# Comprehension of scalar implicatures in five-year-old Dutch children

Faber, M.; Overweg, Jessica; van Hout, Anna Maria Henrica (Angeliek)

*Published in:*
Groninger Arbeiten zur germanistischen Linguistik

**Comprehension of scalar implicatures in five-year-old Dutch children: *some* but not *all* five-year old children draw implicatures**

**Myrthe Faber, Jessica Overweg and Angeliek van Hout**

**University of Groningen**

## 1 Introduction

### *1.1 Scalar implicatures*

This study focuses on the conversational implicature known as the scalar implicature[1]. Scalar implicatures may arise whenever expressions like *some* and *most* are used. These expressions form entailment scales, of which the members are ordered depending on the informational strength of utterances containing them, such as <*all, most, some*> (Horn, 1976). Propositions with *all* are more informative than *most*, which in turn are more informative than *some* (with informativity measured in terms of licensing entailments). In the case of *some*, *most* and *all*, *all* logically entails *some* and *most*, given their quantitative properties. When all items are in a box, some items are also in the box, and it is also true that most items are in the box. When the scalar expression *some* is used, the two different readings in (1) are possible:

(1) Readings of *some*
  a. Pragmatic reading: 'some but not all'
  b. Logical reading: 'some if not all' (Horn, 1972)

'Some but not all' is the pragmatically enriched reading. This reading excludes *all* while the semantic meaning, the non-enriched reading, is compatible with *all*. For instance, consider example (2).

(2) Some elephants have trunks.

Under a logical reading the statement in (2) is felicitous but under a pragmatic reading it is infelicitous. When the statement in (2) is taken literally, it is true because in the present world elephants generally have trunks. It logically follows that at least some elephants have trunks. But this statement can also be considered (pragmatically) infelicitous because it is underinformative. It would be more informative to use the quantifier *all* instead of *some*. According to the traditional Gricean account of such examples, if the speaker uses the underinformative or weak form on a scale, the listener may assume that the speaker is not in a position to use the informative or strong form on the scale (Grice, 1975). By using *some* instead of *all*

[1] This research is part of the EU-funded COST A33 project "Crosslinguistically Robust Stages of Children's Linguistic Performance, with Applications to the Diagnosis of Specific Language Impairment" (P.I. U. Sauerland, 2006-2010). Researchers from twenty-five different countries participate. The goal is to provide a cross-linguistically uniform picture of 5-year-olds' knowledge of grammar, which can serve as the basis for further research into clinical markers for the detection of SLI. The COST research themes include pronouns, quantification, implicatures, passives, tense and aspect, and questions.
http://www.zas.gwz-berlin.de/index.php?id=47&L=1

Groninger Arbeiten zur germanistischen Linguistik 51 (2010), 55-74
Center for Language and Cognition Groningen
http://gagl.eldoc.ub.rug.nl

a listener interprets the meaning in (2) with a pragmatic reading, namely as "not all elephants have trunks". This inference is called a *scalar implicature*.

With regard to the computation of a scalar implicature, the quantifier *most* behaves like *some*. For instance, consider example (3).

(3) Most elephants have trunks.

Example (3) is true in case the number of elephants who have trunks is larger than the number of elephants who do not (i.e., if more than half of the elephants have trunks). This literal meaning can be upper bounded by a scalar implicature ("more than half but not all") (Horn, 1972).

### *1.2 Acquisition research on scalar implicatures*

Noveck (2001) discovered that children systematically lack scalar implicatures. He used a statement evaluation task to test the computation of scalar implicatures in 8- and 10-year-old French-speaking children and found that children have difficulty accessing the pragmatically strengthened meaning of the weak scalar item *some*. Instead, they interpret an underinformative statement such as (2) with a logical reading, i.e., according to its literal meaning. Most adults consider this statement infelicitous because it is underinformative. They interpret the statement according to its pragmatic reading. Both observations also seem to hold cross-linguistically in a uniform fashion: in addition to Noveck's research on French, several other languages such as Greek, Italian and English have been investigated (respectively Papafragou & Musolino, 2003; Guasti, 2005; Katsos & Bishop, 2011; among others).

Papafragou and Musolino (2003) showed that children's performance with implicatures improves under certain conditions. In their experiment they tested two groups of five-year-old children; one group without a training phase and one group with a training phase. In this training phase children were trained to detect infelicitous statements. During the test phase a Truth Value Judgment Task was used, where all children were introduced to a puppet. The puppet said things that were true, but infelicitous. The children were asked to judge descriptions of acted-out stories given by the puppet. As a result of the training phase, the computation of scalar implicatures triggered by *some* improved from 12.5% (group without training phase) to 52.5% (group with training phase) for the five-year-old-children, even though this was still lower than in the adult group (92.5%). According to Papafragou and Musolino the introduction of training and clear informativeness expectations in the critical trials made children more likely to judge underinformative statements as bad descriptions of what happened. However, even in these contexts, children only showed sensitivity to scalar implicatures about half of the time.

Guasti et al. (2005) tested seven-year-old Italian speaking children using a statement evaluation task (replication of Noveck, 2001). In their experiment, one of two groups of children participated in a training session before the test phase. Their results were rather similar to the results of Papafragou and Musolino: the rejection rate of pragmatically underinformative statements was 12% for children without training, and rose to 52% with training. However, the children's responses in the group with training session formed a bimodal distribution, suggesting that the training had a strong effect on some children, but not on all children. Also, they showed that the effect of training is only temporary. When children were tested one week later (without training phase), the rejection rate of pragmatically underinformative statements dropped from 52% to 22%.

In addition to children, Guasti et al. (2005) tested adults using the same statement evaluation task. Although the level of performance by adults was superior to that of children, their performance on underinformative statements was not at ceiling. Adults rejected underinformative sentences only 50% of the time. To test whether the results of the children and adults were due to experimental design, Guasti et al. used a truth value judgment task. Their results with this task were strikingly different; 75% of the children and 83% of the adults rejected underinformative statements.

Pouscoulous et al. (2007) also showed that the type of task influences comprehension of scalar implicatures. They showed that the task is easier for children when using action-based judgments instead of meta-linguistic judgments. An action-based task would spare younger children's cognitive resources that they may in turn use to compute the scalar implicature. The results show that four-year-old and five-year-old children are able to compute an implicature in an action-based task (respectively 68% and 73%). Seven-year-old children (83%) show almost adult-like behavior (86%) on this task.

With regard to cross-linguistic comparability, the choice of the particular scalar expression used in the task has been shown to influence the results. For example, many languages have several ways to translate the English quantifier *some*. Pouscoulous et al. (2007) showed that French translations *quelques* and *certains* affect implicature drawing in children but not in adults. *Certains* is linguistically more complex because it is a partitive, while *quelques* is simpler. Nine-year-old French-speaking children are more likely to produce implicatures in an action-based task with the less complex quantifier *quelques* (100% of the time) compared to the more complex quantifier *certains* (58% of the time). This difference in implicature computation is attributed to the complexity of *certains*. Children have fewer resources left to compute the implicature.

