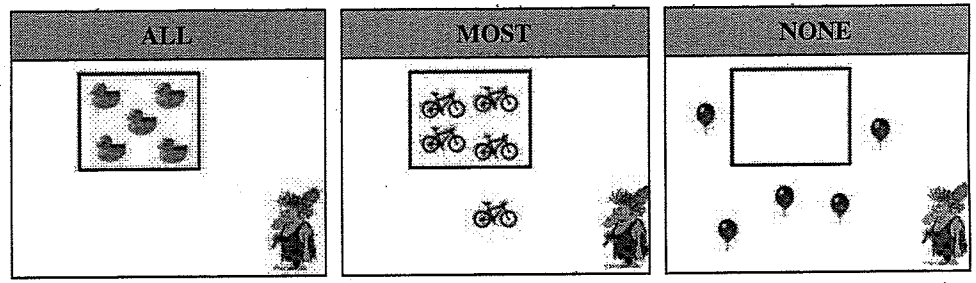

Figure 1: Examples of visual contexts

There were 16 underinformative test items, 4 per quantifier. Underinformative test sentences of the positive scale were created by combining test sentences using *einige* 'some' or *die meisten* 'most' with a so-called visual ALL-background. Underinformative test sentences of the negative scale were created by combining test sentences using *einige nicht* 'some not' or *die meisten nicht* 'most not' with a visual NONE-background. Examples are displayed in Table1.

Test item	Visual context	Pragmatic answer	Logical answer
Positive scale			
Here are some of the X in the box.	ALL in box	no	yes
Here are most of the X in the box.	ALL in box	no	yes
Negative scale			
Here are some of the X not in the box.	NONE in box	no	yes
Here are most of the X not in the box.	NONE in box	no	yes

Table 1: Example items in the underinformative test condition

To assess children's comprehension of the core meaning of the quantifiers used, 22 test sentences were combined with visual contexts that resulted in logically true and false combinations. There were 3 each for *some* and *most* and their negated counterparts *some not* and *most not*, and 5 each for *all* and *none*. Table 2 illustrates the control condition for *most*. Because in these semantic control conditions SIs do not change the truth values, it was not relevant whether subjects generated scalar implicatures in these conditions.

Test item	Visual context	Pragmatic answer	Logical answer
Positive scale			
Here are most of the X in the box	MOST in box	yes	yes
Here are most of the X in the box	NONE in box	no	no
Negative scale			
Here are most of the X not in the box	MOST NOT in box	yes	yes
Here are most of the X not in the box	ALL in box	no	no

Table 2: Example items in the semantic control condition (using most)

In order to control for a yes bias (Wason, 1961), control items had to be answered with No more often than with Yes.

3.1.3 Hypotheses

We assume that pragmatics and specifically informativeness is not dependent on language specific aspects (von Fintel & Matthewson, 2008). Therefore, based on the results from previous studies in other languages we tested the following four hypotheses:

(H1) German-speaking five-year-old children compute fewer scalar implicatures than the adult control group.

(H2) In the semantic control conditions, children and adults do not differ, since the core meaning of the quantifiers used (*some, most, all*) are already acquired by age 5.

(H3) The rate of scalar implicatures computed is higher for the positive than for the negative scales in both groups, since interpretation of negation leads to higher task demands resulting in fewer pragmatic readings.

(H4) Both, in the adult and the children group, the rate of scalar implicatures computed for the quantifiers *some* and *most* does not differ, because they belong to the same scale.

3.2 Results

In the semantic control condition (cf. Table 2) children responded correctly in 95.7 % of the trials. There was a tendency for a higher error rate in negated control sentences, which did not reach significance ($p<.19$). Adults performed at ceiling in the semantic control condition (99.8% correct responses). Comparing children and adults' responses, there were no differences with respect to the different quantifiers *some*, *most*, *all*, *some not*, *most not*, *none* in the control condition.