Banga et al. (2009) showed that choice of scalar expression also influences the results of Dutch adults. They investigated the computation of scalar implicatures by Dutch adults using Dutch translations *sommige (*'some') and *enkele* ('some'). Purely quantitative terms such as Dutch *enkele* (85.5%) gave rise to more implicatures than partly qualitative terms such as *sommige* (75%), suggesting that subtle semantic differences between scalar terms affect the rate of implicatures in adults. Surprisingly, when investigating *sommige (*'some') but not *enkele* ('some'), their results show that adults drew implicatures with *sommige* only 41.2% of the time.

The study of Koch et al. (2010) provides further evidence that the overall pattern of acquisition of scalar implicatures also holds for German-speaking children. Children derived significantly fewer implicatures than adults (20% vs. 56%). In underinformative conditions, German-speaking five-year-old children showed a clear preference for logical interpretations of scalar quantifiers such as *some (not)* and *most (not)*. However, their performance on the semantic control conditions showed that children at age five do understand the core meaning of the quantifiers *some (not)* and *most (not)*, where they behave like the adults (respectively, 95.7% and 99.8% correct responses on the semantic control conditions).

In this study we focused on the acquisition of scalar implicatures of five-year-old children in a West-Germanic language, namely Dutch. Not much is known about the acquisition of scalar implicatures in Dutch children. In the next session we report on the research question and predictions that are at issue in this study.

## 2 Research questions

This experiment is designed to assess the understanding of quantifier semantics and the generation of scalar implicatures. It measures the child's ability to evaluate utterances based on their pragmatically enriched meaning, rejecting those that are underinformative with regard to the presented scene. Also, the experiment assesses the child's abilities to interpret and reject utterances in which the quantifier interacts with the scope of a negation (for example "not all" and "all … not")[2]. The first research question is therefore:

(4) How do five-year-old Dutch-speaking children score on a test designed to investigate whether they can draw implicatures?

Furthermore, we will look into these results in more detail analyzing the individual answer patterns. Different patterns may underlie the same group results. For example, Pouscoulous et al. (2007) have shown that four-year-old children have a group score of 68% correct for drawing implicatures in an action based task (note that Pouscoulous et al. do not report on the distribution of their results). Such a group score could be normally distributed, which means that, besides the outliers, each child scores within a normal range that has 68% as its mean. On the other hand, the scores of the group may not be normally distributed but in a different way, such as bimodally. For example, if most children always draw implicatures, while the others never do, this would also lead to a group percentage of 68%. This distribution is similar to what Guasti et al. (2005) found for Italian–speaking seven-year-olds. Examples of a normal and a bimodal distribution are given in Figure 1.

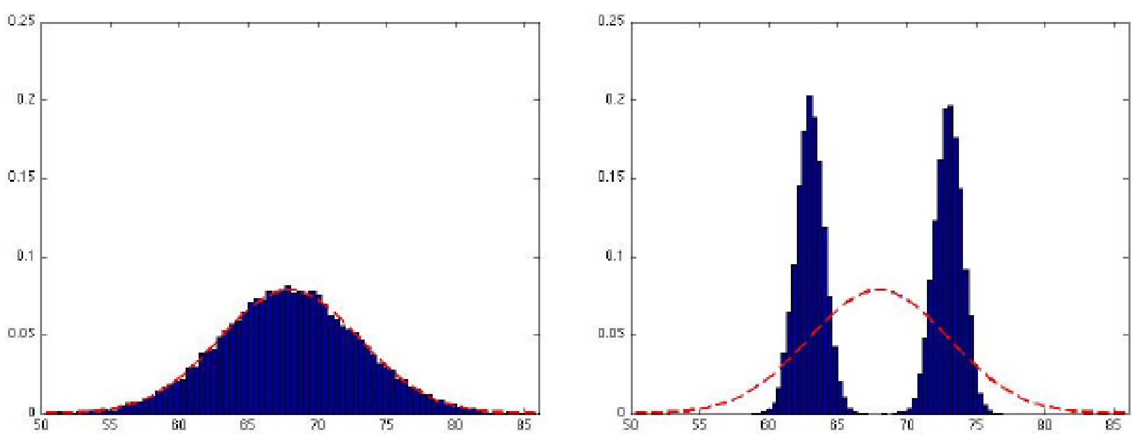



**Figure 1:** Illustration of unimodal (left) and bimodal distribution (right) with mean = 68 and standard deviation = 5.1

The second research question is therefore:

(5) What is the distribution of individual scores underlying the group data?

---

[2] http://cost.zas.gwz-berlin.de/cost/A33-e.pdf

If the subject analyses reveal that the distribution of the data is bimodal, suggesting that some children draw implicatures in most cases, while others do not, a third research question becomes pertinent:

(6) Why do some five-year-old children seem to draw implicatures, while other five-year-olds seem not to?

To answer this question, we will investigate whether (any one of) three factors affect implicature drawing. These factors are: general linguistics abilities (as measured by their scores on the GAPS test), age and the interpretation of the quantifiers used in control conditions. We test three predictions:

(7) a. Children with a higher score on the GAPS test draw more implicatures.
b. Children who are older draw more implicatures.
c. Children who have a higher score on the control items for quantifier semantics draw more implicatures.

The relevance of this study is two-fold. First, by presenting the first study on scalar implicatures in Dutch five-year-old children, we contribute to the development of an (potentially cross-linguistic) assessment tool for diagnosing linguistic delays. Second, we contribute to the fast growing body of literature on scalar implicatures and semantic-pragmatic development by carefully analyzing the distribution of individual scores and possible reasons why children do or do not draw scalar implicatures.

## 3 Method

### *3.1 Participants*

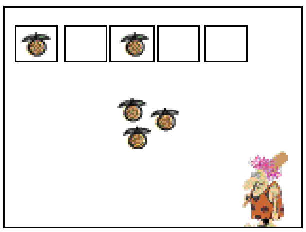
Twenty-four five-year-old children (mean age: 5;3; range: 5;0 to 5;7) and ten adult controls (mean age: 27) participated in this study. The children were recruited from three different kindergartens in Groningen, Haren and Emmen (Netherlands). All participants were native Dutch speakers.

### *3.2 Experimental design*

To assess the understanding of quantifiers and the child's ability to draw implicatures, a sentence evaluation task was used. This task was modelled along the lines of the *Quantifier Comprehension* task, used by Katsos et al. (2011). Sentences containing quantifiers were judged for a visual arrangement of five boxes that did or did not have (some) objects inside them. In Dutch, the following seven quantifying expressions were tested: *sommige* ('some'), *de meeste* ('the most'), *niet alle* ('not all'), *sommige…niet* ('some…not'), *alle* ('all'), *geen van de* ('none') and *alle…niet* ('all…not'). These quantifiers were combined with one of 25 object nouns to form statements like example (8).

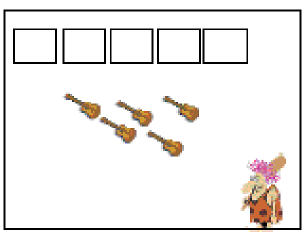
(8) Sommige appels liggen in de dozen.
some apples lie in the boxes
'Some [of the] apples are in the boxes.'

These statements were uttered in the context of one of the following four visual arrangements: 'all-arrangement' (where all five out of five objects are in the boxes), 'subset-arrangement' (where two of the objects are in the boxes), 'most-arrangement' (where four out of five objects are in the boxes), and 'none-arrangement' (where none of the objects are in the boxes). Depending on the quantifier and the arrangement of objects in the boxes in the visual display, the sentences are either true, false or underinformative. The dependent variable was the child's judgement. In our analysis, we focused on the quantifiers *sommige* ('some') and *de meeste* ('most'), because they trigger implicatures. These quantifiers were presented in three conditions: true, false and underinformative. The underinformative condition was the critical condition in which implicatures can be drawn; the other two were used as control conditions to investigate whether children fully understand the quantifiers on target. For each quantifier there were six items in the critical condition. Of the items in the control conditions, three were true and three were false. Figures 2, 3 and 4 illustrate examples of the three conditions for *sommige* ('some').



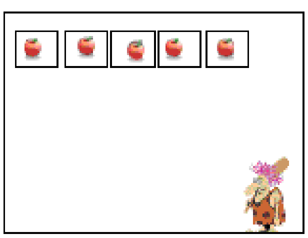
*Sommige sinaasappels liggen in de dozen.*
'Some oranges are in the boxes.'

**Figure 2:** Example control condition True: *sommige* ('some') in subset-arrangement



*Sommige gitaren liggen in de dozen.*
'Some guitars are in the boxes.'

**Figure 3:** Example control condition False: *sommige* ('some') in none-arrangement



*Sommige appels liggen in de dozen***.**
'Some apples are in the boxes.'

**Figure 4:** Example critical condition Underinformative**:** *sommige* ('some') in all-arrangement

For the quantifier *de meeste* ('most'), the true condition was presented with a most-arrangement, false condition with a sub-set arrangement and the underinformative (critical) condition with an all-arrangement.

We also used the understanding of *alle* ('all') and *geen* ('none') as additional control conditions. Both of these conditions consist of six items, of which three are true and three are false. For both the quantifiers *alle* ('all') and *geen* ('none'), the false condition is presented with a sub-set arrangement. The true conditions were an all-arrangement for *alle* ('all') and a none-arrangement for *geen* ('none'). The order of the items in the experiment was pseudo-randomized and the items were presented in six blocks. Also, the experiment was presented in three different lists to avoid any confounding factors due to the order of the items. For an overview of the used items, we refer to the appendix of this paper.

The most important independent variable is age. To keep this variable constant on the group level, the children that were tested in this study are between 5;0 and 5;7 years old. For the subject analyses, age was measured in months. Finally, we used the Dutch version of the Grammar and Phonology Screening (GAPS)[3] test to assess the child's general linguistic abilities. The GAPS test is a short screening test that can be used to asses the grammatical abilities and pre-reading skills of children between 3;6 and 6;6 years old. The two key abilities that are tested using GAPS are whether a child knows how to use grammatical rules to create sentences (sentence repetition) and whether the child knows the rules underlying consonant clusters (non-word repetition).

### *3.3 Procedure*

The experiment was presented to the child on a laptop, using Microsoft PowerPoint. The experimenter first introduced a fictional character, the Cave girl, and explained the task: "We are going to play a game on the computer with a friend of mine. Shall I introduce you to my friend? Her name is Cave girl!" Cave girl says in a pre-recorded voice in Dutch: "Hello! I am the Cave girl. I know quite a lot of Dutch and I would like to learn more and speak Dutch like you do! Will you help me?". Then, the experimenter says to the participant: "We are going to help the Cave girl to say how many things are in the boxes. Here are the boxes [POINTS TO THE BOXES ON THE TOP OF THE SCREEN], and here are the items." The experimenter asked the child to name all of the items, to make sure that the child understand all the nouns used in the experiment. When all the items are named, the experimenter continues: "We are going to see some pictures, and the Cave girl will say how many things are in the boxes. If she says it right, you tell her "that was right!". If she says it wrong, you tell her "that was wrong!". And to help her learn, you could tell her why it was wrong. So, are you ready?". Once the participant says that s/he is ready, Cave girl also asserts "OK! I am ready too." In the warming-up part of the experiment, the experimenter tells the child "Now, the Cave girl has to learn to use words like 'one', 'two', 'three', 'four', and 'five'". Before the main part of the experiment starts, the experimenter tells the participant "Now, the Cave girl has to learn to use words like 'all', 'none' and 'some'". The warming-up phase at the beginning of the experiment is designed to make sure the child is able to (a) count up to five, and (b) accept and reject utterances with number words.

### *3.4 Scoring and coding*

The responses given during the experiment were recorded using a voice recorder. The answers were scored on a score form by (one of) the experimenter(s). When a participant rejected an utterance, the experimenter asked why. These justifications for rejection were used to evaluate whether the participant actually drew the relevant implicature in the critical conditions, or whether the participant had other (trivial) reasons unrelated to falsity or informativeness for rejecting the utterance. When the justifications for rejecting an utterance were trivial, the answer was coded as "not

[3] See Gardner et al. (2006) for a more detailed description of the GAPS test.

drawing the implicature”. In this case the final answer was changed from ‘no’ to ‘yes’. However, if children seemed unsure of the final answer, but gave the right reasoning, showing that they actually drew the implicature, this was coded as “being aware of underinformativeness / drawing the implicature”. In this case the final answer was changed from ‘yes’ to ‘no’. Note however, that this could not be done systematically, because the children were only asked for a justification when they rejected the utterance. A few examples are given in Figure 5:

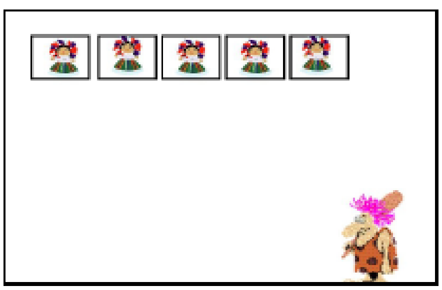
<u>Not drawing the implicature (irrelevant reason):</u>

*Sommige auto’s liggen in de dozen.*
‘Some (of the) cars are in the boxes.

Subject 7; age 5;03 :
“Nee, die gaan in een garage”
‘No, they go in a garage’

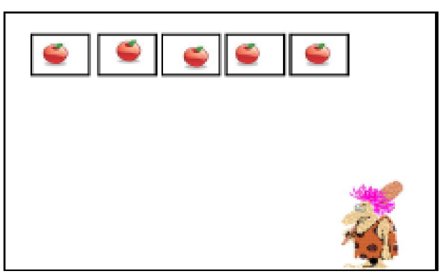
Subject 10; age 5;01 :
“Nee, te groot”
‘No, too big’

<u>Aware of underinformativeness / drawing implicature:</u>



*Sommige poppen liggen in de dozen.*
‘Some (of the) dolls are in the boxes.’

Subject 20; age 5;3 :
“Allen! Niet goed. Eh, ik bedoel wel goed.”
‘All of them! Not right. Eh, I mean right’



*De meeste appels liggen in de dozen.*
‘Most of the apples are in the boxes.’

Subject 20; age 5;3 :
“Niet, allen, niet goed. Eh, ik bedoel wel goed.”
‘No, all of them, not right. Eh, right I mean.’