With respect to the underinformative test condition, significant differences were observed between children and adults. Generally, children derived significantly fewer implicatures than adults (20% vs. 56%, $p<.01$). For the quantifiers individually, children computed significantly fewer SIs than adults for *some* ($p<.05$), *most* ($p<.001$), and *most not* ($p<.01$). Results for *some not* did not reach significance ($p=.14$). Results for individual quantifiers are displayed in Figure 2.

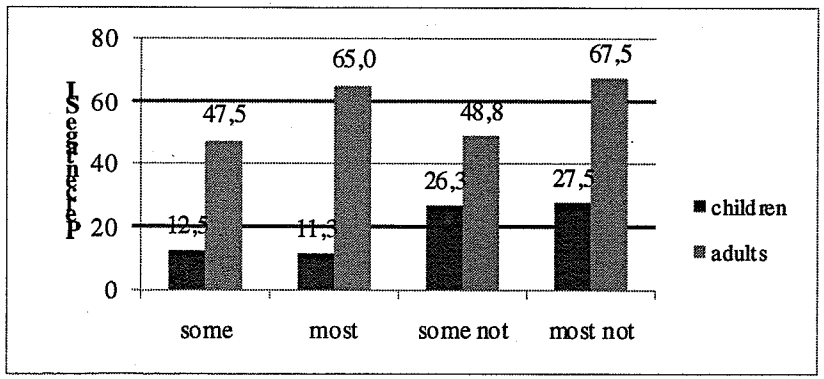

Figure 2: Percentage of SIs in underinformative condition

Both within the children and the adult group, the proportion of scalar implicatures derived for *some* and *most* did not differ significantly (children: $p>.94$, adults: $p>.23$). When comparing positive and negated quantifiers, children derived more SIs for the negated than for the positive scalar quantifiers, but the difference did not reach significance. In the adult group, the percentage of SIs derived for negative and positive quantifiers was the same (cf. Figure 3).

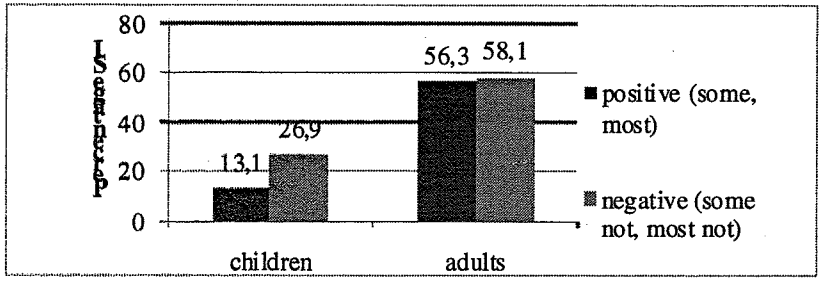

Figure 3: Percentage of SIs in negative and positive scale

A subject-wise analysis of the proportion of scalar implicatures derived in the underinformative test condition revealed a bimodal distribution in both the children and the adult group, i.e. either a subject derives at least ¾ of all scalar implicatures or none. Within the children group, only one child derived scalar implicatures in three or four items with all quantifiers, and 15 children derived only one or none scalar implicature. In the adult group, nine subjects derived three out of four scalar implicatures and at least six derived only one or none. These

distributions differed significantly from chance (Pearson's χ^2, $p<0.01$).

3.3 Discussion

Providing evidence from German this study supports the view that children and adults differ in the extent to which they derive scalar implicatures. Performance on the semantic control condition showed that children at age five understand the core meaning of the quantifiers *some (not)* and *most (not)* and behave like the adults. However, as could be seen from children's responses in the underinformative test condition, five-year-olds did not derive as many scalar implicatures as the adults did. Thus, hypotheses (H1) and (H2) can be confirmed. The non-significant difference between children and adults in the underinformative condition of the quantifier *some not* might be due to the big variance in the adult group. Furthermore, the results of the Pearson's χ^2-tests indicate that the German-speaking children did not give random responses in the underinformative test condition; they either consistently derived SIs or not at all. In addition, the present results suggest that children's interpretation pattern is not limited to the scalar quantifier *some*, but also holds for *most* and for the negated quantifiers *some not* and *most not*, although these quantifiers arguably differ in complexity.