**Figure 5:** Examples of explanations in children’s answers

## 4 Results

### *4.1 Results underinformative items*

Twenty-four children participated in this study. Four children were excluded from analysis. Two of them were excluded because they did not understand the experiment properly. They showed a ‘yes’-bias when all items were in boxes and a ‘no’-bias otherwise, regardless of the quantifier used. Two other children were excluded based on their scores on the control items for *alle* ('all') and *geen* ('no') (less than ten out of twelve items correct). In total, the scores of twenty children were used in the main analysis.

The test items involve *sommige* ('some') and *de meeste* ('most') in the all-arrangement). The mean group score on these underinformative items for *sommige* ('some') is 3.8 (out of a maximum score of six), or 63% rejection. The mean group score on the underinformative test items for *de meeste* ('most') is 2.75 (out of a maximum score of six), or 46% rejection. Children thus seem to draw implicatures more often in the *sommige* ('some') condition compared to the *de meeste* ('most') condition. The mean group score of the adults on these underinformative items for *sommige* ('some') is 4.3 (out of a maximum score of six), or 68% rejection. The mean group score on the underinformative test items for *de meeste* ('most') is 3.6 (out of a maximum score of six), or 60% rejection. Like the children, the adults also seem to draw implicatures more often in the *sommige* ('some') condition than in the *de meeste* ('most') condition.

*4.2 Distribution of results*

The individual scores of the children were used to look into the distribution the implicatures ('no'-answers). The histogram in Figures 6 and 7 represent the number of children who gave the target 'no'-answer 0, 1, 2, 3, 4, 5 or 6 times. The maximum score is six pragmatic target answers per child.

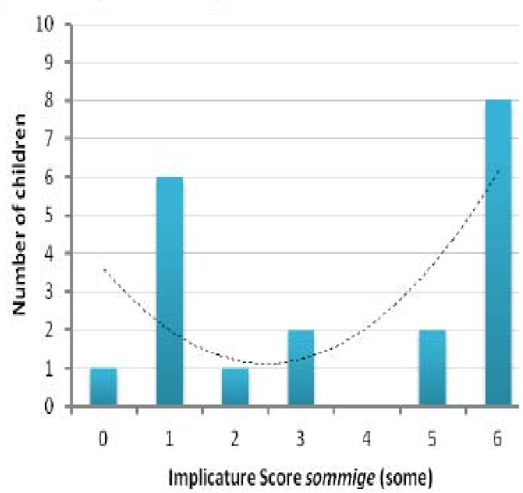



**Figure 6:** Distribution of number of implicatures for *sommige* ('some') in all-arrangement

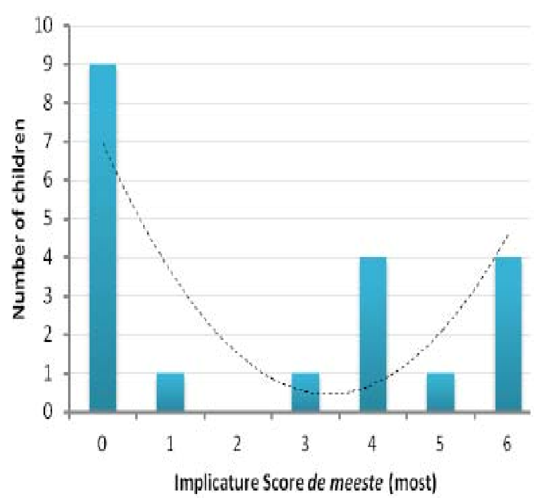



**Figure 7**: Distribution of number of implicatures for *de meeste* ('most') in all-arrangement

Figure 6 and 7 illustrate that most children were consistent with their answers and answered either "yes" or "no" to all or almost all of the test items. Only a few children were not consistent. The children on the left-hand side of the histograms are those who hardly ever or never draw an implicature and hence answer "yes"; the children on the right-hand side draw implicatures in most or all of the cases and hence answer "no". This bimodal distribution is found for both *sommige* ('some') and *de meeste* ('most').

Furthermore, we calculated the total number of implicatures drawn by each participant (children and adults) in the critical *sommige* ('some') and *de meeste* ('most') conditions together. These total numbers were then binned in three categories: zero to three implicatures, four to eight implicatures and nine to twelve implicatures. Figure 8 shows the percentage of children and the percentage of adults in each bin.

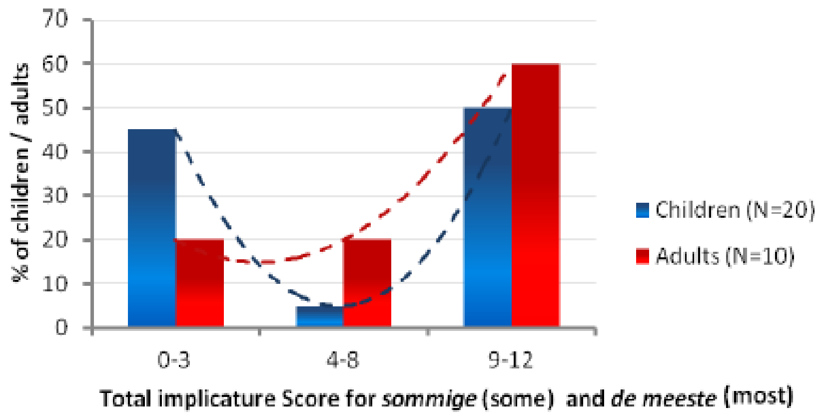



**Figure 8:** Distribution (in %) of implicatures drawn for *sommige* ('some') and *de meeste* ('most') in the all-arrangement

Figure 8 illustrates that the distribution of answers given by the adults is different from the distribution of answers given by the children. The children are on either end of the continuum: they only draw three or fewer implicatures, or they draw nine or more implicatures (only one child drew five out of twelve implicatures, consisting of three pragmatic answers in the *sommige* ('some') condition and two pragmatic answers in the *de meeste* ('most') condition). Adults on the other hand do not show a similarly clear bimodal distribution; rather, the data shows a skewed distribution.

Another observation is that there is a correlation between the scores in the critical conditions for *sommige* ('some') and *de meeste* ('most') for the children ($R^2$= 0.8276, $p < 0.0001$). This means that the children who have high scores on the *sommige* ('some') condition also have high scores on the *de meeste* ('most') condition. This is illustrated in Figure 9.

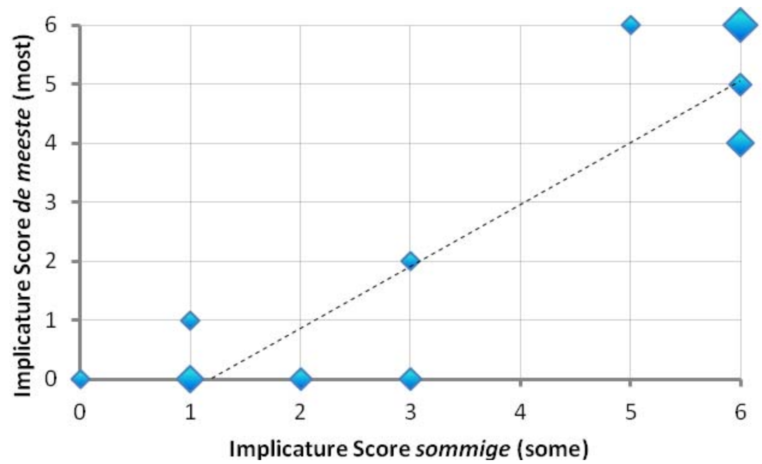



**Figure 9:** Scatterplot of the number of implicatures drawn by the children in the *sommige* ('some') and *de meeste* ('most') condition. The trendline is shown in dashes.

### *4.3 Hypotheses*

Given that the distribution of the data is bimodal, suggesting that some children draw implicatures in most cases, while others do not, the third research question becomes pertinent: *Why do some five-year-old children seem to draw implicatures, while other five-year-olds seem not to?* To answer this question, we further analyzed the data, looking into correlations between the implicature score and three other measures: general linguistic abilities, age and whether the child fully understands the quantifier.

#### *4.3.1 Linguistic abilities*

To test the hypothesis that children with a higher score on the GAPS test draw more implicatures, we calculated the correlation coefficient to analyze the relationship between general linguistic abilities (for which we used the GAPS-test as our measure) and whether a child draws an implicature or not. To see how well children performed on the implicature condition, we used the total implicature score; i.e. the 12 items of the *sommige* ('some') and *de meeste* ('most') test conditions together. In Figure 10, the score on the GAPS test is plotted on the y-axis (note that only eleven out of the twenty children were assessed using the GAPS). The maximal obtainable score on the GAPS test is 20 points. The total implicature score is on the x-axis; the higher the number, the more implicatures the child drew. Figure 10 illustrates that there is no relationship between GAPS and drawing an implicature; the correlation coefficient is near zero ($R^2 = 0.08$, $p = 0.39$).

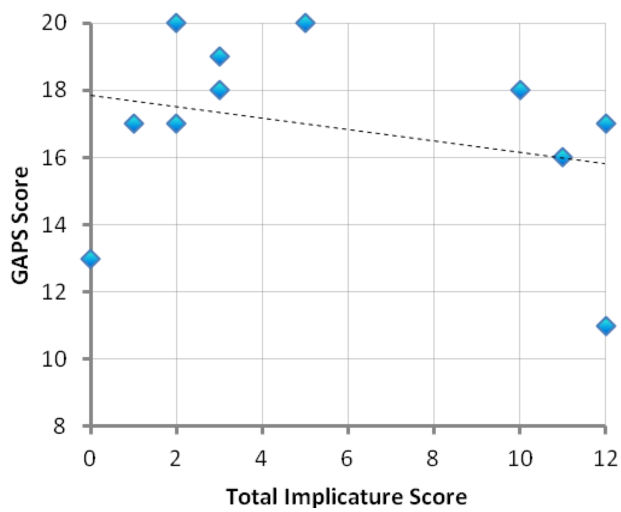



**Figure 10:** Scatterplot of the GAPS score and the total number of implicatures drawn in the *sommige* ('some') and *de meeste* ('most') condition. The trendline is shown in dashes.

#### *4.3.2 Age*

Similarly, we calculated the correlation coefficient to analyze the relationship between age and whether a child draws an implicature. In Figure 11, the age (years, months) is on the y-axis. The total implicature score is on the x-axis. Figure 11 illustrates that there is no relationship between age and whether a child draws an implicature ($R^2$= 0.09, p = 0.18).

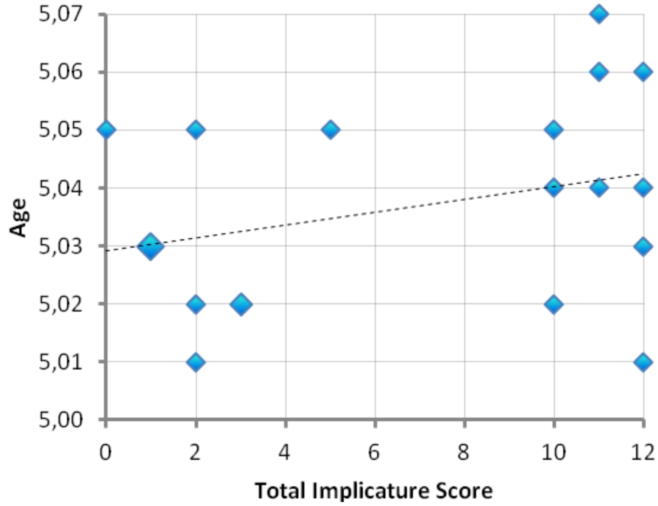



**Figure 11:** Scatterplot of the age of the child and the total number of implicatures drawn in the *sommige* ('some') and *de meeste* ('most') condition. The trendline is shown in dashes.

*4.3.3 Quantifier understanding*

Furthermore, we calculated the correlation coefficient to analyze the relationship between the understanding of each quantifier and whether a child draws an implicature or not. In figure 11, the score on the six control items for *sommige* ('some') is on the y-axis. In figure 12, the score on the six control items for *de meeste* ('most') is on the y-axis. Both figures show the number of implicatures on the x-axis; figure 12 for the *sommige* ('some') condition and figure 13 for the *de meeste* ('most') condition. Note that in both conditions, children were not at ceiling with regard to the control items.

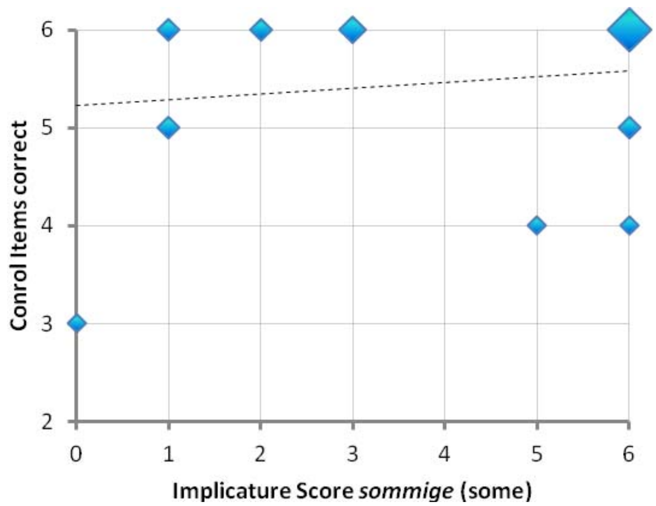



**Figure 12:** Scatterplot of the number of control items correct for *sommige* ('some') and the number of implicatures drawn for *sommige* ('some') in an all-arrangement. The trendline is shown in dashes.