Since quantifiers on the positive and negative scale did not differ in the number of scalar implicatures derived, the presence of sentential negation does not seem to impede children's ability to compute scalar implicatures. Hypothesis (H3) is therefore not supported. Whether this result is due to semantics of negation or due to the task design will need further investigation.

The number of scalar implicatures derived with *some* and *most* was the same in both subject groups, supporting Hypothesis (H4). Thus, an extension of the scale *<all, some>* to *most* did not influence the availability of implicatures. Note, however, that in the present design the visual context did not vary the number of elements depicted as being inside or outside of the box in the underinformative condition (cf. Table 1). It remains to be seen whether variation of these proportions result in differences wrt the computation of SIs.

4. Conclusion

This study provides first evidence that the overall pattern of acquisition of scalar implicatures reported previously also holds for German-speaking children. In underinformative conditions German-speaking five-year-old

children showed a clear preference for logical interpretations of scalar quantifiers such as *some* and *most*. The inclusion of a larger scale by using *most* and by using negated quantifiers showed that children's ability to compute scalar implicatures is robust and did not change depending on the specific quantifier used. Note that due to the low number of control items (cf. Pouscoulous, 2007) and due to the experimental setting including a visual context and a background story (cf. Papafragou & Musolino, 2003), the task demands in the present study were not very high. Nevertheless, the proportion of scalar implicatures computed was relatively low in both the adult and the children group. To support the derivation of scalar implicatures for existential quantifier 'some' in further experiments, other translations of the *some* could be used and as well partitive constructions such as *einige von den* 'some of the'. As reported by Pouscoulous et al. (2007) for French, the choice of *quelques* instead of *certain* as a translation of *some* led to an increase in scalar implicatures in French-speaking children. Whether this finding also holds for other translations of *most* needs further investigations. Finally, further crosslinguistic studies are required to investigate the influence of negation in the derivation and in the time course of development of scalar implicatures.

Acknowledgment

We thank Uli Sauerland and the Working Group 5 of the COST Action A33 for their cooperation.

References

Barker, C. (1998). Partitives, double genitives and anti-uniqueness. *Natural Language and Linguistic Theory*, 16, 679-717.

von Fintel, K. & Matthewson, L. (2008). Universals in semantics. *The Linguistic Review* 25(1–2), 139–201.

Gamut, L.T.F. (1991). Logic, Language, and Meaning, Volume 1. Un. of Chicago Press.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5)*. Göttingen: Hogrefe.

Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini A. & Mernoni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognition Processes*, 20 (5), 667-696.

Horn, (1976). On the semantic properties of the logical operators in English. Bloomington, IN: Indiana University Linguistics Club.

Katsos, N. (2008). The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature. *Synthese* 165 (3), 385-401.

Katsos, N. (2009). Evaluating under-informative utterances with context-dependent and context-independent scales: experimental and theoretical implications. In Sauerland, U. & Yatsushiro, K., *Experimental Semantics and Pragmatics*. Basingstoke: Palgrave Studies in Pragmatics, Language & Cognition. 51-73.

Koch, C. (2008). Konversationelle *Implikaturen – Logik oder Pragmatik?* Unpublished BA Thesis, University of Frankfurt.

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

Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition* 86 (3), 253-282.

Papafragou, A. & Tantalou, N. (2004). Children's Computation of Implicatures. Language Acquisition 12 (1), 71-82.

Pouscoulous, N., Noveck, I. A., Politzer, G. & Bastide, A. (2007). A Developmental Investigation of Processing Costs in Implicature Production. *Language Acquisition* 14 (4), 347-375.

Verbuk, A. (2007). Acquisition of scalar implicatures. *Electronic Doctoral Dissertations for UMass Amherst.* Paper AAI3290054.

Wason, P. C. (1961). Responses to affirmative and negative statements. *British Journal of Psychology*, 52, 133-142.