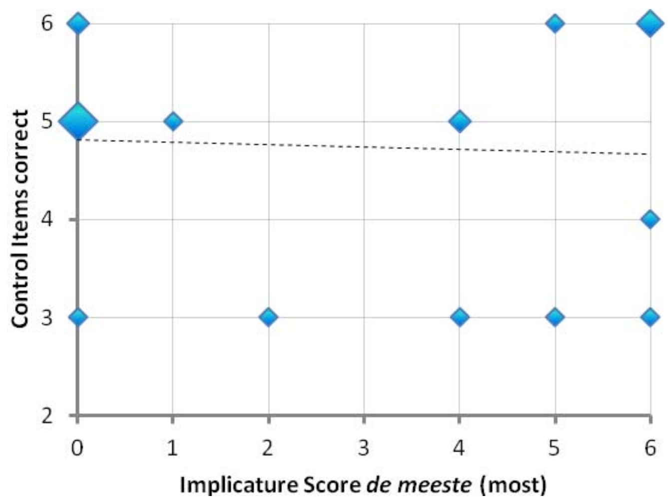



**Figure 13:** Scatterplot of the number of control items correct for *de meeste* ('most') and the number of implicatures drawn for *de meeste* ('most') in an all-arrangement. The trendline is shown in dashes.

As Figure 12 illustrates, there is no correlation between the score on the control items for *sommige* ('some') and drawing an implicature with *sommige* ('some'), $R^2 < 0.01$, $p = 0.52$. Similarly, figure 13 illustrates that there is no correlation between the score on the control items for *de meeste* ('most') and drawing an implicature with *de meeste* ('most'), $R^2 < 0.01$, $p = 0.81$.

*4.4 Results error analysis*

We also analyzed the errors children made in the control conditions for *sommige* ('some') and *de meeste* ('most'). In our experiment, the control conditions for *sommige* ('some') were the none-arrangement (false) and subset-arrangement (true). We hypothesize that children make errors in the control condition where *sommige* ('some') is presented in a subset-arrangement, suggesting that they have a different understanding of what *sommige* ('some') means, but do understand that it is different from none (note that this hypothesis also challenges a 'yes'-bias, because we predict that most of the errors will be in the true condition). We found that 92% of the errors are indeed in the condition where *sommige* ('some') is presented in a subset-arrangement (total number of errors = 13). A one-sided binomial test indicates that this finding is highly significant (p = .0017).

Furthermore, we analyzed the errors children made in the control items for *de meeste* ('most'). These control items consisted of *de meeste* ('most') in most-arrangement (TRUE) and *de meeste* ('most') in a subset-arrangement (FALSE). Here, we hypothesize that children will make most errors in the condition where *de meeste* ('most') is presented in a subset-arrangement, suggesting that they do not fully understand what *de meeste* ('most') means, because they perceive a subset-arrangement as most. We found that 71% of the errors are in the subset-arrangement (total number of errors = 24). A one-sided binomial test shows that this result is highly significant (p = .0320).

These analyses of errors in the critical condition for *sommige* ('some') and *de meeste* ('most) revealed that high percentage of the errors was made in control items in the arrangement where the quantifier is TRUE. We looked into the justifications given for rejection by the children to find out why they reject the utterances. Most children reported that the utterance with the quantifier *sommige* ('some') is wrong in a subset-arrangement because "there are (only) two objects in the boxes". They reject the quantifier *de meeste* ('most') in a most-arrangement for similar reasons, for example because "four objects are in the boxes" or "one object is not in the boxes".

## 5 Discussion

### *5.1 Lexical knowledge*

Our results show that drawing an implicature is not affected by general linguistics abilities, age and the interpretation of quantifier semantics. This raises an interesting issue: how can it be that children who do not fully understand a quantifier are still able to draw implicatures? We suggest that understanding that *some* and *most* differ from *all* (and *none*) is sufficient lexical information to draw an implicature. One might be able to draw an implicature if one knows that *some* and *most* are on a scale of informativeness, without knowing the core meaning. Our data illustrates this in the following way: the control items for *all* are presented in an all-arrangement (TRUE) and a subset-arrangement (FALSE). The control items for *none* are presented in a none-arrangement (TRUE) and a subset-arrangement (FALSE). Because we only selected children who score ten or more points out of twelve points on these control items, we know that the children in our data know the difference between *all* and *some* and *none* and *some*. However, this does not necessarily entail that children know exactly what *some* or *most* means. As we have seen in the error analysis, the children made most errors where the utterances with the quantifiers *some* and *most* were TRUE. This finding might be due to the difference in control condition where the utterance is TRUE, since *de meeste* ('most') was tested in a subset-arrangement while *sommige* ('some') was tested in a none-arrangement. The children in our data do know the difference between *none* and *most* (based on the fact that they do know the difference between *none* and *some*). If *de meeste* ('most') will be tested in a none-arrangement, the number of errors for *de meeste* ('most') will most probably be lower. Conversely, if *sommige* ('some') would be tested in a most-arrangement, this would probably generate more errors.

On the other hand, the justifications children gave for rejecting the utterances where the quantifiers are suppose to be TRUE, tell that they associated the quantifiers with a specific quantity of objects, suggesting they have problems with understanding the core meaning of the quantifiers *some* and *most*. According to Katsos et al. (2011) these errors are not related to logical complexity of the quantifiers but to the vagueness. Mastering the logical meaning of *some* requires understanding that the quantifier is true in a wide range of circumstances (in this case ranging from two or more to five) while *all* is true in only one situation, namely five out of five objects are in the boxes. Further research is required to learn more about comprehension of the quantifiers *sommige* ('some') and *de meeste* ('most') in five-year-old children.

### *5.2 Acquisition of informativeness*

Our results reflect a bimodal distribution: some children never or only very rarely draw an implicature, whereas other children draw an implicature most of the time. The question is how this pattern can be explained. Our results showed that children who draw implicatures with the quantifier *sommige* ('some') also draw implicatures with the quantifier *de meeste* ('most'). This finding suggests that once children acquire the ability to draw implicatures for one quantifier, it carries over to another quantifier (bearing in mind that *most* is overall harder than *some*). An important question that rises, is how informativeness develops. Guasti et al. (2005) report that achieving competence with the Gricean maxim of informativeness involves a categorical transition from underinformativeness to informativeness without fluctuation between the two stages. This may also explain why our results reflect a bimodal distribution: once children achieve competence with the maxim of informativeness they draw an implicature most of the time.

Another question that rises, is how informativeness can be mastered. Katsos et al. (2011) suggest that informativeness is mastered across the board, regardless of the specific quantifier mentioned. This is in line with the results of our data, given the fact that children who draw implicatures with one quantifier also draw implicatures with another quantifier. Katsos et al. also suggest that children cannot be expected to be informative and derive the scalar implicatures associated with a quantifier if they have not yet acquired what the quantifier means and discovered where it would enter into a scale of informativeness. However, this is not in line with our study, since the children who drew implicatures most of the times did not show a better understanding of the quantifier (see 5.1). More research is needed to answer the question how informativeness is mastered, and what the role of quantifier understanding is in implicature drawing.

### *5.3 Task effect: sentence evaluation tasks*

An important methodological issue is that because of the nature of the task, participants are forced to choose between acceptance and rejection. Even though the experiment was designed as a sentence evaluation task, the binary choice pushes it towards a truth-value judgment task (TVJT). In a sentence evaluation task subjects are asked to judge whether or not the sentence is (the most) appropriate (way) to describe a situation. However, in a TVJT subjects have to judge whether the proposition expressed by the sentence is true or false. An advantage of the truth-value judgment task is that the task can be used to access children's full range of grammatical competence, by allowing the researcher to test directly for various possible interpretations (Chien & Wexler, 1987). However, one of the objections against this kind of task is that because the utterance is true in terms of truth-value, children may not reject the utterance. A truth-value judgment task allows to see all of the possible interpretations that a child allows (van Hout et al., 2010), as opposed to other experimental designs that show which interpretation is preferred (Avrutin & Baauw, 2005). Examples are sentence-to-picture matching tasks and act-out tasks, which both show that children are informative comprehenders (Katsos & Smith, 2010). Although the choice of task in this study limits the conclusions that can be drawn about the actual understanding of informativeness in five-year-olds, we will argue below that there are other conclusions that can be drawn.

### *5.4 Pragmatic Tolerance Hypothesis*

Recent studies (Katsos & Bishop, 2011; Katsos & Smith, 2010) have shown that there is a production-comprehension asymmetry of underinformative utterances in five-year-old English speaking children. In production over 90% of these children are fully informative speakers, while in comprehension fewer than 30% of the underinformative utterances were rejected. In addition, these children were 100% able to reject false utterances, indicating that this pattern cannot be explained by an inability to reject utterances in general. So children seem to fail to comprehend pragmatic aspects of meaning, even though they do not seem to have any difficulty to produce pragmatically appropriate utterances as speakers (Davies & Katsos, 2010). This is in line with Reinhart's (2004) Interface Asymmetry account, which predicts no delays to arise in production while comprehension is comprised. In comprehension hearers must compare interpretations through an operation called reference-set computation but in production this is superfluous because the speaker "always knows which meaning he or she intends". According to Reinhart, comprehension is comprised due to working memory limitations concerning reference-set computation. Another possible explanation is that children have different expectations about the required level of informativeness in a specific situation than adults, and that children therefore do not see a need for rejecting underinformative utterances (Papafragou & Musolino, 2003). This view is expanded by the Pragmatic Tolerance hypothesis (Katsos & Smith, 2010; Davies & Katsos, 2010; Katsos & Bishop, 2011). The hypothesis predicts that "in certain cases, children are in fact pragmatically competent both as speakers and as comprehenders and what develops with age is their meta-linguistic awareness about accepting or rejecting pragmatically infelicitous utterances as comprehenders" (Davies & Katsos, 2010; p.1956). We argue that the age of 5;0 to 5;7 might be a critical age for the development of this meta-linguistic awareness. The bimodal distribution might reflect that some, but not all, children already have this meta-linguistic awareness.

Furthermore, the Pragmatic Tolerance Hypothesis is supported by the hypothesis that low child performance in comprehension is an artefact of the task with which underinformativeness has been studied (Katsos & Bishop, 2011). Katsos and Bishop hypothesize that the violation of pragmatics - unlike the violation of semantics - does not lead to utterances that are logically false, meaning that underinformative utterances are still in a sense "right". Children therefore may not reject underinformative utterance in a binary choice task, because rejection is the response appropriate to falsity. They also find that adults differentiate between false and underinformative utterances in an indirect way. While adults straightforwardly rejected semantically false utterances, many underinformative utterances were "objected to by accepting that the utterance was true, but highlighting that something important was not mentioned" (Davies & Katsos, 2010: 1960). Similar patterns were found in our data. First, we found that adults indirectly differentiate between false and underinformative utterances. Adults often came up with answers like "yes, it's true, but actually all of the items are in the boxes", indicating that something important was not mentioned. Second, we found that also children sometimes indirectly differentiate between false and underinformative utterances. The children were less specific, but we did see them hesitate and occasionally come up with answers like "all, not right, eh I mean right". This result might indicate that these children did not reject because rejection is the response to falsity. The pragmatic answer (rejection) is argued to reflect the insight in the logical entailment (when it is true that all items are in the boxes, it is also true that some items are in the boxes), as well as the insight in a scale (when there is a stronger term on the scale to describe the situation, use that term). The logical answer (acceptance) on the other hand does not necessarily rule out the pragmatic reasoning. Because of the experimental design and time limitations, the experimenters only asked for the reason for rejection, not for the reason for acceptance. Therefore, we may have overlooked important meta-linguistic reasoning underlying acceptance.

## 6 Conclusions

This study focused on the acquisition of scalar implicatures in five-year-old Dutch children. We investigated whether acceptance or rejection of an underinformative utterance depends on lexical knowledge of quantifiers, age or general language development. We have found that none of these three factors affects implicature drawing. We suggest that understanding that *some* and *most* differ from *all* (and *none*) is for children sufficient lexical information to draw an implicature. Children do not have to understand the core meaning of the quantifiers *some* and *most* to draw an implicature.

Furthermore, we found that children between 5;0 and 5;7 years old show different patterns of acceptance and rejection, resulting in a bimodal distribution. We argue that children may have different expectations about the informativeness of an utterance than adults (cf. Papafragou & Musolino, 2003). What develops with age is their meta-linguistic awareness about accepting or rejecting pragmatically infelicitous utterances (cf. Davies & Katsos, 2010). Our results imply that around the age of 5;0-5;7, there may be a critical period for the development of a child's meta-linguistic abilities.

Finally, we found that once children acquire the ability to draw implicatures for one quantifier, it carries over to another quantifier, suggesting that children achieve competence with the maxim of informativeness and become more informative once they draw implicatures with one quantifier. This implies that informativeness is mastered across the board, regardless of a specific quantifier (Katsos et al., 2011).

## Acknowledgments

This work forms part of a larger study on implicatures as developed in Katsos et al. (in prep.). The quantitative *er* experiment was designed and developed by Napoleon Katsos and others in the COST A33 Implicature group. We thank the audiences at the COST A33 Final Conference *Let the children speak* and the Acquisition Lab at the University of Groningen for their questions and feedback. We thank Napoleon Katsos for valuable comments and suggestions on this paper.

## References

Avrutin, Sergey, and Sergio Baauw. 2005. Truth Value Judgment Task vs. Picture Selection Task, Workshop Experimental Methodology in Language Acquisition Research, Universiteit Utrecht, November.

Banga, Arina; Heutinck, Ingeborg; Berends, Sanne, and Petra Hendriks. 2009. Some implicatures reveal semantic differences. In: Bert Botma & Jacqueline van Kampen (Eds), *Linguistics in the Netherlands 2009*. John Benjamins, Amsterdam. 1-13.

Chien, Yu-Chin, and Ken Wexler. 1987. A comparison between Chinese-speaking and English-speaking children's acquisition of reflexives and pronouns. Paper presented in the 12th Annual Boston University Conference on Language Development.

Davies, Catherine, and, Napoleon Katsos. 2010. Over-informative children: Production/ comprehension asymmetry or tolerance to pragmatic violations? *Lingua (Special Issue on Asymmetries in Child Language)*, 120: 1956-72.

Gardner, Hilary; Froud, Karen; McClelland, Alastair, and Heather van der Lely. 2006. The development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language difficulties in young children. *International Journal of Language & Communication Disorders*, 41: 513-540.

Grice, Paul. 1975. Logic and conversation. In: P. Cole & J. L. Morgan (eds), *Syntax and semantics: Speech acts,* 3: 41–58.

Guasti, Maria Theresa; Chiercia, Gennaro; Crain, Stephen; Foppolo, Francesca; Gualmini, Andrea, and Luisa Meroni. 2005. Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes,* 20: 667-696.

Hout, van, Angeliek; Harrigan, Kaitlyn, and Jill de Villiers. 2010. Asymmetries in the acquisition of definite and indefinite NPs. *Lingua,* 120: 1973-1990.

Horn, Laurence. 1972. On the Semantic Properties of Logical Operators in English. UCLA dissertation. Distributed by Indiana University Linguistics Club, 1976.

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

Katsos, Napoleon & COST Consortium. In prep. To what extent does the development of conceptual categories depend on language?

Katsos, Napoleon, and Dorothy Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120, 67-81.

Katsos, Napoleon; Andrés Roqueta, Clara; Estevan, Rosa Ana Clemente, and Chris Cummins. 2011. Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition*, 119, 43-57.

Katsos, Napoleon, and Nafsika Smith. 2010. Pragmatic Tolerance and speaker-comprehender asymmetries. In: Franich, K., Iserman, K. M., and Keil, L. L. (eds.), *Proceedings of the 34th Annual Boston Conference in Language Development*. Cascadilla Press, MA. 221-232.

Koch, Corinna; Schulz, Petra, and Napoleon Katsos. 2010. Do children compute some or most scalar implicatures? – Evidence from German. In: A. Castro, J. Costa, M. Lobo & F. Pratas (2010) (eds.), *Language Acquisition and Development*. Proceedings of GALA 2009. Cambridge: Cambridge Scholars Press.

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

Papafragou, Anna, and Julien Musolino. 2003. Scalar implicatures: experiments at the semantics pragmatics interface. *Cognition,* 86: 253-282.

Pouscoulous, Nausicaa; Noveck, Ira; Politzer, Guy, and Anne Bastide. 2007. A developmental investigation of processing costs in implicature production. *Language Acquisition,* 14: 347-375.

Reinhart, Tanja. 2004. The processing cost of reference-set computation: acquisition of stress shift and focus. *Language Acquisition*, 12: 109-155.

## Appendix Item list

The first column presents the utterance. In the second column you can find the arrangement of how many objects were inside the boxes and the third column encodes which response is correct.

### Numerals

| | | |
|---|---|---|
| Twee appels liggen in de dozen<br>'Two apples are in the boxes' | ARR: two | RIGHT |
| Drie appels liggen in de dozen<br>'Three apples are in the boxes' | ARR: four | WRONG |
| Eén appel ligt in de dozen<br>'One apple is in the boxes' | ARR: one | RIGHT |
| Vier appels liggen in de dozen<br>'Four apples are in the boxes' | ARR: three | WRONG |
| Vijf appels liggen in de dozen<br>'Five apples are in the boxes ' | ARR: five | RIGHT |

### *Alle* ('All')

| | | |
|---|---|---|
| Alle ballen liggen in de dozen<br>'All the balls are in the boxes' | ARR: all | RIGHT |
| Alle dino's liggen in de dozen<br>'All the dinosaurs are in the boxes' | ARR: all | RIGHT |
| Alle pennen liggen in de dozen<br>'All the pens are in the boxes' | ARR: all | RIGHT |
| Alle boterhammen liggen in de dozen<br>'All the sandwiches are in the boxes' | ARR: two | WRONG |
| Alle schoenen liggen in de dozen<br>'All the shoes are in the boxes' | ARR: two | WRONG |
| Alle T-shirts liggen in de dozen<br>'All the t-shirts are in the boxes' | ARR: two | WRONG |

### *Meeste* ('Most')

| | | |
|---|---|---|
| De meeste appels liggen in de dozen<br>'Most of the apples are in the boxes' | ARR: five | WRONG |
| De meeste gitaren liggen in de dozen<br>'Most of the guitars are in the boxes' | ARR: five | WRONG |
| De meeste teddy beren liggen in de dozen<br>'Most of the teddy-bears are in the boxes' | ARR: five | WRONG |
| De meeste treinen liggen in de dozen<br>'Most of the trains are in the boxes' | ARR: five | WRONG |
| De meeste sinaasappels liggen in de dozen<br>'Most of the oranges are in the boxes' | ARR: five | WRONG |
| De meeste aardbeien liggen in de dozen<br>'Most of the strawberries are in the boxes' | ARR: five | WRONG |
| De meeste vazen liggen in de dozen<br>'Most of the vases are in the boxes' | ARR: four | RIGHT |
| De meeste schoenen liggen in de dozen<br>'Most of the shoes are in the boxes' | ARR: four | RIGHT |
| De meeste poppen liggen in de dozen<br>'Most of the dolls are in the boxes' | ARR: four | RIGHT |
| De meeste boterhammen liggen in de dozen<br>'Most of the sandwiches are in the boxes' | ARR: two | WRONG |
| De meeste fietsen liggen in de dozen<br>'Most of the bicycles are in the boxes' | ARR: two | WRONG |

| | | |
|---|---|---|
| De meeste auto's liggen in de dozen<br>'Most of the cars are in the boxes' | ARR: two | WRONG |

***Sommige* ('Some')**

| | | |
|---|---|---|
| Sommige klokken liggen in de dozen<br>'Some of the clocks are in the boxes' | ARR: five | WRONG |
| Sommige bananen liggen in de dozen<br>'Some of the bananas are in the boxes' | ARR: five | WRONG |
| Sommige auto's liggen in de dozen<br>'Some of the cars are in the boxes' | ARR: five | WRONG |
| Sommige telefoons liggen in de dozen<br>'Some of the telephones are in the boxes' | ARR: five | WRONG |
| Sommige poppen liggen in de dozen<br>'Some of the dolls are in the boxes' | ARR: five | WRONG |
| Sommige bloemen liggen in de dozen<br>'Some of the flowers are in the boxes' | ARR: five | WRONG |
| Sommige sinaasappels liggen in de dozen<br>'Some of the oranges are in the boxes' | ARR: two | RIGHT |
| Sommige pennen liggen in de dozen<br>'Some of the pens are in the boxes' | ARR: two | RIGHT |
| Sommige peren liggen in de dozen<br>'Some of the pears are in the boxes' | ARR: two | RIGHT |
| Sommige rokjes liggen in de dozen<br>'Some of the skirts are in the boxes' | ARR: none | WRONG |
| Sommige appels liggen in de dozen<br>'Some of the apples are in the boxes' | ARR: none | WRONG |
| Sommige aardbeien liggen in de dozen<br>'Some of the strawberries are in the boxes' | ARR: none | WRONG |

***Geen* ('None')**

| | | |
|---|---|---|
| Geen klok ligt in de dozen<br>'None of the clocks are in the boxes' | ARR: none | RIGHT |
| Geen banaan ligt in de dozen<br>'None of the bananas are in the boxes' | ARR: none | RIGHT |
| Geen bal ligt in de dozen<br>'None of the balls are in the boxes' | ARR: none | RIGHT |
| Geen auto ligt in de dozen<br>'None of the cars are in the boxes' | ARR: two | WRONG |
| Geen ballon ligt in de dozen<br>'None of the balloons are in the boxes' | ARR: two | WRONG |
| Geen rokje ligt in de dozen<br>'None of the skirts are in the boxes' | ARR: two | WRONG |